Shape constrained nonparametric estimation in the Cox model

PROEFSCHRIFT

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

dinsdag **3 september 2013** om *15:00 uur*

door

Gabriela Florentina NANE

Master of Science in Applied Mathematics
geboren te Boekarest, Roemenië

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. G. Jongbloed

Copromotor: Dr. H. P. Lopuhaä

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. G. Jongbloed,	Technische Universiteit Delft, promotor
Dr. H. P. Lopuhaä,	Technische Universiteit Delft, copromotor
Prof. dr. J. A. Wellner,	University of Washington, Seattle
Prof. dr. I. van Keilegom,	Université catholique de Louvain
Prof. dr. A. H. Zwinderman,	Academic Medical Center, University of Amsterdam
Prof. dr. P. Groeneboom,	Technische Universiteit Delft
Prof. dr. F. H. J. Redig,	Technische Universiteit Delft
Prof. dr. ir. A. W. Heemink,	Technische Universiteit Delft, reservelid

Shape constrained nonparametric estimation in the Cox model
Dissertation at Delft University of Technology

ISBN: 978-94-6186-200-6

Printed by: CPI Wöhrmann Print Service, Zutphen, The Netherlands

To my family

Acknowledgements

Just like almost everything else in life, this thesis is the result of hard work and mere chance. The hard work was both scientific and non-scientific and was not only of myself, but also of supervisors, friends and family. I am indebted to all of them and hereby want to show my deepest appreciation and consideration. Furthermore, I had the chance to be in the right place at the right time, as much as I had the chance to meet the right people at the right time. Apparently, the "right" time for me started almost seven years ago, when I came to The Netherlands, and the "right" place turned out to be Delft. I will express now my gratitude to all my "right" people, and I hope I don't forget anyone.

Chronologically, my first "right" people are Roger Cooke, Arnold Heemink and Dorota Kurowicka, who I thank for giving me the opportunity to follow the applied master programme in "Risk and Environmental Modeling" at TU Delft. I am especially indebted to Roger Cooke for the opportunity to join an exciting project with NASA Langley Research Institute regarding aviation safety, which introduced me to a true applied setting. His constant enthusiasm and great desire to build bridges between theory and applications were truly inspirational for me.

Four and a half years ago, I embarked on a beautiful and intense academic journey that pursued a PhD degree. I want to express my deepest gratitude to my daily supervisor Rik Lopuhaä and promotor Geurt Jongbloed, for giving me this great opportunity, and for their considerable help and support throughout this journey, and especially while doing the most recent part of my research and writing my thesis. Thank you for encouraging me to present my research at various meetings, seminars and conferences. I thank Geurt for his prompt implication in various scientific and non-scientific matters, and of course for the very instructive conversations in Dutch. I am most deeply indebted to Rik without whom finishing this PhD would not have been possible. His patience and his guidance into rigorous reasoning and writing a rigorous proof made it possible for me to complete a theoretical research. When needed in terms of time constraints, this guidance grew into a collaboration, especially towards the end of my first pregnancy. I am also grateful for his valuable advice, both academic and personal. Rik, thank you for everything!

I want to thank all the members of my dissertation committee for accepting to take part in the whole process of my PhD defense and for reading my thesis. I am especially thankful to Jon Wellner for his comments on the thesis, but also for his help with one of the proofs of my first paper.

I want to thank all my friends, colleagues and acquaintances from TU Delft and outside the university, for your help with various matters and for spending lots of cheerful time together during these years, for sharing helpful information or even for a nice chat in Dutch. Special thanks go to Lidka Burzała-Kowalczyk and Birgit Witte for sharing some great time inside and outside our office and for providing me with any information and material that I needed, sometimes even before I realized I needed them. I want to express my appreciations to Cindy Bosman and Carl Schneider and Roniet Sharabi, for all their help and assistance with administrative and computer problems. I also want to thank my friends Anca Hanea, Dana and Remus Hanea and Samira Safaei Farahani. During those years, I was so lucky to have Anca so closely and that she was always willing to listen to my (un)funded and mostly incoherent worries, or to discuss parental issues. Thank you! Special gratitude goes to our (wedding) godparents Dana and Remus, for being our family "mentors", for being next to us in the good and hard moments, even now when 1600 km set us apart. I thank Samira for get togethers and discussions, but also for the best recipe of cooked rice in the world.

Finally, I have reached the hardest part of this thesis, where I try to show my gratitude, love and appreciations to my dearest, by finding as less shallow words as possible. I dedicate this thesis to my family. The years I have spent doing my PhD were not only prolific academic-wise, but also personally. I have gained not only a wonderful husband, but also four parents, a brother and a sister. I want to thank all my parents for their love, constant help, support and encouragement. I am truly grateful that my dearest brother is close by us and I thank him for that!

My last words are dedicated to the love of my life. I thank you for constantly having more faith in me than myself. I thank you for your endless patience and for your relaxed attitude that calms my choleric personality. I thank you for constructively supporting all my decisions. I thank you for the wonderful family that we have and that we expect, and for being a wonderful father. I thank you for teaching me how to be a better me. I thank you for being you!

Tina Nane
Delft, July 2013

CONTENTS

CHAPTER 1

INTRODUCTION

The occurrence of time-related events of interest are thoroughly investigated in statistics. For medical research, the event is mainly death, which led to the name survival analysis, as well as the onset (relapse) of a disease, time to the first heart attack, etc. Consequently, the event of interest is frequently named survival time, and the probability that the event of interest occurs later than some specified time point t is called the survival or the survivorship function at time t. In economics, these studies are part of what is referred to as duration analysis, in sociology they are called event history analysis while in engineering, the field is called reliability theory, which studies the life history of machines.

The main interest in survival analysis is to study the distribution of the event times of interest. Although the common statistical approach would be to estimate the distribution or the density function, a more natural approach is to focus on the rate of occurrence, since one is interested in studying the event of interest across time. The rate of occurrence is commonly referred to as the hazard function; some researchers suggestively called it "force of mortality". In reliability theory it is called the failure rate.

Furthermore, a peculiar characteristic in survival analysis is that the data is recorded over a (pre)specified period of time, and hence not all events are observed, and some subjects can be lost to follow-up during the study. Such observations are called censored. There are various censoring schemes, including right or left censoring, and interval censoring. In this thesis, the events of interest are assumed to be right censored, which implies that for some subjects, the event of interest occurs after the end of the study. The censoring mechanism is usually assumed to be independent of the event of interest, and moreover, to be non-informative.

It is commonly of interest to analyze how different characteristics of subjects, such as age, sex, or undergoing treatment affect the distribution of the event of interest. These characteristics are referred to as covariates and for time-independent covariates, this information is usually recorded at the beginning of the study. Within survival analysis, the most popular method to investigate the hazard function, while accounting for covariates is the Cox model. In the Cox model, the hazard of a subject given a set of covariates can be expressed in terms of a baseline hazard, for which all the covariates are zero, weighted by an exponential function of the covariates.

Even though the baseline hazard can be left completely unspecified, in practice, it is often reasonable to assume a qualitative shape. This can be done by assuming the baseline hazard to be monotone, for example, as suggested by Cox himself. Various studies have indicated that a monotonicity constraint should be imposed on the baseline hazard, which complies with the medical expertise. Time to death, infection or development of a disease of interest are observed to have a nondecreasing baseline hazard in most studies. Nevertheless, the survival time after a successful medical treatment is usually modeled using a nonincreasing baseline hazard. Therefore, it would be highly desirable to provide estimates that incorporate the shape restrictions of the baseline hazard while preserving the flexible semiparametric setting of the Cox model.

In this thesis, shape constrained estimators of the baseline hazard and density function within the Cox model will be defined and their asymptotic properties will be examined. In addition to point estimates, interval estimates of a monotone baseline hazard function will be provided, based on a likelihood ratio method, along with testing at a fixed point. Finally, kernel smoothed monotone baseline hazard estimates will be considered and their asymptotic properties will be investigated.

1.1 MONOTONE HAZARD ESTIMATION

This thesis focuses on monotone hazard estimation. This section introduces the concept of monotone hazard function or failure rate, discusses acknowledged estimators for monotone hazards in different models, as well as asymptotic properties of these estimators.

For a random variable X, the hazard function is defined as

$$\lambda(x) = \lim_{\Delta x \to 0} \frac{P(x \leqslant X \leqslant x + \Delta x | X > x)}{\Delta x}.$$

If the distribution of X is assumed to be absolutely continuous, then the hazard function can also be expressed as

$$\lambda(x) = \frac{f(x)}{1 - F(x)},$$

where f is the density function, and F is the cumulative distribution function of the random variable X. It is frequently of interest to consider another characteristic of the event time distribution, namely the cumulative hazard function, which is defined as

$$\Lambda(x) = \int_0^x \lambda(u) \mathrm{d}u. \tag{1.1.1}$$

The cumulative hazard function can also be expressed in terms of the cumulative distribution function in the following way

$$\Lambda(x) = -\log[1 - F(x)]. \qquad (1.1.2)$$

1.1.1 CHARACTERIZATION OF THE ESTIMATORS

Estimating distributions with monotone hazard functions has received considerable attention starting with the forefront work of GRENANDER (1956), who derived a maximum likelihood estimator for a distribution function with a nondecreasing failure rate. BARLOW *et al.* (1963) provided additional properties for distributions with monotone failure rates. Furthermore, MARSHALL & PROSCHAN (1965) showed that the maximum likelihood estimator of a nondecreasing failure rate is a right-continuous step function that is 0 before the first observation and jumps to infinity at the largest observation.

MAXIMUM LIKELIHOOD ESTIMATOR

To illustrate this, let $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ be the ordered observations. The likelihood function is

$$L = \prod_{i=1}^{n} f(X_{(i)}) = \prod_{i=1}^{n} \lambda(X_{(i)})[1 - F(X_{(i)})].$$

Then, by (1.1.1) and (1.1.2), the loglikelihood, written as a function of the hazard rate λ, is given by

$$l(\lambda) = \sum_{i=1}^{n} \left[\log \lambda(X_{(i)}) - \int_0^{X_{(i)}} \lambda(u) \mathrm{d}u \right]. \qquad (1.1.3)$$

Since $\lambda(X_{(n)})$ can be chosen arbitrarily large, maximizing over nondecreasing λ bounded by some $M > 0$ will be considered first. This translates to maximizing the loglikelihood function in (1.1.3) over $0 \leqslant \lambda(X_{(1)}) \leqslant \cdots \leqslant \lambda(X_{(n)}) \leqslant M$. It can be easily seen that the loglikelihood function is maximized by minimizing the second term in the sum and hence choosing the hazard to be constant between observations, and moreover, to take the minimal possible value on each interval $[X_{(i)}, X_{(i+1)}]$. Let $\lambda(X_{(i)}) = \lambda_i$. The maximization problem reduces then to maximizing the following objective function

$$\sum_{i=1}^{n-1} \left[\log \lambda_i - (n-i)[X_{(i+1)} - X_{(i)}]\lambda_i \right], \qquad (1.1.4)$$

subject to $0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_{n-1} \leqslant M$, as given by (7.4.4) in ROBERTSON *et al.* (1988). The resulting estimator $\hat{\lambda}_n^M(x)$ is a right-continuous step function of the following form

$$\hat{\lambda}_n^M(x) = \begin{cases} 0 & x < X_{(1)}, \\ \hat{\lambda}_i & X_{(i)} \leqslant x < X_{(i+1)}, \text{ for } i = 1, 2, \ldots, n-1, \\ M & x \geqslant X_{(n)}. \end{cases}$$

Moreover, by Theorem 1.4.4 in ROBERTSON *et al.* (1988),

$$\hat{\lambda}_i = \min_{i \leqslant t \leqslant n-1} \max_{1 \leqslant s \leqslant i} \frac{t-s+1}{\sum_{j=s}^{t}(n-j)\left[X_{(j+1)} - X_{(j)}\right]},$$

for $i = 1, 2, \ldots, n-1$. Finally, letting $M \to \infty$ yields the maximum likelihood estimator (MLE) $\hat{\lambda}_n$.

GRENANDER-TYPE ESTIMATOR

GRENANDER (1956) derived the maximum likelihood estimator of a nonincreasing density and additionally showed that the left-hand slope of the least concave majorant of the empirical distribution function coincides with the nonincreasing density maximum likelihood estimator. While this may not necessarily hold for other estimators in different (censoring) models, it advanced an attractive recipe of producing monotone estimators, usually referred to as Grenander-type estimators. Thus, for constructing an estimator of a nonincreasing function, one has to consider an estimator of the integrated function, and take slopes of the least concave majorant of that estimator. Similarly, to obtain an estimator of a nondecreasing function, one has to take slopes of the greatest convex minorant of the estimator of the integrated function.

To illustrate this method for a nondecreasing hazard function, consider an estimator of the cumulative hazard function. By (1.1.2), a natural estimator, referred to as the empirical cumulative hazard function, is

$$\Lambda_n(x) = -\log[1 - F_n(x)], \tag{1.1.5}$$

with F_n the empirical distribution function. Let $\widetilde{\Lambda}_n$ be the greatest convex minorant of Λ_n, which is the greatest convex function lying below the empirical cumulative hazard function. We then define the estimator $\tilde{\lambda}_n$ as the left-hand slope of the greatest convex minorant $\widetilde{\Lambda}_n$ of the empirical cumulative hazard function Λ_n. The Grenander-type estimator has thus a convenient graphical representation. Specifically, construct the cusum diagram consisting of the points

$$P_j = (X_{(j)}, \Lambda_n(X_{(j)})),$$

for $j = 1, 2, \ldots, n-1$ and $P_0 = (0, 0)$. Then, $\tilde{\lambda}_n$ is the left-hand slope of the greatest convex minorant of this cusum diagram. Unlike in the monotone density case, the Grenander-type estimator $\tilde{\lambda}_n$ is different from the maximum likelihood estimator $\hat{\lambda}_n$.

To see this, we will make use of results from ROBERTSON *et al.* (1988) and notice that the objective function in (1.1.4) can also be written as

$$\sum_{i=1}^{n-1} (g_i \log \lambda_i - \lambda_i)\, w_i,$$

where, for $i = 1, 2, \ldots, n-1$,

$$w_i = (n-i)[X_{(i+1)} - X_{(i)}],$$

and

$$g_i = \frac{1}{(n-i)[X_{(i+1)} - X_{(i)}]}.$$

As a result of Theorem 1.5.1 and 1.2.1 in ROBERTSON *et al.* (1988), the value $\hat{\lambda}_i$ can also be represented as the left derivative at P_i of the greatest convex minorant of the cumulative sum diagram consisting of the points

$$P_i = \left(\frac{1}{n} \sum_{j=1}^{i} w_j, \frac{1}{n} \sum_{j=1}^{i} w_j g_j \right),$$

for $i = 1, 2, \ldots, n-1$, and $P_0 = (0, 0)$. This shows that this cusum diagram differs from the cusum diagram of the Grenander-type estimator.

EXAMPLE 1. (U.S. power nuclear plants)
Failure data in the nuclear industry has been investigated by KVAM *et al.* (2002), under the setting of a nondecreasing failure rate in an imperfect repair model. The interest of the study is the estimation of the reliability of two large and repairable components of the nuclear power plants. The event times of interest are thus the failure times for groups of emergency diesel generators and motor-driven pumps. For the study of emergency diesel generators, failure data were collected between 1976 and 1978, in three U.S. nuclear power plants: Calvert Cliffs, Big Rock Point and Zion I and II. This example focuses only on the 24 emergency diesel generators failure times at Calvert Cliffs nuclear power plant. A maximum likelihood estimator is proposed by KVAM *et al.* (2002) to estimate the failure time distribution of plants.

Figure 1.1 shows the Grenander-type and the maximum likelihood estimator of the assumed nondecreasing failure rate (hazard function) at the Calvert Cliffs nuclear power plant. Note that the failure times are expressed in days. Furthermore, it is worth mentioning that the authors provide no underlying motivation for the increasing failure rate assumption, apart from the general consensus that "a significant proportion of working components in industry [...] are known to have an increasing failure rate".

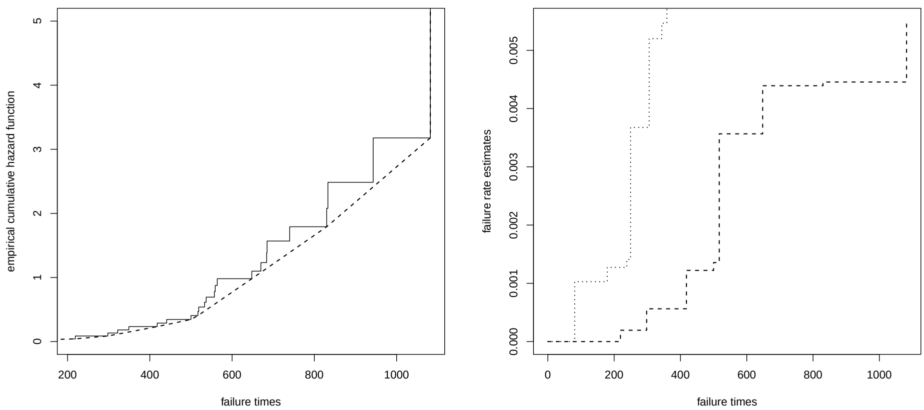

FIGURE 1.1: Left panel: The empirical cumulative hazard function (solid line) along with its GCM (dashed line). Right panel: The corresponding Grenander-type (dashed line) and maximum likelihood (dotted line) estimators of a nondecreasing failure rate of the Calvert Cliffs nuclear power plant.

RIGHT CENSORING MODEL

Within the right censoring model, suppose that the observed data consist of the following pairs $(T_1, \Delta_1), (T_2, \Delta_2), \dots, (T_n, \Delta_n)$, where the generic follow-up time is defined as $T = \min(X, C)$, with X denoting the event time and C the censoring time, and where $\Delta = \{X \leqslant C\}$ is the censoring indicator. Suppose that the distribution function F of the event time X is absolutely continuous with density f and that the censoring time C has an absolutely continuous distribution function G with density g. The event time X and the censoring time C are assumed to be independent. Moreover, the censoring mechanism is assumed to be non-informative, which implies that the distributions F and G share no parameters. For the sake of simplicity, we will further use the same notation for the estimators in the right censoring model as in the case of no covariates.

Monotone maximum likelihood hazard estimation within the right censoring model has been initially studied by PADGETT & WEI (1980) and MYKYTYN & SANTNER (1981). The Grenander-type estimator has been proposed and its asymptotic properties have been investigated by HUANG & WELLNER (1995), who also provided the asymptotic distribution of the maximum likelihood estimator of a monotone hazard. The Grenander-type estimator is defined as the left-hand slope of the greatest convex minorant of the Nelson-Aalen estimator of the cumulative hazard function,

$$\Lambda_n(t) = \sum_{T_i \leqslant t} \frac{d_i}{n_i}, \tag{1.1.6}$$

where d_i is the number of events at time T_i and n_i is the number of observations greater than T_i (the number of individuals at risk at time T_i). The Nelson-Aalen estimator is different from the natural correspondent in the right-censoring model of the cumulative hazard estimator in (1.1.5), which is sometimes referred to as the Kaplan-Meier estimator (of the cumulative hazard function).

Let $T_{(1)} \leqslant T_{(2)} \leqslant \cdots \leqslant T_{(n)}$ be the ordered observations and $\Delta_{(1)}, \Delta_{(2)}, \ldots, \Delta_{(n)}$ their corresponding censoring indicators. The likelihood function is then given by

$$L = \prod_{i=1}^{n} [f(T_{(i)})(1 - G(T_{(i)}))]^{\Delta_{(i)}} [g(T_{(i)})(1 - F(T_{(i)}))]^{1-\Delta_{(i)}}.$$

Since the censoring mechanism is assumed to be independent of the event times and non-informative, the resulting function to be maximized is the following (pseudo)loglikelihood

$$l = \sum_{i=1}^{n} \Big[\Delta_{(i)} \log f(T_{(i)}) + (1 - \Delta_{(i)}) \log[1 - F(T_{(i)})] \Big],$$

which can be re-written in terms of the hazard and cumulative hazard function

$$l = \sum_{i=1}^{n} \Big[\Delta_{(i)} \log \lambda(T_{(i)}) - \Lambda(T_{(i)}) \Big].$$

Evidently, this (pseudo)loglikelihood function differs from the loglikelihood in (1.1.3). Following the same reasoning as in the case of no censoring, it can be shown that the maximum likelihood estimator of a nondecreasing hazard function is obtained by maximizing the following objective function

$$\sum_{i=1}^{n-1} \Big\{ \Delta_i \log \lambda_i - (n - i)[T_{(i+1)} - T_{(i)}]\lambda_i \Big\}, \tag{1.1.7}$$

where $\lambda_i = \lambda(T_{(i)})$. As in the case of no censoring, one can obtain the maximum likelihood estimator following results in ROBERTSON *et al.* (1988). It is noteworthy that the above objective function differs slightly from the objective function in HUANG & WELLNER (1995) and consequently the two resulting estimators are different. However, it can be easily shown that the two estimators are asymptotically equivalent.

1.1.2 Asymptotic distribution via inverse processes

The pointwise asymptotic distribution of the maximum likelihood estimator $\hat{\lambda}_n$ for the uncensored data has been derived by PRAKASA RAO (1970). For a fixed x_0 in the interior of the support of the distribution and under mild conditions,

$$n^{1/3}\left(\frac{1-F(x_0)}{4\lambda(x_0)\lambda'(x_0)}\right)^{1/3}\left[\hat{\lambda}_n(x_0)-\lambda(x_0)\right] \xrightarrow{d} \underset{t\in\mathbb{R}}{\operatorname{argmin}}\{\mathbb{W}(t)+t^2\}, \tag{1.1.8}$$

where $\mathbb{W}$ is standard two-sided Brownian motion originating from zero and the argmin function represents the supremum of times at which the minimum is attained. In his proof, PRAKASA RAO (1970) followed the same approach as in deriving the asymptotic distribution of the nonincreasing density.

GROENEBOOM (1985) derived the asymptotic distribution of a nonincreasing density in a more elegant manner and his proof relied on the equivalence between the maximum likelihood estimator and the left-hand slope of the least concave minorant of the empirical distribution function. His novel approach introduced and made use of a so-called inverse process, defined in terms of the empirical distribution function. Even though, as mentioned before, the equivalence between the two estimators might not necessarily hold, his approach produced a general method for acquiring the asymptotic distribution of a monotone estimator. His method also applies to the maximum likelihood estimator, for example, since, as shown in the previous section, the maximum likelihood estimator of a nondecreasing function can be represented as the left-hand slope of the greatest convex minorant of a given cusum diagram. In this respect, the inverse process is represented in terms of the processes defining the cusum diagram.

To the author's best knowledge, the asymptotic distribution of the Grenander-type hazard estimator in the case of no censoring has not been derived. Groeneboom's approach is illustrated below in the case of right censoring, by reproducing the results of HUANG & WELLNER (1995). Consider the inverse process

$$U_n(a) = \underset{x\in[0,T_{(n)}]}{\operatorname{argmin}}\{\Lambda_n(x)-ax\}, \tag{1.1.9}$$

for $a>0$, where Λ_n is the Nelson-Aalen estimator in (1.1.6). The following switching relationship holds with probability one

$$U_n(a) \geqslant x \Leftrightarrow \tilde{\lambda}_n(x) \leqslant a.$$

Figure 1.2 below exhibits a clear graphical representation of the switching relationship. The 100 event times were generated from a Weibull distribution with shape parameter 3/2 and the censoring times were assumed to be uniform $[0,1]$.

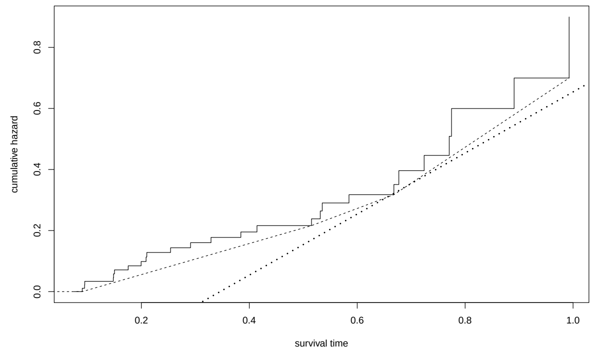

FIGURE 1.2: Nelson-Aalen estimator (solid line) along with its greatest convex minorant (dashed line). The dotted line represents a line with slope a.

This relationship enables the derivation of the asymptotic distribution of $\tilde{\lambda}_n$ through the more tractable asymptotic distribution of the inverse process U_n, since, for a fixed x_0,

$$P\left(n^{1/3}[\tilde{\lambda}_n(x_0) - \lambda(x_0)] > x\right) = P\left(\tilde{\lambda}_n(x_0) > \lambda(x_0) + n^{-1/3}x\right)$$
$$= P\left(U_n(\lambda(x_0) + n^{-1/3}x) < x_0\right) = P\left(n^{1/3}[U_n(\lambda(x_0) + n^{-1/3}x) - x_0] < 0\right).$$

By the definition of the inverse process in (1.1.9), and given that the argmin is invariant under addition of and multiplication with positive constants, it can be derived that

$$n^{1/3}[U_n(\lambda(x_0) + n^{-1/3}x) - x_0] = \underset{x \in I_n(x_0)}{\operatorname{argmin}}\{\mathbb{Z}_n(x) - ax\}, \tag{1.1.10}$$

where $I_n(x_0) = [-n^{1/3}x_0, n^{1/3}(T_{(n)} - x_0)]$. The process

$$\mathbb{Z}_n(x) = n^{2/3}\left[\Lambda_n(x_0 + n^{-1/3}x) - \Lambda_n(x_0) - n^{-1/3}\lambda(x_0)x\right]$$

is usually referred to as the local process. The weak convergence of $\mathbb{Z}_n$ as a process in the space of all locally bounded real-valued functions endowed with the topology of uniform convergence on compact intervals follows then from the Hungarian embedding result in BURKE *et al.* (1988),

$$\mathbb{Z}_n(x) \xrightarrow{d} \mathbb{Z}(x) = \mathbb{W}\left(\frac{\lambda(x_0)}{H(x_0)}x\right) + \frac{1}{2}\lambda(x_0)x^2,$$

where $H(t) = P(T \leqslant t)$ is the distribution function of the observed data. An extension of the argmax continuous mapping theorem in KIM & POLLARD (1990) gives that

$$n^{1/3}[U_n(\lambda(x_0) + n^{-1/3}x) - x_0] \xrightarrow{d} \underset{x \in \mathbb{R}}{\operatorname{argmin}}\{\mathbb{Z} - ax\},$$

which finally leads to the asymptotic distribution of the nondecreasing estimator $\tilde{\lambda}_n$,

$$n^{1/3}\left(\frac{1-H(x_0)}{4\lambda(x_0)\lambda'(x_0)}\right)^{1/3}\left[\tilde{\lambda}_n(x_0)-\lambda(x_0)\right]\xrightarrow{d}\underset{t\in\mathbb{R}}{\operatorname{argmin}}\{\mathbb{W}(t)+t^2\},$$

In the case of no censoring, $H(t)$ reduces to $F(t)$ and the above asymptotic distribution coincides with the asymptotic distribution in (1.1.8).

Strong pointwise convergence of the monotone hazard estimators has been investigated as well and results are available. A proof of the pointwise strong consistency of the Grenander-type estimator can be found in GROENEBOOM & JONGBLOED (2013), in the case of no censoring, while in the right censoring setting, the Grenander-type estimator is shown to be strongly consistent by HUANG & WELLNER (1995).

1.1.3 THE LIKELIHOOD RATIO METHOD

Ensuing inference can be pursued, by testing the hypothesis that the underlying monotone hazard has a particular value ϑ_0, at a fixed point x_0. To this end, a likelihood ratio test of $H_0 : \lambda(x_0) = \vartheta_0$ versus $H_1 : \lambda(x_0) \neq \vartheta_0$ can be used. Within shape restricted problems, this approach was initially employed for monotone distributions in the current status model, by BANERJEE & WELLNER (2001). The authors focused on deriving the limiting distribution of the likelihood ratio test under the null hypothesis, and obtained what the authors referred to as a fixed universal distribution, defined in terms of slopes of the greatest convex minorant of the two-sided Brownian motion plus a parabola. These findings were followed by a rapid stream of research (e.g., see BANERJEE & WELLNER, 2001; BANERJEE, 2007; BANERJEE, 2008), which revealed that the likelihood ratio method could be extended straightforwardly to other shape constrained settings.

In the right censoring model, the limiting distribution of the likelihood ratio test has been derived by BANERJEE (2008). Since the objective function in (1.1.7) differs slightly from the objective function in HUANG & WELLNER (1995), BANERJEE (2008) considers estimators which are different from the ones described below. Nonetheless, the asymptotic distribution of the likelihood ratio test can be easily shown to be the same.

The maximum likelihood estimator $\hat{\lambda}_n = (\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_{n-1})$ is considered to be the unrestricted estimator, that is obtained by maximizing (1.1.7) over all $0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_{n-1}$, where $\lambda_i = \lambda(T_{(i)})$. Let now m such that $T_{(m)} < x_0 < T_{(m+1)}$. Then, the constrained estimator $\hat{\lambda}_n^0 = (\hat{\lambda}_1^0, \hat{\lambda}_2^0, \dots, \hat{\lambda}_{n-1}^0)$, the maximum likelihood estimator under the null hypothesis $H_0 : \lambda(x_0) = \vartheta_0$, is obtained by maximizing (1.1.7) over all $0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_m \leqslant \vartheta_0 \leqslant \lambda_{m+1} \leqslant \cdots \leqslant \lambda_{n-1}$. In line with the reasoning for the unconstrained estimator, it can be argued that the constrained estimator has to be a nondecreasing step function that is zero for $x < T_{(1)}$, constant on the interval $[T_{(i)}, T_{(i+1)})$, for $i = 1, 2, \dots, n-1$, is equal to ϑ_0 on the interval $[x_0, T_{(m+1)})$, and can be chosen arbitrarily large for $x \geqslant T_{(n)}$. The objective function to be maximized reduces then to

$$\sum_{i=1}^{m-1}\left\{\Delta_i \log\lambda_i - (n-i)[T_{(i+1)}-T_{(i)}]\lambda_i\right\} + \Delta_m \log\lambda_m - (n-m)\left[x_0 - T_{(m)}\right]\lambda_m + \sum_{i=m+1}^{n-1}\left\{\Delta_i \log\lambda_i - (n-i)[T_{(i+1)} - T_{(i)}]\lambda_i\right\}.$$

The likelihood ratio statistic for testing $H_0 : \lambda(x_0) = \vartheta_0$ is thus

$$\begin{aligned} 2\log\xi_n(\vartheta_0) = & 2\left\{\sum_{i=1}^{n-1}\left\{\Delta_i \log\hat{\lambda}_i - (n-i)[T_{(i+1)} - T_{(i)}]\hat{\lambda}_i\right\}\right. \\ & - \sum_{i=1}^{m-1}\left\{\Delta_i \log\hat{\lambda}_i^0 - (n-i)[T_{(i+1)} - T_{(i)}]\hat{\lambda}_i^0\right\} \\ & - \Delta_m \log\hat{\lambda}_m^0 + (n-m)[x_0 - T_{(m)}]\hat{\lambda}_m^0 \\ & \left. - \sum_{i=m+1}^{n-1}\left\{\Delta_i \log\hat{\lambda}_i^0 - (n-i)(T_{(i+1)} - T_{(i)})\hat{\lambda}_i^0\right\}\right\}. \end{aligned}$$

The asymptotic distribution of $2\log\xi_n(\vartheta_0)$ is defined in terms of slopes of the process

$$\mathbb{X}(t) = \mathbb{W}(t) + t^2,$$

where $\mathbb{W}$ is standard two-sided Brownian motion starting from zero. More specifically, let g be the left-hand slope of the greatest convex minorant of the process $\mathbb{X}$, which will be denoted by G. Moreover, the constrained analogue is defined as follows: for $t \leqslant 0$, construct the GCM of $\mathbb{X}$, that will be denoted by G^L and take its left-hand slope at point t, denoted by $D_L(\mathbb{X})(t)$. When the slope exceeds zero, replace it by zero. In the same manner, for $t > 0$, denote the GCM of $\mathbb{X}$ by G^R and its slope at point t by $D_R(\mathbb{X})(t)$. Replace the slope by zero when it decreases below zero. This slope process will be denoted by g^0. Then,

$$2\log\xi_n(\vartheta_0) \xrightarrow{d} \mathbb{D},$$

where $\mathbb{D} = \int\left[(g(u))^2 - (g^0(u))^2\right] \mathrm{d}u$.

Furthermore, confidence sets for $\lambda(x_0)$ can be derived, based on the likelihood ratio method. More specifically, it will be used that inverting the family of tests can yield, in turn, pointwise confidence intervals for the hazard function. Let $2\log\xi_n(\vartheta)$ denote the likelihood ratio for testing $H_0 : \lambda(x_0) = \vartheta$ versus $H_1 : \lambda(x_0) \neq \vartheta$. A $100(1-\alpha)\%$ confidence interval is then obtained by inverting the likelihood ratio test $2\log\xi_n(\vartheta)$ for different values of ϑ, namely

$$\left\{\vartheta : 2\log\xi_n(\vartheta) \leqslant q(\mathbb{D}, 1-\alpha)\right\},$$

where $q(\mathbb{D}, 1-\alpha)$ is the $(1-\alpha)^{th}$ quantile of the distribution $\mathbb{D}$. Quantiles of $\mathbb{D}$, based on discrete approximations of Brownian motion, are provided in BANERJEE & WELLNER (2005).

Another method of constructing pointwise confidence intervals is based on the asymptotic distribution, at a fixed point x_0, of the nonparametric maximum likelihood estimator $\hat{\lambda}_n$, derived in HUANG & WELLNER (1995). Recall that

$$\begin{aligned} n^{1/3}\left[\hat{\lambda}_n(x_0) - \lambda(x_0)\right] &\xrightarrow{d} \left(\frac{4\lambda(x_0)\lambda'(x_0)}{1-H(x_0)}\right)^{1/3} \underset{t\in\mathbb{R}}{\operatorname{argmin}}\{\mathbb{W}(t)+t^2\} \\ &\equiv C(x_0)\underset{t\in\mathbb{R}}{\operatorname{argmin}}\{\mathbb{W}(t)+t^2\}, \end{aligned}$$

where $C(x_0)$ depends on x_0 and on the underlying parameters. An estimator $\hat{C}_n(x_0)$ of $C(x_0)$ will then yield a $100(1-\alpha)\%$ confidence interval for $\lambda(x_0)$

$$\left[\hat{\lambda}_n(x_0) - n^{-1/3}\hat{C}_n(x_0)q(\mathbb{Z}, 1-\alpha/2), \hat{\lambda}_n(x_0) + n^{-1/3}\hat{C}_n(x_0)q(\mathbb{Z}, 1-\alpha/2)\right],$$

where $q(\mathbb{Z}, 1-\alpha/2)$ is the $(1-\alpha/2)^{th}$ quantile of the distribution $\mathbb{Z}$. These quantiles have been computed in GROENEBOOM & WELLNER (2001). Nonetheless, this method entails estimating the nuisance parameter, and more specifically, estimating the derivative of the hazard function $\lambda'(x_0)$. One option would be to kernel smooth the NPMLE $\hat{\lambda}_n$, which will be investigated in the following section, for uncensored data.

1.1.4 SMOOTH HAZARD ESTIMATES

As emphasized in GROENEBOOM & JONGBLOED (2013), there are various approaches to construct smooth shape constrained estimators. It essentially depends on the order of operations, i.e., first isotonize and then smooth or first smooth and then isotonize. Moreover, these approaches depend on the method of isotonization, whether it involves a maximum likelihood or a Grenander-type estimation. Results involving only kernel smoothing will be detailed further.

For uncensored data, GROENEBOOM & JONGBLOED (2013) propose a smooth monotone hazard estimator, that kernel smooths the Grenander-type hazard estimator $\tilde{\lambda}_n$. More specifically, consider

$$\tilde{\lambda}_n^{SG}(x) = \int k_b(x-u)\,\mathrm{d}\tilde{\Lambda}_n(u) = \int k_b(x-u)\tilde{\lambda}_n(u)\mathrm{d}u,$$

where $\tilde{\Lambda}_n$ is the greatest convex minorant of the empirical hazard function Λ_n in (1.1.5), and for a bandwidth $b = b_n > 0$, $k_b(u) = (1/b)k(u/b)$ is the scaled version of kernel density k with compact support.

The monotonicity of $\tilde{\lambda}_n^{SG}$ follows from the monotonicity of $\tilde{\lambda}_n$. This method provides a straightforward estimate of the derivative of the hazard, $\tilde{\lambda}_n^{SG'}(x) = \int k_b(x -$

$u)\mathrm{d}\tilde{\lambda}_n(u)$, which can be used, for example, to obtain confidence intervals based on the asymptotic distribution of the Grenander-type estimator $\tilde{\lambda}_n$.

Moreover, by Theorem 3.1 in GROENEBOOM & JONGBLOED (2013), for a bandwidth b_n such that $n^{1/5}b_n \to \nu \in (0,\infty)$, as $n \to \infty$, and x_0 in the interior of the support,

$$n^{2/5}\left[\tilde{\lambda}_n^{SG}(x_0) - \lambda(x_0)\right] \xrightarrow{d} \mathcal{N}(\mu_0(\nu), \sigma_0^2(\nu)),$$

where

$$\mu_0(\nu) = \frac{1}{2}\nu^2\lambda''(x_0)\int u^2k(u)\mathrm{d}u, \qquad \sigma_0^2(\nu) = \frac{\lambda^2(x_0)}{\nu f(x_0)}\int k^2(u)\mathrm{d}u.$$

The inconsistency of the Grenander-type estimator at the boundaries is inherited by the smoothed estimator. This problem can be solved by using a boundary kernel.

1.2 COX PROPORTIONAL HAZARDS MODEL

The Cox proportional hazards model is one of the most popular approaches to model right-censored time to event data in the presence of covariates. Let the observed data consist of independent identically distributed triplets (T_i, Δ_i, Z_i), with $i = 1, 2, \ldots, n$, where T_i denotes the follow-up time, with a corresponding censoring indicator Δ_i and covariate vector $Z_i \in \mathbb{R}^p$. A generic follow-up time is defined by $T = \min(X, C)$, where X represents the event time and C is the censoring time. Accordingly, $\Delta = \{X \leqslant C\}$, where $\{\cdot\}$ denotes the indicator function. The event time X and censoring time C are assumed to be conditionally independent given Z, and the censoring mechanism is assumed to be non-informative. The covariate vector $Z \in \mathbb{R}^p$ is assumed to be time invariant.

Within the Cox model, the distribution of the event time is related to the corresponding covariate by

$$\lambda\left(x|z\right) = \lambda_0(x)\,\mathrm{e}^{\beta_0' z}, \tag{1.2.1}$$

where $\lambda\left(x|z\right)$ is the hazard function for an individual with covariate vector $z \in \mathbb{R}^p$, λ_0 represents the baseline hazard function and $\beta_0 \in \mathbb{R}^p$ is the vector of the underlying regression coefficients. Conditionally on $Z = z$, the event time X is assumed to be a non-negative random variable with an absolutely continuous distribution function $F(x|z)$ with density $f(x|z)$. The same assumptions hold for the censoring variable C and its distribution function G. The distribution function of the follow-up time T is denoted by H. We will assume the following conditions, which are commonly employed in deriving large sample properties of Cox proportional hazards estimators (e.g., see TSIATIS, 1981).

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

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

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

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

where $|\cdot|$ denotes the Euclidean norm.

The first assumption (A1) implies that, with positive probability, at the end of the study there is at least one subject alive, i.e. there is at least one event of interest that has not been observed. The second assumption (A2) concerns the covariates. Numerous studies of the Cox model assume bounded covariates. In this thesis, the results are proven under a milder condition, that the covariates have a bounded second moment, for β in a neighborhood of the underlying regression parameter β_0.

EXAMPLE 2. (Bone marrow transplant data)
Patients receiving bone marrow transplant as a treatment for leukemia frequently develop complications. The causes may be the presence of a certain virus or that the transplanted (grafted) immune cells attack the host tissue, which is known as graft-versus-host disease (GVHD). As a consequence, one might expect that the development of those complications would increase the risk of patients dying in remission or the risk of leukemic relapse. The data was provided by the Medical College of Wisconsin, Division of Biostatistics, Department of Population Health and the sample included 137 patients from 4 hospitals. The data is available in KLEIN & MOESCHBERGER (1997). The observed time is the time to death or on study time, which is right-censored. There is data on 17 covariates, representing factors that could influence the successfulness of the transplant. Out of these covariates, it is worthwhile mentioning the type of leukemia the patients have, whether they develop acute or chronic GVHD, the age and sex of the patient and donor, the waiting time for the transplant, the hospital where the transplant took place, whether the donor or patient has the CMV virus and a certain drug (MTX) usage. It is of interest to investigate what factors have the most important impact and what kind of patient groups are more exposed to the risk of death or relapse.

1.2.1 THE BRESLOW ESTIMATOR

Let $\Lambda(x|z) = -\log[1-F(x|z)]$ be the conditional cumulative hazard function given $Z = z$. Then, from (4.2.1) it follows that $\Lambda(x|z) = \Lambda_0(x)\exp(\beta_0' z)$, where $\Lambda_0(x) = \int_0^x \lambda_0(u)\,du$ denotes the baseline cumulative hazard function. When G has a density g, then together with the relation $\lambda = f/(1-F)$, the likelihood becomes

$$\prod_{i=1}^n \lambda(T_i \mid Z_i)^{\Delta_i} \exp\left[-\Lambda(T_i \mid Z_i)\right] \times \prod_{i=1}^n \left[1 - G(T_i \mid Z_i)\right]^{\Delta_i} g(T_i \mid Z_i)^{1-\Delta_i}.$$

The term with g does not involve the baseline distribution and can be treated as a constant term. Therefore, one essentially needs to maximize

$$\prod_{i=1}^{n} \lambda(T_i \mid Z_i)^{\Delta_i} \exp\left[-\Lambda(T_i \mid Z_i)\right] = \prod_{i=1}^{n} \left[\lambda_0(T_i) e^{\beta_0' Z_i}\right]^{\Delta_i} \exp\left[-e^{\beta_0' Z_i} \Lambda_0(T_i)\right]. \quad (1.2.2)$$

This leads to the following (pseudo) loglikelihood, written as a function of $\beta \in \mathbb{R}^p$ and λ_0,

$$\sum_{i=1}^{n} \left[\Delta_i \log \lambda_0(T_i) + \Delta_i \beta' Z_i - e^{\beta' Z_i} \Lambda_0(T_i)\right]. \quad (1.2.3)$$

Let $X_{(1)} < X_{(2)} < \cdots < X_{(m)}$ denote the ordered, observed survival times. COX (1972) introduced the proportional hazards model and focused on estimating the underlying regression coefficients of the covariates. He later showed (COX, 1975) that his proposed estimator $\hat{\beta}_n$ is the maximizer of the partial likelihood function

$$\prod_{l=1}^{m} \frac{e^{\beta' Z_l}}{\sum_{j=1}^{n} \{T_j \geqslant X_{(l)}\} e^{\beta' Z_j}}. \quad (1.2.4)$$

The asymptotic properties of the maximum partial likelihood $\hat{\beta}_n$ were broadly studied by TSIATIS (1981), ANDERSEN *et al.* (1993), OAKES (1977) and SLUD (1982), among others.

Moreover, different functionals of the lifetime distribution are commonly investigated and the (cumulative) hazard function is of particular interest. In the discussion following COX's (1972) paper, Breslow focused on estimating the baseline cumulative hazard function, $\Lambda_0(x) = \int_0^x \lambda_0(u)\, du$, and proposed

$$\Lambda_n(x) = \sum_{i \mid X_{(i)} \leqslant x} \frac{d_i}{\sum_{j=1}^{n} \{T_j \geqslant X_{(i)}\} e^{\hat{\beta}_n' Z_j}}, \quad (1.2.5)$$

as a nonparametric maximum likelihood estimator of the baseline cumulative hazard function Λ_0, where d_i is the number of events at $X_{(i)}$ and $\hat{\beta}_n$ is the maximum partial likelihood estimator of the regression coefficients. The estimator Λ_n is most commonly referred to as the Breslow estimator. Under the assumption of a piecewise constant baseline hazard function and assuming that all the censoring times are shifted to the preceding observed event time, Breslow showed that the maximum partial likelihood estimator $\hat{\beta}_n$ along with the baseline cumulative hazard estimator Λ_n can be obtained by jointly maximizing the full loglikelihood function, via a profile likelihood method.

For β fixed, Breslow maximized the (pseudo) loglikelihood function in (1.2.3) and obtained

$$\lambda_i = \frac{d_i}{[X_{(i)} - X_{(i-1)}] \sum_{j=1}^{n} \{T_j \geqslant X_{(i)}\} e^{\beta' Z_j}}.$$

Substituting λ_i in the (pseudo) likelihood function in (1.2.3) yields exactly the partial likelihood function in (1.2.4). Therefore, substituting β by $\hat{\beta}_n$ gives the estimates $\hat{\lambda}_i^B$ of the baseline hazard function. Nonetheless, the estimates $\hat{\lambda}_i^B$ were not regarded to provide a baseline hazard estimator, but as an intermediate step in computing the cumulative baseline hazard estimator Λ_n. BURR (1994) proved that the baseline hazard estimator $\hat{\lambda}^B = (\hat{\lambda}_1^B, \hat{\lambda}_2^B, \dots, \hat{\lambda}_n^B)$ is inconsistent. Unfortunately, this estimator is frequently referred to in the literature as the (baseline hazard) Breslow estimator.

It is noteworthy that in the case of no covariates, i.e., $\beta = 0$, the Breslow estimator reduces to the Nelson-Aalen estimator in (1.1.6). Figure 1.3 depicts the Breslow estimator for the bone marrow transplant data in Example 2.

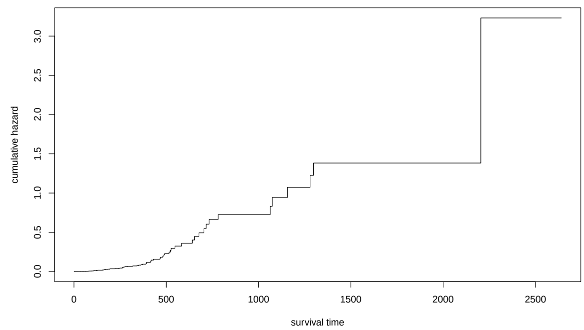

FIGURE 1.3: The Breslow estimator of the baseline cumulative hazard function for the bone marrow transplant data in Example 2.

Asymptotic properties of the Breslow estimator, such as consistency and the asymptotic distribution, were derived by TSIATIS (1981) and ANDERSEN *et al.* (1993). For an overview of the Breslow estimator, see LIN (2007).

1.3 OUTLINE OF THE THESIS

The research compiled in this thesis focuses on shape constrained nonparametric estimation within the Cox model. It is of main interest to estimate the baseline hazard function under the assumption of monotonicity and investigate the estimators' asymptotic properties. The standard piecewise constant isotonic estimators are provided, as well as smoothed versions of these estimators. Furthermore, along with baseline hazard point estimates, interval estimates obtained through the likelihood ratio method are examined. Finally, a baseline nonincreasing density estimator together with its asymptotic properties is also included in the thesis. This research amounts in four distinct papers, which are reproduced in the following four chapters.

Chapter 2 contains the first paper, which investigates the nonparametric estimation of a monotone baseline hazard λ_0 and a nonincreasing baseline density f_0 within the Cox model. Two estimators of a nondecreasing baseline hazard function are proposed, the nonparametric maximum likelihood estimator and a Grenander-type estimator, defined as the left-hand slope of the greatest convex minorant of the Breslow estimator Λ_n. It is demonstrated that the two estimators are strongly consistent and asymptotically equivalent and the common limit distribution at a fixed point is derived. Estimators of a nonincreasing baseline hazard are considered as well and their asymptotic properties are acquired in a similar manner. Furthermore, a Grenander-type estimator for a nonincreasing baseline density is defined as the left-hand slope of the least concave majorant of an estimator of the baseline cumulative distribution function, derived from the Breslow estimator. This estimator is shown to be strongly consistent, and its asymptotic distribution at a fixed point is derived.

GROENEBOOM's (1985) approach of deriving the asymptotic distribution of a monotone estimator makes use of an inverse process, but also of a Hungarian embedding type of result for the integrated estimator. To the author's best knowledge, there is no such result available in the literature for the Breslow estimator. The lack of a Hungarian embedding for the Breslow estimator poses serious problems in deriving the asymptotic distribution of shape constrained baseline hazard estimators. Nonetheless, this problem can be circumvented by using a linearization result of the Breslow estimator together with the theory in KIM & POLLARD (1990). The linearization result of the Breslow estimator is the subject of the second paper in Chapter 3.

The second paper thus provides an asymptotic linear representation of the Breslow estimator of the baseline cumulative hazard function in the Cox model. The representation consists of an average of independent random variables and a term involving the difference between the maximum partial likelihood estimator and the underlying regression parameter. The order of the remainder term is shown to be arbitrarily close to n^{-1}. This result extends the result in KOSOROK (2008), showing that the Breslow estimator is asymptotically linear with a given influence function and a remainder term of order $o_p(n^{-1/2})$, while relaxing the strong assumption of bounded covariates.

Chapter 4 contains the third paper, that considers a likelihood ratio method for testing whether a monotone baseline hazard function in the Cox model has a particular value at a fixed point. The characterization of the estimators involved is provided both in the nondecreasing and the nonincreasing setting. These characterizations facilitate the derivation of the asymptotic distribution of the likelihood ratio test, which is identical in the nondecreasing and in the nonincreasing case. The asymptotic distribution of the likelihood ratio test enables, via inversion, the construction of pointwise confidence intervals. Simulations show that these confidence intervals exhibit comparable coverage probabilities with the confidence intervals based on the asymptotic distribution of the nonparametric maximum likelihood estimator of a monotone baseline hazard function.

The fourth paper, in Chapter 5, focuses on estimating the baseline hazard function λ_0, under the assumption that λ_0 is nondecreasing and smooth. The estimators are obtained by kernel smoothing the maximum likelihood and Grenander-type estimator of a nondecreasing baseline hazard function. Three different estimators are proposed for a nondecreasing baseline hazard, depending on when the smoothing step takes place. With this respect, a smoothed maximum likelihood estimator (SMLE) is proposed, as well as a smoothed Grenander-type (SG) and a Grenander-type smoothed (GS) estimator. The pointwise and uniform strong consistency of the three smooth estimators is investigated.

REFERENCES

ANDERSEN, P. K., BORGAN, O., GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes.* Springer. New York.

BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Annals of Statistics*, **35:** 931–956.

BANERJEE, M. (2008). Estimating monotone, unimodal and U-shaped failure rates using asymptotic pivots. *Statistica Sinica*, **18:** 467–492.

BANERJEE, M. & WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Annals of Statistics*, **29:** 1699–1731.

BANERJEE, M. & WELLNER, J. A. (2005). Score statistics for current status data: comparisons with likelihood ratio and Wald statistics. *International Journal of Biostatistics*, **1** Art. 3, 29 pp. (electronic).

BARLOW, R. E., MARSHALL, A. W. & PROSCHAN, F. (1963). Properties of probability distributions with monotone hazard rate. *Annals of Mathematical Statistics*, **34:** 375–389.

BURKE, M. D., CSÖRGŐ, S. & HORVÁTH, L. (1988). A correction to and improvement of: "Strong approximations of some biometric estimates under random censorship" [Z. Wahrsch. Verw. Gebiete **56** (1981), no. 1, 87–112]. *Probability Theory and Related Fields*, **79:** 51–57.

BURR, D. (1994). On Inconsistency of Breslow's estimator as an estimator of the hazard rate in the Cox model. *Biometrics*, **50:** 1142–1145.

COX, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, **34:**, 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika*, **62:** 269–276.

GRENANDER, U. (1956). On the theory of mortality measurement, II. *Skandinavisk Aktuarietidskrift*, **39:** 125–153.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer II*, 539–555.

GROENEBOOM, P. & JONGBLOED, G. (2013). Smooth and non-smooth estimates of a monotone hazard. *IMS Collection. From Probability to Statistics and Back: High-Dimensional Models and Processes,* **9:** 174–196.

GROENEBOOM, P. & WELLNER, J. A. (2001). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics*, **10:** 388–400.

HUANG, J. & WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Statistics,* **22:** 3–33.

KIM, J. & POLLARD, D. (1990). Cube root asymptotics. *Annals of Statistics,* **18:** 191–219.

KLEIN, J. P. & MOESCHBERGER, M. L. (1997). *Survival Analysis Techniques for Censored and Truncated Data*. Springer. New York.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric inference*. Springer. New York.

KVAM, P. H., SINGH, H. & WHITAKER, L. R. (2002). Estimating distributions with increasing failure rate in an imperfect repair model. *Lifetime Data Analysis,* **8:** 53–67.

LIN, D. Y. (2007). On the Breslow estimator. *Lifetime Data Analysis,* **13:** 471–480.

MARSHALL, A. W. & PROSCHAN, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Annals of Mathematical Statistics*, **36:** 69–77.

MYKYTYN, S. W. & SANTNER, T. J. (1981). Maximum likelihood estimation of the survival function based on censored data under hazard rate assumptions. *Communications in Statistics*, **10:** 1369–1387.

OAKES, D. (1977). The asymptotic information in censored survival data. *Biometrika*, **64:** 441–448.

PADGETT, W. J. & WEI, L. J. (1980). Maximum likelihood estimation of a distribution function with increasing failure rate based on censored observations. *Biometrika*, **67:** 470–474.

PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Annals of Mathematical Statistics*, **41:** 507–519.

ROBERTSON, T., WRIGHT, F. T. & DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. John Wiley & Sons, New York.

SLUD, E. V. (1982). Consistency and efficiency of inferences with the partial likelihood. *Biometrika*, **69:** 547–552.

TSIATIS, A. (1981). A large sample study of Cox's regression model. *Annals of Statistics*, **9:** 93–108.

Copies of the four papers

CHAPTER **2**

SHAPE CONSTRAINED NONPARAMETRIC ESTIMATORS OF THE BASELINE DISTRIBUTION IN COX PROPORTIONAL HAZARDS MODEL [1]

We investigate nonparametric estimation of a monotone baseline hazard and a nonincreasing baseline density within the Cox model. Two estimators of a nondecreasing baseline hazard function are proposed. We derive the nonparametric maximum likelihood estimator and consider a Grenander-type estimator, defined as the left-hand slope of the greatest convex minorant of the Breslow estimator. We demonstrate that the two estimators are strongly consistent and asymptotically equivalent and derive their common limit distribution at a fixed point. Both estimators of a nonincreasing baseline hazard and their asymptotic properties are acquired in a similar manner. Furthermore, we introduce a Grenander-type estimator of a nonincreasing baseline density, defined as the left-hand slope of the least concave majorant of an estimator of the baseline cumulative distribution function, derived from the Breslow estimator. We show that this estimator is strongly consistent and derive its asymptotic distribution at a fixed point.

2.1 INTRODUCTION

Shape constrained nonparametric estimation dates back to the 1950s. The milestone paper of GRENANDER (1956) introduced the maximum likelihood estimator of a nonincreasing density, while PRAKASA RAO (1969) derived its asymptotic distribution at a fixed point. Similarly, the maximum likelihood estimator of a monotone hazard function has been proposed by MARSHALL & PROSCHAN (1965) and its asymptotic distribution was determined by PRAKASA RAO (1970). Other estimators have been proposed and despite the high interest and applicability, the difficulty in the derivation of the asymptotics was a major drawback. Shape constrained estimation was revived by GROENEBOOM (1985), who proposed an alternative for Prakasa

[1] By Lopuhaä, H. P. and Nane, G. F. (2013). *Scandinavian Journal of Statistics*, doi: 10.1002/sjos.12008.

Rao's bothersome type of proof. Groeneboom's approach employs a so-called inverse process and makes use of the Hungarian embedding (KOMLÓS *et al*, 1975). Once such an embedding is available, it enables the derivation of the asymptotic distribution of the considered estimator. This is the case, for example, when estimating a monotone density or hazard function from right-censored observations, as proposed by HUANG & ZHANG (1994) and HUANG & WELLNER (1995). Another setting for deriving the asymptotic distribution, that does not require a Hungarian embedding, was later provided by the limit theorems KIM & POLLARD (1990). Their cube root asymptotics are based on a functional limit theorem for empirical processes.

The present chapter treats the estimation of a monotone baseline hazard and a nonincreasing baseline density in the Cox model. Ever since the model was introduced (see COX, 1972) and in particular, since the asymptotic properties of the proposed estimators were first derived by TSIATIS (1981), the Cox model is the classical survival analysis framework for incorporating covariates in the study of a lifetime distribution. The hazard function is of particular interest in survival analysis, as it represents an important feature of the time course of a process under study, e.g., death or a certain disease. The main reason lies in its ease of interpretation and in the fact that the hazard function takes into account ageing, while, for example, the density function does not. Times to death, infection or development of a disease of interest in most survival analysis studies are observed to have a nondecreasing baseline hazard. Nevertheless, the survival time after a successful medical treatment is usually modeled using a nonincreasing hazard. An example of nonincreasing hazard is presented in COOK *et al.* (1998), where the authors concluded that the daily risk of pneumonia decreases with increasing duration of stay in the intensive care unit.

CHUNG & CHANG (1994) consider a maximum likelihood estimator of a nondecreasing baseline hazard function in the Cox model, adopting the convention that each censoring time is equal to its preceding observed event time. They prove consistency, but no distributional theory is available. We consider a maximum likelihood estimator $\hat{\lambda}_n$ of a monotone baseline hazard function, which imposes no extra assumption on the censoring times. This estimator differs from the one of CHUNG & CHANG (1994) and has a higher likelihood. Furthermore, we introduce a Grenander-type estimator of a monotone baseline hazard function based on the well-known baseline cumulative hazard estimator, the Breslow estimator Λ_n (COX, 1972). The nondecreasing baseline hazard estimator $\tilde{\lambda}_n$ is defined as the left-hand slope of the greatest convex minorant (GCM) of Λ_n. Similarly, a nonincreasing baseline estimator is characterized as the left-hand slope of the least concave majorant (LCM) of Λ_n. It is noteworthy that, just as in the no covariates case (see HUANG & WELLNER, 1995), the two monotone estimators are different, but are shown to be asymptotically equivalent. Additionally, we introduce a nonparametric estimator of a nonincreasing baseline density. An estimator F_n of the baseline distribution function is based on the Breslow estimator and next, the baseline density estimator $\tilde{f}_n$ is defined as the left-hand slope of the LCM of F_n. The treatment of the maximum likelihood estimator of a nonincreasing baseline density is much more complex and is deferred to another paper. For the remaining three estimators, we show that they converge at rate $n^{1/3}$ and we establish their limit distribution. Since, to the authors best knowledge, there does not exist a Hungarian embedding for the Breslow estimator, our results are based on the theory of KIM & POLLARD (1990) and an argmax continuous mapping theorem of HUANG & WELLNER (1995).

Chapter 2 is organized as follows. In Section 2.2, we introduce the model and state our assumptions. The formal characterization of the maximum likelihood estimator $\hat{\lambda}_n$ is given in Lemmas 2.1 and 2.2. Our main results concerning the asymptotic properties of the proposed estimators are gathered in Section 2.3. Section 2.4 is devoted to proving the strong consistency results of the paper. The strong uniform consistency of the Breslow estimator in Theorem 2.9 and of the baseline cumulative distribution estimator F_n in Corollary 2.10, emerge as necessary results. These results are preceded by three preparatory lemmas, that establish properties of functionals in terms of which derivations thereof can be expressed. In order to prepare the application of results from KIM & POLLARD (1990), in Section 2.5 we introduce the inverses of the estimators in terms of minima and maxima of random processes and obtain the limiting distribution of these processes. Finally, in Section 2.6, we derive the asymptotic distribution of the estimators, at a fixed point.

2.2 DEFINITIONS AND ASSUMPTIONS

Let the observed data consist of independent identically distributed triplets (T_i, Δ_i, Z_i), with $i = 1, 2, \ldots, n$, where T_i denotes the follow-up time, with a corresponding censoring indicator Δ_i and covariate vector $Z_i \in \mathbb{R}^p$. A generic follow-up time is defined by $T = \min(X, C)$, where X represents the event time and C is the censoring time. Accordingly, $\Delta = \{X \leqslant C\}$, where $\{\cdot\}$ denotes the indicator function. The event time X and censoring time C are assumed to be conditionally independent given Z, and the censoring mechanism is assumed to be non-informative. The covariate vector $Z \in \mathbb{R}^p$ is assumed to be time invariant.

Within the Cox model, the distribution of the event time is related to the corresponding covariate by

$$\lambda\left(x|z\right) = \lambda_0(x)\, e^{\beta_0' z}, \tag{2.2.1}$$

where $\lambda\left(x|z\right)$ is the hazard function for an individual with covariate vector $z \in \mathbb{R}^p$, λ_0 represents the baseline hazard function and $\beta_0 \in \mathbb{R}^p$ is the vector of the underlying regression coefficients. Conditionally on $Z = z$, the event time X is assumed to be a nonnegative random variable with an absolutely continuous distribution function $F(x|z)$ with density $f(x|z)$. The same assumptions hold for the censoring variable C and its distribution function G. The distribution function of the follow-up time T is denoted by H. We will assume the following conditions, which are commonly employed in deriving large sample properties of Cox proportional hazards estimators (e.g., see

TSIATIS, 1981).

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

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

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

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

where $|\cdot|$ denotes the Euclidean norm.

2.2.1 Nondecreasing baseline hazard

Let $\Lambda(x|z) = -\log[1 - F(x|z)]$ be the cumulative hazard function. Then, from (2.2.1) it follows that $\Lambda(x|z) = \Lambda_0(x)\exp(\beta_0' z)$, where $\Lambda_0(x) = \int_0^x \lambda_0(u)\,\mathrm{d}u$ denotes the baseline cumulative hazard function. When G has a density g, then together with the relation $\lambda = f/(1-F)$, the likelihood becomes

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

The term with g does not involve the baseline distribution and can be treated as a constant term. Therefore, one essentially needs to maximize

$$\prod_{i=1}^n \lambda(T_i \mid Z_i)^{\Delta_i} \exp\Big[-\Lambda(T_i \mid Z_i)\Big] = \prod_{i=1}^n \Big[\lambda_0(T_i)\mathrm{e}^{\beta_0' Z_i}\Big]^{\Delta_i} \exp\Big[-\mathrm{e}^{\beta_0' Z_i}\Lambda_0(T_i)\Big].$$

This leads to the following (pseudo) loglikelihood, written as a function of $\beta \in \mathbb{R}^p$ and λ_0,

$$\sum_{i=1}^n \Big[\Delta_i \log \lambda_0(T_i) + \Delta_i \beta' Z_i - \mathrm{e}^{\beta' Z_i}\Lambda_0(T_i)\Big]. \qquad (2.2.2)$$

Remark. Note that if the censoring distribution is discrete, the likelihood of (T, Δ, Z) can still be written as

$$[f(T \mid Z)(1 - G(T \mid Z))]^{\Delta}[g(T \mid Z)(1 - F(T \mid Z))]^{1-\Delta},$$

where $g(y|z) = P(C = y|Z = z)$, which will lead to the same expression as in (2.2.2). However, as we will make use of other results in the literature that are established under the assumption of an absolutely continuous censoring distribution (e.g., from

TSIATIS, 1981), we do not further investigate the behavior of our estimators in the case of a discrete censoring distribution.

For $\beta \in \mathbb{R}^p$ fixed, we first consider maximum likelihood estimation of a nondecreasing λ_0. This requires the maximization of (2.2.2) over all nondecreasing λ_0. Let $T_{(1)} < T_{(2)} < \cdots < T_{(n)}$ be the ordered follow-up times and, for $i = 1, 2, \ldots, n$, let $\Delta_{(i)}$ and $Z_{(i)}$ be the censoring indicator and covariate vector corresponding to $T_{(i)}$. The characterization of the maximizer $\hat{\lambda}_n(x; \beta)$ can be described by means of the processes

$$W_n(\beta, x) = \int \left(e^{\beta' z} \int_0^x \{u \geqslant s\} \, \mathrm{d}s \right) \mathrm{d}P_n(u, \delta, z), \tag{2.2.3}$$

and

$$V_n(x) = \int \delta \{u < x\} \, \mathrm{d}P_n(u, \delta, z), \tag{2.2.4}$$

with $\beta \in \mathbb{R}^p$ and $x \geqslant 0$, where P_n is the empirical measure of the (T_i, Δ_i, Z_i), and is given by the following lemma.

LEMMA 2.1. *For a fixed $\beta \in \mathbb{R}^p$, let W_n and V_n be defined in* (2.2.3) *and* (2.2.4). *Then, the NPMLE $\hat{\lambda}_n(x; \beta)$ of a nondecreasing baseline hazard function λ_0 is of the form*

$$\hat{\lambda}_n(x; \beta) = \begin{cases} 0 & x < T_{(1)}, \\ \hat{\lambda}_i & T_{(i)} \leqslant x < T_{(i+1)}, \text{ for } i = 1, 2, \ldots, n-1, \\ \infty & x \geqslant T_{(n)}, \end{cases}$$

where $\hat{\lambda}_i$ is the left derivative of the greatest convex minorant at the point P_i of the cumulative sum diagram consisting of the points

$$P_j = \Big(W_n(\beta, T_{(j+1)}) - W_n(\beta, T_{(1)}), V_n(T_{(j+1)}) \Big),$$

for $j = 1, 2, \ldots, n-1$ and $P_0 = (0, 0)$. Furthermore,

$$\hat{\lambda}_i = \max_{1 \leqslant s \leqslant i} \min_{i \leqslant t \leqslant n-1} \frac{\sum_{j=s}^t \Delta_{(j)}}{\sum_{j=s}^t \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^n e^{\beta' Z_{(l)}}}, \tag{2.2.5}$$

for $i = 1, 2, \ldots, n-1$.

PROOF. Similar to MARSHALL & PROSCHAN (1965) and Section 7.4 in ROBERTSON *et al.* (1988), since $\lambda_0(T_{(n)})$ can be chosen arbitrarily large, we first consider the maximization over nondecreasing λ_0 bounded by some $M > 0$. When we increase the value of λ_0 on an interval $(T_{(i-1)}, T_{(i)})$, the terms $\lambda_0(T_{(i)})$ in (2.2.2) are not changed, whereas terms with $\Lambda_0(T_{(i)})$ will decrease the loglikelihood. Since λ_0 must be nondecreasing, we conclude that the solution is a nondecreasing step function, that is zero for $x < T_{(1)}$, constant on $[T_{(i)}, T_{(i+1)})$, for $i = 1, 2, \dots, n-1$, and equal to M, for $x \geqslant T_{(n)}$. Consequently, for $\beta \in \mathbb{R}^p$ fixed, the (pseudo) loglikelihood reduces to

$$\begin{aligned} L_\beta(\lambda_0) &= \sum_{i=1}^{n-1} \Delta_{(i)} \log \lambda_0(T_{(i)}) - \sum_{i=2}^{n} \mathrm{e}^{\beta' Z_{(i)}} \sum_{j=1}^{i-1} \left[T_{(j+1)} - T_{(j)} \right] \lambda_0(T_{(j)}) \\ &= \sum_{i=1}^{n-1} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}. \end{aligned} \tag{2.2.6}$$

Maximization over $0 \leqslant \lambda_0(T_{(1)}) \leqslant \cdots \leqslant \lambda_0(T_{(n-1)}) \leqslant M$ will then have a solution $\hat{\lambda}_n^M(x;\beta)$ and by letting $M \to \infty$, we obtain the NPMLE $\hat{\lambda}_n(x;\beta)$ for λ_0.

First, notice that the loglikelihood function in (2.2.6) can also be written as

$$\sum_{i=1}^{n-1} \left[s_i \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \right] w_i, \tag{2.2.7}$$

where, for $i = 1, 2, \dots, n-1$,

$$w_i = \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}},$$

and

$$s_i = \frac{\Delta_{(i)}}{\left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}}}.$$

As mentioned above, we first maximize over nondecreasing λ_0 bounded by some M. Since M can be chosen arbitrarily large, the problem of maximizing (2.2.7) over $0 \leqslant \lambda_0(T_{(1)}) \leqslant \cdots \leqslant \lambda_0(T_{(n-1)}) \leqslant M$ can be identified with the problem solved in Example 1.5.7 in ROBERTSON *et al.* (1988). The existence of $\hat{\lambda}_n^M$ is therefore immediate and is given by

$$\hat{\lambda}_n^M(x;\beta) = \begin{cases} 0 & x < T_{(1)}, \\ \hat{\lambda}_i & T_{(i)} \leqslant x < T_{(i+1)}, \text{ for } i = 1, 2, \dots, n-1, \\ M & x \geqslant T_{(n)}, \end{cases}$$

where, as a result of Theorems 1.5.1 and 1.2.1 in ROBERTSON *et al.* (1988), the value $\hat{\lambda}_i$ is the left derivative at P_i of the GCM of the cumulative sum diagram (CSD) consisting of the points

$$P_i = \left(\frac{1}{n} \sum_{j=1}^{i} w_j, \frac{1}{n} \sum_{j=1}^{i} w_j s_j \right), \qquad i = 1, 2, \dots, n-1,$$

and $P_0 = (0,0)$. It follows that

$$\frac{1}{n}\sum_{j=1}^{i} w_j = \sum_{j=1}^{i}\left[T_{(j+1)} - T_{(j)}\right]\frac{1}{n}\sum_{l=1}^{n}\{T_l \geqslant T_{(j+1)}\}e^{\beta' Z_l}$$
$$= \int_{T_{(1)}}^{T_{(i+1)}}\int\{u \geqslant s\}e^{\beta' z}\,dP_n(u,\delta,z)\,ds = W_n(\beta, T_{(i+1)}) - W_n(\beta, T_{(1)}).$$

For the y-coordinate of the CSD, notice that

$$\frac{1}{n}\sum_{j=1}^{i} w_j s_j = \frac{1}{n}\sum_{j=1}^{i}\Delta_{(j)} = \frac{1}{n}\sum_{j=1}^{n}\{T_j \leqslant T_{(i)}, \Delta_j = 1\} = V_n(T_{(i+1)}).$$

By letting $M \to \infty$, we obtain the NPMLE $\hat{\lambda}_n(\beta, x)$ for λ_0. The max-min formula in (2.2.5) follows from Theorem 1.4.4 in ROBERTSON *et al.* (1988). ▫

REMARK. From the characterization given in Lemma 2.1, it can be seen that the GCM of the CSD only changes slope at points corresponding to uncensored observations, which means that $\hat{\lambda}_n(x;\beta)$ is constant between successive uncensored follow-up times. Moreover, similar to the reasoning in the proof of Lemma 2.1, it follows that $\hat{\lambda}_n(x;\beta)$ maximizes (2.2.2). The reason to provide the characterization in Lemma 2.1 in terms of all follow-up times is that this facilitates the treatment of the asymptotics for this estimator. Finally, for the solution $\hat{\lambda}_n^M(x;\beta)$, on the interval $[T_{(n)}, \tau_H)$, in principle one could take any value between $\hat{\lambda}_{n-1}$ and M. This means that for $\hat{\lambda}_n(x;\beta)$, on the interval $[T_{(n)}, \tau_H)$, one could take any value larger than $\hat{\lambda}_{n-1}$.

In practice, one also has to estimate β_0. The standard choice is $\hat{\beta}_n$, the maximizer of the partial likelihood function

$$\prod_{l=1}^{m}\frac{e^{\beta' Z_j}}{\sum_{j=1}^{n}\{T_j \geqslant X_{(i)}\}e^{\beta' Z_j}},$$

as proposed by COX (1972, 1975), where $X_{(1)} < X_{(2)} < \cdots < X_{(m)}$ denote the ordered, observed event times. Since the maximum partial likelihood estimator $\hat{\beta}_n$ for β_0 is asymptotically efficient under mild conditions and because the amount of information on β_0 lost through lack of knowledge of λ_0 is usually small (e.g., see EFRON, 1977; OAKES, 1977; SLUD, 1982), we do not pursue joint maximization of (2.2.2) over nondecreasing λ_0 and β_0. We simply replace β in $\hat{\lambda}_n(x;\beta)$ by $\hat{\beta}_n$, and we propose $\hat{\lambda}_n(x) = \hat{\lambda}_n(x;\hat{\beta}_n)$ as our estimator of λ_0.

Note that $\hat{\lambda}_n$ is different from the estimator derived by CHUNG & CHANG (1994), where each censoring time is taken equal to the preceding observed event time. This leads to a CSD that is slightly different from the one in Lemma 2.1. However, it can be shown that both estimators have the same asymptotic behavior. Furthermore, if we take all regression coefficients equal to zero, the model coincides with the ordinary random censorship model with a nondecreasing hazard function as considered in HUANG & WELLNER (1995). The characterization in Lemma 2.1, with $\beta_0 = 0$, differs slightly from the one in Theorem 3.2 in HUANG & WELLNER (1995). Their estimator seems to be the result of maximization of the loglikelihood corresponding to (2.2.2) over left-continuous λ_0 that are constant between follow-up times. Although this estimator does not maximize the loglikelihood corresponding to (2.2.2) over all nondecreasing λ_0, the asymptotic distribution will turn out to be the same as that of $\hat{\lambda}_n$, for the special case of no covariates. The computation of joint maximum likelihood estimates for β and λ_0 is considered by HUI & JANKOWSKI (2012), who also developed an R package to compute the estimates.

To illustrate the computation of the estimator described in Lemma 2.1, consider an artificial survival dataset consisting of 10 follow-up times, with only $T_{(2)}, T_{(5)}, T_{(6)}$, and $T_{(8)}$ being observed event times. In Figure 2.1 we illustrate the construction of the proposed estimator and compare the resulting estimate with the one suggested by CHUNG & CHANG (1994). In order to compare the CSD of both estimates, the coordinates of the CSD described in Lemma 2.1 have been multiplied with a factor n, which obviously leads to the same slopes. Figure 2.1 displays the points of the CSD (bullet points) and the GCM (solid curve) in the left panel. The horizontal segments are generated by $(nW_n(\hat{\beta}_n, x) - nW_n(\hat{\beta}_n, T_{(1)}), nV_n(x))$ for $x \geqslant T_{(1)}$. Note that the process nV_n has a jump of size 1 right after a point P_j that corresponds to an observed event time. Taking left derivatives then yield jumps of $\hat{\lambda}_n$ only at observed event times. The right panel of Figure 2.1 displays the corresponding graph of $\hat{\lambda}_n$ (solid curve). The jumps of $\hat{\lambda}_n$ in the right panel correspond to the changes of slope of the GCM at the points P_1, P_4 and P_7 in the left panel and occur at the event times $T_{(2)}, T_{(5)}$, and $T_{(8)}$. The height of the horizontal segments in the right panel corresponds to the slopes of the GCM in the left panel. For comparison we have added the CSD (star points) and the corresponding GCM (dashed curve) of the estimator derived by CHUNG & CHANG (1994) in the left panel and the resulting estimator in the right panel (dashed curve). Note that shifting the censoring times back to the nearest previous event time, as suggested in CHUNG & CHANG (1994), pushes points in the CSD, that correspond to event times, to the left. As a consequence this yields steeper slopes in the left panel and hence a larger estimate of the hazard in the right panel.

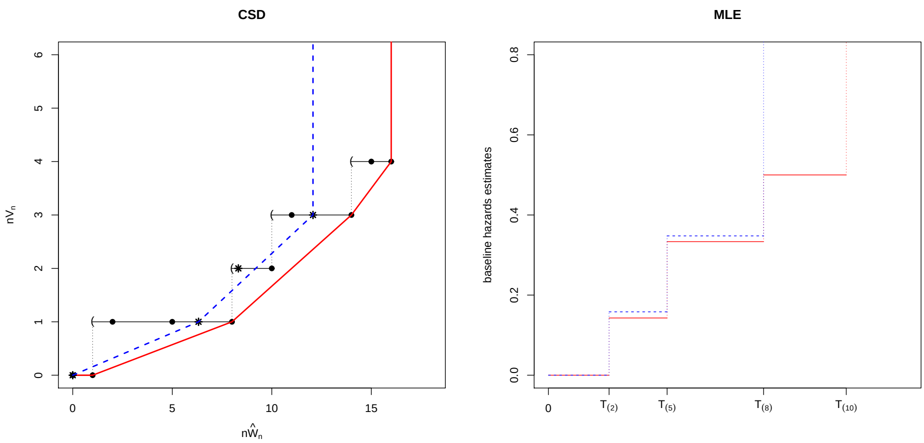

FIGURE 2.1: The cumulative sum diagrams along with their GCM (left panel) and the corresponding estimates of a nondecreasing baseline hazard (right panel). Bullet points and solid curve correspond to the estimator in Lemma 2.1; star points and dashed curve correspond to the estimator in CHUNG & CHANG (1994).

Another possibility to estimate a nondecreasing hazard is to construct a Grenander-type estimator, i.e., consider an unconstrained estimator Λ_n of the cumulative hazard Λ_0 and take the left slope of the GCM as an estimator of λ_0. Several isotonic estimators are of this form (e.g., see GRENANDER, 1956; BRUNK, 1958; HUANG & WELLNER, 1995; DUROT, 2007). Breslow (COX, 1972) proposed

$$\Lambda_n(x) = \sum_{i|X_{(i)} \leqslant x} \frac{d_i}{\sum_{j=1}^{n}\{T_j \geqslant X_{(i)}\}e^{\hat{\beta}_n' Z_j}}, \tag{2.2.8}$$

as an estimator of Λ_0, where d_i is the number of events at $X_{(i)}$ and $\hat{\beta}_n$ is the maximum partial likelihood estimator of the regression coefficients. The estimator Λ_n is most commonly referred to as the Breslow estimator. In the case of no covariates, i.e., $\beta = 0$, the NPMLE estimate of a nondecreasing hazard rate has been illustrated in HUANG & WELLNER (1995).

Following the derivations in TSIATIS (1981), it can be inferred that

$$\lambda_0(x) = \frac{\mathrm{d}H^{uc}(x)/\mathrm{d}x}{\mathbb{E}\left[\{T \geqslant x\}\exp(\beta_0' Z)\right]}, \tag{2.2.9}$$

where $H^{uc}(x) = \mathbb{P}(T \leqslant x, \Delta = 1)$ is the sub-distribution function of the uncensored observations. Consequently, it can be derived that

$$\Lambda_0(x) = \int \frac{\delta\{u \leqslant x\}}{\mathbb{E}\left[\{T \geqslant x\}\exp(\beta_0' Z)\right]} \, \mathrm{d}P(u, \delta, z), \tag{2.2.10}$$

where P is the underlying probability measure corresponding to the distribution of (T, Δ, Z). From (A1), it follows that $\Lambda_0(\tau_H) < \infty$. In view of the above expression, an intuitive baseline cumulative hazard estimator is obtained by replacing the expectations in (2.2.10) by averages and by plugging in $\hat{\beta}_n$, which yields exactly the Breslow estimator in (2.2.8). As a Grenander-type estimator of a nondecreasing hazard, we propose the left-hand derivative $\tilde{\lambda}_n$ of the greatest convex minorant $\widetilde{\Lambda}_n$ of Λ_n. This estimator is different from $\hat{\lambda}_n$ for finite samples, but we will show that both estimators are asymptotically equivalent. For the special case of no covariates, this coincides with the results in HUANG & WELLNER (1995).

2.2.2 NONINCREASING BASELINE HAZARD

A completely similar characterization is provided for the NPMLE of a nonincreasing baseline hazard function. As in the nondecreasing case, one can argue that the loglikelihood is maximized by a decreasing step function that is constant on $(T_{(i-1)}, T_{(i)}]$, for $i = 1, 2, \ldots, n$, where $T_{(0)} = 0$. In this case, the loglikelihood reduces to

$$L_\beta(\lambda_0) = \sum_{i=1}^{n} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i)} - T_{(i-1)} \right] \sum_{l=i}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\},$$

which is maximized over all $\lambda_0(T_{(1)}) \geqslant \cdots \geqslant \lambda_0(T_{(n)}) \geqslant 0$. The solution is characterized by the following lemma. The proof of this lemma is completely similar to that of Lemma 2.1.

LEMMA 2.2. *For a fixed $\beta \in \mathbb{R}^p$, let W_n be defined in* (2.2.3) *and let*

$$Y_n(x) = \int \delta\{u \leqslant x\} \, \mathrm{d}\mathbb{P}_n(u, \delta, z). \tag{2.2.11}$$

Then the NPMLE $\hat{\lambda}_n(x; \beta)$ of a nonincreasing baseline hazard function λ_0 is given by

$$\hat{\lambda}_n(x; \beta) = \hat{\lambda}_i \qquad \text{for } x \in (T_{(i-1)}, T_{(i)}],$$

for $i = 1, 2, \ldots, n$, where $\hat{\lambda}_i$ is the left derivative of the least concave majorant (LCM) at the point P_i of the cumulative sum diagram consisting of the points

$$P_j = \Big(W_n(\beta, T_{(j)}), Y_n(T_{(j)}) \Big),$$

for $j = 1, 2, \ldots, n$ and $P_0 = (0, 0)$. Furthermore,

$$\hat{\lambda}_i = \max_{1 \leqslant s \leqslant i} \min_{i \leqslant t \leqslant n} \frac{\sum_{j=s}^{t} \Delta_{(j)}}{\sum_{j=s}^{t} \left[T_{(j)} - T_{(j-1)} \right] \sum_{l=j}^{n} e^{\beta' Z_{(l)}}},$$

for $i = 1, 2, \ldots, n$.

Analogous to the nondecreasing case, for $x \geqslant T_{(n)}$, one can choose for $\hat{\lambda}_n(x;\beta)$ any value smaller than $\hat{\lambda}_n$. As before, we propose $\hat{\lambda}_n(x) = \hat{\lambda}_n(x;\hat{\beta}_n)$ as an estimator of λ_0, where $\hat{\beta}_n$ denotes the maximum partial likelihood estimator of β_0. Similar to the nondecreasing case, the Grenander-type estimator $\tilde{\lambda}_n$ of a nonincreasing λ_0 is defined as the left-hand slope of the LCM of the Breslow estimator Λ_n, defined in (2.2.8).

An illustration of the NPMLE of a nonincreasing baseline hazard function can be found in van GELOVEN *et al.* (2012), who investigated the hazard of patients with acute coronary syndrome. Previous clinical trials indicated a decreasing risk pattern, which the authors confirmed by a test based on a bootstrap procedure. The above estimate has been computed for 1200 patients undergoing early or selective invasive strategies, that were monitored for five years, and their performance was evaluated by means of a simulation experiment. The R code is available in the online version of their paper.

2.2.3 NONINCREASING BASELINE DENSITY

Suppose one is interested in estimating a nonincreasing baseline density $f_0(\cdot) = f(\cdot|z=0)$. One might argue that this problem is of less interest, because the monotonicity assumption assumed for $z = 0$ may no longer hold if one transforms the covariates by $a + bz$, whereas the Cox model essentially remains unchanged. Whereas the estimator of the baseline hazard remains monotone under such transformations, this may no longer hold for the estimator of the baseline density. Despite this drawback, we feel that the estimation of a nonincreasing baseline density may be of interest.

In this case, the corresponding baseline distribution function F_0 is concave and it relates to the baseline cumulative hazard function Λ_0 as follows

$$F_0(x) = 1 - \mathrm{e}^{-\Lambda_0(x)}. \tag{2.2.12}$$

Hence, a natural estimator of the baseline distribution function is

$$F_n(x) = 1 - \mathrm{e}^{-\Lambda_n(x)}, \tag{2.2.13}$$

where Λ_n is the Breslow estimator, defined in (2.2.8). A Grenander-type estimator $\tilde{f}_n$ of a nonincreasing baseline density is defined as the left-hand slope of the LCM of F_n. Recall that Λ_n depends on $\hat{\beta}_n$ and $Z_1, Z_2, \ldots, Z_n$, and therefore the same holds for F_n and $\tilde{f}_n$.

The derivation of the NPMLE for f_0 is much more complex than the previous estimators and its treatment is postponed to a future manuscript. In the special case of no covariates, the NPMLE $\hat{f}_n$ has first been derived in HUANG & ZHANG (1994). In HUANG & WELLNER (1995) a different characterization has been provided for $\hat{f}_n$ in terms of a self-induced cusum diagram and it was shown that $\hat{f}_n$ and $\tilde{f}_n$, the

Grenander-type estimator defined as the left-hand slope of the least concave majorant of the Kaplan-Meier estimator are asymptotically equivalent.

2.3 Main results

In this section, we state our main results. The proofs are postponed to subsequent sections. The next theorem provides pointwise consistency of the proposed estimators at a fixed point x_0 in the interior of the support. Note that the results below imply that if x_0 is a point of continuity of λ_0, then $\hat{\lambda}_n(x_0) \to \lambda_0(x_0)$ with probability one, and likewise for the other estimators.

Theorem 2.3. *Assume that (A1) and (A2) hold.*

(i) Suppose that λ_0 is nondecreasing on $[0,\infty)$ and let $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ be the estimators defined in Section 2.2.1. Then, for any $x_0 \in (0,\tau_H)$,

$$\lambda_0(x_0-) \leqslant \liminf_{n\to\infty} \hat{\lambda}_n(x_0) \leqslant \limsup_{n\to\infty} \hat{\lambda}_n(x_0) \leqslant \lambda_0(x_0+),$$
$$\lambda_0(x_0-) \leqslant \liminf_{n\to\infty} \tilde{\lambda}_n(x_0) \leqslant \limsup_{n\to\infty} \tilde{\lambda}_n(x_0) \leqslant \lambda_0(x_0+),$$

with probability one, where the values $\lambda_0(x_0-)$ and $\lambda_0(x_0+)$ denote the left and right limit at x_0.

(ii) Suppose that λ_0 is nonincreasing on $[0,\infty)$ and let $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ be the estimators defined in Section 2.2.2. Then, for any $x_0 \in (0,\tau_H)$,

$$\lambda_0(x_0+) \leqslant \liminf_{n\to\infty} \hat{\lambda}_n(x_0) \leqslant \limsup_{n\to\infty} \hat{\lambda}_n(x_0) \leqslant \lambda_0(x_0-),$$
$$\lambda_0(x_0+) \leqslant \liminf_{n\to\infty} \tilde{\lambda}_n(x_0) \leqslant \limsup_{n\to\infty} \tilde{\lambda}_n(x_0) \leqslant \lambda_0(x_0-),$$

with probability one.

(iii) Suppose that f_0 is nonincreasing on $[0,\infty)$ and let $\tilde{f}_n$ be the estimator defined in Section 2.2.3. Then, for any $x_0 \in (0,\tau_H)$,

$$f_0(x_0+) \leqslant \liminf_{n\to\infty} \tilde{f}_n(x_0) \leqslant \limsup_{n\to\infty} \tilde{f}_n(x_0) \leqslant f_0(x_0-),$$

with probability one, where $f_0(x_0-)$ and $f_0(x_0+)$ denote the left and right limit at x_0.

The following two theorems yield the asymptotic distribution of the monotone constrained baseline hazard estimators. In order to keep notations compact, it becomes useful to introduce

$$\Phi(\beta,x) = \int \{u \geqslant x\}\, \mathrm{e}^{\beta' z}\, \mathrm{d}P(u,\delta,z), \tag{2.3.1}$$

for $\beta \in \mathbb{R}^p$ and $x \in \mathbb{R}$, where P is the underlying probability measure corresponding to the distribution of (T, Δ, Z). Furthermore, by the argmin function we mean the supremum of times at which the minimum is attained. Note that the limiting distribution and the rate of convergence coincide with the results commonly obtained for isotonic estimators and differ from the corresponding quantities in the traditional central limit theorem. The limiting distribution, usually referred to as the Chernoff distribution, has been tabulated in GROENEBOOM & WELLNER (2001).

THEOREM 2.4. *Assume (A1) and (A2) and let $x_0 \in (0, \tau_H)$. Suppose that λ_0 is nondecreasing on $[0,\infty)$ and continuously differentiable in a neighborhood of x_0, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) > 0$. Moreover, suppose that $H^{uc}(x)$ and $x \mapsto \Phi(\beta_0, x)$ are continuously differentiable in a neighborhood of x_0, where H^{uc} is defined below* (2.2.9) *and Φ is defined in* (2.3.1). *Let $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ be the estimators defined in Section 2.2.1. Then,*

$$n^{1/3}\left(\frac{\Phi(\beta_0, x_0)}{4\lambda_0(x_0)\lambda_0'(x_0)}\right)^{1/3}\left[\hat{\lambda}_n(x_0) - \lambda_0(x_0)\right] \xrightarrow{d} \underset{t\in\mathbb{R}}{\operatorname{argmin}}\{\mathbb{W}(t) + t^2\}, \tag{2.3.2}$$

where $\mathbb{W}$ is standard two-sided Brownian motion originating from zero. Furthermore,

$$n^{1/3}\left[\tilde{\lambda}_n(x_0) - \hat{\lambda}_n(x_0)\right] \xrightarrow{p} 0, \tag{2.3.3}$$

so that the convergence in (2.3.2) *also holds with $\hat{\lambda}_n$ replaced by $\tilde{\lambda}_n$.*

The next theorem establishes the same results as in Theorem 2.4, for the nonincreasing case.

THEOREM 2.5. *Assume (A1) and (A2) and let $x_0 \in (0, \tau_H)$. Suppose that λ_0 is nonincreasing on $[0,\infty)$ and continuously differentiable in a neighborhood of x_0, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) < 0$. Moreover, suppose that $H^{uc}(x)$ and $x \mapsto \Phi(\beta_0, x)$ are continuously differentiable in a neighborhood of x_0, where H^{uc} is defined below* (2.2.9) *and Φ is defined in* (2.3.1). *Let $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ be the estimators defined in Section 2.2.2. Then,*

$$n^{1/3}\left|\frac{\Phi(\beta_0, x_0)}{4\lambda_0(x_0)\lambda_0'(x_0)}\right|^{1/3}\left[\hat{\lambda}_n(x_0) - \lambda_0(x_0)\right] \xrightarrow{d} \underset{t\in\mathbb{R}}{\operatorname{argmin}}\{\mathbb{W}(t) + t^2\}, \tag{2.3.4}$$

where $\mathbb{W}$ is standard two-sided Brownian motion originating from zero. Furthermore,

$$n^{1/3}\left[\tilde{\lambda}_n(x_0) - \hat{\lambda}_n(x_0)\right] \xrightarrow{p} 0,$$

so that the convergence in (2.3.4) *also holds with $\hat{\lambda}_n$ replaced by $\tilde{\lambda}_n$.*

In the special case of no covariates, i.e., $\beta_0 \equiv 0$, it follows that $\Phi(\beta_0, x_0) = 1 - H(x_0)$, so that with the above results we recover Theorems 2.2 and 2.3 in HUANG & WELLNER (1995). If, in addition, one specializes to the case of no censoring, i.e., $\Phi(\beta_0, x_0) = 1 - H(x_0) = 1 - F(x_0)$, we recover Theorems 6.1 and 7.1 in PRAKASA RAO (1970). The asymptotic distribution of the baseline density estimator is provided by the next theorem.

THEOREM 2.6. *Assume (A1) and (A2) and let $x_0 \in (0, \tau_H)$. Suppose that f_0 is non-increasing on $[0,\infty)$ and continuously differentiable in a neighborhood of x_0, with $f_0(x_0) \neq 0$ and $f_0'(x_0) < 0$. Let F_0 be the baseline distribution function and suppose that $H^{uc}(x)$ and $x \mapsto \Phi(\beta_0, x)$ are continuously differentiable in a neighborhood of x_0, where H^{uc} is defined below* (2.2.9) *and Φ is defined in* (2.3.1). *Let $\tilde{f}_n$ be the estimator defined in Section 2.2.3. Then,*

$$n^{1/3} \left| \frac{\Phi(\beta_0, x_0)}{4 f_0(x_0) f_0'(x_0)[1 - F_0(x_0)]} \right|^{1/3} \left[\tilde{f}_n(x_0) - f_0(x_0) \right] \xrightarrow{d} \underset{t\in\mathbb{R}}{\operatorname{argmin}}\{\mathbb{W}(t) + t^2\},$$

where $\mathbb{W}$ is standard two-sided Brownian motion originating from zero.

In the special case of no covariates, it follows that

$$\frac{\Phi(\beta_0, x_0)}{1 - F_0(x_0)} = \frac{1 - H(x_0)}{1 - F(x_0)} = 1 - G(x_0),$$

so that the above result recovers Theorem 2.1 in HUANG & WELLNER (1995). If, in addition, one specializes to the case of no censoring, i.e., $G(x_0) = 0$, we recover Theorem 6.3 in PRAKASA RAO (1969), and the corresponding result in GROENEBOOM (1985).

2.4 CONSISTENCY

The strong pointwise consistency of the proposed estimators will be proven using arguments similar to those in ROBERTSON *et al.* (1988) and HUANG & WELLNER (1985). First, define

$$\Phi_n(\beta, x) = \int \{u \geqslant x\}\, \mathrm{e}^{\beta' z}\, \mathrm{d}P_n(u, \delta, z), \tag{2.4.1}$$

for $\beta \in \mathbb{R}^p$ and $x \geqslant 0$ and note that the Breslow estimator in (2.2.8) can also be represented as

$$\Lambda_n(x) = \int \frac{\delta\{u \leqslant x\}}{\Phi_n(\hat{\beta}_n, u)}\, \mathrm{d}P_n(u, \delta, z), \qquad x \geqslant 0. \tag{2.4.2}$$

To establish consistency of the estimators, we first obtain some properties of Φ_n and Φ, as defined in (2.4.1) and (2.3.1) and their first and second partial derivatives, which by the dominated convergence theorem and conditions (A1) and (A2) are given by

$$\begin{aligned}
D^{(1)}(\beta,x) &= \frac{\partial \Phi(\beta,x)}{\partial \beta} = \int \{u \geqslant x\}\, z\, \mathrm{e}^{\beta' z}\, \mathrm{d}P(u,\delta,z) \in \mathbb{R}^p,\\
D_n^{(1)}(\beta,x) &= \frac{\partial \Phi_n(\beta,x)}{\partial \beta} = \int \{u \geqslant x\}\, z\, \mathrm{e}^{\beta' z}\, \mathrm{d}P_n(u,\delta,z) \in \mathbb{R}^p,\\
D^{(2)}(\beta,x) &= \frac{\partial^2 \Phi(\beta,x)}{\partial \beta^2} = \int \{u \geqslant x\}\, zz'\, \mathrm{e}^{\beta' z}\, \mathrm{d}P(u,\delta,z) \in \mathbb{R}^p \times \mathbb{R}^p,\\
D_n^{(2)}(\beta,x) &= \frac{\partial^2 \Phi_n(\beta,x)}{\partial \beta^2} = \int \{u \geqslant x\}\, zz'\, \mathrm{e}^{\beta' z}\, \mathrm{d}P_n(u,\delta,z) \in \mathbb{R}^p \times \mathbb{R}^p.
\end{aligned}$$

In order to prove consistency, we need uniform bounds on Φ and its derivatives. These are provided by the next lemma.

LEMMA 2.7. *Suppose that (A2) holds for some $\varepsilon > 0$. Then, for any $0 < M < \tau_H$,*

(i)

$$0 < \inf_{x \leqslant M} \inf_{|\beta-\beta_0|\leqslant\varepsilon} |\Phi(\beta,x)| \leqslant \sup_{x\in\mathbb{R}} \sup_{|\beta-\beta_0|\leqslant\varepsilon} |\Phi(\beta,x)| < \infty.$$

(ii) For any sequence β_n^*, such that $\beta_n^* \to \beta_0$ almost surely,

$$0 < \liminf_{n\to\infty} \inf_{x\leqslant M} |\Phi_n(\beta_n^*,x)| \leqslant \limsup_{n\to\infty} \sup_{x\in\mathbb{R}} |\Phi_n(\beta_n^*,x)| < \infty,$$

with probability one.

(iii) For $i = 1,2$,

$$\sup_{x\in\mathbb{R}} \sup_{|\beta-\beta_0|\leqslant\varepsilon} |D^{(i)}(\beta,x)| < \infty.$$

(iv) For $i = 1,2$ and for any sequence β_n^*, such that $\beta_n^* \to \beta_0$ almost surely,

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

with probability one.

PROOF. First, for every $x \leqslant M$ and $\beta \in \mathbb{R}^p$,

$$0 < \Phi(\beta,M) \leqslant \Phi(\beta,x) \tag{2.4.3}$$

and for every $x \in \mathbb{R}$ and $|\beta - \beta_0| \leqslant \varepsilon$,

$$\Phi(\beta,x) \leqslant \Phi(\beta,0) \leqslant \sup_{|\beta-\beta_0|\leqslant\varepsilon} \mathbb{E}\left[\mathrm{e}^{\beta' Z}\right] < \infty. \tag{2.4.4}$$

Hence, by dominated convergence, for every $x \leqslant M$, the function $\beta \mapsto \Phi(\beta, x)$ is continuous and therefore attains a minimum on the set $|\beta-\beta_0| \leqslant \varepsilon$. Together with (2.4.3) and (2.4.4), this proves (i).

To show (ii), note that similar to (2.4.3) and (2.4.4), for every $x \in [0, M]$ and $\beta \in \mathbb{R}^p$,

$$\Phi_n(\beta, M) \leqslant \Phi_n(\beta, x) \tag{2.4.5}$$

and for every $x \in \mathbb{R}$ and $\beta \in \mathbb{R}^p$,

$$\Phi_n(\beta, x) \leqslant \Phi_n(\beta, 0). \tag{2.4.6}$$

Choose $\varepsilon > 0$ from (A2) and let $\delta = \varepsilon/2\sqrt{p}$. Strong consistency of β_n^* yields that, for n sufficiently large,

$$\beta_{0j} - \delta \leqslant \beta_{nj}^* \leqslant \beta_{0j} + \delta, \quad \text{for all } j = 1, 2, \dots, p,$$

with probability one. Next, consider all subsets $I_k = \{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, p\} = I$. Then, for each I_k fixed, on each event

$$\bigcap_{j \in I_k} \{Z_{ij} \geqslant 0\} \bigcap_{l \in I \setminus I_k} \{Z_{il} < 0\}, \quad \text{where } Z_i = (Z_{i1}, \dots, Z_{ip})' \in \mathbb{R}^p,$$

we have

$$\sum_{j \in I_k} (\beta_{0j} - \delta) Z_{ij} + \sum_{l \in I \setminus I_k} (\beta_{0j} + \delta) Z_{il} \leqslant \beta_n^{*'} Z \leqslant \sum_{j \in I_k} (\beta_{0j} + \delta) Z_{ij} + \sum_{l \in I \setminus I_k} (\beta_{0j} - \delta) Z_{il}.$$

Define $\alpha_k, \gamma_k \in \mathbb{R}^p$ with coordinates

$$\alpha_{kj} = \begin{cases} \beta_{0j} - \delta, & j \in I_k, \\ \beta_{0j} + \delta, & j \in I \setminus I_k, \end{cases} \quad \text{and} \quad \gamma_{kj} = \begin{cases} \beta_{0j} + \delta, & j \in I_k, \\ \beta_{0j} - \delta, & j \in I \setminus I_k. \end{cases}$$

Then $|\beta_0 - \alpha_k| \leqslant \varepsilon$ and $|\beta_0 - \gamma_k| \leqslant \varepsilon$ and together with (2.4.5) and (2.4.6), we find that for every $x \leqslant M$,

$$\min_{I_k \subseteq I} \left\{ \frac{1}{n} \sum_{i=1}^{n} \{T_i \geqslant M\} \mathrm{e}^{\alpha_k' Z_i} \right\} \leqslant \Phi_n(\beta_n^*, x) \tag{2.4.7}$$

and for every $x \in \mathbb{R}$,

$$\Phi_n(\beta_n^*, x) \leqslant \max_{I_k \subseteq I} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{e}^{\gamma_k' Z_i} \right\}. \tag{2.4.8}$$

By (A2) and the law of large numbers,

$$\min_{I_k \subseteq I} \left\{ \frac{1}{n} \sum_{i=1}^{n} \{T_i \geqslant M\} \mathrm{e}^{\alpha_k' Z_i} \right\} \to \min_{I_k \subseteq I} \mathbb{E}\left[\{T \geqslant M\} \mathrm{e}^{\alpha_k' Z} \right] > 0,$$

with probability one and similarly,

$$\max_{I_k \subseteq I} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{e}^{\gamma_k' Z_i} \right\} \to \max_{I_k \subseteq I} \mathbb{E}\left[\mathrm{e}^{\gamma_k' Z}\right] \leqslant \sup_{|\beta-\beta_0|\leqslant\varepsilon} \mathbb{E}\left[\mathrm{e}^{\beta' Z}\right] < \infty, \tag{2.4.9}$$

with probability one. This proves (ii).

To prove (iii), it suffices to show that the inequalities hold componentwise. For this, notice that for the jth element of the vector $D^{(1)}$,

$$\sup_{x\in\mathbb{R}} \sup_{|\beta-\beta_0|\leqslant\varepsilon} \left| \mathbb{E}\left[\{T \geqslant x\} Z_j\, \mathrm{e}^{\beta' Z}\right]\right| \leqslant \sup_{|\beta-\beta_0|\leqslant\varepsilon} \mathbb{E}\left[|Z_j| \mathrm{e}^{\beta' Z}\right] < \infty,$$

by (A2). Completely analogous, a similar inequality can be shown for each element of $D^{(2)}$.

Finally, to prove (iv), note that similar to (2.4.8) and (2.4.9), for the jth component of $D_n^{(1)}$, we can write

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

with probability one, as n tends to infinity. Likewise, a similar result can be obtained for each element of $D_n^{(2)}$. ◻

Obviously, we will approximate $\Phi_n(\hat{\beta}_n, x)$ and $\Phi_n(\beta_0, x)$ by $\Phi(\beta_0, x)$. According to the law of large numbers, Φ_n will converge to Φ, for β and x fixed. However, we need uniform convergence at proper rates. This is established by the following lemma.

LEMMA 2.8. *Suppose that condition (A2) holds and $\hat{\beta}_n \to \beta_0$, with probability one. Then,*

$$\sup_{x\in\mathbb{R}} \left| \Phi_n(\hat{\beta}_n, x) - \Phi(\beta_0, x) \right| \to 0,$$

with probability one. Moreover,

$$\sqrt{n} \sup_{x\in\mathbb{R}} \left| \Phi_n(\beta_0, x) - \Phi(\beta_0, x) \right| = \mathcal{O}_p(1). \tag{2.4.10}$$

PROOF. For all $x \in \mathbb{R}$, write

$$|\Phi_n(\hat{\beta}_n, x) - \Phi(\beta_0, x)| \leqslant |\Phi_n(\hat{\beta}_n, x) - \Phi_n(\beta_0, x)| + |\Phi_n(\beta_0, x) - \Phi(\beta_0, x)|.$$

For the second term on the right hand side, consider the class of functions

$$\mathscr{G} = \left\{ g(u, z; x) : x \in \mathbb{R} \right\},$$

where for each $x \in \mathbb{R}$ and $\beta_0 \in \mathbb{R}^p$ fixed,

$$g(u,z;x) = \{u \geqslant x\}\exp(\beta_0' z)$$

is a product of an indicator and a fixed function. It follows that $\mathscr{G}$ is a VC-subgraph class (e.g., see Lemma 2.6.18 in van der VAART & WELLNER, 1996) and its envelope $G = \exp(\beta_0' z)$ is square integrable under condition (A2). Standard results from empirical process theory (van der VAART & WELLNER, 1996) yield that the class of functions $\mathscr{G}$ is Glivenko-Cantelli, i.e.,

$$\sup_{x\in\mathbb{R}} |\Phi_n(\beta_0, x) - \Phi(\beta_0, x)| = \sup_{g\in\mathscr{G}} \left| \int g(u,z;x)\,\mathrm{d}(P_n - P)(u,\delta,z) \right| \to 0, \qquad (2.4.11)$$

with probability one. Moreover, $\mathscr{G}$ is a Donsker class, i.e.,

$$\sqrt{n} \int g(u,z;x)\,\mathrm{d}(P_n - P)(u,\delta,z) = \mathscr{O}_p(1),$$

so that (2.4.10) follows by continuous mapping theorem. Finally, by Taylor expansion and the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}\sup_{x\in\mathbb{R}} \left|\Phi_n(\hat\beta_n, x) - \Phi_n(\beta_0, x)\right| &= \sup_{x\in\mathbb{R}} \left| (\hat\beta_n - \beta_0)' D_n^{(1)}(\beta^*, x) \right| \\ &\leqslant |\hat\beta_n - \beta_0| \sup_{x\in\mathbb{R}} \left| D_n^{(1)}(\beta^*, x) \right|,\end{aligned}$$

for some β^*, for which $|\beta^* - \beta_0| \leqslant |\hat\beta_n - \beta_0|$. Together with (2.4.11), from the strong consistency of $\hat\beta_n$ (e.g., see Theorem 3.1 in TSIATIS, 1981) and Lemma 2.7, the lemma follows. □

The previous results can be used to prove a first step in the direction of proving Theorem 2.3, i.e., suitable uniform approximation of Λ_n and F_n by Λ_0 and F_0. Strong uniform consistency of Λ_n and process convergence of $\sqrt{n}(\Lambda_n - \Lambda_0)$ has been established by KOSOROK (2008), under the stronger assumption of bounded covariates. Weak consistency has been derived or mentioned before, see for example PRENTICE & KALBFLEISCH (2003).

THEOREM 2.9. *Under the assumptions (A1) and (A2), for all* $0 < M < \tau_H$*,*

$$\sup_{x\in[0,M]} \left|\Lambda_n(x) - \Lambda_0(x)\right| \to 0,$$

with probability one and

$$\sqrt{n} \sup_{x\in[0,M]} \left|\Lambda_n(x) - \Lambda_0(x)\right| = \mathscr{O}_p(1).$$

PROOF. From the expression for the baseline cumulative hazard function in (2.2.10) together with (2.3.1) and (2.4.2), it follows that

$$\begin{aligned}\sup_{x\in[0,M]}\left|\Lambda_n(x)-\Lambda_0(x)\right| &\leqslant \sup_{x\in[0,M]}\left|\int \delta\{u\leqslant x\}\left(\frac{1}{\Phi_n(\hat{\beta}_n,u)}-\frac{1}{\Phi_n(\beta_0,u)}\right)\,\mathrm{d}P_n(u,\delta,z)\right|\\ &+\sup_{x\in[0,M]}\left|\int \delta\{u\leqslant x\}\left(\frac{1}{\Phi_n(\beta_0,u)}-\frac{1}{\Phi(\beta_0,u)}\right)\,\mathrm{d}P_n(u,\delta,z)\right|\\ &+\sup_{x\in[0,M]}\left|\int \frac{\delta\{u\leqslant x\}}{\Phi(\beta_0,u)}\,\mathrm{d}\left(P_n-P\right)(u,\delta,z)\right|\\ &=A_n+B_n+C_n.\end{aligned}$$

Starting with the first term on the right hand side, note that

$$A_n\leqslant\frac{|\hat{\beta}_n-\beta_0|}{\Phi_n(\hat{\beta}_n,M)\Phi_n(\beta_0,M)}\sup_{x\in\mathbb{R}}\left|D_n^{(1)}(\beta^*,x)\right| \tag{2.4.12}$$

for some $|\beta^*-\beta_0|\leqslant|\hat{\beta}_n-\beta_0|$. According to Lemma 2.7, the right hand side is bounded by $C|\hat{\beta}_n-\beta_0|$, for some $C>0$. Since $\hat{\beta}_n$ is strongly consistent and $|\hat{\beta}_n-\beta_0|=\mathcal{O}_p(n^{-1/2})$, (e.g., see Theorems 3.1 and 3.2 in TSIATIS, 1981), it follows that $A_n\to 0$ almost surely and $A_n=\mathcal{O}_p(n^{-1/2})$. Similarly,

$$B_n\leqslant\frac{1}{\Phi_n(\beta_0,M)\Phi(\beta_0,M)}\sup_{x\in\mathbb{R}}\left|\Phi_n(\beta_0,x)-\Phi(\beta_0,x)\right|. \tag{2.4.13}$$

From Lemmas 2.7 and 2.8, it follows that $B_n\to 0$ almost surely and $B_n=\mathcal{O}_p(n^{-1/2})$. For the last term C_n, consider the class of functions $\mathcal{H}=\{h(u,\delta;x):x\in[0,M]\}$, where for each $x\in[0,M]$, with $M<\tau_H$ and $\beta_0\in\mathbb{R}^p$ fixed,

$$h(u,\delta;x)=\frac{\delta\{u\leqslant x\}}{\Phi(\beta_0,u)}.$$

The function h is a product of indicators and a fixed uniformly bounded monotone function. Similar to the arguments given in the proof of Lemma 2.8, it follows that the class $\mathcal{H}$ is Glivenko-Cantelli, i.e.,

$$\sup_{h\in\mathcal{H}}\left|\int h(u,\delta;\cdot)\mathrm{d}(P_n-P)(u,\delta,z)\right|\to 0,$$

almost surely, which gives the first statement of the lemma. Moreover, $\mathcal{H}$ is a Donsker class and hence the second statement of the lemma follows by continuous mapping theorem. This completes the proof. □

Strong uniform consistency of F_n follows immediately from the strong consistency of the Breslow estimator established in Theorem 2.9, and is stated in the next corollary.

COROLLARY 2.10. *Under the assumptions (A1) and (A2) and for all* $0 < M < \tau_H$,

$$\sup_{x \in [0,M]} |F_n(x) - F_0(x)| \to 0,$$

with probability one.

PROOF. The proof is straightforward and follows immediately from Theorem 2.9, relations (2.2.12) and (2.2.13), together with the fact that $|e^{-y} - 1| \leqslant 2|y|$, as $y \to 0$. □

Note that the estimators in Theorem 2.3 (i) of the baseline hazard are essentially the slopes of the GCM of V_n. For this reason, as a final preparation for the proof of Theorem 2.3, we establish uniform convergence of the GCM of V_n by the following lemma. This lemma is completely similar to Lemma 4.3 in HUANG & WELLNER (1995).

LEMMA 2.11. *Assume that* Λ_0 *is convex on* $[0, \tau_H]$ *and that conditions (A1) and (A2) hold. Let* $\hat\beta_n$ *be the maximum partial likelihood estimator and define*

$$\widehat{W}_n(x) = W_n(\hat\beta_n, x) - W_n(\hat\beta_n, T_{(1)}), \qquad x \geqslant T_{(1)}, \tag{2.4.14}$$

where W_n *is defined in* (2.2.3). *Let* $\big(\widehat{W}_n(x), \widehat{V}_n(x)\big)$ *be the GCM of* $\big(\widehat{W}_n(x), V_n(x)\big)$, *for* $x \in [T_{(1)}, T_{(n)}]$, *where* V_n *is defined in* (2.2.4). *Then*

$$\sup_{x \in [T_{(1)}, T_{(n)}]} \left|\widehat{V}_n(x) - V(x)\right| \to 0, \tag{2.4.15}$$

with probability one, where $V(x) = H^{uc}(x)$, *as defined just below* (2.2.9).

PROOF. By Glivenko-Cantelli,

$$\sup_{x \in [T_{(1)}, T_{(n)}]} \left|V_n(x) - V(x)\right| \to 0, \tag{2.4.16}$$

almost surely, because of the continuity of V. Furthermore,

$$W_n(\hat\beta_n, T_{(1)}) = \int_0^{T_{(1)}} \Phi_n(\hat\beta_n, s)\, ds = T_{(1)} \Phi_n(\hat\beta_n, T_{(1)}) \to 0, \tag{2.4.17}$$

almost surely, since $\Phi_n(\hat\beta_n, s)$ is bounded uniformly according to Lemma 2.7 and $T_{(1)} \to 0$ with probability one, by the Borel-Cantelli lemma. Moreover, if we define

$$W(\beta, x) = \int \left(e^{\beta' z} \int_0^x \{u \geqslant s\}\, ds \right) dP(u, \delta, z), \tag{2.4.18}$$

then we can write

$$W_0(x) = W(\beta_0, x) = \int_0^x \Phi(\beta_0, s)\,\mathrm{d}s, \tag{2.4.19}$$

where Φ is defined in (2.3.1). It follows that

$$\begin{aligned} \sup_{x\in[T_{(1)},T_{(n)}]} \left|\widehat{W}_n(x) - W_0(x)\right| &\leqslant \sup_{x\in[T_{(1)},T_{(n)}]} \left| \int_0^x \left(\Phi_n(\hat{\beta}_n, s) - \Phi(\beta_0, s)\right)\,\mathrm{d}s \right|, \\ &\leqslant \tau_H \sup_{x\in\mathbb{R}} \left|\Phi_n(\hat{\beta}_n, x) - \Phi(\beta_0, x)\right| \to 0, \end{aligned} \tag{2.4.20}$$

with probability one, by Lemma 2.8.

Take $\widehat{W}_n^{-1}$ to be the inverse of $\widehat{W}_n$, which is well defined on $[0, \widehat{W}_n(T_{(n)})]$, since $\widehat{W}_n$ is strictly monotone on $[T_{(1)}, T_{(n)}]$. We first extend $\widehat{W}_n$ to $[T_{(1)}, \infty)$ and $\widehat{W}_n^{-1}$ to $[0, \infty)$. Define $\widehat{W}_n(t) = \widehat{W}_n(T_{(n)}) + (t - T_{(n)})$, for all $t \geqslant T_{(n)}$, so that $\widehat{W}_n^{-1}(y) = T_{(n)} + (y - \widehat{W}_n(T_{(n)}))$, for $y \geqslant \widehat{W}_n(T_{(n)})$. Similarly, take W_0^{-1} to be the inverse of W_0, which is well-defined since W_0 is strictly monotone on $[0, \tau_H]$ and extend W_0 and W_0^{-1} to $[0, \infty)$, by defining $W_0(t) = W_0(\tau_H) + (t - \tau_H)$, for all $t \geqslant \tau_H$, so that $W_0^{-1}(y) = \tau_H + (y - W_0(\tau_H))$, for $y \geqslant W_0(\tau_H)$. It follows that the extension $W_0^{-1}(y)$ is uniformly continuous on $[0, \infty)$. Immediate derivations give that

$$\sup_{0\leqslant y\leqslant \widehat{W}_n(T_{(n)})} \left|\widehat{W}_n^{-1}(y) - W_0^{-1}(y)\right| \to 0, \tag{2.4.21}$$

with probability one. Furthermore, it can be inferred that

$$\begin{aligned} \delta_n = & \sup_{y\in[0,\widehat{W}_n(T_{(n)})]} \left|V_n \circ \widehat{W}_n^{-1}(y) - V \circ W_0^{-1}(y)\right| \\ \leqslant & \sup_{y\in[0,\widehat{W}_n(T_{(n)})]} \left|(V_n - V) \circ \widehat{W}_n^{-1}(y)\right| + \sup_{y\in[0,\widehat{W}_n(T_{(n)})]} \left|V \circ \widehat{W}_n^{-1}(y) - V \circ W_0^{-1}(y)\right| \\ \leqslant & \sup_{t\in[T_{(1)},T_{(n)}]} \left|V_n(t) - V(t)\right| + \sup_{y\in[0,\widehat{W}_n(T_{(n)})]} \left|V \circ \left(\widehat{W}_n^{-1}(y) - W_0^{-1}(y)\right)\right| \\ \to & \, 0, \end{aligned}$$

almost surely, by (2.4.16), (2.4.21), and the continuity of V. According to (2.2.9) and (2.4.19), λ_0 can also be represented as

$$\lambda_0(x) = \frac{\mathrm{d}V(x)/\mathrm{d}x}{\mathrm{d}W_0(x)/\mathrm{d}x}, \tag{2.4.22}$$

which is well-defined for $x \in [0, \tau_H)$, since Φ is bounded away from zero, by Lemma 2.7. Taking $x = W_0^{-1}(y)$, gives that

$$\frac{\mathrm{d}V\left(W_0^{-1}(y)\right)}{\mathrm{d}y} = \lambda_0\left(W_0^{-1}(y)\right), \quad y \in [0, W_0(\tau_H)).$$

Therefore, convexity of Λ_0 implies convexity of $V \circ W_0^{-1}$ and subsequently of $V \circ W_0^{-1} - \delta_n$. Moreover, from the definition of δ_n, it follows that for every $y \in [0, \widehat{W}_n(T_{(n)})]$,

$$V \circ W_0^{-1}(y) - \delta_n \leqslant V_n \circ \widehat{W}_n^{-1}(y).$$

As $\widehat{V}_n \circ \widehat{W}_n^{-1}(y)$ is the greatest convex function below $V_n \circ \widehat{W}_n^{-1}(y)$, we must have

$$V \circ W_0^{-1}(y) - \delta_n \leqslant \widehat{V}_n \circ \widehat{W}_n^{-1}(y) \leqslant V_n \circ \widehat{W}_n^{-1}(y),$$

for each $y \in [0, \widehat{W}_n(T_{(n)})]$. Re-writing the above inequalities leads to

$$-\delta_n \leqslant \widehat{V}_n \circ \widehat{W}_n^{-1}(y) - V \circ W_0^{-1}(y) \leqslant V_n \circ \widehat{W}_n^{-1}(y) - V \circ W_0^{-1}(y) \leqslant \delta_n.$$

Taking the supremum over $[0, \widehat{W}_n(T_{(n)})]$ then yields

$$\sup_{y \in [0, \widehat{W}_n(T_{(n)})]} \left| \widehat{V}_n \circ \widehat{W}_n^{-1}(y) - V \circ W_0^{-1}(y) \right| \to 0, \tag{2.4.23}$$

with probability one. From (2.4.21), (2.4.23) and the continuity of V, we conclude that

$$\begin{aligned} \sup_{t \in [T_{(1)}, T_{(n)}]} \left| \widehat{V}_n(t) - V(t) \right| &= \sup_{y \in [0, \widehat{W}_n(T_{(n)})]} \left| \left(\widehat{V}_n - V \right) \circ \widehat{W}_n^{-1}(y) \right| \\ &\leqslant \sup_{y \in [0, \widehat{W}_n(T_{(n)})]} \left| \widehat{V}_n \circ \widehat{W}_n^{-1}(y) - V \circ W_0^{-1}(y) \right| \\ &\quad + \sup_{y \in [0, \widehat{W}_n(T_{(n)})]} \left| V \circ W_0^{-1}(y) - V \circ \widehat{W}_n^{-1}(y) \right| \to 0, \end{aligned}$$

with probability one. □

Obviously, in the nonincreasing case, similar to (2.4.15), one can show

$$\sup_{x \in [0, T_{(n)}]} \left| \widehat{Y}_n(x) - V(x) \right| \to 0, \tag{2.4.24}$$

almost surely, where $\big(W_n(\hat{\beta}_n, x), \widehat{Y}_n(x)\big)$ is the LCM of $\big(W_n(\hat{\beta}_n, x), Y_n(x)\big)$, with Y_n defined in (2.2.11). We are now in the position to prove Theorem 2.3, which establishes strong pointwise consistency of the estimators.

PROOF. [Proof of Theorem 2.3] First consider the second statement of case (i). Since $\widetilde{\Lambda}_n$ is convex on the open interval $(0, \tau_H)$, it admits in every point $x_0 \in (0, \tau_H)$ a finite left and a right derivative, denoted by $\widetilde{\Lambda}_n^-$ and $\widetilde{\Lambda}_n^+$ respectively. Moreover, for any fixed $x_0 \in (0, \tau_H)$ and for sufficiently small $\delta > 0$, it follows that

$$\frac{\widetilde{\Lambda}_n(x_0) - \widetilde{\Lambda}_n(x_0 - \delta)}{\delta} \leqslant \widetilde{\Lambda}_n^-(x_0) \leqslant \widetilde{\Lambda}_n^+(x_0) \leqslant \frac{\widetilde{\Lambda}_n(x_0 + \delta) - \widetilde{\Lambda}_n(x_0)}{\delta}.$$

When $n \to \infty$, then for any $0 < M < \tau_H$,

$$\sup_{x \in [0,M]} \left| \widetilde{\Lambda}_n(x) - \Lambda_0(x) \right| \leqslant \sup_{x \in [0,M]} \left| \Lambda_n(x) - \Lambda_0(x) \right|. \qquad (2.4.25)$$

This is a variation of Marshall's lemma and can be proven similar to (7.2.3) in ROBERTSON *et al.* (1988) or Lemma 4.1 in HUANG & WELLNER (1995). By convexity of Λ_0 and the fact that $\widetilde{\Lambda}_n$ is the greatest convex function below Λ_n, one must have

$$\Lambda_0(x) - \delta_n \leqslant \widetilde{\Lambda}_n(x) \leqslant \Lambda_n(x),$$

where $\delta_n = \sup_{x \in [0,M]} |\Lambda_0(x) - \Lambda_n(x)|$, which yields inequality (2.4.25). From (2.4.25) and Theorem 2.9, by first letting $n \to \infty$ and then $\delta \to 0$, we find

$$\lambda_0(x_0-) \leqslant \liminf_{n\to\infty} \widetilde{\Lambda}_n^-(x_0) \leqslant \limsup_{n\to\infty} \widetilde{\Lambda}_n^-(x_0) \leqslant \limsup_{n\to\infty} \widetilde{\Lambda}_n^+(x_0) \leqslant \lambda_0(x_0+).$$

Because $\tilde{\lambda}_n(x_0) = \widetilde{\Lambda}_n^-(x_0)$, this proves that $\tilde{\lambda}_n$ is a strongly consistent estimator.

For $\hat{\lambda}_n$, first note that since $\widehat{V}_n$ is convex on the open interval $(0, \tau_H)$, it admits in every point $x_0 \in (0, \tau_H)$ a finite left and a right derivative, denoted by $\widehat{V}_n^-$ and $\widehat{V}_n^+$ respectively, where

$$\widehat{V}_n^-(x) = \lim_{\delta \downarrow 0} \frac{\widehat{V}_n(x) - \widehat{V}_n(x-\delta)}{\widehat{W}_n(x) - \widehat{W}_n(x-\delta)},$$
$$\widehat{V}_n^+(x) = \lim_{\delta \downarrow 0} \frac{\widehat{V}_n(x+\delta) - \widehat{V}_n(x)}{\widehat{W}_n(x+\delta) - \widehat{W}_n(x)}.$$

For any fixed $x \in (0, \tau_H)$ and for sufficiently small $\delta > 0$, it follows that

$$\frac{\widehat{V}_n(x_0) - \widehat{V}_n(x_0-\delta)}{\widehat{W}_n(x_0) - \widehat{W}_n(x_0-\delta)} \leqslant \widehat{V}_n^-(x_0) \leqslant \widehat{V}_n^+(x_0) \leqslant \frac{\widehat{V}_n(x_0+\delta) - \widehat{V}_n(x_0)}{\widehat{W}_n(x_0+\delta) - \widehat{W}_n(x_0)}.$$

By (2.4.19) and (2.4.20), and letting $n \to \infty$, we obtain

$$\frac{V(x_0) - V(x_0-\delta)}{W_0(x_0) - W_0(x_0-\delta)} \leqslant \liminf_{n\to\infty} \widehat{V}_n^-(x_0) \leqslant \limsup_{n\to\infty} \widehat{V}_n^+(x_0) \leqslant \frac{V(x_0+\delta) - V(x_0)}{W_0(x_0+\delta) - W_0(x_0)}.$$

Furthermore, by letting $\delta \to 0$, together with (2.4.22), we get

$$\lambda_0(x_0-) \leqslant \liminf_{n\to\infty} \widehat{V}_n^-(x_0) \leqslant \limsup_{n\to\infty} \widehat{V}_n^-(x_0) \leqslant \limsup_{n\to\infty} \widehat{V}_n^+(x_0) \leqslant \lambda_0(x_0+),$$

which completes the proof of (i), since $\hat{\lambda}_n(x_0) = \widehat{V}_n^-(x_0)$. The proofs of (ii) and (iii) are completely analogous, using (2.4.24) and Corollary 2.10. □

2.5 Inverse processes

To obtain the limit distribution of the estimators, we follow the approach proposed by GROENEBOOM (1985). For each proposed estimator, we define an inverse process and establish its asymptotic distribution. The asymptotic distribution of the estimators then emerges via the switching relationships. The inverse processes are defined in terms of some local processes and this section is devoted to acquire the weak convergence of these local processes. Furthermore, the inverse processes need to be bounded in probability. This result, along with the limiting distribution of the inverse processes and hence of the estimators are deferred to Section 2.6.

In order to keep the exposition brief, we do not treat all five separate cases in detail, but we confine ourselves to the most important ones, as the other cases can be handled similarly. In the case of a nondecreasing λ_0, the distribution of the NPMLE $\hat{\lambda}_n$ can be obtained through the study of the inverse process

$$\widehat{U}_n^{\lambda}(a) = \underset{x \in [T_{(1)}, T_{(n)}]}{\operatorname{argmin}} \left\{ V_n(x) - a\widehat{W}_n(x) \right\}, \tag{2.5.1}$$

for $a > 0$, where V_n and $\widehat{W}_n$ have been defined in (2.2.4) and (2.4.14). Succeedingly, for a given $a > 0$, the switching relationship holds, i.e., $\widehat{U}_n^{\lambda}(a) \geqslant x$ if and only if $\hat{\lambda}_n(x) \leqslant a$ with probability one, so that after scaling, it follows that

$$n^{1/3}\left[\hat{\lambda}_n(x_0) - \lambda_0(x_0)\right] > a \Leftrightarrow n^{1/3}\left[\widehat{U}_n^{\lambda}(\lambda_0(x_0) + n^{-1/3}a) - x_0\right] < 0, \tag{2.5.2}$$

for $0 < x_0 < \tau_H$, with probability one. A similar relationship holds for $\tilde{\lambda}_n$ and the corresponding inverse process

$$\widetilde{U}_n^{\lambda}(a) = \underset{x \in [0, T_{(n)}]}{\operatorname{argmin}} \left\{ \Lambda_n(x) - ax \right\}. \tag{2.5.3}$$

For the nonincreasing density estimator $\tilde{f}_n$, we consider the inverse process

$$\widetilde{U}_n^{f}(a) = \underset{x \in [0, T_{(n)}]}{\operatorname{argmax}} \left\{ F_n(x) - ax \right\}, \tag{2.5.4}$$

where argmax denotes the largest location of the maximum. In this case, instead of (2.5.2), we have

$$n^{1/3}\left[\tilde{f}_n(x_0) - f_0(x_0)\right] > a \Leftrightarrow n^{1/3}\left[\widetilde{U}_n^{f}(f_0(x_0) + n^{-1/3}a) - x_0\right] > 0, \tag{2.5.5}$$

Similarly, in the case of estimating a nonincreasing λ_0, we consider the inverse processes $\widehat{U}_n^{\lambda}$ and $\widetilde{U}_n^{\lambda}$ defined with argmax instead of argmin in (2.5.1) and (2.5.3) and we have switching relations similar to (2.5.5).

From the definition of the inverse process in (2.5.3) and given that the argmin is invariant under addition of and multiplication with positive constants, it can be derived that

$$n^{1/3}\left[\widetilde{U}_n^{\lambda}(\lambda_0(x_0)+n^{-1/3}a)-x_0\right]=\underset{x\in I_n(x_0)}{\operatorname{argmin}}\left\{\widetilde{\mathbb{Z}}_n^{\lambda}(x)-ax\right\} \tag{2.5.6}$$

where $I_n(x_0)=[-n^{1/3}x_0,n^{1/3}(T_{(n)}-x_0)]$ and

$$\widetilde{\mathbb{Z}}_n^{\lambda}(x)=n^{2/3}\left[\Lambda_n(x_0+n^{-1/3}x)-\Lambda_n(x_0)-n^{-1/3}\lambda_0(x_0)x\right]. \tag{2.5.7}$$

Likewise, $n^{1/3}\left[\widehat{U}_n^{\lambda}(\lambda_0(x_0)+n^{-1/3}a)-x_0\right]$ is equal to

$$\underset{x\in I_n'(x_0)}{\operatorname{argmin}}\left\{\widehat{\mathbb{Z}}_n^{\lambda}(x)-\frac{n^{1/3}a}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+n^{-1/3}x)-\widehat{W}_n(x_0)\right]\right\}, \tag{2.5.8}$$

where $I_n'(x_0)=[-n^{1/3}(x_0-T_{(1)}),n^{1/3}(T_{(n)}-x_0)]$ and

$$\begin{aligned}\widehat{\mathbb{Z}}_n^{\lambda}(x)=\frac{n^{2/3}}{\Phi(\beta_0,x_0)}\Bigg(&V_n(x_0+n^{-1/3}x)-V_n(x_0)\\&-\lambda_0(x_0)\left[\widehat{W}_n(x_0+n^{-1/3}x)-\widehat{W}_n(x_0)\right]\Bigg),\end{aligned} \tag{2.5.9}$$

and similarly

$$n^{1/3}\left[\widetilde{U}_n^{f}(f_0(x_0)+n^{-1/3}a)-x_0\right]=\underset{x\in I_n(x_0)}{\operatorname{argmax}}\{\widetilde{\mathbb{Z}}_n^{f}(x)-ax\}, \tag{2.5.10}$$

where

$$\widetilde{\mathbb{Z}}_n^{f}(x)=n^{2/3}\left[F_n(x_0+n^{-1/3}x)-F_n(x_0)-n^{-1/3}f_0(x_0)x\right]. \tag{2.5.11}$$

In the case of estimating a nonincreasing λ_0, we consider the argmax of the processes (2.5.9) and (2.5.7). Before investigating the asymptotic behavior of the above processes, we first need to establish the following technical lemma. It provides a sufficient bound on the order of shrinking increments of an empirical process that we will encounter later on.

LEMMA 2.12. *Assume (A1) and (A2). Let $x_0\in(0,\tau_H)$ fixed and suppose that*

$$H^{uc} \text{ is continuously differentiable in a neighborhood of } x_0. \tag{2.5.12}$$

Then, for any $k = 1, 2, \ldots$,

$$\sup_{|x|\leqslant k}\left|\int \delta\Big(\{u \leqslant x_0 + n^{-1/3}x\} - \{u \leqslant x_0\}\Big)\left(\frac{1}{\Phi_n(\beta_0,u)} - \frac{1}{\Phi(\beta_0,u)}\right)\mathrm{d}(P_n - P)(u,\delta,z)\right|$$

is of the order $\mathcal{O}_p(n^{-7/6}\log n)$.

Proof. Take $0 \leqslant x \leqslant k$ and consider the class of functions

$$\mathscr{F}_n = \left\{f_n(u,\delta,z;x) : 0 \leqslant x \leqslant k\right\}, \tag{2.5.13}$$

where for each $0 \leqslant x \leqslant k$,

$$f_n(u,\delta,z;x) = \delta\{x_0 < u \leqslant x_0 + n^{-1/3}x\}\left(\frac{1}{\Phi_n(\beta_0,u)} - \frac{1}{\Phi(\beta_0,u)}\right).$$

Correspondingly, consider the class $\mathscr{G}_{n,k,\alpha}$ consisting of functions

$$g(u,\delta,z;y,\Psi) = \delta\{x_0 < u \leqslant x_0 + y\}\left(\frac{1}{\Psi(u)} - \frac{1}{\Phi(\beta_0,u)}\right).$$

where $0 \leqslant y \leqslant n^{-1/3}k$ and Ψ is nonincreasing left continuous, such that

$$\Psi(x_0 + n^{-1/3}k) \geqslant K \quad \text{and} \quad \sup_{u\in\mathbb{R}}\left|\Psi(u) - \Phi(\beta_0,u)\right| \leqslant \alpha,$$

where $K = \Phi(\beta_0, (x_0 + \tau_H)/2)/2$. Then, for any $\alpha > 0$ and $k = 1, 2, \ldots$,

$$P\left(\mathscr{F}_n \subset \mathscr{G}_{n,k,\alpha}\right) \to 1,$$

by Lemma 2.8. Furthermore, the class $\mathscr{G}_{n,k,\alpha}$ has envelope

$$G(u,\delta,z) = \delta\{x_0 < u \leqslant x_0 + n^{-1/3}k\}\frac{\alpha}{K^2},$$

for which it follows from (2.5.12), that

$$\begin{aligned}\|G\|_{P,2}^2 &= \int G(u,\delta,z)^2\,\mathrm{d}P(u,\delta,z)\\ &= \frac{\alpha^2}{K^4}P(x_0 < T \leqslant x_0 + n^{-1/3}k, \Delta = 1) = \mathcal{O}(\alpha^2 k n^{-1/3}).\end{aligned}$$

Since the functions in $\mathscr{G}_{n,k,\alpha}$ are sums and products of bounded monotone functions, its entropy with bracketing satisfies

$$\log N_{[\,]}(\varepsilon, \mathscr{G}_{n,k,\alpha}, L_2(P)) \lesssim \frac{1}{\varepsilon},$$

see e.g., Theorem 2.7.5 in van der VAART & WELLNER (1996) and Lemma 9.25 in KOSOROK (2008), and hence, for any $\delta > 0$, the bracketing integral

$$J_{[\,]}(\delta, \mathscr{G}_{n,k,\alpha}, L_2(P)) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon \|G\|_2, \mathscr{G}_{n,k,\alpha}, L_2(P))}\, \mathrm{d}\varepsilon < \infty.$$

By Theorem 2.14.2 in van der VAART & WELLNER (1996), we have

$$\begin{aligned}\mathbb{E}\left\| \sqrt{n} \int g(u,\delta,z;y,\Psi) \mathrm{d}(P_n - P)(u,\delta,z) \right\|_{\mathscr{G}_{n,k,\alpha}} &\leqslant J_{[\,]}(1, \mathscr{G}_{n,k,\alpha}, L_2(P)) \|G\|_{P,2} \\ &= \mathscr{O}(\alpha k^{1/2} n^{-1/6}),\end{aligned}$$

where $\|\cdot\|_{\mathscr{F}}$ denotes the supremum over the class of functions $\mathscr{F}$. Now, according to (2.4.10)

$$(\log n)^{-1} \sqrt{n} \sup_{x \in \mathbb{R}} \left| \Phi_n(\beta_0, x) - \Phi(\beta_0, x) \right| \to 0,$$

in probability. Therefore, if we choose $\alpha = n^{-1/2} \log n$, this gives

$$\mathbb{E}\left\| \int g(u,\delta,z;y,\Psi) \mathrm{d}(P_n - P)(u,\delta,z) \right\|_{\mathscr{G}_{n,k,\alpha}} = \mathscr{O}(k^{1/2} n^{-7/6} \log n)$$

and hence by the Markov inequality, this proves the lemma for the case $0 \leqslant x \leqslant k$. The argument for $-k \leqslant x \leqslant 0$ is completely similar. □

Our approach in deriving the asymptotic distribution of the monotone estimators involves application of results from KIM & POLLARD (1990). To this end, we first determine the limiting processes of (2.5.9), (2.5.7) and (2.5.11).

LEMMA 2.13. *Suppose that (A1) and (A2) hold. Assume* (2.5.12) *and that*

$$\lambda_0 \text{ is continuously differentiable in a neighborhood of } x_0. \tag{2.5.14}$$

Moreover, assume that

$$x \mapsto \Phi(\beta_0, x) \text{ is continuously differentiable in a neighborhood of } x_0. \tag{2.5.15}$$

Then, for any $k = 1, 2, \ldots,$

$$\sup_{|x| \leqslant k} \left| \widetilde{\mathbb{Z}}_n^\lambda(x) - \widehat{\mathbb{Z}}_n^\lambda(x) \right| \to 0,$$

in probability, where the processes $\widetilde{\mathbb{Z}}_n^\lambda$ *and* $\widehat{\mathbb{Z}}_n^\lambda$ *are defined in* (2.5.7) *and* (2.5.9), *respectively.*

PROOF. We will prove that for any $k = 1, 2, \ldots,$

$$\sup_{x\in[0,k]} \left|\widetilde{\mathbb{Z}}_n^\lambda(x) - \widehat{\mathbb{Z}}_n^\lambda(x)\right| \to 0,$$

in probability, since the result for $-k \leqslant x \leqslant 0$ follows completely analogous. Write

$$\begin{aligned}
&\Phi(\beta_0, x_0)\left(\widetilde{\mathbb{Z}}_n^\lambda(x) - \widehat{\mathbb{Z}}_n^\lambda(x)\right)\\
&= n^{2/3}\int \delta\left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi_n(\hat\beta_n,u)} - 1\right)\mathrm{d}P_n(u,\delta,z)\\
&\quad - n^{2/3}\lambda_0(x_0)\int_{x_0}^{x_0+n^{-1/3}x}\left[\Phi(\beta_0,x_0) - \Phi_n(\hat\beta_n,s)\right]\mathrm{d}s\\
&= n^{2/3}\int \delta\left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi_n(\hat\beta_n,u)} - \frac{\Phi(\beta_0,x_0)}{\Phi_n(\beta_0,u)}\right)\mathrm{d}P_n(u,\delta,z)\\
&\quad + n^{2/3}\int \delta\left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi_n(\beta_0,u)} - 1\right)\mathrm{d}P_n(u,\delta,z)\\
&\quad - n^{2/3}\lambda_0(x_0)\int_{x_0}^{x_0+n^{-1/3}x}\left[\Phi(\beta_0,x_0) - \Phi_n(\beta_0,s)\right]\mathrm{d}s\\
&\quad - n^{2/3}\lambda_0(x_0)\int_{x_0}^{x_0+n^{-1/3}x}\left[\Phi_n(\beta_0,s) - \Phi_n(\hat\beta_n,s)\right]\mathrm{d}s\\
&= A_{n1}(x) + A_{n2}(x) + A_{n3}(x) + A_{n4}(x).
\end{aligned}$$

We will show that the supremum of all four terms on the right hand side tend to zero in probability. Similar to (2.4.12), according to Lemma 2.7,

$$|A_{n1}(x)| \leqslant C|\hat\beta_n - \beta_0| n^{2/3}\int \left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\mathrm{d}P_n(u,\delta,z),$$

for some $C > 0$. Since, $|\hat\beta_n - \beta_0| = \mathcal{O}_p(n^{-1/2})$ and

$$\int \left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\mathrm{d}(P_n - P)(u,\delta,z) = \mathcal{O}_p(n^{-2/3}x^{1/2}) + \mathcal{O}_p(n^{-1/3}x),$$

it follows that

$$|A_{n1}(x)| = \mathcal{O}_p(n^{-1/2}x^{1/2}) + \mathcal{O}_p(n^{-1/6}x), \tag{2.5.16}$$

and likewise, $|A_{n4}(x)| = \mathcal{O}_p(n^{-1/6}x)$. Furthermore, write

$$\begin{aligned}
&A_{n2}(x)\\
&=n^{2/3}\int \delta\left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi_n(\beta_0,u)} - \frac{\Phi(\beta_0,x_0)}{\Phi(\beta_0,u)}\right)\mathrm{d}\left(P_n - P\right)(u,\delta,z)\\
&\quad + n^{2/3}\int \delta\left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi(\beta_0,u)} - 1\right)\mathrm{d}\left(P_n - P\right)(u,\delta,z)\\
&\quad + n^{2/3}\int \delta\left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi_n(\beta_0,u)} - \frac{\Phi(\beta_0,x_0)}{\Phi(\beta_0,u)}\right)\mathrm{d}P(u,\delta,z)\\
&\quad + n^{2/3}\int \delta\left\{x_0 < u \leqslant x_0 + n^{-1/3}x\right\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi(\beta_0,u)} - 1\right)\mathrm{d}P(u,\delta,z)\\
&\quad = B_{n1}(x) + B_{n2}(x) + B_{n3}(x) + B_{n4}(x).
\end{aligned}$$

According to Lemma 2.12,

$$\sup_{0\leqslant x\leqslant k}\left|B_{n1}(x)\right| = \mathcal{O}_p(n^{-1/2}\log n). \tag{2.5.17}$$

For the term B_{n2}, consider the class $\mathscr{F}$ consisting of functions

$$f(u,\delta,z;x) = \delta\{x_0 < u \leqslant x_0 + n^{-1/3}x\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi(\beta_0,u)} - 1\right),$$

where $0 \leqslant x \leqslant k$, with envelope

$$F(u) = \delta\{x_0 < u \leqslant x_0 + n^{-1/3}k\}\left(\frac{\Phi(\beta_0,x_0)}{\Phi(\beta_0,x_0+n^{-1/3}k)} - 1\right).$$

Then, the $L_2(P)$ norm of the envelope satisfies

$$\|F\|_{P,2}^2 = \left(\frac{\Phi(\beta_0,x_0)}{\Phi(\beta_0,x_0+n^{-1/3}k)} - 1\right)^2\left[H^{uc}(x_0+n^{-1/3}k) - H^{uc}(x_0)\right] = \mathcal{O}(n^{-1}),$$

according to (2.5.12) and Lemma 2.7, so that by arguments similar as in the proof of Lemma 2.12,

$$\sup_{0\leqslant x\leqslant k}|B_{n2}(x)| = \mathcal{O}_p(n^{-1/3}). \tag{2.5.18}$$

For the term B_{n3}, similar to the treatment of the right hand side of (2.4.13), it follows that

$$|B_{n3}(x)| \leqslant n^{2/3}\mathcal{O}_p(n^{-1/2})\left|H^{uc}(x_0+n^{-1/3}x) - H^{uc}(x_0)\right| = \mathcal{O}_p(n^{-1/6}x), \tag{2.5.19}$$

by condition (2.5.12). Next, we combine $B_{n4}(x)$ with $A_{n3}(x)$. First write

$$\begin{aligned} A_{n3}(x) &= n^{2/3}\lambda_0(x_0)\int_{x_0}^{x_0+n^{-1/3}x}\left[\Phi_n(\beta_0,s)-\Phi(\beta_0,s)\right]\,\mathrm{d}s \\ &\qquad + n^{2/3}\lambda_0(x_0)\int_{x_0}^{x_0+n^{-1/3}x}\left[\Phi(\beta_0,s)-\Phi(\beta_0,x_0)\right]\mathrm{d}s \\ &= C_{n1}(x)+C_{n2}(x). \end{aligned}$$

As for C_{n1},

$$|C_{n1}(x)| \leqslant n^{1/3}x\lambda_0(x_0)\sup_{x\in\mathbb{R}}\left|\Phi_n(\beta_0,x)-\Phi(\beta_0,x)\right| = \mathcal{O}_p(n^{-1/6}x), \tag{2.5.20}$$

according to Lemma 2.8. Finally, using (2.2.9) and (2.3.1),

$$\begin{aligned} B_{n4}(x)+C_{n2}(x) &= n^{2/3}\int_{x_0}^{x_0+n^{-1/3}x}\left[\Phi(\beta_0,x_0)-\Phi(\beta_0,u)\right]\lambda_0(u)\,\mathrm{d}u \\ &\qquad + n^{2/3}\lambda_0(x_0)\int_{x_0}^{x_0+n^{-1/3}x}\left[\Phi(\beta_0,s)-\Phi(\beta_0,x_0)\right]\,\mathrm{d}s \\ &= n^{2/3}\int_{x_0}^{x_0+n^{-1/3}x}\left[\Phi(\beta_0,s)-\Phi(\beta_0,x_0)\right]\left[\lambda_0(s)-\lambda_0(x_0)\right]\,\mathrm{d}s \\ &= \mathcal{O}_p(n^{-1/3}x), \end{aligned} \tag{2.5.21}$$

by conditions (2.5.15) and (2.5.14). We conclude that

$$\Phi(\beta_0,x_0)\left|\widetilde{\mathbb{Z}}_n^\lambda(x)-\widehat{\mathbb{Z}}_n^\lambda(x)\right| = \mathcal{O}_p(n^{-1/2}x^{1/2})+\mathcal{O}_p(n^{-1/6}x)+\mathcal{O}_p(n^{-1/3}), \tag{2.5.22}$$

and after taking the supremum over $[0,k]$, the lemma follows. □

To find the limit process of $\widehat{\mathbb{Z}}_n^\lambda$, we will apply results from KIM & POLLARD (1990). The limit distribution for $\widetilde{\mathbb{Z}}_n^\lambda$ will then follow directly from Lemma 2.13. Let $\mathbf{B}_{loc}(\mathbb{R})$ be the space of all locally bounded real functions on $\mathbb{R}$, equipped with the topology of uniform convergence on compact domains.

LEMMA 2.14. *Assume (A1) and (A2) and let* $0 < x_0 < \tau_H$. *Suppose that* (2.5.12), (2.5.14) *and* (2.5.15) *hold. Then the processes* $\widehat{\mathbb{Z}}_n^\lambda$ *and* $\widetilde{\mathbb{Z}}_n^\lambda$ *defined in* (2.5.9) *and* (2.5.7) *converge in distribution to the process*

$$\mathbb{Z}(x) = \mathbb{W}\left(\frac{\lambda_0(x_0)}{\Phi(\beta_0,x_0)}x\right)+\frac{1}{2}\lambda_0'(x_0)x^2, \tag{2.5.23}$$

in $\boldsymbol{B}_{loc}(\mathbb{R})$, *where* $\mathbb{W}$ *is standard two-sided Brownian motion originating from zero.*

PROOF. We will apply Theorem 4.7 in KIM & POLLARD (1990). To this end, write the process $\widehat{\mathbb{Z}}_n^\lambda$ in (2.5.9) as

$$\widehat{\mathbb{Z}}_n^\lambda(x) = -n^{2/3}P_n g(\cdot, n^{-1/3}x) + n^{2/3}R_n(x), \tag{2.5.24}$$

for $x \in [-n^{1/3}(x_0 - T_{(1)}), n^{1/3}(T_{(n)} - x_0)]$, where for $Y = (T, \Delta, Z)$ and $\vartheta \in [-x_0, \tau_H - x_0]$,

$$\begin{aligned} g(Y,\vartheta) &= -g_1(Y,\vartheta) + g_2(Y,\vartheta), \\ g_1(Y,\vartheta) &= \left(\{T < x_0 + \vartheta\} - \{T < x_0\}\right) \frac{\Delta}{\Phi(\beta_0, x_0)} \\ g_2(Y,\vartheta) &= \frac{\lambda_0(x_0)e^{\beta_0' Z}}{\Phi(\beta_0, x_0)} \int_{x_0}^{x_0+\vartheta} \{T \geqslant s\}\, ds. \end{aligned} \tag{2.5.25}$$

Furthermore,

$$R_n(x) = \frac{-\lambda_0(x_0)}{\Phi(\beta_0, x_0)} \Bigg[\left(\widehat{W}_n(x_0 + n^{-1/3}x) - W_{n0}(x_0 + n^{-1/3}x)\right) - \left(\widehat{W}_n(x_0) - W_{n0}(x_0)\right) \Bigg],$$

where $W_{n0}(x) = W_n(\beta_0, x)$, with W_n defined in (2.2.3). For all $k = 1, 2, \ldots$, consider

$$|R_n(x)| \leqslant \frac{\lambda_0(x_0)}{\Phi(\beta_0, x_0)} \int \left|\{s \leqslant x_0 + n^{-1/3}x\} - \{s \leqslant x_0\}\right| \left|\Phi_n(\hat\beta_n, s) - \Phi_n(\beta_0, s)\right| ds,$$

which by similar reasoning as in (2.4.12) gives that

$$|R_n(x)| = \mathcal{O}_p(n^{-5/6}x), \tag{2.5.26}$$

by Lemma 2.7. Hence, the process $x \mapsto n^{2/3}R_n(x)$ tends to zero in $\mathbf{B}_{loc}(\mathbb{R})$. It is sufficient then to demonstrate that $-n^{2/3}P_n g(\cdot, n^{-1/3}x)$ converges to $\mathbb{Z}(x)$ in $\mathbf{B}_{loc}(\mathbb{R})$. To this end, we will show that the conditions of Lemma 4.5 and 4.6 in KIM & POLLARD (1990) hold. Condition (i) of Lemma 4.5 is trivially fulfilled, since $\vartheta_0 = 0$ is an interior point of $[-x_0, \tau_H - x_0]$. Moreover, observe that for all $\vartheta \in [-x_0, \tau_H - x_0]$, from (2.2.9) and (2.3.1), we have

$$Pg(\cdot, \vartheta) = \frac{-1}{\Phi(\beta_0, x_0)} \int_{x_0}^{x_0+\vartheta} \left[\lambda_0(u) - \lambda(x_0)\right] \Phi(\beta_0, u)\, du. \tag{2.5.27}$$

Thus, by (2.5.15) and (2.5.14),

$$\begin{aligned} \frac{\partial Pg(\cdot, \vartheta)}{\partial \vartheta} &= -\frac{\Phi(\beta_0, x_0 + \vartheta)}{\Phi(\beta_0, x_0)} \left\{\lambda_0(x_0 + \vartheta) - \lambda_0(x_0)\right\} \\ \frac{\partial^2 Pg(\cdot, \vartheta)}{\partial \vartheta^2} &= -\left(\frac{\partial \Phi(\beta_0, x_0 + \vartheta)}{\partial \vartheta}\right) \frac{\lambda_0(x_0 + \vartheta) - \lambda_0(x_0)}{\Phi(\beta_0, x_0)} - \frac{\Phi(\beta_0, x_0 + \vartheta)}{\Phi(\beta_0, x_0)} \lambda_0'(x_0 + \vartheta). \end{aligned}$$

It follows that $Pg(\cdot,\vartheta)$ is twice differentiable at $\vartheta_0 = 0$, its unique maximizing value, with second derivative $-\lambda_0'(x_0) < 0$, which establishes condition (iii) of Lemma 4.5 in KIM & POLLARD (1990). Next, compute

$$H(s,t) = \lim_{\alpha\to\infty} \alpha Pg(\cdot,s/\alpha)g(\cdot,t/\alpha),$$

for finite s and t. Write

$$\alpha Pg(\cdot,s/\alpha)g(\cdot,t/\alpha) = \alpha P\Big(- g_1(\cdot,s/\alpha) + g_2(\cdot,s/\alpha)\Big)\Big(- g_1(\cdot,t/\alpha) + g_2(\cdot,t/\alpha)\Big)$$

and compute the four terms separately. For all s and t,

$$\begin{aligned}\alpha P\big|g_1(\cdot,s/\alpha)g_2(\cdot,t/\alpha)\big| \\ \leqslant \frac{\lambda_0(x_0)t}{\Phi^2(\beta_0,x_0)}\mathbb{E}\Big[|\{T < x_0 + s/\alpha\} - \{T < x_0\}|\mathrm{e}^{\beta_0' Z}\Big] \to 0,\end{aligned} \tag{2.5.28}$$

as $\alpha \to \infty$. Completely analogous, it follows that

$$\lim_{\alpha\to\infty} \alpha Pg_2(\cdot,s/\alpha)g_2(\cdot,t/\alpha) = 0, \tag{2.5.29}$$

for all s and t. Finally, consider the limit for $\alpha Pg_1(\cdot,s/\alpha)g_1(\cdot,t/\alpha)$. For $s,t \geqslant 0$,

$$\begin{aligned}\alpha Pg_1(\cdot,s/\alpha)g_1(\cdot,t/\alpha) &= \frac{\alpha}{\Phi^2(\beta_0,x_0)}\int \delta\{x_0 \leqslant u < x_0 + (s\wedge t)/\alpha\}\,\mathrm{d}P(u,\delta,z) \\ &= \frac{\alpha}{\Phi^2(\beta_0,x_0)}\int_{x_0}^{x_0+(s\wedge t)/\alpha} \lambda_0(u)\Phi(\beta_0,u)\,\mathrm{d}u \\ &= \frac{1}{\Phi^2(\beta_0,x_0)}\int_0^{s\wedge t} \lambda_0(x_0+v/\alpha)\Phi(\beta_0,x_0+v/\alpha)\,\mathrm{d}v,\end{aligned}$$

by (2.2.9) and (2.3.1). Therefore, by the continuity of λ_0 and Φ,

$$\lim_{\alpha\to\infty} \alpha Pg_1(\cdot,s/\alpha)g_1(\cdot,t/\alpha) = \frac{\lambda_0(x_0)}{\Phi(\beta_0,x_0)}(s\wedge t). \tag{2.5.30}$$

A similar reasoning applies for $s,t < 0$ and $Pg_1(\cdot,s/\alpha)g_1(\cdot,t/\alpha) = 0$, when s and t have opposite signs. Hence, condition (ii) of Lemma 4.5 in KIM & POLLARD (1990) is verified, with

$$H(s,t) = \frac{\lambda_0(x_0)}{\Phi(\beta_0,x_0)}(|s|\wedge|t|),$$

for $st \geqslant 0$ and $H(s,t) = 0$, for $st < 0$. Note that $H(s,t)$ is the covariance kernel of the centered Gaussian process in (2.5.23). For condition (iv) of Lemma 4.5 in KIM & POLLARD (1990), it needs to be shown that for each t and $\varepsilon > 0$

$$\lim_{\alpha\to\infty} \alpha Pg(\cdot,t/\alpha)^2\{|g(\cdot,t/\alpha)| > \alpha\varepsilon\} = 0. \tag{2.5.31}$$

In view of (2.5.28) and (2.5.29), it suffices to show that

$$\lim_{\alpha\to\infty} \alpha P g_1(\cdot,t/\alpha)^2\{|g(\cdot,t/\alpha)|>\alpha\varepsilon\}=0.$$

Moreover, since g_1 is bounded uniformly for $\vartheta\in[-x_0,\tau_H-x_0]$, by Lemma 2.7,

$$\{|g(\cdot,t/\alpha)|>\alpha\varepsilon\}\leqslant\{|g_2(\cdot,t/\alpha)|>\alpha\varepsilon/2\}\leqslant\frac{2}{\alpha\varepsilon}|g_2(\cdot,t/\alpha)|,$$

for α sufficiently large. By (2.5.28), it follows that

$$\begin{aligned}\alpha P g_1(\cdot,t/\alpha)^2\{|g(\cdot,t/\alpha)|>\alpha\varepsilon\}&\leqslant\frac{2}{\varepsilon}P g_1(\cdot,t/\alpha)^2\left|g_2(\cdot,t/\alpha)\right|\\&\leqslant\frac{2}{\varepsilon\Phi(\beta_0,M)}P\left|g_1(\cdot,t/\alpha)g_2(\cdot,t/\alpha)\right|\to0.\end{aligned}$$

Hence, all conditions of Lemma 4.5 in KIM & POLLARD (1990) are satisfied.

To continue with verifying the conditions of Lemma 4.6 in KIM & POLLARD (1990), consider the class of functions $\mathscr{G}=\{g(\cdot,\vartheta):\vartheta\in[-x_0,\tau_H-x_0]\}$ and the classes

$$\mathscr{G}_R=\left\{g(\cdot,\vartheta)\in\mathscr{G}:|\vartheta|\leqslant R\right\},\tag{2.5.32}$$

for any $R>0$, R in a neighborhood of zero. Since the functions in $\mathscr{G}_R$ are the difference of $g_1(\cdot,\vartheta)$, which is an the product of indicators, and $g_2(\cdot,\vartheta)$, which is the product of a fixed function and a linear function, it follows that $\mathscr{G}_R$ is a VC-subgraph class of functions, and hence it is uniformly manageable, which proves condition (i) of Lemma 4.6 in KIM & POLLARD (1990). Furthermore, choose as an envelope for $\mathscr{G}_R$,

$$G_R=G_{R1}+G_{R2},\tag{2.5.33}$$

where

$$\begin{aligned}G_{R1}(T,\Delta,Z)&=\frac{\{x_0-R\leqslant T<x_0+R\}}{\Phi(\beta_0,x_0)},\\G_{R2}(T,\Delta,Z)&=\frac{2R\lambda_0(x_0)}{\Phi(\beta_0,x_0)}\mathrm{e}^{\beta_0'Z}.\end{aligned}\tag{2.5.34}$$

Calculations completely analogous to (2.5.28), (2.5.29) and (2.5.30), with $1/R$ playing the role of $\alpha\to\infty$, yield that $PG_R^2=\mathscr{O}(R)$, as $R\to0$. This proves condition (ii) of Lemma 4.6 in KIM & POLLARD (1990). To show condition (iii) of Lemma 4.6 in KIM & POLLARD (1990), first note that

$$P|g(\cdot,\vartheta_1)-g(\cdot,\vartheta_2)|\leqslant P|g_1(\cdot,\vartheta_1)-g_1(\cdot,\vartheta_2)|+P|g_2(\cdot,\vartheta_1)-g_2(\cdot,\vartheta_2)|.$$

Now,

$$P|g_1(\cdot,\vartheta_1)-g_1(\cdot,\vartheta_2)|=\frac{1}{\Phi(\beta_0,x_0)}\left|H^{uc}(x_0+\vartheta_1)-H^{uc}(x_0+\vartheta_2)\right|=\mathscr{O}(|\vartheta_1-\vartheta_2|),$$

according to (2.5.12). Analogously,

$$P|g_2(\cdot,\vartheta_1) - g_2(\cdot,\vartheta_2)| \leqslant \frac{\lambda_0(x_0)}{\Phi(\beta_0,x_0)}|\vartheta_1 - \vartheta_2|\mathbb{E}\left[e^{\beta_0' Z}\right] = \mathcal{O}(|\vartheta_1 - \vartheta_2|),$$

by (A2), which proves condition (iii) of Lemma 4.6 in KIM & POLLARD (1990). Finally, to establish condition (iv) of Lemma 4.6 in KIM & POLLARD (1990), we have to show that for each $\varepsilon > 0$, there exists $K > 0$ such that

$$PG_R^2\{G_R > K\} < \varepsilon R,$$

for R near zero. The proof of this is completely analogous to proving (2.5.31), with $1/R$ playing the role $\alpha \to \infty$. This shows that all conditions of Theorem 4.7 in KIM & POLLARD (1990) are fulfilled, from which we conclude that the process $-n^{2/3}P_n g(\cdot, n^{-1/3}x)$ converges in distribution to the process

$$-\mathbb{W}\left(\frac{\lambda_0(x_0)}{\Phi(\beta_0,x_0)}x\right) + \frac{1}{2}\lambda_0'(x_0)x^2 \stackrel{d}{=} \mathbb{W}\left(\frac{\lambda_0(x_0)}{\Phi(\beta_0,x_0)}x\right) + \frac{1}{2}\lambda_0'(x_0)x^2.$$

Together with (2.5.24) and (2.5.26), this proves the weak convergence of $\widehat{\mathbb{Z}}_n^\lambda$. Weak convergence of $\widetilde{\mathbb{Z}}_n^\lambda$ is then immediate, by Lemma 2.13. □

As a consequence, we obtain the limiting distribution of the process in (2.5.8).

LEMMA 2.15. *Assume (A1) and (A2) and suppose that* (2.5.12), (2.5.14) *and* (2.5.15) *hold. Let* $0 < x_0 < \tau_H$ *and* $a > 0$ *fixed and let* $\widehat{\mathbb{Z}}_n^\lambda$ *and* $\widehat{W}_n$ *be defined in* (2.5.9) *and* (2.4.14). *Then, the process*

$$\widehat{\mathbb{Z}}_n^\lambda(x) - \frac{n^{1/3}a}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0 + n^{-1/3}x) - \widehat{W}_n(x_0)\right]$$

converges weakly to

$$\mathbb{Z}(x) - ax = \mathbb{W}\left(\frac{\lambda_0(x_0)}{\Phi(\beta_0,x_0)}x\right) + \frac{1}{2}\lambda_0'(x_0)x^2 - ax,$$

in $\boldsymbol{B}_{loc}(\mathbb{R})$*, where* $\mathbb{W}$ *is standard two-sided Brownian motion originating from zero.*

PROOF. In view of Lemma 2.14, it suffices to show that for any $k = 1, 2, \ldots,$

$$\sup_{|x| \leqslant k}\left|n^{1/3}\left[\widehat{W}_n(x_0 + n^{-1/3}x) - \widehat{W}_n(x_0)\right] - \Phi(\beta_0,x_0)x\right| \to 0, \qquad (2.5.35)$$

almost surely. This is immediate, since similar to (2.4.20), together with the monotonicity of $\Phi(\beta_0, u)$, one has, for $x \geqslant 0$,

$$\begin{aligned}
&\left| n^{1/3} \left[\widehat{W}_n(x_0 + n^{-1/3}x) - \widehat{W}_n(x_0) \right] - \Phi(\beta_0, x_0)x \right| \\
&\quad \leqslant n^{1/3} \int_{x_0}^{x_0+n^{-1/3}x} \left| \Phi_n(\hat{\beta}_n, u) - \Phi(\beta_0, x_0) \right| \mathrm{d}u \\
&\quad \leqslant |x| \sup_{u\in\mathbb{R}} \left| \Phi_n(\hat{\beta}_n, u) - \Phi(\beta_0, u) \right| + |\Phi(\beta_0, x_0 + n^{-1/3}x) - \Phi(\beta_0, x_0)| \\
&\quad = o(x) + \mathcal{O}(n^{-1/3}x),
\end{aligned} \tag{2.5.36}$$

almost surely, using Lemma 2.8 and (2.5.15). The case $x < 0$ can be treated likewise. □

Finally, the next lemma provides the limit process of $\widetilde{\mathbb{Z}}_n^f$.

LEMMA 2.16. *Assume (A1) and (A2). Let* $x_0 \in (0, \tau_H)$ *and suppose that* (2.5.12), (2.5.14) *and* (2.5.15) *hold. Then the process* $\widetilde{\mathbb{Z}}_n^f$ *defined in* (2.5.11) *converges in distribution to the process*

$$\mathbb{Z}^f(x) = \mathbb{W}\left(\frac{f_0(x_0)(1 - F_0(x_0))}{\Phi(\beta_0, x_0)} x \right) + \frac{1}{2} f_0'(x_0) x^2. \tag{2.5.37}$$

in $\boldsymbol{B}_{loc}(\mathbb{R})$*, where* $\mathbb{W}$ *is standard two-sided Brownian motion originating from zero.*

PROOF. From (2.5.7), we have $\Lambda_n(x_0+n^{-1/3}x)-\Lambda_n(x_0) = n^{-2/3}\widetilde{\mathbb{Z}}_n^\lambda(x)+n^{-1/3}\lambda_0(x_0)x$, so that by (2.2.13),

$$\begin{aligned}
\widetilde{\mathbb{Z}}_n^f(x) &= n^{2/3} \left[-\mathrm{e}^{-\Lambda_n(x_0+n^{-1/3}x)} + \mathrm{e}^{-\Lambda_n(x_0)} - n^{-1/3} f_0(x_0)x \right] \\
&= n^{2/3} \left[-\mathrm{e}^{-\Lambda_n(x_0)} \left(\mathrm{e}^{-n^{-2/3}\widetilde{\mathbb{Z}}_n^\lambda(x) - n^{-1/3}\lambda_0(x_0)x} - 1 \right) - n^{-1/3} f_0(x_0)x \right].
\end{aligned} \tag{2.5.38}$$

Because $\mathrm{e}^{-y} - 1 = -y + y^2/2 + o(y^2)$, for $y \to 0$ and $\sup_{x\in\mathbb{R}} |\widetilde{\mathbb{Z}}_n^\lambda(x)| = \mathcal{O}_p(1)$, according to Lemma 2.14, it follows that

$$\begin{aligned}
\mathrm{e}^{-n^{-2/3}\widetilde{\mathbb{Z}}_n^\lambda(x) - n^{-1/3}\lambda_0(x_0)x} - 1 = &- n^{-2/3}\widetilde{\mathbb{Z}}_n^\lambda(x) - n^{-1/3}\lambda_0(x_0)x + \frac{1}{2} n^{-2/3}\lambda_0(x_0)^2 x^2 \\
&+ \mathcal{O}_p(n^{-4/3}) + \mathcal{O}_p(n^{-1}x) + o_p(n^{-2/3}x^2).
\end{aligned}$$

Similarly, from Theorem 2.9, we have that $\mathrm{e}^{-\Lambda_n(x_0)} = \mathrm{e}^{-\Lambda_0(x_0)} + \mathcal{O}_p(n^{-1/2})$. Since

$$\mathrm{e}^{-\Lambda_0(x_0)}\lambda_0(x_0) = [1 - F_0(x_0)]\lambda_0(x_0) = f_0(x_0),$$

from (2.5.38), we find that

$$\widetilde{\mathbb{Z}}_n^f(x) = [1-F_0(x_0)]\widetilde{\mathbb{Z}}_n^\lambda(x) - \frac{1}{2}[1-F_0(x_0)]\lambda_0(x_0)^2x^2 \\ + \mathcal{O}_p(n^{-1/2}) + \mathcal{O}_p(n^{-1/6}x) + o_p(x^2). \tag{2.5.39}$$

According to Lemma 2.14, the process $[1-F_0(x_0)]\widetilde{\mathbb{Z}}_n^\lambda(x) - \frac{1}{2}[1-F_0(x_0)]\lambda_0(x_0)^2x^2$ converges weakly to

$$[1-F_0(x_0)]\mathbb{W}\left(\frac{\lambda_0(x_0)}{\Phi(\beta_0,x_0)}x\right) + \frac{1}{2}[1-F_0(x_0)]\lambda_0'(x_0)x^2 - \frac{1}{2}[1-F_0(x_0)]\lambda_0^2(x_0)x^2,$$

which has the same distribution as the process in (2.5.37), by means of Brownian scaling and the fact that

$$\lambda_0' = \left(\frac{f_0}{1-F_0}\right)' = \frac{(1-F_0)f_0' + f_0^2}{(1-F_0)^2} = \frac{f_0'}{1-F_0} + \lambda_0^2. \tag{2.5.40}$$

Hence, for any $k = 1, 2, \ldots$, it follows from (2.5.39) that

$$\sup_{|x|\leqslant k} |\widetilde{\mathbb{Z}}_n^f(x) - \mathbb{Z}^f(x)| = o_p(1),$$

which finishes the proof. □

2.6 LIMIT DISTRIBUTION

The last step in deriving the asymptotic distribution of the estimators is to find the limiting distribution of the inverse processes $\widetilde{U}_n^\lambda$, $\widehat{U}_n^\lambda$ and $\widetilde{U}_n^f$ defined in (2.5.3), (2.5.1) and (2.5.4) and of the versions of $\widetilde{U}_n^\lambda$ and $\widehat{U}_n^\lambda$ in the case of a nonincreasing hazard, by applying Theorem 2.7 in KIM & POLLARD (1990). This requires the inverse processes to be bounded in probability.

LEMMA 2.17. *Assume (A1) and (A2) and let $x_0 \in (0, \tau_H)$. Suppose that λ_0 is monotone and suppose that f_0 is nondecreasing. Suppose that* (2.5.14) *and* (2.5.15) *hold, with $\lambda_0(x_0) \neq 0$. Then, for each $\varepsilon > 0$ and $M_1 > 0$, there exists $M_2 > 0$ such that,*

$$\mathbb{P}\left(\max_{|a|\leqslant M_1} n^{1/3}\left|\widehat{U}_n^\lambda(\lambda_0(x_0)+n^{-1/3}a) - x_0\right| > M_2\right) < \varepsilon \tag{2.6.1}$$

$$\mathbb{P}\left(\max_{|a|\leqslant M_1} n^{1/3}\left|\widetilde{U}_n^\lambda(\lambda_0(x_0)+n^{-1/3}a) - x_0\right| > M_2\right) < \varepsilon \tag{2.6.2}$$

$$\mathbb{P}\left(\max_{|a|\leqslant M_1} n^{1/3}\left|\widetilde{U}_n^f(f_0(x_0)+n^{-1/3}a) - x_0\right| > M_2\right) < \varepsilon, \tag{2.6.3}$$

for n sufficiently large.

PROOF. The proof of the lemma follows closely the lines of proof of Lemma 5.3 in GROENEBOOM & WELLNER (1992) (see also Lemma 7.1 in HUANG & WELLNER, 1995). First consider (2.6.1) in case λ_0 is nondecreasing. It will be shown that

$$\mathbb{P}\left(\max_{|a|\leqslant M_1} n^{1/3}\left[\hat{U}_n^\lambda(\lambda_0(x_0)+n^{-1/3}a)-x_0\right]>M_2\right)<\varepsilon, \tag{2.6.4}$$

as the other part can be proved similarly. Because $\hat{U}_n^\lambda(a)$ is nondecreasing, the probability in (2.6.4) is equal to

$$\mathbb{P}\left(n^{1/3}\left[\hat{U}_n(\lambda_0(x_0)+n^{-1/3}M_1)-x_0\right]>M_2\right).$$

The relationship between the inverse process $\hat{U}_n^\lambda$ and the process $\hat{\mathbb{Z}}_n^\lambda$ defined in (2.5.9), together with the fact that $\hat{\mathbb{Z}}_n^\lambda(0)=0$, implies that

$$\begin{aligned}&\mathbb{P}\left(n^{1/3}\left[\hat{U}_n^\lambda(\lambda_0(x_0)+n^{-1/3}M_1)-x_0\right]>M_2\right)\\ &\leqslant \mathbb{P}\left(\hat{\mathbb{Z}}_n^\lambda(x)-\frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\hat{W}_n(x_0+n^{-1/3}x)-\hat{W}_n(x_0)\right]\leqslant 0,\ \text{for some } x\geqslant M_2\right).\end{aligned} \tag{2.6.5}$$

By condition (2.5.14), there exists $M_0>0$ such that, for any $x\in[T_{(1)},T_{(n)}]$ with $|x-x_0|\leqslant M_0$, $\lambda_0'(x)>0$ and $\lambda_0'(x)$ is close to $\lambda_0'(x_0)$. Take $n^{-1/3}x\leqslant M_0$. From (2.5.24) and (2.5.35),

$$\begin{aligned}&\hat{\mathbb{Z}}_n^\lambda(x)-\frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\hat{W}_n(x_0+n^{-1/3}x)-\hat{W}_n(x_0)\right]\\ &=-n^{2/3}P_n g(\cdot,n^{-1/3}x)-M_1x+\hat{R}_n(x),\end{aligned} \tag{2.6.6}$$

where $\hat{R}_n(x)=\mathcal{O}_p(n^{-1/6}x)+o(x)+\mathcal{O}(n^{-1/3}x)$, by (2.5.26) and (2.5.36). Furthermore, for $0<R\leqslant M_0$, consider the class of functions $\mathcal{G}_R$ defined in (2.5.32) along with its envelope defined in (2.5.33). It has been determined in the proof of Lemma 2.14 that $\mathcal{G}_R$ is uniformly manageable for its envelope G_R and that $PG_R^2=\mathcal{O}(R)$, for $0<R\leqslant M_0$. Thus, Lemma 4.1 in KIM & POLLARD (1990) states that for each $\delta>0$, there exist random variables $S_n=\mathcal{O}_p(1)$ such that

$$|P_n g(\cdot,n^{-1/3}x)-Pg(\cdot,n^{-1/3}x)|\leqslant \delta n^{-2/3}x^2+n^{-2/3}S_n^2, \tag{2.6.7}$$

for $n^{-1/3}x\leqslant M_0$. Choose $\delta=\lambda_0'(x_0)/8$ in the above inequality. It will result that

$$-n^{2/3}(P_n-P)g(\cdot,n^{-1/3}x)\geqslant -\frac{1}{8}\lambda_0'(x_0)x^2-S_n^2.$$

Furthermore, by (2.5.14), (2.5.15) and (2.5.27),

$$
\begin{aligned}
-n^{2/3}Pg(\cdot,n^{-1/3}x) = \frac{x^2}{2\Phi(\beta_0,x_0)} \Big\{ & \lambda_0'(x_0+\vartheta_n)\Phi(\beta_0,x_0+\vartheta_n) \\
& + \left[\lambda_0(x_0+\vartheta_n)-\lambda_0(x_0)\right]\Phi'(\beta_0,x_0+\vartheta_n)\Big\}
\end{aligned}
\tag{2.6.8}
$$

for $|\vartheta_n| \leqslant n^{-1/3}x \leqslant M_0$, where $\Phi'(\beta_0,x) = \partial\Phi(\beta_0,x)/\partial x$. From the choice of M_0 and since $\lambda_0'(x_0) > 0$, we can find $K > 0$ such that for any $x > K$,

$$
-n^{2/3}Pg(\cdot,n^{-1/3}x) - M_1 x \geqslant \frac{1}{4}\lambda_0'(x_0)x^2,
$$

for n sufficiently large. We conclude that

$$
\begin{aligned}
&\widehat{\mathbb{Z}}_n^{\lambda}(x) - \frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+n^{-1/3}x)-\widehat{W}_n(x_0)\right] \\
&\quad = -n^{2/3}P_n g(\cdot,n^{-1/3}x) - M_1 x + \widehat{R}_n(t) \\
&\quad = -n^{2/3}(P_n-P)g(\cdot,n^{-1/3}x) - n^{2/3}Pg(\cdot,n^{-1/3}x) - M_1 x + \widehat{R}_n(x) \\
&\quad \geqslant \frac{1}{8}\lambda_0'(x_0)x^2 - S_n^2 + \widehat{R}_n(x),
\end{aligned}
$$

where $\widehat{R}_n(x) = \mathcal{O}_p(n^{-1/6}x) + o(x) + \mathcal{O}(n^{-1/3}x)$ and the $\mathcal{O}_p$, $\mathcal{O}$ and o terms do not depend on x. It follows that for $x \geqslant M_2 > K$,

$$
\widehat{\mathbb{Z}}_n^{\lambda}(x) - \frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+n^{-1/3}x)-\widehat{W}_n(x_0)\right] \geqslant \frac{1}{8}\lambda_0'(x_0)x^2 - S_n^2 + o_P(1), \tag{2.6.9}
$$

where the o_P term does not depend on x. Then, M_2 can be chosen such that

$$
\mathbb{P}\left(S_n^2 \geqslant \frac{1}{8}\lambda_0'(x_0)M_2^2 + o_P(1)\right) < \varepsilon,
$$

for n sufficiently large. We find that

$$
\begin{aligned}
&\mathbb{P}\left(\widehat{\mathbb{Z}}_n^{\lambda}(x) - \frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+n^{-1/3}x)-\widehat{W}_n(x_0)\right] \leqslant 0, \text{ for some } M_2 \leqslant x \leqslant n^{1/3}M_0\right) \\
&\quad \leqslant \mathbb{P}\left(\frac{1}{8}\lambda_0'(x_0)x^2 - S_n^2 + o_P(1) \leqslant 0, \text{ for some } M_2 \leqslant x \leqslant n^{1/3}M_0\right) \\
&\quad \leqslant \mathbb{P}\left(S_n^2 \geqslant \frac{1}{8}\lambda_0'(x_0)x^2 + o_P(1), \text{ for some } M_2 \leqslant x \leqslant n^{1/3}M_0\right) \leqslant \varepsilon,
\end{aligned}
$$

for n sufficiently large.

For $n^{-1/3}x > M_0$, we first show that

$$\begin{aligned}
&\widehat{\mathbb{Z}}_n^{\lambda}(x) - \frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+n^{-1/3}x) - \widehat{W}_n(x_0)\right] \\
&\qquad \geqslant \widehat{\mathbb{Z}}_n(n^{1/3}M_0/2) - \frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+M_0/2) - \widehat{W}_n(x_0)\right],
\end{aligned} \tag{2.6.10}$$

with large probability, for n sufficiently large. Then,

$$\mathbb{P}\left(\widehat{\mathbb{Z}}_n(n^{1/3}M_0/2) - \frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+M_0/2) - \widehat{W}_n(x_0)\right] \leqslant 0\right)$$

can be bounded with the argument above. Lemma 2.11 and (2.4.16) yield that $\widehat{V}_n(x_0+M_0/2) = V_n(x_0+M_0/2) + o(1)$, with probability one and by definition $V_n(x_0+n^{-1/3}x) \geqslant \widehat{V}_n(x_0+n^{-1/3}x)$, for all $x_0+n^{-1/3}x \in [T_{(1)}, T_{(n)}]$. This implies that

$$\begin{aligned}
&V_n(x_0+n^{-1/3}x) - V_n(x_0+M_0/2) \\
&\qquad \geqslant \widehat{V}_n(x_0+n^{-1/3}x) - \widehat{V}_n(x_0+M_0/2) + o(1), \\
&\qquad \geqslant \hat{\lambda}_n(x_0+M_0/2)\left[\widehat{W}_n(x_0+n^{-1/3}x) - \widehat{W}_n(x_0+M_0/2)\right] + o(1),
\end{aligned} \tag{2.6.11}$$

using the convexity of $\widehat{V}_n$. To show (2.6.10), note that by definition (2.5.9),

$$\begin{aligned}
&\widehat{\mathbb{Z}}_n^{\lambda}(x) - \frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+n^{-1/3}x) - \widehat{W}_n(\hat{\beta}_n,x_0)\right] \\
&\qquad - \left\{\widehat{\mathbb{Z}}_n^{\lambda}(n^{1/3}M_0/2) - \frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\left[\widehat{W}_n(x_0+M_0/2) - \widehat{W}_n(x_0)\right]\right\} \\
&\quad = \frac{n^{2/3}}{\Phi(\beta_0,x_0)}\left\{V_n(x_0+n^{-1/3}x) - V_n(x_0+M_0/2)\right. \\
&\qquad \left. - \left[\lambda_0(x_0) + n^{-1/3}M_1\right]\left[\widehat{W}_n(x_0+n^{-1/3}x) - \widehat{W}_n(x_0+M_0/2)\right]\right\},
\end{aligned}$$

and furthermore,

$$\begin{aligned}
&\frac{n^{2/3}}{\Phi(\beta_0,x_0)}\Big\{V_n(x_0+n^{-1/3}x)-V_n(x_0+M_0/2)\\
&\qquad -\Big[\lambda_0(x_0)+n^{-1/3}M_1\Big]\Big(\widehat{W}_n(x_0+n^{-1/3}x)-\widehat{W}_n(x_0+M_0/2)\Big)\Big\}\\
&\qquad\geqslant\frac{n^{2/3}}{\Phi(\beta_0,x_0)}\Big\{\Big[\hat{\lambda}_n(x_0+M_0/2)-\lambda_0(x_0)-n^{-1/3}M_1\Big]\\
&\qquad\qquad\times\Big[\widehat{W}_n(x_0+n^{-1/3}x)-\widehat{W}_n(x_0+M_0/2)\Big]+o(1)\Big\}\\
&\qquad=\frac{n^{2/3}}{\Phi(\beta_0,x_0)}\Big\{\Big[\lambda_0(x_0+M_0/2)-\lambda_0(x_0)-n^{-1/3}M_1+o(1)\Big]\\
&\qquad\qquad\times\Big[W_0(x_0+n^{-1/3}x)-W_0(x_0+M_0/2)+o(1)\Big]+o(1)\Big\}>0,
\end{aligned}$$

for n sufficiently large, using (2.4.20) and the fact that λ_0 and W_0 are strictly increasing and $n^{-1/3}x>M_0$. It follows that

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\mathbb{Z}}_n^\lambda(x)-\frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\Big[\widehat{W}_n(x_0+n^{-1/3}x)-\widehat{W}_n(x_0)\Big]\leqslant 0,\ \text{for some } x>n^{1/3}M_0\right)\\
&\quad\leqslant\mathbb{P}\left(\widehat{\mathbb{Z}}_n^\lambda(n^{1/3}M_0/2)-\frac{n^{1/3}M_1}{\Phi(\beta_0,x_0)}\Big[\widehat{W}_n(x_0+M_0/2)-\widehat{W}_n(x_0)\Big]\leqslant 0\right)\leqslant\varepsilon.
\end{aligned}$$

This completes the proof of (2.6.4). The other part of (2.6.1) for a nondecreasing λ_0 is proven similarly.

For (2.6.2), in case of a nondecreasing λ_0, by the same reasoning that leads to (2.6.5) we first have

$$\begin{aligned}
&\mathbb{P}\Big(n^{1/3}\Big[\widetilde{U}_n^\lambda(\lambda_0(x_0)+n^{-1/3}M_1)-x_0\Big]>M_2\Big)\\
&\qquad\qquad\leqslant\mathbb{P}\Big(\widetilde{\mathbb{Z}}_n^\lambda(x)-M_1x\leqslant 0,\ \text{for some } x\geqslant M_2\Big).
\end{aligned}$$

Moreover, by (2.5.22),

$$\widetilde{\mathbb{Z}}_n^\lambda(x)=\widehat{\mathbb{Z}}_n^\lambda(x)+\mathcal{O}_p(n^{-1/2}x^{1/2})+\mathcal{O}_p(n^{-1/6}x)+\mathcal{O}_p(n^{-1/3}),$$

where the $\mathcal{O}_p$ terms do not depend on x. Similar to (2.6.9), one obtains

$$\widetilde{\mathbb{Z}}_n^\lambda(x)-M_1x\geqslant\frac{1}{8}\lambda_0'(x_0)x^2-S_n^2+o_p(1),$$

for $M_2\leqslant x\leqslant n^{1/3}M_0$, where the o_p-term does not depend on x, which yields

$$\mathbb{P}\Big(\widetilde{\mathbb{Z}}_n^\lambda(x)-M_1x\leqslant 0,\ \text{for some } M_2\leqslant x\leqslant n^{1/3}M_0\Big)\leqslant\varepsilon.$$

In the case $x > n^{1/3}M_0$, similar to (2.6.11), Theorem 2.9 and (2.4.25) yield

$$\begin{aligned} \Lambda_n(x_0 + n^{-1/3}x) - \Lambda_n(x_0 + M_0/2) \\ \geqslant \widetilde{\Lambda}_n(x_0 + n^{-1/3}x) - \widetilde{\Lambda}_n(x_0 + M_0/2) + o(1) \\ \geqslant \tilde{\lambda}_n(x_0 + M_0/2)(n^{-1/3}x - M_0/2) + o(1). \end{aligned} \tag{2.6.12}$$

This leads to

$$\widetilde{\mathbb{Z}}_n^{\lambda}(x) - M_1 x \geqslant \widetilde{\mathbb{Z}}_n^{\lambda}(n^{1/3}M_0/2) - M_1 n^{1/3}M_0/2,$$

from which we conclude

$$\mathbb{P}\left(\widetilde{\mathbb{Z}}_n^{\lambda}(x) - M_1 x \leqslant 0, \text{ for some } x > n^{1/3}M_0\right) \leqslant \varepsilon.$$

This completes one part of the proof of (2.6.2) for a nondecreasing λ_0. The other part is shown similarly.

For (2.6.3), using that $\widetilde{U}_n^f$ is nonincreasing, similar to (2.6.5), we first have

$$\begin{aligned} \mathbb{P}\Big(n^{1/3}\Big[\widetilde{U}_n^f(f_0(x_0) + n^{-1/3}M_1) - x_0\Big] > M_2\Big) \\ \leqslant \mathbb{P}\left(\widetilde{\mathbb{Z}}_n^f(x) - M_1 x \geqslant 0, \text{ for some } x \geqslant M_2\right), \end{aligned}$$

Next, according to (2.5.39), (2.5.22) and (2.5.36), we obtain

$$\begin{aligned} \widetilde{\mathbb{Z}}_n^f(x) - M_1 x = &- [1 - F_0(x_0)]n^{2/3}(\mathbb{P}_n - P)g(\cdot, n^{-1/3}x) \\ &- [1 - F_0(x_0)]n^{2/3}Pg(\cdot, n^{-1/3}x) - \frac{1}{2}[1 - F_0(x_0)]\lambda_0(x_0)^2x^2 - M_1 x \\ &+ \mathcal{O}_p(n^{-1/3}) + \mathcal{O}_p(n^{-1/2}x^{1/2}) + o_p(x) + o_p(x^2), \end{aligned}$$

where the $\mathcal{O}_p$ and o_p terms do not depend on x and where $Pg(\cdot, n^{-1/3}x)$ is given in (2.6.8). Now, choose $\delta = -f'(x_0)/(8[1 - F_0(x_0)]) > 0$ in (2.6.7), so that according to Lemma 4.1 in KIM & POLLARD (1990),

$$-[1 - F_0(x_0)]n^{2/3}(\mathbb{P}_n - P)g(\cdot, n^{-1/3}x) \leqslant -\frac{1}{8}f_0'(x_0)x^2 + S_n^2,$$

for $n^{-1/3}x \leqslant M_0$ and $S_n^2 = \mathcal{O}_p(1)$. Furthermore, from (2.6.8) together with (2.5.40), it follows that we can find a $K > 0$ such that for any $x > K$,

$$-[1 - F_0(x_0)]n^{2/3}Pg(\cdot, n^{-1/3}x) - \frac{1}{2}[1 - F_0(x_0)]\lambda_0(x_0)^2x^2 - M_1 x < \frac{1}{4}f_0'(x_0)x^2,$$

for n sufficiently large. Similar to (2.6.9) we have for $x \geqslant M_2 \geqslant K$,

$$\widetilde{\mathbb{Z}}_n^f(x) - M_1 x \leqslant \left[\frac{1}{8}f_0'(x_0) + o_p(1)\right]x^2 + S_n^2 + o_p(1),$$

where the o_p terms do not depend on x, which leads to

$$\mathbb{P}\left(\widetilde{\mathbb{Z}}_n^f(x) - M_1 x \geqslant 0, \text{ for some } M_2 \leqslant x \leqslant n^{1/3} M_0\right) \leqslant \varepsilon,$$

for n sufficiently large. In the case $x > n^{1/3}M_0$, first, similar to (2.4.25), we can obtain that for any $0 < M < \tau_H$,

$$\sup_{x\in[0,M]} \left|\widetilde{F}_n(x) - F_0(x)\right| \leqslant \sup_{x\in[0,M]} \left|F_n(x) - \Lambda_0(x)\right|,$$

which then similar to (2.6.12) together with Corollary 2.10 yields

$$\begin{aligned} F_n(x_0 + n^{-1/3}x) - F_n(x_0 + M_0/2) & \\ &\leqslant \widetilde{F}_n(x_0 + n^{-1/3}x) - \widetilde{F}_n(x_0 + M_0/2) + o(1) \\ &\leqslant \tilde{f}_n(x_0 + M_0/2)(n^{-1/3}x - M_0/2) + o(1). \end{aligned} \tag{2.6.13}$$

This leads to

$$\widetilde{\mathbb{Z}}_n^f(x) - M_1 x \leqslant \widetilde{\mathbb{Z}}_n^f(n^{1/3}M_0/2) - M_1 n^{1/3} M_0/2,$$

from which we conclude

$$\mathbb{P}\left(\widetilde{\mathbb{Z}}_n^\lambda(x) - M_1 x \geqslant 0, \text{ for some } x > n^{1/3} M_0\right) \leqslant \varepsilon.$$

This completes one part of the proof of (2.6.3). The other part is shown similarly.

Finally, the proof of (2.6.1) and (2.6.2) in the case of a nonincreasing λ_0 is along the lines of the proof of (2.6.3), combined with arguments used for the proof of (2.6.1) and (2.6.2) in the nondecreasing case. □

Hereafter, the continuous mapping theorem from KIM & POLLARD (1990) will be applied to the inverse processes in (2.5.1), (2.5.3) and (2.5.4), in order to derive the limiting distribution of the considered estimators. Let $\mathbb{C}_{max}(\mathbb{R})$ denote the subset of $\mathbf{B}_{loc}(\mathbb{R})$ consisting of continuous functions f for which $f(t) \to -\infty$, when $|t| \to \infty$ and f has an unique maximum.

PROOF.[Proof of Theorem 2.4] The aim is to apply Theorem 2.7 in KIM & POLLARD (1990) and Theorem 6.1 in HUANG & WELLNER (1995). Since Theorem 2.7 from KIM & POLLARD (1990) applies to the argmax of processes on the whole real line, we extend the process

$$\hat{Z}_n^\lambda(a,x) = \hat{\mathbb{Z}}_n^\lambda(x) - \frac{n^{1/3}a}{\Phi(\beta_0, x_0)}\left[\hat{W}_n(x_0 + n^{-1/3}x) - \hat{W}_n(x_0)\right]$$

from (2.5.8) for $x \in [n^{1/3}(T_{(1)} - x_0), n^{1/3}(T_{(n)} - x_0)]$, to the whole real line. Define $\hat{Z}_n^\lambda(a,x) = \hat{Z}_n^\lambda(a, n^{1/3}(T_{(1)} - x_0))$, for $x < n^{1/3}(T_{(1)} - x_0)$ and $\hat{Z}_n^\lambda(a,x) =$

$\widehat{Z}_n^\lambda(a, n^{1/3}(T_{(n)} - x_0)) + 1$, for $x > n^{1/3}(T_{(n)} - x_0)$. Then, $\widehat{Z}_n^\lambda(a,x) \in \mathbf{B}_{loc}(\mathbb{R})$ and according to (2.5.8),

$$n^{1/3}\left[\widehat{U}_n^\lambda(\lambda_0(x_0) + n^{-1/3}a) - x_0\right] = \underset{x\in\mathbb{R}}{\operatorname{argmin}}\left\{\widehat{Z}_n^\lambda(a,x)\right\} = \underset{x\in\mathbb{R}}{\operatorname{argmax}}\left\{-\widehat{Z}_n^\lambda(a,x)\right\}.$$

By Lemma 2.14, for any a fixed, the process $-\widehat{Z}_n^\lambda(a,x)$ converges weakly to the process $-\mathbb{Z}(x) + ax \in \mathbb{C}_{max}(\mathbb{R})$ with probability one, where $\mathbb{Z}$ has been defined in (2.5.23). Lemma 2.17 ensures the boundedness in probability of $n^{1/3}\{\widehat{U}_n^\lambda(\lambda_0(x_0) + n^{-1/3}a) - x_0\}$. Consequently, by Theorem 2.7 in KIM & POLLARD (1990), it follows that

$$n^{1/3}\left[\widehat{U}_n^\lambda(\lambda_0(x_0) + n^{-1/3}a) - x_0\right] \xrightarrow{d} \underset{x\in\mathbb{R}}{\operatorname{argmax}}\left\{-\mathbb{Z}(x) + ax\right\} = \underset{x\in\mathbb{R}}{\operatorname{argmin}}\left\{\mathbb{Z}(x) - ax\right\}.$$

The same argument applies to the process $\widetilde{\mathbb{Z}}_n^\lambda(x) - ax$ from (2.5.6), for $x \in [-n^{1/3}x_0, n^{1/3}(T_{(n)} - x_0)]$, which we extend to the whole real line in a similar fashion. Furthermore, if we fix $a, b \in \mathbb{R}$, it will follow that

$$\left(\widehat{Z}_n^\lambda(a,x), \widetilde{\mathbb{Z}}_n^\lambda(x) - bx\right) \xrightarrow{d} \left(\mathbb{Z}(x) - ax, \mathbb{Z}(x) - bx\right),$$

by Lemma 2.15 and Lemma 2.14. Hence, the first condition of Theorem 6.1 in HUANG & WELLNER (1995) is verified. The second condition is provided by Lemma 2.17, whereas the third condition is given by (2.5.6) and (2.5.8). Therefore, by Theorem 6.1 in HUANG & WELLNER (1995),

$$\left(n^{1/3}\left[\widehat{U}_n^\lambda(\lambda_0(x_0) + n^{-1/3}a) - x_0\right], n^{1/3}\left[\widetilde{U}_n^\lambda(\lambda_0(x_0) + n^{-1/3}b) - x_0\right]\right) \xrightarrow{d} \left(U^\lambda(a), U^\lambda(b)\right),$$

where

$$U^\lambda(a) = \sup\left\{t : \mathbb{W}\left(\frac{\lambda_0(x_0)}{\Phi(\beta_0, x_0)}\, t\right) + \frac{1}{2}\lambda_0'(x_0)t^2 - at \text{ is minimal}\right\}.$$

Additional computations show that $U^\lambda(a) \stackrel{d}{=} U^\lambda(0) + a/\lambda_0'(x_0)$ and therefore, by the definition of the inverse processes in (2.5.1) and (2.5.3),

$$\mathbb{P}\left(n^{1/3}\left[\hat{\lambda}_n(x_0) - \lambda_0(x_0)\right] > a, n^{1/3}\left[\tilde{\lambda}_n(x_0) - \lambda_0(x_0)\right] > b\right)$$
$$\to \mathbb{P}(U^\lambda(a) < 0, U^\lambda(b) < 0) = \mathbb{P}(-\lambda_0'(x_0)U^\lambda(0) > a, -\lambda_0'(x_0)U^\lambda(0) > b),$$

as $n \to \infty$. This implies that

$$\left(n^{1/3}\left[\hat{\lambda}_n(x_0) - \lambda_0(x_0)\right], n^{1/3}\left[\tilde{\lambda}_n(x_0) - \lambda_0(x_0)\right]\right) \xrightarrow{d} \left(-\lambda_0'(x_0)U^\lambda(0), -\lambda_0'(x_0)U^\lambda(0)\right),$$

which proves (2.3.3). To establish the limiting distribution, define

$$A(x) = \left(\frac{\Phi(\beta_0, x)}{4\lambda_0(x)\lambda_0'(x)} \right)^{1/3},$$

and note that

$$n^{1/3} A(x_0) \left[\hat{\lambda}_n(x_0) - \lambda_0(x_0) \right] \xrightarrow{d} A(x_0)\lambda_0'(x_0) U^{\lambda}(0) \stackrel{d}{=} \underset{t\in\mathbb{R}}{\operatorname{argmin}} \left\{ \mathbb{W}(t) + t^2 \right\},$$

by Brownian scaling and the fact that the distribution of $U^{\lambda}(0)$ is symmetric around zero. □

PROOF.[Proof of Theorem 2.5] The proof of Theorem 2.5 is completely analogous to that of Theorem 2.4. The inverse processes to be considered in this case are

$$\widehat{U}_n^{\lambda}(a) = \underset{x\in[0,T_{(n)}]}{\operatorname{argmax}} \left\{ Y_n(x) - aW_n(\hat{\beta}_n, x) \right\},$$
$$\widetilde{U}_n^{\lambda}(a) = \underset{x\in[0,T_{(n)}]}{\operatorname{argmax}} \left\{ \Lambda_n(x) - ax \right\},$$

for $a > 0$, where W_n, Y_n and Λ_n have been defined in (2.2.3), (2.2.11) and (2.2.8) and $\hat{\beta}_n$ is the maximum partial likelihood estimator. By the same arguments as used in the proof of Theorem 2.4, the limiting distribution is expressed in terms of

$$\underset{t\in\mathbb{R}}{\operatorname{argmax}} \left\{ \mathbb{W}(t) - t^2 \right\} \stackrel{d}{=} \underset{t\in\mathbb{R}}{\operatorname{argmax}} \left\{ -\mathbb{W}(t) - t^2 \right\} = \underset{t\in\mathbb{R}}{\operatorname{argmin}} \{ \mathbb{W}(t) + t^2 \},$$

by properties of Brownian motion. □

PROOF.[Proof of Theorem 2.6] Completely similar to the reasoning in the proof of Theorem 2.4, we obtain

$$n^{1/3} \left[\widetilde{U}_n^{f}(f_0(x_0) + n^{-1/3}a) - x_0 \right] \xrightarrow{d} U^f(a),$$

where

$$U^f(a) = \sup \left\{ t : \mathbb{W}\left(\frac{f_0(x_0)[1 - F_0(x_0)]}{\Phi(\beta_0, x_0)} t \right) + \frac{1}{2} f_0'(x_0) t^2 - at \text{ is maximal} \right\}.$$

As before, by Brownian scaling, $U^f(a) \stackrel{d}{=} U^f(0) + a/f_0'(x_0)$ and together with (2.5.5) we obtain

$$\mathbb{P}\left(n^{1/3} \left[\tilde{f}_n(x_0) - f_0(x_0) \right] < a \right) \to \mathbb{P}\left(-f_0'(x_0) U^f(0) < a \right).$$

Similar to the proof of Theorem 2.4, with

$$A(x) = \left| \frac{\Phi(\beta_0, x)}{4 f_0(x) f_0'(x) [1 - F_0(x)]} \right|^{1/3},$$

we conclude that $n^{1/3} A(x_0) [\tilde{f}_n(x_0) - f_0(x_0)]$ converges in distribution to

$$A(x_0) f_0'(x_0) U^f(0) = \underset{t \in \mathbb{R}}{\operatorname{argmax}} \{\mathbb{W}(t) - t^2\} \stackrel{d}{=} \underset{t \in \mathbb{R}}{\operatorname{argmin}} \{\mathbb{W}(t) + t^2\},$$

using Brownian scaling and the fact that the distribution of $U^f(0)$ is symmetric around zero. □

REFERENCES

BRUNK, H. D. (1958). On the estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics*, **29:** 437–454.

CHUNG, D. & CHANG, M. N. (1994). An isotonic estimator of the baseline hazard function in Cox's regression model under order restriction. *Statistics & Probability Letters*, **21:** 223–228.

COOK, D. J., WALTER, S. D., COOK, R. J., GRIFFITH, L. E., GUYATT, G. H., LEASA, D., JAESCHKE, R. Z. & BRUN-BUISSON, C. (1998). Incidence of and risk factors for ventilator-associated pneumonia in critically ill patients. *Annals of Internal Medicine*, **129:** 433–440.

COX, D. R. (1972). Regression models and life-tabels (with discussion). *Journal of the Royal Statistical Society*, **34:** 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika*, **62:** 269–276.

DUROT, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Annals of Statistics*, **35:** 1080–1104.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72:** 557–565.

van GELOVEN, N., MARTIN, I., DAMMAN, P., de WINTER, R. J., THIJSSEN, J. G. & LOPUHAÄ, H. P. (2013). Estimation of a decreasing hazard of patients with myocardial infarction. *Statistics in Medicine*, **32:** 1223–1238.

GRENANDER, U. (1956). On the theory of mortality measurement, part II. *Skandinavisk Aktuarietidskrift*, **39:** 125–153.

GROENEBOOM, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer II*, 539–555.

GROENEBOOM, P. & WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, vol. 19 of *DMV Seminar*. Birkhäuser Verlag, Basel.

GROENEBOOM, P. & WELLNER, J. A. (2001). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics*, **10:** 388–400.

HUANG, J. & WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Statistics*, **22:** 3–33.

HUANG, Y. & ZHANG, C. H. (1994). Estimating a monotone density from censored observations. *Annals of Statistics*, **22:** 1256–1274.

HUI, R. & JANKOWSKI, H. (2012). CPHshape: estimating a shape constrained baseline hazard in Cox proportional hazards model. *Submitted for publication.*

KIM, J. & POLLARD, D. (1990). Cube root asymptotics. *Annals of Statistics*, **18:** 191–219.

KOMLÓS, J., MAJOR, P. & TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **32:** 111–131.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer. New York.

MARSHALL, A. W. & PROSCHAN, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Annals of Mathematical Statistics*, **36:** 69–77.

OAKES, D. (1977). The asymptotic information in censored survival data. *Biometrika*, **64:** 441– 448.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A*, **31:** 23–36.

PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Annals of Mathematical Statistics*, **41:** 507–519.

PRENTICE, R. L. & KALBFLEISCH, J. D. (2003). Mixed discrete and continuous Cox regression model. *Lifetime Data Analysis*, **9:** 195–210.

ROBERTSON, T., WRIGHT, F. T. & DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference.* John Wiley& Sons, New York.

SLUD, E. V. (1982). Consistency and efficiency of inferences with the partial likelihood. *Biometrika*, **69:** 547–552.

TSIATIS, A. (1981). A large sample study of Cox's regression model. *Annals of Statistics*, **9:** 93–108.

van der VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer. New York.

CHAPTER **3**

AN ASYMPTOTIC LINEAR REPRESENTATION FOR THE BRESLOW ESTIMATOR [1]

We provide an asymptotic linear representation of the Breslow estimator of the baseline cumulative hazard function in the Cox model. Our representation consists of an average of independent random variables and a term involving the difference between the maximum partial likelihood estimator and the underlying regression parameter. The order of the remainder term is arbitrarily close to n^{-1}.

3.1 INTRODUCTION

The proportional hazards model is one of the most popular approaches to model right-censored time to event data in the presence of covariates. COX (1972) introduced this semiparametric model and focused on estimating the underlying regression coefficients of the covariates. His estimator was later shown by COX (1975) to be a maximum partial likelihood estimator and its asymptotic properties were broadly studied (TSIATIS, 1981; ANDERSEN *et al.*, 1993; OAKES, 1977; SLUD, 1982). Different functionals of the lifetime distribution are commonly investigated and the (cumulative) hazard function is of particular interest. In the discussion following the Cox's (1972) paper, Breslow proposed a nonparametric maximum likelihood estimator for the baseline cumulative hazard function. Asymptotic properties of the Breslow estimator, such as consistency and the asymptotic distribution, were derived by TSIATIS (1981) and ANDERSEN *et al.* (1993). For an overview of the Breslow estimator, see LIN (2007).

Estimators in unconditional censorship models such as the Kaplan–Meier and Nelson–Aalen estimators have received considerable attention, especially in the 1980s. Established large sample properties include consistency and asymptotic normality (BRESLOW & CROWLEY, 1974), rate of strong uniform consistency (CSÖRGŐ & HORVÁTH, 1983), strong approximation or Hungarian embedding (BURKE *et al.*, 1981), and linearization results (LO & SINGH, 1985). LO & SINGH (1985) expressed the difference between the Kaplan–Meier estimator and the underlying dis-

[1] By Lopuhaä, H. P. and Nane, G. F. (2013). *Communications in Statistics - Theory and Methods*, **42:** 1314–1324.

tribution function in terms of a sum of independent identically distributed random variables, almost surely, with a remainder term of the order $n^{-3/4}(\log n)^{3/4}$, with n denoting the sample size; this rate was later improved to $n^{-1}\log n$ by LO *et al.* (1989). To our knowledge, a strong approximation result for the Breslow estimator is unavailable in the literature. KOSOROK (2008) establishes a representation of the Breslow estimator in terms of counting processes. Although this can be turned into an asymptotic linear representation similar to the one in LO & SINGH (1985), the covariates are assumed to be in a bounded set and the remainder term is only shown to be of the order $o_p(n^{-1/2})$.

In this chapter, we derive a similar linearization result for the Breslow estimator, i.e., we prove that the difference between the estimator Λ_n and the cumulative baseline hazard function Λ_0 can be represented as a sum of independent random variables and a term involving the difference between the regression parameter and its maximum partial likelihood estimator. However, we allow unbounded covariates and we show that the remainder term is of the order $n^{-1}a_n^{-1}$, where a_n may be any sequence tending to zero. As a_n can be chosen to converge to zero arbitrarily slowly, this means that the order of the remainder term is arbitrarily close to n^{-1}. The proof is based on empirical process theory, which allows the extension of our result to related semi-parametric models, such as marginal regression models. Our main motivation is isotonic estimation of the baseline distribution in the Cox model. An example is the Grenander-type estimator $\tilde{\lambda}_n$ for a nondecreasing baseline hazard λ_0, considered in Chapter 2, which is defined as the left-hand slope of the greatest convex minorant of the Breslow estimator. The limit behavior of $\tilde{\lambda}_n$ at a fixed point t_0 essentially follows from the limit behavior of the process

$$t \mapsto n^{2/3}\left\{(\Lambda_n - \Lambda_0)\left(t_0 + n^{-1/3}t\right) - (\Lambda_n - \Lambda_0)(t_0)\right\}.$$

In the absence of a strong approximation result for the process $\Lambda_n - \Lambda_0$, an alternative to obtain the limit process is to apply the results in KIM & POLLARD (1990) to the linear representation of $\Lambda_n - \Lambda_0$, provided that the remaining terms in the representation are of order smaller than $n^{-2/3}$. This cannot be ensured by the representation in KOSOROK (2008), whereas the order $n^{-1}a_n^{-1}$ can be chosen sufficiently small, for suitable choices of a_n. Another application of our linear representation is that, together with a linear representation of the maximum partial likelihood estimator, a central limit theorem can be established for $\Lambda_n - \Lambda_0$. Moreover, such a representation may also provide a means to estimate the variance of the Breslow estimator, by using plug-in estimators. A linear representation of the partial maximum likelihood estimator can be deduced from results in TSIATIS (1981) or KOSOROK (2008).

The chapter is organized as follows. The Cox model and the Breslow estimator are introduced in Section 3.2. Section 3.3 is devoted to the main result of the paper and its proof as well as two preparatory lemmas.

3.2 BACKGROUND, NOTATION, AND ASSUMPTIONS

Let X denote a positive random variable representing the survival time of a population of interest. The random variable C denotes the censoring time. Now, define $T = \min(X, C)$ as the generic follow-up time and $\Delta = \{X \leqslant C\}$ as its corresponding indicator, where $\{\cdot\}$ denotes the indicator function. Suppose that at the beginning of the study, extra information such as sex, age, status of a disease, etc. is recorded for each subject as covariates. Let Z denote a p-dimensional covariate vector. Therefore, suppose we observe the following independent, identically distributed triplets (T_i, Δ_i, Z_i), with $i = 1, 2, \ldots, n$. The censoring mechanism is assumed to be non-informative. Moreover, given the covariate Z, the survival time X is assumed to be independent of the censoring time C. The p-dimensional covariate vector Z is assumed to be time invariant and non-degenerate.

In the Cox model, the distribution of the survival time is related to the corresponding covariate by

$$\lambda\left(x \mid z\right) = \lambda_0(x)\,\mathrm{e}^{\beta_0' z}, \quad x \in \mathbb{R}^+,$$

where $\lambda\left(x \mid z\right)$ is the hazard function for a subject with covariate vector $z \in \mathbb{R}^p$, λ_0 represents the underlying baseline hazard function, and $\beta_0 \in \mathbb{R}^p$ is the vector of the underlying regression coefficients. Conditionally on $Z = z$, the survival time X is assumed to be a nonnegative random variable, with an absolutely continuous distribution function $F(x \mid z)$ with density $f(x \mid z)$. The same assumptions hold for the censoring variable C and its distribution function G. Let H be the distribution function of the follow-up time T and let $\tau_H = \inf\{t : H(t) = 1\}$ be the end point of the support of H. Moreover, let τ_F and τ_G be the end points of the support of F and G, respectively. We employ the usual assumptions for deriving large sample properties of Cox proportional hazards estimators (TSIATIS, 1981):

(A1) $\tau_H = \tau_G < \tau_F$.

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

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

where $|\cdot|$ denotes the Euclidean norm.

Let $X_{(1)} < X_{(2)} < \cdots < X_{(m)}$ denote the ordered, observed survival times. COX (1972, 1975) introduced the proportional hazards model and proposed the partial likelihood estimator $\hat{\beta}_n$ as an estimator for the underlying regression coefficients β_0. Breslow (COX, 1972) focused on estimating the baseline cumulative hazard function, $\Lambda_0(x) = \int_0^x \lambda_0(u)\,\mathrm{d}u$, and proposed

$$\Lambda_n(x) = \sum_{i \mid X_{(i)} \leqslant x} \frac{d_i}{\sum_{j=1}^n \{T_j \geqslant X_{(i)}\}\,\mathrm{e}^{\hat{\beta}_n' Z_j}}, \tag{3.2.1}$$

as an estimator for Λ_0, where d_i is the number of events at $X_{(i)}$ and $\hat{\beta}_n$ is the partial maximum likelihood estimator of the regression coefficients. The estimator Λ_n is most commonly referred to as the Breslow estimator. Under the assumption of a piecewise constant baseline hazard function and assuming that all the censoring times are shifted to the preceding observed survival time, Breslow showed that the partial maximum likelihood estimator $\hat{\beta}_n$ along with the baseline cumulative hazard estimator Λ_n can be obtained by jointly maximizing the full loglikelihood function.

Let

$$\begin{aligned} \Phi(\beta, x) &= \int \{u \geqslant x\} \, e^{\beta' z} \, dP(u, \delta, z), \\ \Phi_n(\beta, x) &= \int \{u \geqslant x\} \, e^{\beta' z} \, dP_n(u, \delta, z), \end{aligned} \tag{3.2.2}$$

where P is the underlying probability measure corresponding to the distribution of (T, Δ, Z) and P_n is the empirical measure of the triplets (T_i, Δ_i, Z_i), for $i = 1, 2, \ldots, n$. Furthermore, let $H^{uc}(x) = \mathbb{P}(T \leqslant x, \Delta = 1)$ be the sub-distribution function of the uncensored observations. Then, using the derivations in TSIATIS (1981), it can be deduced that

$$\lambda_0(u) = \frac{dH^{uc}(u)/du}{\Phi(\beta_0, u)}. \tag{3.2.3}$$

Consequently, it can be derived that

$$\Lambda_0(x) = \int \frac{\delta\{u \leqslant x\}}{\Phi(\beta_0, u)} \, dP(u, \delta, z). \tag{3.2.4}$$

From (A1) it follows that $\Lambda_0(\tau_H) < \infty$. An intuitive baseline cumulative hazard function estimator is obtained by replacing Φ in (3.2.4) by Φ_n and by plugging in $\hat{\beta}_n$, which yields exactly the Breslow estimator in (3.2.1),

$$\Lambda_n(x) = \int \frac{\delta\{u \leqslant x\}}{\Phi_n(\hat{\beta}_n, u)} \, dP_n(u, \delta, z). \tag{3.2.5}$$

KOSOROK (2008) established strong uniform consistency for the Breslow estimator and the process convergence of $\sqrt{n}(\Lambda_n - \Lambda_0)$, yet under the strong assumption of bounded covariates. Using standard empirical processes methods, LOPUHAÄ & NANE (2013) establish strong uniform consistency at rate $n^{-1/2}$ for the Breslow estimator under the relatively mild conditions (A1) and (A2).

3.3 ASYMPTOTIC REPRESENTATION

The following two lemmas will be used in proving the main result of the paper.

LEMMA 3.1. *Suppose that condition (A2) holds and let* Φ_n *and* Φ *be defined in* (3.2.2). *With* $\varepsilon > 0$ *taken from (A2), for* $|\beta - \beta_0| < \varepsilon$, *let*

$$\begin{aligned} D^{(1)}(\beta, x) &= \frac{\partial \Phi(\beta, x)}{\partial \beta} = \int \{u \geqslant x\}\, z\, e^{\beta' z}\, \mathrm{d}P(u, \delta, z) \in \mathbb{R}^p, \\ D_n^{(1)}(\beta, x) &= \frac{\partial \Phi_n(\beta, x)}{\partial \beta} = \int \{u \geqslant x\}\, z\, e^{\beta' z}\, \mathrm{d}P_n(u, \delta, z) \in \mathbb{R}^p. \end{aligned} \tag{3.3.1}$$

Then,

$$\begin{aligned} \sqrt{n} \sup_{x \in \mathbb{R}} \left| \Phi_n(\beta_0, x) - \Phi(\beta_0, x) \right| &= \mathcal{O}_p(1), \\ \sqrt{n} \sup_{x \in \mathbb{R}} \left| D_n^{(1)}(\beta_0, x) - D^{(1)}(\beta_0, x) \right| &= \mathcal{O}_p(1). \end{aligned} \tag{3.3.2}$$

PROOF. Consider the class of functions $\mathcal{G} = \{g(u, z; x) : x \in \mathbb{R}\}$, where, for each $x \in \mathbb{R}$ and $\beta_0 \in \mathbb{R}^p$ fixed,

$$g(u, z; x) = \{u \geqslant x\} \exp(\beta_0' z)$$

is a product of an indicator and a fixed function. It follows that $\mathcal{G}$ is a Vapnik–Červonenkis (VC)-subgraph class (Lemma 2.6.18 in van der VAART & WELLNER, 1996) and its envelope $G = \exp(\beta_0' z)$ is square integrable under condition (A2). Standard results from empirical process theory (van der VAART & WELLNER, 1996) yield that the class of functions $\mathcal{G}$ is a Donsker class, i.e.,

$$\sqrt{n} \int g(u, z; x)\, \mathrm{d}(P_n - P)(u, \delta, z) = \mathcal{O}_p(1),$$

so that the first statement in (3.3.2) follows by the continuous mapping theorem. To prove the second statement, it suffices to consider each jth coordinate, for $j = 1, 2, \dots, p$, fixed. In this case, we deal with the class $\mathcal{G}_j = \{g_j(u, z; x) : x \in \mathbb{R}\}$, where

$$g_j(u, z; x) = \{u \geqslant x\} z_j \exp(\beta_0' z).$$

From here the argument is exactly the same, which proves the lemma. □

LEMMA 3.2. *Assume (A1) and (A2). Then, for all* $M \in (0, \tau_H)$,

$$a_n n \sup_{x \in [0, M]} \left| \int \delta \{u \leqslant x\} \left(\frac{1}{\Phi_n(\beta_0, u)} - \frac{1}{\Phi(\beta_0, u)} \right) \mathrm{d}(P_n - P)(u, \delta, z) \right| = \mathcal{O}_p(1),$$

for any sequence $a_n = o(1)$.

PROOF. Consider the class of functions $\mathscr{F}_n = \{f_n(u,\delta,z;x) : 0 \leqslant x \leqslant M\}$, where

$$f_n(u,\delta,z;x) = \delta\{u \leqslant x\}\left(\frac{1}{\Phi_n(\beta_0,u)} - \frac{1}{\Phi(\beta_0,u)}\right).$$

Correspondingly, consider the class $\mathscr{G}_{n,M,\alpha}$ consisting of functions

$$g(u,\delta,z;y,\Psi) = \delta\{u \leqslant y\}\left(\frac{1}{\Psi(u)} - \frac{1}{\Phi(\beta_0,u)}\right),$$

where $0 \leqslant y \leqslant M$ and Ψ is nonincreasing left continuous, such that

$$\Psi(M) \geqslant K, \qquad \sup_{u\in[0,M]} \left|\Psi(u) - \Phi(\beta_0,u)\right| \leqslant \alpha,$$

where $K = \Phi(\beta_0,M)/2$. Then, for any $\alpha > 0$, we have $\mathbb{P}(\mathscr{F}_n \subset \mathscr{G}_{n,M,\alpha}) \to 1$, by Lemma 3.1. Furthermore, the class $\mathscr{G}_{n,M,\alpha}$ has envelope $G(u,\delta,z) = \alpha/K^2$. Since the functions in $\mathscr{G}_{n,M,\alpha}$ are products of indicators and a difference of bounded monotone functions, its entropy with bracketing satisfies

$$\log N_{[\,]}(\varepsilon, \mathscr{G}_{n,M,\alpha}, L_2(P)) \lesssim \frac{1}{\varepsilon},$$

see e.g., Theorem 2.7.5 in van der VAART & WELLNER (1996) and Lemma 9.25 in KOSOROK (2008). Hence, for any $\delta > 0$, the bracketing integral

$$J_{[\,]}(\delta, \mathscr{G}_{n,M,\alpha}, L_2(P)) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon\|G\|_2, \mathscr{G}_{n,M,\alpha}, L_2(P))}\,\mathrm{d}\varepsilon < \infty.$$

By Theorem 2.14.2 in van der VAART & WELLNER (1996), we have

$$\mathbb{E}\left\|\sqrt{n}\int g(u,\delta,z;y,\Psi)\mathrm{d}(P_n - P)(u,\delta,z)\right\|_{\mathscr{G}_{n,M,\alpha}} \leqslant J_{[\,]}(1, \mathscr{G}_{n,M,\alpha}, L_2(P))\|G\|_{P,2} = \mathcal{O}(\alpha),$$

where $\|\cdot\|_{\mathscr{F}}$ denotes the supremum over the class of functions $\mathscr{F}$. Now, let $a_n = o(1)$. Then, according to (3.3.2),

$$a_n\sqrt{n}\sup_{x\in\mathbb{R}}\left|\Phi_n(\beta_0,x) - \Phi(\beta_0,x)\right| = o_p(1).$$

Therefore, if we choose $\alpha = n^{-1/2}a_n^{-1}$, this gives

$$\mathbb{E}\left\|\int g(u,\delta,z;y,\Psi)\mathrm{d}(P_n - P)(u,\delta,z)\right\|_{\mathscr{G}_{n,M,\alpha}} = \mathcal{O}((na_n)^{-1})$$

and hence, by the Markov inequality, this proves the lemma. □

The asymptotic linear representation of the Breslow estimator is provided by the next theorem.

Theorem 3.3. *Assume (A1) and (A2). Let* Φ *and* $D^{(1)}$ *be defined in* (3.2.2) *and* (3.3.1). *Then, for all* $M \in (0, \tau_H)$ *and* $x \in [0, M]$,

$$\Lambda_n(x) - \Lambda_0(x) = \frac{1}{n}\sum_{i=1}^{n} \xi(T_i, \Delta_i, Z_i; x) + (\hat{\beta}_n - \beta_0)' A_0(x) + R_n(x),$$

where $\hat{\beta}_n$ *is the maximum partial likelihood estimator,*

$$A_0(x) = \int_0^x \frac{D^{(1)}(\beta_0, u)}{\Phi(\beta_0, u)} \lambda_0(u)\, \mathrm{d}u \tag{3.3.3}$$

and

$$\xi(t, \delta, z; x) = -\mathrm{e}^{\beta_0' z} \int_0^{x \wedge t} \frac{\lambda_0(u)}{\Phi(\beta_0, u)}\, \mathrm{d}u + \frac{\delta\{t \leqslant x\}}{\Phi(\beta_0, t)}$$

and R_n *is such that*

$$\sup_{x \in [0,M]} \left| R_n(x) \right| = \mathcal{O}_p(n^{-1} a_n^{-1}),$$

for any sequence $a_n = o(1)$.

Proof. For $\beta \in \mathbb{R}^p$, define

$$\Lambda_n(\beta, x) = \int \frac{\delta\{u \leqslant x\}}{\Phi_n(\beta, u)}\, \mathrm{d}P_n(u, \delta, z).$$

Hence, the Breslow estimator in (3.2.5) can also be written as $\Lambda_n(\hat{\beta}_n, x)$. For $x \in [0, M]$, consider the following decomposition

$$\Lambda_n(x) - \Lambda_0(x) = T_{n1}(x) + T_{n2}(x),$$

where $T_{n1}(x) = \Lambda_n(\hat{\beta}_n, x) - \Lambda_n(\beta_0, x)$ and $T_{n2}(x) = \Lambda_n(\beta_0, x) - \Lambda_0(x)$.

For the term T_{n1}, first notice that a Taylor expansion of $\Lambda_n(\cdot, x)$ around β_0 yields that

$$\Lambda_n(\hat{\beta}_n, x) - \Lambda_n(\beta_0, x) = -\big(\hat{\beta}_n - \beta_0\big)' A_n(x) + \frac{1}{2}\big(\hat{\beta}_n - \beta_0\big)' R_{n1}(x) \big(\hat{\beta}_n - \beta_0\big), \tag{3.3.4}$$

where the vector A_n and matrix R_{n1} are given by

$$A_n(x) = \int \delta\{u \leqslant x\} \frac{D_n^{(1)}(\beta_0, u)}{\Phi_n^2(\beta_0, u)}\, \mathrm{d}P_n(u, \delta, z), \tag{3.3.5}$$

$$R_{n1}(x) = \int \delta\{u \leqslant x\} \frac{2 D_n^{(1)}(\beta^*, u) D_n^{(1)}(\beta^*, u)' - D_n^{(2)}(\beta^*, u) \Phi_n(\beta^*, u)}{\Phi_n^3(\beta^*, u)}\, \mathrm{d}P_n(u, \delta, z),$$

for some $|\beta^* - \beta_0| \leqslant |\hat{\beta}_n - \beta_0|$, with $D_n^{(1)}$ as defined in (3.3.1) and

$$D_n^{(2)}(\beta, x) = \frac{\partial^2 \Phi_n(\beta, x)}{\partial \beta^2} = \int \{u \geqslant x\}\, zz'\, \mathrm{e}^{\beta' z}\, \mathrm{d}P_n(u, \delta, z) \in \mathbb{R}^p \times \mathbb{R}^p.$$

We define $D^{(2)}(\beta, x)$ similarly, with P_n replaced by P.
According to (A2), we have $|D^{(1)}(\beta_0, x)| \leqslant \mathbb{E}\left[|Z|\exp(\beta_0' Z)\right] < \infty$, for all $x \in \mathbb{R}$, and similarly

$$|D_n^{(1)}(\beta_0, x)| \leqslant \frac{1}{n}\sum_{i=1}^{n} |Z_i| \mathrm{e}^{\beta_0' Z_i} \to \mathbb{E}\left[|Z|\mathrm{e}^{\beta_0' Z}\right] < \infty,$$

with probability one. Likewise, $|D^{(2)}(\beta_0, x)| < \infty$ and

$$|D_n^{(2)}(\beta_0, x)| \leqslant \frac{1}{n}\sum_{i=1}^{n} |Z_i|^2 \mathrm{e}^{\beta_0' Z_i} \to \mathbb{E}\left[|Z|^2\mathrm{e}^{\beta_0' Z}\right] < \infty,$$

with probability one. Furthermore, for all $x \in [0, M]$,

$$0 < \Phi(\beta_0, M) \leqslant \Phi(\beta_0, x) \leqslant \Phi(\beta_0, 0) = \mathbb{E}\left[\mathrm{e}^{\beta_0' Z}\right] < \infty$$

and $\Phi_n(\beta_0, M) \leqslant \Phi_n(\beta_0, x) \leqslant \Phi_n(\beta_0, 0)$, where $\Phi_n(\beta_0, M) \to \Phi(\beta_0, M)$ and $\Phi_n(\beta_0, 0) \to \Phi(\beta_0, 0)$, with probability one. It follows that there exist constants $K_1, K_2 > 0$, such that for all $x \in [0, M]$,

$$|D^{(1)}(\beta_0, x)| \leqslant K_2, \quad |D^{(2)}(\beta_0, x)| \leqslant K_2, \quad K_1 \leqslant \Phi(\beta_0, x) \leqslant K_2 \tag{3.3.6}$$

and for n sufficiently large,

$$|D_n^{(1)}(\beta_0, x)| \leqslant K_2, \quad |D_n^{(2)}(\beta_0, x)| \leqslant K_2, \quad K_1 \leqslant \Phi_n(\beta_0, x) \leqslant K_2, \tag{3.3.7}$$

with probability one. According to (3.2.3),

$$\frac{\delta}{\Phi(\beta_0, u)}\, \mathrm{d}P(u, \delta, y) = \frac{\mathrm{d}H^{uc}(u)}{\Phi(\beta_0, u)} = \lambda_0(u)\, \mathrm{d}u, \tag{3.3.8}$$

so that A_0, as defined in (3.3.3), is equal to

$$A_0(x) = \int \delta\{u \leqslant x\} \frac{D^{(1)}(\beta_0, u)}{\Phi^2(\beta_0, u)}\, \mathrm{d}P(u, \delta, z) \in \mathbb{R}^p,$$

Then, for the A_n term in (3.3.4), it can be deduced that

$$\begin{aligned} \sup_{0 \leqslant x \leqslant M} |A_n(x) - A_0(x)| \leqslant{}& \sup_{0 \leqslant u \leqslant M} \left| \frac{D_n^{(1)}(\beta_0, u)}{\Phi_n^2(\beta_0, u)} - \frac{D^{(1)}(\beta_0, u)}{\Phi^2(\beta_0, u)} \right| \\ &+ \sup_{0 \leqslant x \leqslant M} \left| \int \delta\{u \leqslant x\} \frac{D^{(1)}(\beta_0, u)}{\Phi^2(\beta_0, u)} \mathrm{d}(P_n - P)(u, \delta, z) \right|. \end{aligned}$$

By (3.3.6) and (3.3.7), the first term on the right hand side is bounded by

$$\frac{1}{K_1^2} \sup_{0\leqslant x\leqslant M} \left| D_n^{(1)}(\beta_0, x) - D^{(1)}(\beta_0, x) \right| + \frac{2K_2^2}{K_1^4} \sup_{0\leqslant x\leqslant M} \left| \Phi_n(\beta_0, x) - \Phi(\beta_0, x) \right|,$$

which is of the order $\mathcal{O}_p(n^{-1/2})$, by Lemma 3.1. For the second term on the right hand side, for each $j = 1, \ldots, p$, fixed, consider the class $\mathcal{G}_j = \{g_j(u, \delta; x) : x \in [0, M]\}$, consisting of functions

$$g_j(u, \delta; x) = \delta\{u \leqslant x\} \frac{D_j^{(1)}(\beta_0, u)}{\Phi^2(\beta_0, u)},$$

where $D_j^{(1)}$ denotes the jth coordinate of $D^{(1)}$. Now, each $g_j(u, \delta; x)$ is the product of indicators and a fixed uniformly bounded function. Standard results from empirical process theory (van der VAART & WELLNER, 1996) give that the class $\mathcal{G}_j$ is Donsker. As in the proof of Lemma 3.1, we find that for every $j = 1, \ldots, p$,

$$\sqrt{n} \sup_{0\leqslant x\leqslant M} \left| \int g_j(u, \delta; x) \, \mathrm{d}(P_n - P)(u, \delta, z) \right| = \mathcal{O}_p(1).$$

It follows that

$$\sup_{0\leqslant x\leqslant M} |A_n(x) - A_0(x)| = \mathcal{O}_p(n^{-1/2}).$$

and we can conclude that

$$(\hat{\beta}_n - \beta_0)' A_n(x) = (\hat{\beta}_n - \beta_0)' A_0(x) + R_{n2}(x),$$

where $R_{n2}(x) = \mathcal{O}_p(n^{-1})$, uniformly for $x \in [0, M]$, since $\hat{\beta}_n - \beta_0 = \mathcal{O}_p(n^{-1/2})$ (TSIATIS, 1981). For the term containing R_{n1}, first observe that, according to (3.3.7), for n sufficiently large,

$$\sup_{u\in[0,M]} \left| \frac{2D_n^{(1)}(\beta^*, u) D_n^{(1)}(\beta^*, u)' - D_n^{(2)}(\beta^*, u)\Phi_n(\beta^*, u)}{\Phi_n^3(\beta^*, u)} \right| = \mathcal{O}(1),$$

almost surely, so that

$$\sup_{0\leqslant x\leqslant M} \left| \frac{1}{2} (\hat{\beta}_n - \beta_0)' R_{n1}(x) (\hat{\beta}_n - \beta_0) \right| = \mathcal{O}_p(n^{-1}).$$

Concluding,

$$T_{n1}(x) = (\hat{\beta}_n - \beta_0)' A_0(x) + \mathcal{O}_p(n^{-1}), \tag{3.3.9}$$

uniformly in $x \in [0, M]$. Proceeding with T_{n2}, write

$$T_{n2}(x) = \Lambda_n(\beta_0, x) - \Lambda_0(x) = B_n(x) + C_n(x) + R_{n3}(x) + R_{n4}(x),$$

where

$$
\begin{aligned}
B_n(x) &= \int \delta\{u \leqslant x\} \frac{\Phi(\beta_0,u) - \Phi_n(\beta_0,u)}{\Phi^2(\beta_0,u)} \,\mathrm{d}P(u,\delta,z),\\
C_n(x) &= \int \frac{\delta\{u \leqslant x\}}{\Phi(\beta_0,u)} \,\mathrm{d}(P_n - P)(u,\delta,z),\\
R_{n3}(x) &= \int \delta\{u \leqslant x\} \left(\frac{1}{\Phi_n(\beta_0,u)} - \frac{1}{\Phi(\beta_0,u)} \right) \mathrm{d}(P_n - P)(u,\delta,z),\\
R_{n4}(x) &= \int \delta\{u \leqslant x\} \frac{[\Phi(\beta_0,u) - \Phi_n(\beta_0,u)]^2}{\Phi^2(\beta_0,u)\Phi_n(\beta_0,u)} \,\mathrm{d}P(u,\delta,z).
\end{aligned}
$$

For the dominating term in T_{n2}, we can write

$$
\begin{aligned}
B_n(x) + C_n(x) &= -\int \delta\{u \leqslant x\} \frac{\Phi_n(\beta_0,u)}{\Phi^2(\beta_0,u)} \,\mathrm{d}P(u,\delta,z) + \int \frac{\delta\{u \leqslant x\}}{\Phi(\beta_0,u)} \,\mathrm{d}P_n(u,\delta,z)\\
&= \frac{1}{n}\sum_{i=1}^{n} \xi(T_i,\Delta_i,Z_i;x),
\end{aligned}
$$

where

$$
\xi(t,\delta,z;x) = -\int \gamma\{u \leqslant x\} \frac{\{t \geqslant u\}\mathrm{e}^{\beta_0' z}}{\Phi^2(\beta_0,u)} \,\mathrm{d}P(u,\gamma,y) + \frac{\delta\{t \leqslant x\}}{\Phi(\beta_0,t)}.
$$

Using (3.3.8), we conclude that

$$
\xi(t,\delta,z;x) = -\mathrm{e}^{\beta_0' z} \int_0^{x \wedge t} \frac{\lambda_0(u)}{\Phi(\beta_0,u)} \,\mathrm{d}u + \frac{\delta\{t \leqslant x\}}{\Phi(\beta_0,t)}.
$$

For the remainder terms, it follows by Lemma 3.2, that for any sequence $a_n = o(1)$,

$$
\sup_{0 \leqslant x \leqslant M} \left| R_{n3}(x) \right| = \mathcal{O}_p(n^{-1} a_n^{-1}). \tag{3.3.10}
$$

To treat R_{n4}, note that

$$
\left| R_{n4}(x) \right| \leqslant \frac{1}{\Phi^2(\beta_0,M)} \frac{1}{\Phi_n(\beta_0,M)} \sup_{x \in \mathbb{R}} |\Phi_n(\beta_0,x) - \Phi(\beta_0,x)|^2,
$$

so that by (3.3.2) and (3.3.7),

$$
\sup_{0 \leqslant x \leqslant M} \left| R_{n4}(x) \right| = \mathcal{O}_p(n^{-1}).
$$

Together with (3.3.9) and (3.3.10), this proves the theorem. ▫

In the special case of no covariates, i.e., $\beta_0 = \hat{\beta}_n = 0$, it follows that

$$\Phi(\beta_0, x) = 1 - H(x),$$

where H is the distribution function of the follow-up times and

$$\begin{aligned}\xi(t, \delta, z; x) &= -\mathrm{e}^{\beta_0' z} \int_0^{x \wedge t} \frac{\lambda_0(u)}{\Phi(\beta_0, u)} \, \mathrm{d}u + \frac{\delta\{t \leqslant x\}}{\Phi(\beta_0, t)} \\ &= -\int_0^{x \wedge t} \frac{\mathrm{d}H^{uc}(u)}{[1 - H(u)]^2} + \frac{\delta\{t \leqslant x\}}{1 - H(t)}.\end{aligned}$$

This means that Theorem 3.3 retrieves a result similar to Lemma 2.1 in LO *et al.* (1989).

The rate at which the error term R_n tends to zero becomes faster as a_n tends to zero more slowly. If $a_n = 1/\log n$, we obtain the same rate as the error term in Lemma 2.1 in LO *et al.* (1989). However, they obtain the order $\mathcal{O}(n^{-1} \log n)$ almost surely, whereas Theorem 3.3, with the choice $a_n = 1/\log n$, only provides this order in probability. Also, the sequence a_n may be chosen to converge to zero arbitrarily slowly. This means that the order $\mathcal{O}_p(n^{-1} a_n^{-1})$ of R_n is arbitrarily close to $\mathcal{O}_p(n^{-1})$.

Using a linear representation of $\hat{\beta}_n - \beta_0$, a full linearization of the Breslow estimator can be obtained. Such a linear representation can be deduced from the proof of Theorem 3.2 in TSIATIS (1981) or from an application of Theorem 2.11 in KOSOROK (2008); see also Section 4.2.1 in KOSOROK (2008). As a consequence, Theorem 1 together with the expansion of $\hat{\beta}_n - \beta_0$ can be used to establish a central limit theorem for the Breslow estimator, as well as to estimate the limiting covariance structure, by using plug-in estimators. For example, the term A_0 in the linear expression can be estimated consistently by A_n in (3.3.5).

REFERENCES

ANDERSEN, P. K., BORGAN, O., GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer. New York.

BRESLOW, N. & CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Annals of Statistics*, **2:** 437–453.

BURKE, M. D., CSÖRGŐ, S. & HORVÁTH, L. (1981). Strong approximations of some biometric estimates under random censorship. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **56:** 87–112.

COX, D. R. (1972). Regression models and life-tabels (with discussion). *Journal of the Royal Statistical Society. Series B*, **34:** 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika*, **62:** 269–276.

CSÖRGŐ, S. & HORVÁTH, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **62:** 411–426.

KIM, J. & POLLARD, D. (1990). Cube root asymptotics. *Annals of Statistics*, **18:** 191–219.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer. New York.

LIN, D. Y. (2007). On the Breslow estimator. *Lifetime Data Analysis*, **13:** 471–480.

LO, S. H., MACK, Y. P. & WANG, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan–Meier estimator. *Probability Theory and Related Fields*, **80:** 461–473.

LO, S. H. & SINGH, K. (1985). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probability Theory and Related Fields* **71:** 455–465.

LOPUHAÄ, H. P. & NANE, G. F. (2013). Shape constrained nonparametric estimators of the baseline distribution in Cox proportional hazards model. *Scandinavian Journal of Statistics*, doi: 10.1002/sjos.12008.

OAKES, D. (1977). The asymptotic information in censored survival data. *Biometrika*, **64:** 441–448.

SLUD, E. V. (1982). Consistency and efficiency of inferences with the partial likelihood. *Biometrika*, **69:** 547–552.

TSIATIS, A. (1981). A large sample study of Cox's regression model. *Annals of Statistics*, **9:** 93–108.

van der VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer. New York.

CHAPTER 4

A LIKELIHOOD RATIO TEST FOR MONOTONE BASELINE HAZARD FUNCTIONS IN THE COX MODEL [1]

We consider a likelihood ratio method for testing whether a monotone baseline hazard function in the Cox model has a particular value at a fixed point. The characterization of the estimators involved is provided both in the nondecreasing and the nonincreasing setting. These characterizations facilitate the derivation of the asymptotic distribution of the likelihood ratio test, which is identical in the nondecreasing and in the nonincreasing case. The asymptotic distribution of the likelihood ratio test enables, via inversion, the construction of pointwise confidence intervals. Simulations show that these confidence intervals exhibit comparable coverage probabilities with the confidence intervals based on the asymptotic distribution of the nonparametric maximum likelihood estimator of a monotone baseline hazard function.

4.1 INTRODUCTION

In survival analysis, using COX (1972) proportional hazards model is the typical choice to account for the effect of covariates on the lifetime distribution. Its attractiveness resides in its form, that allows for efficient estimation of the regression coefficient, while leaving the baseline distribution completely unspecified, see e.g., EFRON (1977), OAKES (1977), and SLUD (1982). The regression coefficient estimator is the well-known maximum partial likelihood estimator (COX, 1972 and 1975). As a response to Cox's paper, Breslow proposed in COX (1972) a different approach, that yields the same maximum partial likelihood estimator, along with an estimator of the baseline cumulative hazard function Λ_0. Impressive amount of research rapidly followed Cox's seminal paper, which primarily focused on deriving the (asymptotic) properties of the maximum partial likelihood estimator of the regression coefficient $\hat{\beta}_n$, as well as of the Breslow estimator Λ_n of the baseline cumulative hazard function.

Even though the baseline hazard λ_0 can be left completely unspecified, in practice,

[1] By Nane, G. F. (2013). Submitted for publication to *Statistica Sinica*.

one might be interested in restricting λ_0 qualitatively. This can be done by assuming the baseline hazard to be monotone, for example, as suggested by Cox himself (COX, 1972). Various studies have indicated that a monotonicity constraint should be imposed on the baseline hazard, which complies with the medical expertise. For an illustration of a nonincreasing baseline hazard estimator in the study of patients with acute coronary syndrome, see van GELOVEN *et al.* (2011). LOPUHAÄ & NANE (2013) propose a nonparametric maximum likelihood estimator and a Grenander-type estimator for estimating a monotone baseline hazard function. The Grenander-type estimator is defined in terms of slopes of the greatest convex minorant of the Breslow estimator Λ_n. The two estimators have been proven to be asymptotically strongly consistent and have been shown to exhibit the same distributional law. Furthermore, at a fixed point x_0, the scaled difference between the maximum likelihood estimator $\hat{\lambda}_n$ and the true baseline hazard λ_0 converges in distribution to the distribution of the minimum of two-sided Brownian motion plus a parabola times a constant depending on the underlying parameters. These results adhere to the general nonparametric shape constrained theory, and, in particular, prolong naturally the findings of HUANG & WELLNER (1995) in the case of the random censorship model with no covariates.

Ensuing inference will be pursued in this chapter, by testing the hypothesis that the underlying monotone baseline hazard has a particular value ϑ_0, at a fixed point x_0. We will use a likelihood ratio test of $H_0 : \lambda_0(x_0) = \vartheta_0$ versus $H_1 : \lambda_0(x_0) \neq \vartheta_0$. Within the shape restricted problems, this approach was initially employed for monotone distributions in the current status model by BANERJEE & WELLNER (2001). The authors focused on deriving the limiting distribution of the likelihood ratio test under the null hypothesis, and to obtain what the authors referred to a fixed universal distribution, defined in terms of slopes of the greatest convex minorant of the two-sided Brownian motion plus a parabola. These findings were followed by a rapid stream of research, see, e.g., BANERJEE & WELLNER (2005), BANERJEE (2007), and BANERJEE (2008), that revealed that the likelihood ratio method could be extended straightforwardly in other shape constrained settings. In this chapter, we carry on this research for the monotone baseline hazard function in the Cox model. In addition to extending directly the results in the right censoring model with no covariates in BANERJEE (2008), we aim to provide a thorough description of the method and detailed proofs for all results.

Furthermore, we will derive confidence sets for $\lambda_0(x_0)$, based on the likelihood ratio method. More specifically, we will use that inverting the family of tests can yield, in turn, pointwise confidence intervals for the baseline hazard function. A more direct method of constructing pointwise confidence intervals is based on the asymptotic distribution, at a fixed point x_0, of the nonparametric maximum likelihood estimator $\hat{\lambda}_n$, derived in Chapter 2. Nonetheless, this entails the bothersome issue of estimating the nuisance parameter, and more specifically, estimating the derivative of the baseline hazard function $\lambda'(x_0)$, since, to the author's best knowledge, there is no available smooth monotone estimator of the baseline hazard function in the Cox model. One option would be to kernel smooth the NPMLE $\hat{\lambda}_n$; however, this would pose extra difficulties, like an appropriate choice of a bandwidth. For a discussion of these issues in the case of right-censoring with no covariates, see BANERJEE (2008).

The chapter is organized as follows. Section 4.2 introduces the Cox model, the notations and the common assumptions. In Section 4.3, we introduce the likelihood ratio method and characterize the maximum likelihood estimator $\hat{\lambda}_n$ of a monotone baseline hazard function and the estimator $\hat{\lambda}_n^0$, for which $\hat{\lambda}_n^0(x_0) = \vartheta_0$, for a fixed x_0 in the interior of the support of the baseline distribution. We provide the characterization of the two estimators in the case of both nondecreasing and nonincreasing baseline hazard functions λ_0. The asymptotic distribution of the likelihood ratio statistic is provided, along with preparatory lemmas, in Section 4.4. Finally, Section 4.5 is devoted to constructing pointwise confidence intervals and comparing them, via simulations, with the conventional confidence intervals based on the asymptotic distribution of the NPMLE $\hat{\lambda}_n$.

4.2 DEFINITIONS AND ASSUMPTIONS

Suppose that the observed data consist of the following independent and identically distributed triplets (T_i, Δ_i, Z_i), with $i = 1, 2, \dots, n$. The event time, denoted by X and commonly referred to as the survival time is subject to random censoring. Thus, $T = \min(X, C)$, where T is the follow-up time and C denotes the censoring time. The indicator $\Delta = \{X \leqslant C\}$ marks whether the follow-up time is an event or a censoring time. Finally, $Z \in \mathbb{R}^p$ denotes the covariate vector of the observed follow-up time T, which is assumed to be time invariant. The event time X and censoring time C are assumed to be conditionally independent, given the covariate vector Z. Furthermore, let F be the distribution function of the non-negative random variable X, G the distribution function of the non-negative random variable C, and H the distribution function of T. The distribution function $F(x|z)$ is assumed to be absolutely continuous, with density $f(x|z)$. Similarly, the distribution function $G(c|z)$ is assumed to be absolutely continuous, with density $g(c|z)$. In addition, $F(x|z)$ and $G(c|z)$ share no parameters, thus the censoring mechanism is assumed to be non-informative.

Let $\lambda(x|z)$ be the hazard function for an individual with covariate vector $z \in \mathbb{R}^p$. The Cox model specifies that

$$\lambda\left(x|z\right) = \lambda_0(x)\, e^{\beta_0' z}, \tag{4.2.1}$$

where λ_0 represents the baseline hazard function, that corresponds to $z = 0$, and $\beta_0 \in \mathbb{R}^p$ is the vector of the underlying regression coefficients. Finally, we consider the following assumptions, that are typically employed when deriving large sample properties of estimators within the Cox model; e.g., see TSIATIS (1981).

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

Then

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

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

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

where $|\cdot|$ denotes the Euclidean norm.

4.3 The likelihood ratio method and the characterization of the estimators

By definition, $\Lambda(x|z) = -\log[1-F(x|z)]$ is the cumulative hazard function. Thus, from (4.2.1), it follows that $\Lambda(x|z) = \Lambda_0(x)\exp(\beta_0' z)$, where $\Lambda_0(x) = \int_0^x \lambda_0(u)\,\mathrm{d}u$ is the baseline cumulative hazard function. Since, for a continuous distribution, $\lambda(t) = f(t)/(1-F(t))$, for $t \geqslant 0$, the full likelihood is given by

$$\prod_{i=1}^n \left\{f(T_i \mid Z_i)\left[1-G(T_i \mid Z_i)\right]\right\}^{\Delta_i} \left\{g(T_i \mid Z_i)\left[1-F(T_i \mid Z_i)\right]\right\}^{1-\Delta_i}$$
$$= \prod_{i=1}^n \lambda(T_i \mid Z_i)^{\Delta_i} \exp\left[-\Lambda(T_i \mid Z_i)\right] \times \prod_{i=1}^n \left[1-G(T_i \mid Z_i)\right]^{\Delta_i} g(T_i \mid Z_i)^{1-\Delta_i}.$$

As the censoring mechanism is assumed to be non-informative, and by (4.2.1), maximizing the full likelihood is the same as maximizing

$$\prod_{i=1}^n \lambda(T_i \mid Z_i)^{\Delta_i} \exp\left[-\Lambda(T_i \mid Z_i)\right] = \prod_{i=1}^n \left[\lambda_0(T_i)\mathrm{e}^{\beta_0' Z_i}\right]^{\Delta_i} \exp\left[-\mathrm{e}^{\beta_0' Z_i}\Lambda_0(T_i)\right],$$

which yields the following (pseudo) loglikelihood function, written as function of $\beta \in \mathbb{R}^p$ and λ_0

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

Let $T_{(1)} < T_{(2)} < \cdots < T_{(n)}$ be the ordered follow-up times and, for $i = 1, 2, \ldots, n$, let $\Delta_{(i)}$ and $Z_{(i)}$ be the censoring indicator and covariate vector corresponding to $T_{(i)}$. Writing the above (pseudo) likelihood as a function of β and λ_0 gives

$$L_\beta(\lambda_0) = \sum_{i=1}^n \left[\Delta_{(i)} \log \lambda_0(T_{(i)}) + \Delta_{(i)}\beta' Z_{(i)} - \mathrm{e}^{\beta' Z_{(i)}} \int_0^{T_{(i)}} \lambda_0(u)\,\mathrm{d}u\right]. \qquad (4.3.1)$$

Following the approach in Chapter 2, we do not proceed with the joint maximization of (4.3.1) over β and monotone λ_0. Alternatively, for $\beta \in \mathbb{R}^p$ fixed, we consider maximum likelihood estimation of a monotone baseline hazard function λ_0 and denote the estimator by $\hat{\lambda}_n(x;\beta)$. Afterwards, we simply replace β by $\hat{\beta}_n$, the maximum partial likelihood estimator (see, e.g., COX (1972) and COX (1975)) of the underlying regression coefficients β_0, due to its commendable asymptotic properties (see, e.g., EFRON, 1977, OAKES, 1977, and SLUD, 1982). The proposed NPMLE is thus $\hat{\lambda}_n(x) = \hat{\lambda}_n(x;\hat{\beta}_n)$ and will be referred to as the unconstrained estimator of a monotone λ_0.

Furthermore, for $\beta \in \mathbb{R}^p$ fixed, we maximize the loglikelihood function $L_\beta(\lambda_0)$ in (4.3.1) over the class of all monotone baseline hazard functions, under the null hypothesis $H_0 : \lambda_0(x_0) = \vartheta_0$, for $x_0 \in (0, \tau_H)$ and $\vartheta_0 \in (0,\infty)$, fixed. We obtain $\hat{\lambda}^0_n(x;\beta)$ and hence propose $\hat{\lambda}^0_n(x) = \hat{\lambda}^0_n(x;\hat{\beta}_n)$ as the constrained NPMLE.

Replacing β by $\hat{\beta}_n$ also in the loglikelihood function (4.3.1) yields the likelihood ratio statistic for testing $H_0 : \lambda_0(x_0) = \vartheta_0$,

$$2\log \xi_n(\vartheta_0) = 2L_{\hat{\beta}_n}(\hat{\lambda}_n) - 2L_{\hat{\beta}_n}(\hat{\lambda}^0_n). \tag{4.3.2}$$

Thus, for computing the likelihood ratio statistic, we need to characterize the unconstrained NPMLE $\hat{\lambda}_n$ and the constrained NPMLE $\hat{\lambda}^0_n$ of a monotone baseline hazard function λ_0.

4.3.1 NONDECREASING BASELINE HAZARD

We consider first maximum likelihood estimation of a nondecreasing baseline hazard function λ_0. Both the unconstrained estimator $\hat{\lambda}_n$ and the constrained estimator $\hat{\lambda}^0_n$ will be characterized in terms of the processes

$$W_n(\beta, x) = \int \left(e^{\beta' z} \int_0^x \{u \geqslant s\}\, ds \right) dP_n(u,\delta,z), \tag{4.3.3}$$

and

$$V_n(x) = \int \delta\{u < x\}\, dP_n(u,\delta,z), \tag{4.3.4}$$

with $\beta \in \mathbb{R}^p$ and $x \geqslant 0$, and where P_n is the empirical measure of the (T_i, Δ_i, Z_i), with $i = 1, 2, \dots, n$. The characterization of the unconstrained estimator $\hat{\lambda}_n(x;\beta)$ has already been provided in Lemma 2.1 in Chapter 2, which we restate below. Furthermore, we provide a closed form of the estimator on blocks of indices on which the estimator is constant. This expression will be useful in deriving the asymptotic distribution of the likelihood ratio statistic.

LEMMA 4.1. *Let $T_{(1)} \leqslant T_{(2)} \leqslant \cdots \leqslant T_{(n)}$ be the ordered follow-up times and consider a fixed $\beta \in \mathbb{R}^p$.*

(i) Let W_n and V_n defined in (4.3.3) *and* (4.3.4). *Then, the NPMLE $\hat{\lambda}_n(x;\beta)$ of a*

nondecreasing baseline hazard function λ_0 *is of the form*

$$\hat{\lambda}_n(x;\beta) = \begin{cases} 0 & x < T_{(1)}, \\ \hat{\lambda}_i & T_{(i)} \leqslant x < T_{(i+1)}, \text{ for } i = 1, 2, \dots, n-1, \\ \infty & x \geqslant T_{(n)}, \end{cases}$$

where $\hat{\lambda}_i$ *is the left derivative of the greatest convex minorant (GCM) at the point* P_i *of the cumulative sum diagram (CSD) consisting of the points*

$$P_j = \Big(W_n(\beta, T_{(j+1)}) - W_n(\beta, T_{(1)}), V_n(T_{(j+1)}) \Big), \tag{4.3.5}$$

for $j = 1, 2, \dots, n-1$ *and* $P_0 = (0, 0)$.

(ii) For $k \geqslant 1$, *let* $B_1, B_2, \dots, B_k$ *be blocks of indices such that* $\hat{\lambda}_n(x;\beta)$ *is constant on each block and* $B_1 \cup B_2 \cup \ldots \cup B_k = \{1, 2, \dots, n-1\}$. *Denote by* $v_{nj}(\beta)$ *the value of* $\hat{\lambda}_n(x;\beta)$ *on block* B_j. *Then,*

$$v_{nj}(\beta) = \frac{\sum_{i \in B_j} \Delta_{(i)}}{\sum_{i \in B_j} \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} e^{\beta' Z_{(l)}}}. \tag{4.3.6}$$

PROOF. The proof of (i) is provided by Lemma 2.1 in Chapter 2. The NPMLE $\hat{\lambda}_n(x;\beta)$ is obtained by maximizing the (pseudo) loglikelihood function in (4.3.1) over all $0 \leqslant \lambda_0(T_{(1)}) \leqslant \lambda_0(T_{(2)}) \leqslant \cdots \leqslant \lambda_0(T_{(n)})$. As argued in Chapter 2, the estimator has to be a nondecreasing step function, that is zero for $x < T_{(1)}$, constant on the interval $[T_{(i)}, T_{(i+1)})$, for $i = 1, 2, \dots, n-1$ and can be chosen arbitrarily large for $x \geqslant T_{(n)}$. Then, for fixed $\beta \in \mathbb{R}^p$, the (pseudo) loglikelihood function in (4.3.1) reduces to

$$\begin{aligned} &\sum_{i=1}^{n-1} \Delta_{(i)} \log \lambda_0(T_{(i)}) - \sum_{i=2}^{n} e^{\beta' Z_{(i)}} \sum_{j=1}^{i-1} \left[T_{(j+1)} - T_{(j)} \right] \lambda_0(T_{(j)}) \\ &\quad = \sum_{i=1}^{n-1} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} e^{\beta' Z_{(l)}} \right\}. \end{aligned} \tag{4.3.7}$$

Let $\lambda_i = \lambda_0(T_{(i)})$, for $i = 1, 2, \dots, n-1$ and $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{n-1})$. Then, finding the NPMLE reduces to maximizing

$$\varphi(\lambda) = \sum_{i=1}^{n-1} \left\{ \Delta_{(i)} \log \lambda_i - \lambda_i \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} e^{\beta' Z_{(l)}} \right\}, \tag{4.3.8}$$

over the set $0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{n-1}$. The NPMLE corresponds thus to a vector $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_{n-1})$ that maximizes φ over $0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_{n-1}$. To prove (ii), we first derive the Fenchel conditions of the estimator. Thus, we will show that the estimator $\hat{\lambda}_n(x;\beta)$ maximizes the (pseudo) loglikelihood function in (4.3.1) over the class of nondecreasing baseline hazard functions if and only if

$$\sum_{j\geqslant i}\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j}-\left[T_{(j+1)}-T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\leqslant 0, \tag{4.3.9}$$

for $i=1,2,\ldots,n-1$, and

$$\sum_{j=1}^{n-1}\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j}-\left[T_{(j+1)}-T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\hat{\lambda}_j=0. \tag{4.3.10}$$

The NPMLE $\hat{\lambda}_n(x;\beta)$ is thus uniquely determined by these Fenchel conditions. The rest of the proof focuses on deriving the Fenchel conditions (4.3.9) and (4.3.10) and on establishing (4.3.6).

First, note that the function φ in (4.3.8) is concave and that the vector of partial derivatives $\nabla\varphi(\lambda)=(\nabla_1\varphi(\lambda),\nabla_2\varphi(\lambda)\ldots,\nabla_{n-1}\varphi(\lambda))$ is given by

$$\nabla\varphi(\lambda)=\left(\frac{\Delta_{(1)}}{\lambda_1}-\left[T_{(2)}-T_{(1)}\right]\sum_{l=2}^{n}\mathrm{e}^{\beta' Z_{(l)}},\ldots,\frac{\Delta_{(n-1)}}{\lambda_{n-1}}-\left[T_{(n)}-T_{(n-1)}\right]\mathrm{e}^{\beta' Z_{(n)}}\right).$$

Define now the functions $g_i(\lambda)=\lambda_{i-1}-\lambda_i$, for $i=1,2,\ldots,n-1$ and $\lambda_0=0$, and the vector $g(\lambda)=(g_1(\lambda),g_2(\lambda),\ldots,g_{n-1}(\lambda))$. Moreover, define the matrix of partial derivatives by

$$G=\left(\frac{\partial g_i(\lambda)}{\partial\lambda_j}\right),\quad \text{for } i=1,2,\ldots,n-1;\ j=1,2,\ldots,n-1. \tag{4.3.11}$$

Let $\tilde{\varphi}(\lambda)=-\varphi(\lambda)$. Then, maximizing (4.3.8) over all $0\leqslant\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_{n-1}$ is equivalent with minimizing $\tilde{\varphi}(\lambda)$ under the restriction that all components of the vector $g(\lambda)$ are negative. An adaptation of the Karush-Kuhn-Tucker theorem (e.g., see Theorem 8.1 in GROENEBOOM, 1998) states that $\hat{\lambda}$ minimizes $\tilde{\varphi}$ over all vectors λ such that $g_i(\lambda)\leqslant 0$, for all $i=1,2,\ldots,n-1$, if and only if the following conditions hold

$$\nabla\tilde{\varphi}(\hat{\lambda})+G^T\alpha=0, \tag{4.3.12}$$

$$g(\hat{\lambda})+w=0, \tag{4.3.13}$$

$$\langle\alpha,w\rangle=0, \tag{4.3.14}$$

for $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_{n-1})$, with $\alpha_i\geqslant 0$, $i=1,2,\ldots,n-1$ and $w=(w_1,w_2,\ldots,w_{n-1})$, with $w_i\geqslant 0$, for $i=1,2,\ldots,n-1$. The first condition (4.3.12), yields that

$$\alpha_i=-\sum_{j\geqslant i}\nabla_j\varphi(\hat{\lambda})=-\sum_{j\geqslant i}\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j}-\left[T_{(j+1)}-T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}. \tag{4.3.15}$$

Since $\alpha_i \geqslant 0$, for all $i = 1,2,\dots,n-1$, condition (4.3.9) is immediate. From (4.3.13), $w = -g(\hat{\lambda}) = (\hat{\lambda}_1-\hat{\lambda}_0, \hat{\lambda}_2-\hat{\lambda}_1, \dots, \hat{\lambda}_{n-1}-\hat{\lambda}_{n-2})$, with $\hat{\lambda}_0 = 0$. Note that the condition $w_i \geqslant 0$ implies that $\hat{\lambda}_{i-1} \leqslant \hat{\lambda}_i$, for all $i = 1,2,\dots,n-1$, which is trivially satisfied. Finally, by (4.3.14),

$$\sum_{i=1}^{n-1} (\hat{\lambda}_i - \hat{\lambda}_{i-1}) \sum_{j \geqslant i} \nabla_j \varphi(\hat{\lambda}) = 0,$$

which re-writes exactly to (4.3.10).

To derive the expression in (4.3.6), we prove first that (4.3.9) and (4.3.10) imply that

$$\sum_{j=1}^{n-1} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} = 0. \tag{4.3.16}$$

Condition (4.3.9) gives that $\sum_{j=1}^{n-1} \nabla_j \varphi(\hat{\lambda}) \leqslant 0$. In addition, as the maximizer $\hat{\lambda}$ is nondecreasing,

$$\begin{aligned} \hat{\lambda}_1 \sum_{j=1}^{n-1} \nabla_j \varphi(\hat{\lambda}) &= - \nabla_2 \varphi(\hat{\lambda}) \hat{\lambda}_2 - \nabla_3 \varphi(\hat{\lambda}) \hat{\lambda}_3 - \dots - \nabla_{n-1} \varphi(\hat{\lambda}) \hat{\lambda}_{n-1} \\ &\quad + \nabla_2 \varphi(\hat{\lambda}) \hat{\lambda}_1 + \nabla_3 \varphi(\hat{\lambda}) \hat{\lambda}_1 + \dots + \nabla_{n-1} \varphi(\hat{\lambda}) \hat{\lambda}_1 \\ &= \sum_{i=2}^{n-1} (\hat{\lambda}_{i-1} - \hat{\lambda}_i) \sum_{j \geqslant i} \nabla_j \varphi(\hat{\lambda}) \geqslant 0. \end{aligned}$$

This shows (4.3.16). Now let $B_1, B_2, \dots, B_k$ be blocks of indices on which $\hat{\lambda}$ is constant such that $B_1 \cup B_2 \cup \dots \cup B_k = \{1,2,\dots,n-1\}$ and let $v_{nj}(\beta)$ be the value of $\hat{\lambda}$ on the block B_j, with $j = 1,2,\dots,k$. If $k = 1$, then the expression of v_{n1} is immediate from (4.3.16). Moreover, observe that, by (4.3.14), $\sum_{i=1}^{n-1} \alpha_i \left(\hat{\lambda}_i - \hat{\lambda}_{i-1} \right) = 0$, and since $\alpha_i \geqslant 0$ and $\hat{\lambda}_i \geqslant \hat{\lambda}_{i-1}$, for any $i = 1,2,\dots,n-1$, it will follow that $\alpha_i = 0$, whenever $\hat{\lambda}_{i-1} < \hat{\lambda}_i$. Hence, for $k \geqslant 2$, there exist $k-1$ α's that are zero. Then (4.3.6) follows by (4.3.15) and (4.3.16). For example, for $k \geqslant 3$, choose any two consecutive α_i that are zero. From (4.3.15), we get that by subtracting these α_i's,

$$\sum_{i \in B_j} \nabla_i \varphi(\hat{\lambda}) = \sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{v_{nj}(\beta)} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} = 0.$$

As $v_{nj}(\beta)$ is constant on B_j, this yields (4.3.6). □

As mentioned beforehand, the proposed unconstrained estimator is thus $\hat{\lambda}_n(x) = \hat{\lambda}_n(x; \hat{\beta}_n)$. Equivalently, on each block of indices B_j, for $j = 1,2,\dots,k$, we propose the estimate $\hat{v}_{nj} = v_{nj}(\hat{\beta}_n)$. Under the null hypothesis $H_0 : \lambda_0(x_0) = \vartheta_0$, the characterization of the constrained maximum likelihood estimator $\hat{\lambda}_n^0$ is provided by the next lemma.

LEMMA 4.2. *Let $x_0 \in (0, \tau_H)$ fixed, such that $T_{(m)} < x_0 < T_{(m+1)}$, for a given $1 \leqslant m \leqslant n-1$. Consider a fixed $\beta \in \mathbb{R}^p$.*

(i) *For $i = 1, 2, \ldots, m$, let $\hat{\lambda}_i^L$ be the left derivative of the GCM at the point P_i^L of the CSD consisting of the points $P_j^L = P_j$, for $j = 1, 2, \ldots, m$, with P_j defined in (4.3.5) and $P_0^L = (0, 0)$. Moreover, for $i = m+1, m+2, \ldots, n-1$, let $\hat{\lambda}_i^R$ be the left derivative of the GCM at the point P_i^R of the CSD consisting of the points $P_j^R = P_j$, for $j = m, m+1, \ldots, n-1$, with P_j defined in (4.3.5). Then, for $\vartheta_0 \in (0, \infty)$, the NPMLE $\hat{\lambda}_n^0(x)$ of a nondecreasing baseline hazard function λ_0, under the null hypothesis $H_0 : \lambda_0 = \vartheta_0$, is of the form*

$$\hat{\lambda}_n^0(x;\beta) = \begin{cases} 0 & x < T_{(1)}, \\ \hat{\lambda}_i^0 & T_{(i)} \leqslant x < T_{(i+1)}, \text{ for } i = 1, 2, \ldots, m-1, m+1, \ldots, n-1, \\ \hat{\lambda}_m^0 & T_{(m)} \leqslant x < x_0, \\ \vartheta_0 & x_0 \leqslant x < T_{(m+1)}, \\ \infty & x \geqslant T_{(n)}, \end{cases} \tag{4.3.17}$$

where $\hat{\lambda}_i^0 = \min(\hat{\lambda}_i^L, \vartheta_0)$, for $i = 1, 2, \ldots, m$, and $\hat{\lambda}_i^0 = \max(\hat{\lambda}_i^R, \vartheta_0)$, for $i = m+1, m+2, \ldots, n-1$.

(ii) *For $k \geqslant 1$, let $B_1^0, B_2^0, \ldots, B_k^0$ be blocks of indices such that $\hat{\lambda}_n^0(x;\beta)$ is constant on each block and $B_1^0 \cup B_2^0 \cup \ldots \cup B_k^0 = \{1, 2, \ldots, n-1\}$. Then, there is one block, say B_r^0, on which $\hat{\lambda}_n^0(x;\beta)$ is equal to ϑ_0, and one block, say B_p^0, that contains m. On all other blocks B_j^0, denote by $v_{nj}^0(\beta)$ the value of $\hat{\lambda}_n^0(x;\beta)$ on block B_j^0. Then,*

$$v_{nj}^0(\beta) = \frac{\sum_{i \in B_j^0} \Delta_{(i)}}{\sum_{i \in B_j^0} \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^n e^{\beta' Z_{(l)}}}, \tag{4.3.18}$$

for $j = 1, \ldots, p-1, p+1, \ldots, k$. On the block B_p^0 that contains m,

$$v_{np}^0(\beta) = \frac{\sum_{i \in B_p^0} \Delta_{(i)}}{\sum_{i \in B_p^0 \setminus \{m\}} \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^n e^{\beta' Z_{(l)}} + [x_0 - T_{(m)}] \sum_{l=m+1}^n e^{\beta' Z_{(l)}}}. \tag{4.3.19}$$

PROOF. We will derive the Karush-Kuhn-Tucker (KKT) conditions, that uniquely determine the constrained NPMLE, and which implicitly provide the characterization in (ii). To prove the lemma, we will show that the estimator proposed in (i) satisfies these conditions.

The constrained NPMLE estimator is obtained by maximizing the objective function (4.3.1) over $0 \leqslant \lambda_0(T_{(1)}) \leqslant \cdots \leqslant \lambda_0(T_{(m)}) \leqslant \vartheta_0 \leqslant \lambda_0(T_{(m+1)}) \leqslant \cdots \leqslant$

$\lambda_0(T_{(n-1)})$. In line with the reasoning for the unconstrained estimator, it can be argued that the constrained estimator has to be a nondecreasing step function that is zero for $x < T_{(1)}$, constant on $[T_{(i)}, T_{(i+1)})$, for $i = 1, 2, \dots, n-1$, is equal to ϑ_0 on the interval $[x_0, T_{(m+1)})$, and can be chosen arbitrarily large for $x \geqslant T_{(n)}$. Therefore, for a fixed $\beta \in \mathbb{R}$, the (pseudo) loglikelihood function in (4.3.1) reduces to

$$\begin{aligned}\sum_{i=1}^{m-1}&\left\{\Delta_{(i)}\log\lambda_0(T_{(i)}) - \lambda_0(T_{(i)})\left[T_{(i+1)} - T_{(i)}\right]\sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}}\right\}\\ &+\Delta_{(m)}\log\lambda_0(T_{(m)}) - \lambda_0(T_{(m)})\left[x_0 - T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\\ &-\vartheta_0\left[T_{(m+1)} - x_0\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\\ &+\sum_{i=m+1}^{n-1}\left\{\Delta_{(i)}\log\lambda_0(T_{(i)}) - \lambda_0(T_{(i)})\left[T_{(i+1)} - T_{(i)}\right]\sum_{l=i+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}.\end{aligned} \tag{4.3.20}$$

By letting $\lambda_i = \lambda_0(T_{(i)})$, for $i = 1, 2, \dots, n-1$ and $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{n-1})$, we then want to maximize

$$\begin{aligned}\varphi^0(\lambda) = \sum_{i=1}^{m-1}&\left\{\Delta_{(i)}\log\lambda_i - \lambda_i\left[T_{(i+1)} - T_{(i)}\right]\sum_{l=i+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\\ &+\Delta_{(m)}\log\lambda_m - \lambda_m\left[x_0 - T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\\ &+\sum_{i=m+1}^{n-1}\left\{\Delta_{(i)}\log\lambda_i - \lambda_i\left[T_{(i+1)} - T_{(i)}\right]\sum_{l=i+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\},\end{aligned} \tag{4.3.21}$$

over the set $0 \leqslant \lambda_1 \leqslant \dots \leqslant \lambda_m \leqslant \vartheta_0 \leqslant \lambda_{m+1} \leqslant \dots \leqslant \lambda_{n-1}$. Let the vector $\hat{\lambda}^c = (\hat{\lambda}_1^c, \hat{\lambda}_2^c, \dots, \hat{\lambda}_{n-1}^c)$ denote the constrained NPMLE under the null hypothesis $H_0 : \lambda_0(x_0) = \vartheta_0$. We will show next that $\hat{\lambda}^c$ maximizes the objective function in (4.3.21) over the class of nondecreasing baseline hazard functions, under the null hypothesis, if and only if the following conditions are satisfied

$$\sum_{j\leqslant i}\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j^c} - \left[T_{(j+1)} - T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\} \geqslant 0, \qquad \text{for } i = 1, 2, \dots, m-1, \tag{4.3.22}$$

$$\begin{aligned}\sum_{j=1}^{m-1}&\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j^c} - \left[T_{(j+1)} - T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\\ &\qquad + \frac{\Delta_{(m)}}{\hat{\lambda}_m^c} - \left[x_0 - T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}} \geqslant 0,\end{aligned} \tag{4.3.23}$$

$$\sum_{j\geqslant i}\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j^c}-\left[T_{(j+1)}-T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\leqslant 0, \tag{4.3.24}$$

for $i=m+1,m+2,\dots,n-1$ and

$$\begin{aligned}\sum_{\substack{j=1\\ j\neq m}}^{n-1}&\left\{\frac{\Delta_{(j)}}{\hat{\lambda}_j^c}-\left[T_{(j+1)}-T_{(j)}\right]\sum_{l=j+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\left(\hat{\lambda}_j^c-\vartheta_0\right)\\ &+\left\{\frac{\Delta_{(m)}}{\hat{\lambda}_m^c}-\left[x_0-T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}\right\}\left(\hat{\lambda}_m^c-\vartheta_0\right)=0.\end{aligned} \tag{4.3.25}$$

The NPMLE $\hat{\lambda}^c$ is thus uniquely determined by these conditions. To prove (i), we will show that $\hat{\lambda}_n^0$ defined in (4.3.17) verifies the Karush-Kuhn-Tucker (KKT) conditions (4.3.22)-(4.3.25). Therefore, $\hat{\lambda}_n^0$ is the unique maximizer of $\varphi^0(\lambda)$ in (4.3.21), over the set $0\leqslant\lambda_1\leqslant\cdots\leqslant\lambda_m\leqslant\vartheta_0\leqslant\lambda_{m+1}\leqslant\cdots\leqslant\lambda_{n-1}$. As it will be seen further, despite bothersome calculations, the distinct form of the likelihood grants a unified framework for deriving the KKT conditions, that uses all the follow-up times, unlike the reasoning in BANERJEE & WELLNER (2001), where the (pseudo) loglikelihood is split and arguments are carried both to the left and to the right of x_0.

Similar to the unconstrained case, observe that the function φ^0 is concave and that the vector of partial derivatives is $\nabla\varphi^0(\lambda)=(\nabla_1\varphi^0(\lambda),\dots,\nabla_m\varphi^0(\lambda),\dots,\nabla_{n-1}\varphi^0(\lambda))$, with

$$\nabla_i\varphi^0(\lambda)=\frac{\Delta_{(i)}}{\lambda_i}-\left[T_{(i+1)}-T_{(i)}\right]\sum_{l=i+1}^{n}\mathrm{e}^{\beta' Z_{(l)}},$$

for $i=1,\dots,m-1,m+1,\dots,n-1$, and

$$\nabla_m\varphi^0(\lambda)=\frac{\Delta_{(m)}}{\lambda_m}-\left[x_0-T_{(m)}\right]\sum_{l=m+1}^{n}\mathrm{e}^{\beta' Z_{(l)}}.$$

Note that the form of $\nabla_m\varphi^0(\lambda)$ differs from the form of $\nabla_i\varphi^0(\lambda)$, for $i=1,2,\dots,m-1,m+1,\dots,n-1$. Moreover, define the vector $g(\lambda)=(g_1(\lambda),g_2(\lambda),\dots,g_{n-1}(\lambda))$, with

$$g_i(\lambda)=\begin{cases}\lambda_i-\lambda_{i+1} & i=1,2,\dots,m-1,\\ \lambda_m-\vartheta_0 & i=m,\\ \vartheta_0-\lambda_{m+1} & i=m+1,\\ \lambda_{i-1}-\lambda_i & i=m+2,\dots,n-1,\end{cases}$$

and consider the matrix of partial derivatives defined in (4.3.11). Computations as in (4.3.15) can be derived to show that condition (4.3.12) yields (4.3.22)-(4.3.24), upon noting that

$$\alpha_i = \begin{cases} \sum_{j\leqslant i} \nabla_j \varphi^0(\hat{\lambda}^c) & i = 1, 2, \ldots, m, \\ -\sum_{j\geqslant i} \nabla_j \varphi^0(\hat{\lambda}^c) & i = m+1, m+2, \ldots, n-1. \end{cases} \tag{4.3.26}$$

Condition (4.3.13) gives that $w = (\hat{\lambda}_2^c - \hat{\lambda}_1^c, \hat{\lambda}_3^c - \hat{\lambda}_2^c, \ldots, \vartheta_0 - \hat{\lambda}_m^c, \hat{\lambda}_{m+1}^c - \vartheta_0, \ldots, \hat{\lambda}_{n-1}^c - \hat{\lambda}_{n-2}^c)$, which together with (4.3.14) and (4.3.26), yields (4.3.25).

Moreover, (4.3.14) gives that

$$\sum_{i=1}^{m-1} \alpha_i \left(\hat{\lambda}_{i+1}^c - \hat{\lambda}_i^c\right) + \alpha_m \left(\vartheta_0 - \hat{\lambda}_m^c\right) + \alpha_{m+1} \left(\hat{\lambda}_{m+1}^c - \vartheta_0\right) + \sum_{m+2}^{n-1} \alpha_i \left(\hat{\lambda}_i^c - \hat{\lambda}_{i-1}^c\right) = 0.$$

Obviously, $\alpha_i = 0$ if $\hat{\lambda}_i^c < \hat{\lambda}_{i+1}^c$, for $i = 1, \ldots, m-1, m+1, \ldots, n-1$ and (4.3.18) can be derived as in the proof of Lemma 4.1. For the block B_p^0 containing m, we get that

$$\sum_{i\in B_p^0\setminus\{m\}} \left\{ \frac{\Delta_{(i)}}{\nu_{np}^0(\beta)} - \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} + \frac{\Delta_{(m)}}{\nu_{np}^0(\beta)} - \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} = 0,$$

which gives exactly (4.3.19). Therefore showing that the estimator $\hat{\lambda}_n^0$ defined in (4.3.17) satisfies the KKT conditions (4.3.22)-(4.3.25) also proves (ii).

Recall that $\hat{\lambda}_n^0$ is $\min(\hat{\lambda}_i^L, \vartheta_0)$, for $i = 1, 2, \ldots, m$, and that $\hat{\lambda}_i^L$ is the unconstrained estimator when considering only the follow-up times $T_{(1)}, T_{(2)}, \ldots, T_{(m)}$. Moreover, $\hat{\lambda}_n^0$ is $\max(\hat{\lambda}_i^R, \vartheta_0)$, for $i = m+1, m+2, \ldots, n-1$, where $\hat{\lambda}_i^R$ can be viewed as the unconstrained estimator when considering only the follow-up times $T_{(m)}, T_{(m+1)}, \ldots, T_{(n-1)}$. Note that (4.3.16) together with (4.3.9) imply that

$$\sum_{j\leqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \geqslant 0, \qquad \text{for } i = 1, 2, \ldots, n-1. \tag{4.3.27}$$

The condition holds for $i = 1, 2, \ldots, m-1$, and, moreover,

$$\sum_{j\leqslant i} \left\{ \frac{\Delta_{(j)}}{\min(\hat{\lambda}_j^L, \vartheta_0)} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \geqslant \sum_{j\leqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j^L} - \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \geqslant 0,$$

for $i = 1, 2, \ldots, m-1$. Therefore, $\min(\hat{\lambda}_i^L, \vartheta)$, for $i = 1, 2, \ldots, m-1$ satisfies (4.3.22). Furthermore, (4.3.27) holds for $i = m$, which implies that

$$\sum_{j=1}^{m-1} \left\{ \frac{\Delta_{(j)}}{\min(\hat{\lambda}_j^L, \vartheta_0)} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} e^{\beta' Z_{(l)}} \right\} + \left\{ \frac{\Delta_{(m)}}{\min(\hat{\lambda}_m^L, \vartheta_0)} - \left[x_0 - T_{(m)} \right] \sum_{l=j+1}^{n} e^{\beta' Z_{(l)}} \right\} \geqslant \sum_{j=1}^{m} \left\{ \frac{\Delta_{(j)}}{\min(\hat{\lambda}_j^L, \vartheta_0)} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} e^{\beta' Z_{(l)}} \right\} \geqslant 0,$$

hence $\hat{\lambda}_n^0$ satisfies (4.3.23) as well. It is straightforward that $\max(\hat{\lambda}_i^R, \vartheta_0)$, for $i = m+1, m+2, \ldots, n-1$ satisfies (4.3.24), since, by definition, $\hat{\lambda}_i^R$ satisfies (4.3.9), for $i = m+1, m+2, \ldots, n-1$, and

$$\sum_{j \geqslant i} \left\{ \frac{\Delta_{(j)}}{\max(\hat{\lambda}_j^R, \vartheta_0)} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} e^{\beta' Z_{(l)}} \right\} \leqslant \sum_{j \geqslant i} \left\{ \frac{\Delta_{(j)}}{\hat{\lambda}_j^R} - \left[T_{(j+1)} - T_{(j)} \right] \sum_{l=j+1}^{n} e^{\beta' Z_{(l)}} \right\} \leqslant 0.$$

Finally, to check if $\hat{\lambda}_n^0$ verifies the condition (4.3.25), we will argue on the blocks of indices on which $\hat{\lambda}_n$, and hence $\hat{\lambda}_i^L$ and $\hat{\lambda}_i^R$ are constant. By (4.3.6), for each block B_j, with $j = 1, 2, \ldots, k$, on which the unconstrained estimator has the constant value $v_{nj}(\beta)$,

$$\sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{v_{nj}(\beta)} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} e^{\beta' Z_{(l)}} \right\} v_{nj}(\beta) = 0,$$

and

$$\sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{v_{nj}(\beta)} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} e^{\beta' Z_{(l)}} \right\} = 0.$$

Then, on each block B_j that does not contain m, we can write

$$\sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{\hat{\lambda}_i} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} e^{\beta' Z_{(l)}} \right\} \hat{\lambda}_i = \vartheta_0 \sum_{i \in B_j} \left\{ \frac{\Delta_{(i)}}{\hat{\lambda}_i} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} e^{\beta' Z_{(l)}} \right\}, \tag{4.3.28}$$

and this holds for $\hat{\lambda}_i^L$, as well as for $\hat{\lambda}_i^R$. It is straightforward that $\min(\hat{\lambda}_i^L, \vartheta_0)$, for $i = 1, 2, \ldots, m$ and $\max(\hat{\lambda}_i^R, \vartheta_0)$, for $i = m+1, m+2, \ldots, n-1$ satisfy this relationship. For the block B_p that contains m, we have

$$\begin{aligned}
&\sum_{i \in B_p \setminus \{m\}} \left\{ \frac{\Delta_{(i)}}{\hat{\lambda}_i^L} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \hat{\lambda}_i^L \\
&\quad + \left\{ \frac{\Delta_{(m)}}{\hat{\lambda}_m^L} - \left[T_{(m+1)} - x_0 \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} - \left[x_0 - T_{(m)} \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \hat{\lambda}_m^L \\
&= \vartheta_0 \sum_{i \in B_p \setminus \{m\}} \left\{ \frac{\Delta_{(i)}}{\hat{\lambda}_i^L} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \\
&\quad + \vartheta_0 \left\{ \frac{\Delta_{(m)}}{\hat{\lambda}_m} - \left[T_{(m+1)} - x_0 \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} - \left[x_0 - T_{(m)} \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}.
\end{aligned}$$

Constraining $\hat{\lambda}_m^L$ to be ϑ_0 on the interval $[x_0, T_{(m+1)})$ yields

$$\begin{aligned}
&\sum_{i \in B_p \setminus \{m\}} \left\{ \frac{\Delta_{(i)}}{\hat{\lambda}_i^L} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \hat{\lambda}_i^L \\
&\quad + \left\{ \frac{\Delta_{(m)}}{\hat{\lambda}_m^L} - \left[x_0 - T_{(m)} \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \hat{\lambda}_m^L \\
&= \vartheta_0 \sum_{i \in B_p \setminus \{m\}} \left\{ \frac{\Delta_{(i)}}{\hat{\lambda}_i^L} - \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\} \\
&\quad + \vartheta_0 \left\{ \frac{\Delta_{(m)}}{\hat{\lambda}_m^L} - \left[x_0 - T_{(m)} \right] \sum_{l=m+1}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}.
\end{aligned} \tag{4.3.29}$$

Once more, for $i \in B_p$, $\min(\hat{\lambda}_i^L, \vartheta_0)$ satisfies this relationship. Summing over all blocks in (4.3.28) and (4.3.29) completes the proof. □

Similar to the unconstrained estimator, we propose $\hat{\lambda}_n^0(x) = \hat{\lambda}_n^0(x; \hat{\beta}_n)$ as the constrained estimator and $\hat{v}_{nj}^0 = v_{nj}^0(\hat{\beta}_n)$, where $\hat{\beta}_n$ is the maximum partial likelihood estimator.

REMARK. As already pointed out in Chapter 2, if we take $\beta_0 = 0$, the characterization of the unconstrained estimator differs slightly from the characterization of the nondecreasing hazard estimator within the ordinary random censorship model, provided in HUANG & WELLNER (1995). Correspondingly, the characterizations in Lemma 4.1 and 4.2, with $\beta_0 = 0$ differ from the characterizations furnished in BANERJEE (2008).

Although the estimators in BANERJEE (2008) do not maximize the (pseudo) loglikelihood function in (4.3.1) (in the absence of covariates and under the null hypothesis) over nondecreasing λ_0, the asymptotic distribution of the likelihood ratio test based on these estimators will coincide with our proposed distribution, in the case of no covariates.

Using the notations in BANERJEE (2008), let $\mathrm{slogcm}(f, I)$ be the left-hand slope of the greatest convex minorant of the restriction of the real-valued function f to the interval I. Denote by $\mathrm{slogcm}(f) = \mathrm{slogcm}(f, \mathbb{R})$. Moreover, let

$$\mathrm{slogcm}^0(f) = \min\left(\mathrm{slogcm}(f, (-\infty, 0]), 0\right) 1_{(-\infty,0]} + \max\left(\mathrm{slogcm}(f, (0, \infty)), 0\right) 1_{(0,\infty)}.$$

Furthermore, for positive constants a and b, define

$$X_{a,b}(t) = a\mathbb{W}(t) + bt^2, \tag{4.3.30}$$

where $\mathbb{W}$ is a standard two-sided Brownian motion originating from zero. Let

$$g_{a,b}(t) = \mathrm{slogcm}(X_{a,b})(t), \tag{4.3.31}$$

the left-hand slope of the GCM $G_{a,b}$ of the process $X_{a,b}$, at point t. The constrained analogous is defined as follows: for $t \leqslant 0$, construct the GCM of $X_{a,b}$, that will be denoted by $G^L_{a,b}$ and take its left-hand slopes at point t, denoted by $D_L(X_{a,b})(t)$. When the slopes exceed zero, replace them by zero. In the same manner, for $t > 0$, denote the GCM of $X_{a,b}$ by $G^R_{a,b}$ and its slopes at point t by $D_R(X_{a,b})(t)$. Replace the slopes by zero when they decrease below zero. This slope process will be denoted by $g^0_{a,b}$, which is thus given by

$$g^0_{a,b}(t) = \begin{cases} \min\left(D_L(X_{a,b})(t), 0\right) & t < 0, \\ 0 & t = 0, \\ \max\left(D_R(X_{a,b})(t), 0\right) & t > 0. \end{cases} \tag{4.3.32}$$

Note that for $t \leqslant 0$, there exists, almost surely $s < 0$ such that $D_L(X_{a,b})(s)$ is strictly positive for any point greater than or equal to s and the left derivative at s is non-positive. Equivalently, for $t > 0$ there exists almost surely $s > 0$ such that $D_R(X_{a,b})(s)$ is strictly negative for any point smaller than or equal to s and the left derivative at s is non-negative. In addition, observe that $g^0_{a,b}(t) = \mathrm{slogcm}^0(X_{a,b})(t)$, as defined and characterized in BANERJEE & WELLNER (2001).

4.3.2 NONINCREASING BASELINE HAZARD

The characterization of the unconstrained and the constrained NPMLE estimators of a nonincreasing baseline hazard function follows analogously to the characterization of the nondecreasing estimators. The unconstrained NPMLE $\hat{\lambda}_n(x; \beta)$ is obtained by maximizing the (pseudo) likelihood function in (4.3.1) over all $\lambda_0(T_{(1)}) \geqslant \lambda_0(T_{(2)}) \geqslant$

$\cdots \geqslant \lambda(T_{(n)}) \geqslant 0$. As derived in Chapter 2, the likelihood is maximized by a nonincreasing step function that is constant on $(T_{(i-1)}, T_{(i)}]$, for $i = 1, 2, \ldots, n$ and where $T_{(0)} = 0$. The (pseudo) loglikelihood in (4.3.1) becomes then

$$\sum_{i=1}^{n} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i)} - T_{(i-1)} \right] \sum_{l=i}^{n} \mathrm{e}^{\beta' Z_{(l)}} \right\}. \tag{4.3.33}$$

The lemmas below provide the characterization of the unconstrained estimator $\hat{\lambda}_n(x;\beta)$ and the constrained estimator $\hat{\lambda}_n^0(x;\beta)$. Their proofs follow by arguments similar to those in the proofs of Lemma 4.1 and Lemma 4.2, as well as the necessary and sufficient conditions that uniquely characterize these estimators..

LEMMA 4.3. *Let $T_{(1)} < T_{(2)} < \cdots < T_{(n)}$ be the ordered follow-up times and consider a fixed $\beta \in \mathbb{R}^p$.*

(i) Let W_n be defined in (4.3.3) *and let*

$$\overline{V}_n(x) = \int \delta\{u \leqslant x\}\, \mathrm{d}\mathbb{P}_n(u, \delta, z). \tag{4.3.34}$$

Then, the NPMLE $\hat{\lambda}_n(x;\beta)$ of a nonincreasing baseline hazard function λ_0 is given by

$$\hat{\lambda}_n(x;\beta) = \begin{cases} \hat{\lambda}_i & T_{(i-1)} < x \leqslant T_{(i)}, \text{ for } i = 1, 2, \ldots, n, \\ 0 & x > T_{(n)}, \end{cases}$$

for $i = 1, 2, \ldots, n$, with $T_{(0)} = 0$ and where $\hat{\lambda}_i$ is the left derivative of the least concave majorant (LCM) at the point P_i of the cumulative sum diagram consisting of the points

$$P_j = \Big(W_n(\beta, T_{(j)}), \overline{V}_n(T_{(j)}) \Big), \tag{4.3.35}$$

for $j = 1, 2, \ldots, n$ and $P_0 = (0, 0)$.

(ii) Let $B_1, B_2, \ldots, B_k$ be blocks of indices such that $\hat{\lambda}_n(x;\beta)$ is constant on each block and $B_1 \cup B_2 \cup \ldots \cup B_k = \{1, 2, \ldots, n\}$. Denote by $v_{nj}(\beta)$, the value of the estimator on block B_j. Then

$$v_{nj}(\beta) = \frac{\sum_{i \in B_j} \Delta_{(i)}}{\sum_{i \in B_j} \left[T_{(i)} - T_{(i-1)} \right] \sum_{l=i}^{n} e^{\beta' Z_{(l)}}}.$$

In fact, for $x \geqslant T_{(n)}$, $\hat{\lambda}_n(x;\beta)$ can take any value smaller than $\hat{\lambda}_n$, the left derivative of the LCM at the point P_n of the CSD. As before, we propose $\hat{\lambda}_n(x) = \hat{\lambda}_n(x;\hat{\beta}_n)$ as the estimator of λ_0 and $\hat{v}_{nj} = v_{nj}(\hat{\beta}_n)$, where $\hat{\beta}_n$ denotes the maximum partial likelihood estimator of β_0. Fenchel conditions as in (4.3.9) and (4.3.10) can be derived analogously.

The NPMLE estimator $\hat{\lambda}_n^0$ maximizes the (pseudo) loglikelihood function in (4.3.33) over the set $\lambda_0(T_{(1)}) \geqslant \cdots \geqslant \lambda_0(T_{(m)}) \geqslant \vartheta_0 \geqslant \lambda_0(T_{(m+1)}) \geqslant \cdots \geqslant \lambda_0(T_{(n)}) \geqslant 0$. It can be argued that the constrained estimator has to be a nonincreasing step function that is constant on $(T_{(i-1)}, T_{(i)}]$, for $i = 1, 2, \ldots, n$, is ϑ_0 on the interval $(T_{(m)}, x_0]$, and is zero for $x \geqslant T_{(n)}$. Hence, the (pseudo) loglikelihood function becomes

$$\begin{aligned}
&\sum_{i=1}^{m} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i)} - T_{(i-1)} \right] \sum_{l=i}^{n} e^{\beta' Z_{(l)}} \right\} \\
&+ \Delta_{(m+1)} \log \lambda_0(T_{(m+1)}) - \vartheta_0 \left[x_0 - T_{(m)} \right] \sum_{l=m+1}^{n} e^{\beta' Z_{(l)}} \\
&- \lambda_0(T_{(m+1)}) \left[T_{(m+1)} - x_0 \right] \sum_{l=m+1}^{n} e^{\beta' Z_{(l)}} \\
&+ \sum_{i=m+2}^{n} \left\{ \Delta_{(i)} \log \lambda_0(T_{(i)}) - \lambda_0(T_{(i)}) \left[T_{(i)} - T_{(i-1)} \right] \sum_{l=i}^{n} e^{\beta' Z_{(l)}} \right\}.
\end{aligned}$$

The characterization of the constrained NPMLE $\hat{\lambda}_n^0$ is provided with the next lemma.

LEMMA 4.4. *Let $x_0 \in (0, \tau_H)$ fixed, such that $T_{(m)} < x_0 < T_{(m+1)}$, for a given $1 \leqslant m \leqslant n-1$. Consider a fixed $\beta \in \mathbb{R}^p$.*

(i) *For $i = 1, 2, \ldots, m$, let $\hat{\lambda}_i^L$ to be the left derivative of the LCM at the point P_i^L of the CSD consisting of the points $P_j^L = P_j$, for $j = 1, 2, \ldots, m$, with P_j defined in* (4.3.35)*, and $P_0^L = (0,0)$. Moreover, for $i = m+1, m+2, \ldots, n$, let $\hat{\lambda}_i^R$ be the left derivative of the LCM at the point P_i^R of the CSD consisting of the points $P_j^R = P_j$, for $j = m, m+1 \ldots, n$, with P_j defined in* (4.3.35)*. Then, the NPMLE $\hat{\lambda}_n^0(x;\beta)$ of a nonincreasing baseline hazard function λ_0, under the null hypothesis $H_0 : \lambda_0 = \vartheta_0$, is given by*

$$\hat{\lambda}_n^0(x;\beta) = \begin{cases} \hat{\lambda}_i^0 & T_{(i-1)} < x \leqslant T_{(i)}, \text{ for } i = 1, 2, \ldots, m, m+2, \ldots, n, \\ \vartheta_0 & T_{(m)} < x \leqslant x_0, \\ \hat{\lambda}_{m+1}^0 & x_0 < x \leqslant T_{(m+1)}, \\ 0 & x > T_{(n)}, \end{cases} \tag{4.3.36}$$

where $T_{(0)} = 0$ and where $\hat{\lambda}_i^0 = \max(\hat{\lambda}_i^L, \vartheta_0)$, for $i = 1, 2, \ldots, m$, and $\hat{\lambda}_i^0 = \min(\hat{\lambda}_i^R, \vartheta_0)$, for $i = m+1, m+2, \ldots, n$.

(ii) *For $k \geqslant 1$, let $B_1^0, B_2^0, \ldots, B_k^0$ be blocks of indices such that $\hat{\lambda}_n^0(x;\beta)$ is constant on each block and $B_1^0 \cup B_2^0 \cup \ldots \cup B_k^0 = \{1, 2, \ldots, n\}$. There is one block, say B_r^0, on which $\hat{\lambda}_n^0(x;\beta)$ is ϑ_0, and one block, say B_p^0, that contains $m+1$. On all other blocks B_j^0, denote by $v_{nj}^0(\beta)$ the value of $\hat{\lambda}_n^0(x;\beta)$ on block B_j^0. Then,*

$$v_{nj}^0(\beta) = \frac{\sum_{i \in B_j^0} \Delta_{(i)}}{\sum_{i \in B_j^0} \left[T_{(j)} - T_{(j-1)} \right] \sum_{l=j}^n e^{\beta' Z_{(l)}}}.$$

On the block B_p^0, that contains $m+1$,

$$v_{np}^0(\beta) = \frac{\sum_{i \in B_p^0} \Delta_{(i)}}{\sum_{i \in B_p^0 \setminus \{m+1\}} \left[T_{(i)} - T_{(i-1)} \right] \sum_{l=i+1}^n e^{\beta' Z_{(l)}} + \left[T_{(m+1)} - x_0 \right] \sum_{l=m+1}^n e^{\beta' Z_{(l)}}}.$$

Evidently, we propose $\hat{\lambda}_n^0(x) = \hat{\lambda}_n^0(x;\hat{\beta}_n)$ as the constrained estimator of a nonincreasing baseline hazard function λ_0, as well as $\hat{v}_{nj}^0 = v_{nj}^0(\hat{\beta}_n)$ on blocks of indices where the estimator is constant. The Fenchel conditions corresponding to (4.3.22)-(4.3.25) can be derived in the same manner as for the constrained estimator in the nondecreasing case.

Let $\text{slolcm}(f, I)$ be the left-hand slope of the LCM of the restriction of the real-valued function f to the interval I. Denote by $\text{slolcm}(f) = \text{slolcm}(f, \mathbb{R})$. For $a, b > 0$, let $\overline{X}_{a,b}(t) = a\mathbb{W}(t) - bt^2$, where $\mathbb{W}$ is a standard two-sided Brownian motion originating from zero. Denote by $L_{a,b}$ the LCM of $\overline{X}_{a,b}$ and let

$$l_{a,b}(t) = \text{slolcm}(\overline{X}_{a,b})(t), \tag{4.3.37}$$

be the left-hand slope of $L_{a,b}$, at point t. Additionally, set

$$\text{slolcm}^0(f) = \max\left(\text{slolcm}(f, (-\infty, 0]), 0\right) 1_{(-\infty,0]} + \min\left(\text{slolcm}(f, (0,\infty)), 0\right) 1_{(0,\infty)}.$$

For $t \leqslant 0$, construct the LCM of $\overline{X}_{a,b}$, that will be denoted by $L_{a,b}^L$ and take its left-hand slope at point t, denoted by $D_L(\overline{X}_{a,b})(t)$. When the slopes fall behind zero, replace them by zero. In the same manner, for $t > 0$, denote the LCM of $\overline{X}_{a,b}$ by $L_{a,b}^R$ and its slope at point t by $D_R(\overline{X}_{a,b})(t)$. Replace the slopes by zero when they exceed zero. This slope process will be denoted by $l_{a,b}^0$, which is thus given by

$$l_{a,b}^0(t) = \begin{cases} \max\left(D_L(\overline{X}_{a,b})(t), 0\right) & t < 0, \\ 0 & t = 0, \\ \min\left(D_R(\overline{X}_{a,b})(t), 0\right) & t > 0. \end{cases} \tag{4.3.38}$$

Observe that $l_{a,b}^0(t) = \text{slolcm}^0(\overline{X}_{a,b})(t)$.

4.4 THE LIMIT DISTRIBUTION

Let $\mathbf{B}_{loc}(\mathbb{R})$ be the space of all locally bounded real functions on $\mathbb{R}$, equipped with the topology of uniform convergence on compact sets. In addition, $\mathbb{C}_{min}(\mathbb{R})$ is defined as the subset of $\mathbf{B}_{loc}(\mathbb{R})$ consisting of continuous functions f for which $f(t) \to \infty$, when $|t| \to \infty$ and f has a unique minimum. Let $\mathscr{L}$ be the space of locally square integrable real-valued functions on $\mathbb{R}$, equipped with the topology of L_2 convergence on compact sets.

For a generic follow-up time T, consider $H^{uc}(x) = \mathbb{P}(T \leqslant x, \Delta = 1)$, the sub-distribution function of the uncensored observations. Moreover, let

$$\Phi(\beta, x) = \int \{u \geqslant x\} \, e^{\beta' z} \, dP(u, \delta, z), \tag{4.4.1}$$

for $\beta \in \mathbb{R}^p$ and $x \in \mathbb{R}$, where P is the underlying probability measure corresponding to the distribution of (T, Δ, Z). For a fixed point $x_0 \in (0, \tau_H)$, define the processes

$$\begin{aligned} X_n(x) &= n^{1/3}\left(\hat{\lambda}_n(x_0 + n^{-1/3}x) - \vartheta_0\right), \\ Y_n(x) &= n^{1/3}\left(\hat{\lambda}_n^0(x_0 + n^{-1/3}x) - \vartheta_0\right). \end{aligned} \tag{4.4.2}$$

The following lemma provides the joint asymptotic distribution of the above processes.

LEMMA 4.5. *Assume (A1) and (A2) and let $x_0 \in (0, \tau_H)$. Suppose that λ_0 is nondecreasing on $[0, \infty)$ and continuously differentiable in a neighborhood of x_0, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) > 0$. Moreover, assume that the functions $x \mapsto \Phi(\beta_0, x)$ and $H^{uc}(x)$, defined in* (4.4.1) *and above* (4.4.1)*, are continuously differentiable in a neighborhood of x_0. Finally, assume that the density of the follow-up times is continuous and bounded away from zero in a neighborhood of x_0. Define*

$$a = \sqrt{\frac{\lambda_0(x_0)}{\Phi(\beta_0, x_0)}} \qquad \text{and} \qquad b = \frac{1}{2}\lambda_0'(x_0). \tag{4.4.3}$$

Then (X_n, Y_n) converges jointly to $(g_{a,b}, g_{a,b}^0)$, in $\mathscr{L} \times \mathscr{L}$, where the processes $g_{a,b}$ and $g_{a,b}^0$ have been defined in (4.3.31) *and* (4.3.32).

PROOF. Note that the processes X_n and Y_n are monotone. By making use of Corollary 2 in HUANG & ZHANG (1994) and the remark above the corollary, it suffices to prove that the finite dimensional marginals of the process (X_n, Y_n) converge to the finite dimensional marginals of the process $(g_{a,b}, g_{a,b}^0)$, in order to prove the lemma.

For $x \geqslant T_{(1)}$, let

$$\hat{W}_n(x) = W_n(\hat{\beta}_n, x) - W_n(\hat{\beta}_n, T_{(1)}),$$

where W_n is defined in (4.3.3), and where $\hat{\beta}_n$ is the maximum partial likelihood estimator. For fixed x_0 and $x \in [-k,k]$, with $0 < k < \infty$, define the process

$$\mathbb{Z}_n(x) = \frac{n^{2/3}}{\Phi(\beta_0, x_0)} \Big\{ V_n(x_0 + n^{-1/3}x) - V_n(x_0) \\ - \lambda_0(x_0)\left[\hat{W}_n(x_0 + n^{-1/3}x) - \hat{W}_n(x_0)\right]\Big\}, \tag{4.4.4}$$

where V_n is defined in (4.3.4). For a and b defined in (4.4.3), $\mathbb{Z}_n$ converges weakly to $X_{a,b}$, as processes in $\mathbf{B}_{loc}(\mathbb{R})$, by Lemma 2.14 in Chapter 2. Define now

$$S_n(x) = \frac{n^{1/3}}{\Phi(\beta_0, x_0)} \left\{ \hat{W}_n(x_0 + n^{-1/3}x) - \hat{W}_n(x_0) \right\}. \tag{4.4.5}$$

From the proof of Lemma 2.15 in Chapter 2, $S_n(x)$ converges almost surely to the deterministic function x, uniformly on every compact set.

Following the approach in GROENEBOOM (1985), LOPUHAÄ & NANE (2013) obtained the asymptotic distribution of the unconstrained maximum likelihood estimator $\hat{\lambda}_n$ by considering the inverse process

$$U_n(z) = \underset{x\in[T_{(1)},T_{(n)}]}{\operatorname{argmin}} \left\{ V_n(x) - z\hat{W}_n(x) \right\}, \tag{4.4.6}$$

for $z > 0$, where the argmin function represents the supremum of times at which the minimum is attained. Since the argmin is invariant under addition of and multiplication with positive constants, it follows that

$$n^{1/3}\left[U_n(\vartheta_0 + n^{-1/3}z) - x_0\right] = \underset{x\in I_n(x_0)}{\operatorname{argmin}} \left\{ \mathbb{Z}_n(x) - S_n(x)z \right\},$$

where $I_n(x_0) = [-n^{1/3}(x_0 - T_{(1)}), n^{1/3}(T_{(n)} - x_0)]$. For $z > 0$, the switching relationship $\hat{\lambda}_n(x) \leqslant z$ holds if and only if $U_n(z) \geqslant x$, with probability one. This translates, in the context of this lemma, to

$$n^{1/3}\left[\hat{\lambda}_n(x_0 + n^{-1/3}x) - \vartheta_0\right] \leqslant z \Leftrightarrow n^{1/3}\left[U_n(\vartheta_0 + n^{-1/3}z) - x_0\right] \geqslant x,$$

for $0 < x_0 < \tau_H$ and $\vartheta_0 > 0$, with probability one. The switching relationship is thus $X_n(x) \leqslant z \Leftrightarrow n^{1/3}\left[U_n(\vartheta_0 + n^{-1/3}z) - x_0\right] \geqslant x$. Hence finding the limiting distribution of $X_n(x)$ resumes to finding the limiting distribution of $n^{1/3}\left[U_n(\vartheta_0 + n^{-1/3}z) - x_0\right]$. By applying Theorem 2.7 in KIM & POLLARD (1990), it follows that, for every $z > 0$,

$$n^{1/3}\left[U_n(\vartheta_0 + n^{-1/3}z) - x_0\right] \xrightarrow{d} U(z),$$

as inferred in the proof of Theorem 2.4 in Chapter 2, where $U(z) = \sup\{t \in \mathbb{R} : X_{a,b}(t) - zt \text{ is minimal}\}$. It will result that, for every $x \in [-k, k]$,

$$\begin{aligned} P\left(X_n(x) \leqslant z\right) &= P\left(n^{1/3}\left[\hat{\lambda}_n(x_0 + n^{-1/3}x) - \vartheta_0\right] \leqslant z\right) \\ &= P\left(n^{1/3}\left[U_n(\vartheta_0 + n^{-1/3}z) - x_0\right] \geqslant x\right) \\ &\to P\left(U(z) \geqslant x\right). \end{aligned}$$

Using the switching relationship on the limiting process, it can be deduced that $U(z) \geqslant x \Leftrightarrow g_{a,b}(x) \leqslant z$, with probability one, and thus $X_n(x) \xrightarrow{d} g_{a,b}(x)$.

In order to prove the same type of result for $Y_n(x)$, consider first the following process

$$\widetilde{Y}_n(x) = n^{1/3}\left(\tilde{\lambda}_n(x_0 + n^{-1/3}x) - \vartheta_0\right), \tag{4.4.7}$$

where, for $x_0 \in (0, \tau_H)$, such that $T_{(m)} < x_0 < T_{(m+1)}$,

$$\tilde{\lambda}_n(x) = \begin{cases} 0 & x < T_{(1)}, \\ \hat{\lambda}_i^L & T_{(i)} \leqslant x < T_{(i+1)}, \text{ for } i = 1, 2, \ldots, m-1 \\ \hat{\lambda}_m^L & T_{(m)} \leqslant x < x_0, \\ 0 & x_0 \leqslant x < T_{(m+1)}, \\ \hat{\lambda}_i^R & T_{(i)} \leqslant x < T_{(i+1)}, \text{ for } i = m+1, m+2, \ldots, n-1 \\ \infty & x \geqslant T_{(n)}, \end{cases}$$

with $\hat{\lambda}_i^L$ and $\hat{\lambda}_i^R$ defined in Lemma 4.2. For this, we have considered up to x_0 an unconstrained estimator which is constructed based on the sample points $T_{(1)}, T_{(2)}, \ldots, T_{(m+1)}$. Moreover, to the right of x_0, we have considered an unconstrained estimator based on the points $T_{(m+1)}, T_{(m+2)}, \ldots, T_{(n)}$. It is not difficult to see that

$$Y_n(x) = \begin{cases} \min\left(\widetilde{Y}_n(x), 0\right) & x < 0, \\ 0 & x = 0, \\ \max\left(\widetilde{Y}_n(x), 0\right) & x > 0. \end{cases} \tag{4.4.8}$$

For $z > 0$, define the inverse processes

$$\begin{aligned} U_n^L(z) &= \underset{x \in [T_{(1)}, T_{(m+1)}]}{\operatorname{argmin}} \left\{V_n(x) - z\widehat{W}_n(x)\right\}, \\ U_n^R(z) &= \underset{x \in [T_{(m+1)}, T_{(n)}]}{\operatorname{argmin}} \left\{V_n(x) - z\widehat{W}_n(x)\right\} \end{aligned}$$

Take $x < x_0$. The switching relationship for $\tilde{\lambda}_n$ is given by $\tilde{\lambda}_n(x) \leqslant z$ if and only if $U_n^L(z) \geqslant x$, with probability one, which gives that

$$n^{1/3}\left[\tilde{\lambda}_n(x_0 + n^{-1/3}x) - \vartheta_0\right] \leqslant z \Leftrightarrow n^{1/3}\left[U_n^L(\vartheta_0 + n^{-1/3}z) - x_0\right] \geqslant x,$$

with probability one. Moreover,

$$n^{1/3}\left[U_n^L(\vartheta_0+n^{-1/3}z)-x_0\right]=\underset{x\in I_n^L(x_0)}{\operatorname{argmin}}\left\{\mathbb{Z}_n(x)-S_n(x)z\right\},$$

where $I_n^L(x_0)=[-n^{1/3}(x_0-T_{(1)}),n^{1/3}(T_{(m+1)}-x_0)]$. Denote by

$$Z_n(z,x)=\mathbb{Z}_n(x)-S_n(x)z.$$

As for the unconstrained estimator, we aim to apply Theorem 2.7 in KIM & POLLARD (1990). As Theorem 2.7 in KIM & POLLARD (1990) applies to the argmax of processes on the whole real line, we extend the above process in the following manner

$$Z_n^-(z,x)=\begin{cases}Z_n(z,-n^{1/3}(x_0-T_{(1)})) & x<-n^{1/3}(x_0-T_{(1)}),\\ Z_n(z,x) & -n^{1/3}(x_0-T_{(1)})\leqslant x\leqslant n^{1/3}(T_{(m+1)}-x_0),\\ Z_n(z,n^{1/3}(T_{(m+1)}-x_0))+1 & x>n^{1/3}(T_{(m+1)}-x_0).\end{cases}$$

Then, $Z_n^-(z,x)\in\mathbf{B}_{loc}(\mathbb{R})$ and

$$n^{1/3}\left[U_n^L(\vartheta_0+n^{-1/3}z)-x_0\right]=\underset{x\in\mathbb{R}}{\operatorname{argmin}}\left\{Z_n^-(z,x)\right\}=\underset{x\in\mathbb{R}}{\operatorname{argmax}}\left\{-Z_n^-(z,x)\right\}.$$

Since $\lambda_0(x_0)=\vartheta_0>0$ and λ_0 is continuously differentiable in a neighborhood of x_0, it follows by a Taylor expansion and by Lemma 2.5 in DEVROYE (1981) that $n^{1/3}(T_{(m+1)}-x_0)=\mathcal{O}_p(n^{-1}\log n)$. Therefore, by virtue of Lemma 2.14 and Lemma 2.15 in Chapter 2, the process $x\mapsto -Z_n^-(z,x)$ converges weakly to $Z^-(x)\in\mathbb{C}_{max}(\mathbb{R})$, for any fixed z, where

$$Z^-(x)=\begin{cases}-X_{a,b}(x)+zx & x\leqslant 0,\\ 1 & x>0,\end{cases}$$

for a and b defined in (4.4.3). Hence, the first condition of Theorem 2.7 in KIM & POLLARD (1990) is verified. The second condition follows directly from Lemma 2.17 in Chapter 2, while the third condition is trivially fulfilled. Thus, for any z fixed,

$$n^{1/3}\left[U_n^L(\vartheta_0+n^{-1/3}z)-x_0\right]\xrightarrow{d}U^-(z),$$

where $U^-(z)=\sup\left\{t\leqslant 0: X_{a,b}(t)-zt \text{ is minimal}\right\}$. Concluding, for $x<0$,

$$\begin{aligned}P\left(\widetilde{Y}_n(x)\leqslant z\right)&=P\left(n^{1/3}\left[\tilde{\lambda}_n(x_0+n^{-1/3}x)-\vartheta_0\right]\leqslant z\right)\\&=P\left(n^{1/3}\left[U_n^L(\vartheta_0+n^{-1/3}z)-x_0\right]\geqslant x\right)\\&\to P\left(U^-(z)\geqslant x\right).\end{aligned}$$

The switching relationship for the limiting process gives that $U^-(z) \geqslant x \Leftrightarrow D_L(X_{a,b})(x) \leqslant z$, with probability one, where $D_L(X_{a,b})(x)$ has been defined as the left-hand slope of the GCM of $X_{a,b}$, at a point $x < 0$. Hence, for $x < 0$,

$$\widetilde{Y}_n(x) \xrightarrow{d} D_L(X_{a,b})(x).$$

Completely analogous, $\widetilde{Y}_n(x) \xrightarrow{d} D_R(X_{a,b})(x)$, for $x > 0$. By continuous mapping theorem and by (4.4.8), it can be concluded that for fixed $x \in [-k, k]$,

$$Y_n(x) \xrightarrow{d} g^0_{a,b}(x),$$

where $g^0_{a,b}$ has been defined in (4.3.32).

Our next objective is to apply Theorem 6.1 in HUANG & WELLNER (1995). The first condition of Theorem 6.1 is trivially fulfilled. The second condition follows by Lemma 2.17 in Chapter 2, while the third condition follows by the definition of the inverse processes. Hence, for fixed x,

$$P\left(X_n(x) \leqslant z, Y_n(x) \leqslant z\right) \to P\left(g_{a,b}(x) \leqslant z, g^0_{a,b}(x) \leqslant z\right),$$

for a and b defined in (4.4.3). The arguments for one dimensional marginal convergence can be extended to the finite dimensional convergence, as in the proof of Theorem 3.6.2 in BANERJEE (2001), by making use of Lemma 3.6.10 in BANERJEE (2001). Hence, we can conclude that the finite dimensional marginals of the process (X_n, Y_n) converge to the finite dimensional marginals of the process $(g_{a,b}, g^0_{a,b})$. This completes the proof. □

By making use of results in Chapter 2, a completely similar result holds in the nonincreasing setting.

LEMMA 4.6. *Assume (A1) and (A2) and let $x_0 \in (0, \tau_H)$. Suppose that λ_0 is nonincreasing on $[0, \infty)$ and continuously differentiable in a neighborhood of x_0, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) < 0$. Moreover, assume that the functions $x \to \Phi(\beta_0, x)$ and $H^{uc}(x)$, defined in* (4.4.1) *and above* (4.4.1)*, are continuously differentiable in a neighborhood of x_0. Then, for a and b defined in* (4.4.3)*, (X_n, Y_n) converge jointly to $\left(l_{a,b}, l^0_{a,b}\right)$ in $\mathscr{L} \times \mathscr{L}$, where the processes $l_{a,b}$ and $l^0_{a,b}$ have been defined in* (4.3.37) *and* (4.3.38).

Subsequently, we state two immediate results, that will be used repeatedly throughout the rest of the paper.

LEMMA 4.7. *Let $x_0 \in (0, \tau_H)$ fixed and let $\overline{D}_n$ be the set on which the unconstrained NPMLE $\hat{\lambda}_n$, defined in Lemma 4.1, differs from the constrained NPMLE $\hat{\lambda}^0_n$, defined in Lemma 4.2. Then, for any $\varepsilon > 0$, there exists $k_\varepsilon > 0$ such that*

$$\liminf_{n\to\infty} P\left(\overline{D}_n \subset [x_0 - n^{-1/3}k_\varepsilon, x_0 + n^{-1/3}k_\varepsilon]\right) \geqslant 1 - \varepsilon.$$

PROOF. The proof of this fact follows by exactly the same reasoning as in the proof of Lemma 2.6 in BANERJEE (2006), preprint for BANERJEE (2007). □

LEMMA 4.8. *Consider the processes X_n and Y_n defined in* (4.4.2). *Then, for every $\varepsilon > 0$ and $k > 0$, there exists an $M > 0$ such that*

$$\limsup_{n\to\infty} P\left(\sup_{x\in[-k,k]} \left|X_n(x)\right| > M\right) \leqslant \varepsilon.$$

Similarly,

$$\limsup_{n\to\infty} P\left(\sup_{x\in[-k,k]} \left|Y_n(x)\right| > M\right) \leqslant \varepsilon.$$

PROOF. The monotonicity of the processes X_n and Y_n yields that

$$\sup_{x\in[-k,k]} \left|X_n(x)\right| = \max\left\{\left|X_n(-k)\right|, \left|X_n(k)\right|\right\},$$
$$\sup_{x\in[-k,k]} \left|Y_n(x)\right| = \max\left\{\left|Y_n(-k)\right|, \left|Y_n(k)\right|\right\}.$$

Assume $|X_n(k)|$ to be the maximum in the above display. Since for fixed k, $X_n(k) \xrightarrow{d} g_{a,b}(k)$, with a and b defined in (4.4.3), it will result that the processes X_n and Y_n in (4.4.2) are, with high probability, uniformly bounded. □

The limiting distribution of the likelihood ratio statistic of a nondecreasing baseline hazard function λ_0 is supplied then by the subsequent theorem.

THEOREM 4.9. *Suppose (A1) and (A2) hold and let $x_0 \in (0, \tau_H)$. Assume that λ_0 is nondecreasing on $[0,\infty)$ and continuously differentiable in a neighborhood of x_0, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) > 0$. Moreover, assume that $H^{uc}(x)$ and $x \to \Phi(\beta_0, x)$, defined in* (4.4.1) *and above* (4.4.1), *are continuously differentiable in a neighborhood of x_0. Let $2\log\xi_n(\vartheta_0)$ be the likelihood ratio statistic for testing $H_0 : \lambda_0(x_0) = \vartheta_0$, as defined in* (4.3.2). *Then,*

$$2\log\xi_n(\vartheta_0) \xrightarrow{d} \mathbb{D},$$

where $\mathbb{D} = \int \left[(g_{1,1}(u))^2 - (g_{1,1}^0(u))^2\right] \mathrm{d}u$, with $g_{1,1}$ and $g_{1,1}^0$ defined in (4.3.31) *and* (4.3.32).

PROOF. By (4.3.7) and (4.3.20), the likelihood ratio statistic $2\log\xi_n(\vartheta_0) = 2L_{\hat\beta_n}(\hat\lambda_n) - 2L_{\hat\beta_n}(\hat\lambda_n^0)$ can be expressed as

$$\begin{aligned}
2\log\xi_n(\vartheta_0) =& 2\sum_{i=1}^{n-1}\Delta_{(i)}\log\hat\lambda_n(T_{(i)}) - 2\sum_{i=1}^{n-1}\Delta_{(i)}\log\hat\lambda_n^0(T_{(i)}) \\
&- 2\sum_{\substack{i=1\\ i\neq m}}^{n-1}\left[T_{(i+1)} - T_{(i)}\right]\left[\hat\lambda_n(T_{(i)}) - \hat\lambda_n^0(T_{(i)})\right]\sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\
&- 2\left[T_{(m+1)} - x_0\right]\left[\hat\lambda_n(T_{(m)}) - \vartheta_0\right]\sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\
&- 2\left[x_0 - T_{(m)}\right]\left[\hat\lambda_n(T_{(m)}) - \hat\lambda_n^0(T_{(m)})\right]\sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}}.
\end{aligned}$$

Let

$$S_n = 2\sum_{i=1}^{n-1}\Delta_{(i)}\log\hat\lambda_n(T_{(i)}) - 2\sum_{i=1}^{n-1}\Delta_{(i)}\log\hat\lambda_n^0(T_{(i)}), \tag{4.4.9}$$

and denote by D_n, the set of indices i on which $\hat\lambda_n(T_{(i)})$ differs from $\hat\lambda_n^0(T_{(i)})$. Hence, expanding both terms of S_n around $\lambda_0(x_0) = \vartheta_0$, we get

$$\begin{aligned}
S_n =& 2\sum_{i\in D_n}\Delta_{(i)}\frac{\hat\lambda_n(T_{(i)}) - \vartheta_0}{\vartheta_0} - 2\sum_{i\in D_n}\Delta_{(i)}\frac{\hat\lambda_n^0(T_{(i)}) - \vartheta_0}{\vartheta_0} \\
&- \sum_{i\in D_n}\Delta_{(i)}\frac{\left[\hat\lambda_n(T_{(i)}) - \vartheta_0\right]^2}{\vartheta_0^2} + \sum_{i\in D_n}\Delta_{(i)}\frac{\left[\hat\lambda_n^0(T_{(i)}) - \vartheta_0\right]^2}{\vartheta_0^2} + R_n,
\end{aligned}$$

with

$$\begin{aligned}
R_n &= \frac{1}{3}\sum_{i\in D_n}\Delta_{(i)}\frac{\left[\hat\lambda_n(T_{(i)}) - \vartheta_0\right]^3}{\left[\hat\lambda_n^*(T_{(i)})\right]^3} - \frac{1}{3}\sum_{i\in D_n}\Delta_{(i)}\frac{\left[\hat\lambda_n^0(T_{(i)}) - \vartheta_0\right]^3}{\left[\hat\lambda_n^{0*}(T_{(i)})\right]^3} \\
&= R_{n,1} - R_{n,2},
\end{aligned}$$

where $\hat\lambda_n^*(T_{(i)})$ is a point between $\hat\lambda_n(T_{(i)})$ and ϑ_0 and $\hat\lambda_n^{0*}(T_{(i)})$ is a point between $\hat\lambda_n^0(T_{(i)})$ and ϑ_0. We want to show that $R_{n,1}$ and $R_{n,2}$, hence R_n converge to zero, in probability. As for the $R_{n,1}$ term, it can be inferred that

$$|R_{n,1}| \leqslant \frac{1}{3}\int \delta\{u\in\overline{D}_n\}\frac{\left|n^{1/3}\left(\hat\lambda_n(u) - \vartheta_0\right)\right|^3}{\left|\hat\lambda_n^*(u)\right|^3}\,\mathrm{d}P_n(u,\delta,z),$$

where $\overline{D}_n$ is the time interval on which $\hat{\lambda}_n$ differs from $\hat{\lambda}_n^0$. Choose now $\varepsilon > 0$ and $\gamma > 0$, and for $x_0 \in (0, \tau_H)$ fixed and $k_\varepsilon > 0$, denote by $I_n = [x_0 - n^{-1/3}k_\varepsilon, x_0 + n^{-1/3}k_\varepsilon]$. We can write $R_{n,1} = R_{n,1}\{\overline{D}_n \subset I_n\} + R_{n,1}\{\overline{D}_n \not\subset I_n\}$. Since, by Lemma 4.7,

$$\mathbb{P}(|R_{n,1}\{\overline{D}_n \not\subset I_n\}| > \gamma) \leqslant \mathbb{P}(\overline{D}_n \not\subset I_n) < \varepsilon,$$

we will further focus on bounding $|R_{n,1}\{\overline{D}_n \subset I_n\}|$. By Lemma 4.7 and by Lemma 4.8, there exists $k_\varepsilon > 0$ such that $\sup_{x\in[-k_\varepsilon,k_\varepsilon]} \left|\hat{\lambda}_n(x_0 + n^{-1/3}x) - \vartheta_0\right|$ is $\mathcal{O}_p(n^{-1/3})$. Furthermore, since

$$\sup_{x\in[-k_\varepsilon,k_\varepsilon]} \left|\hat{\lambda}_n^*(x_0 + n^{-1/3}x) - \vartheta_0\right| \leqslant \sup_{x\in[-k_\varepsilon,k_\varepsilon]} \left|\hat{\lambda}_n(x_0 + n^{-1/3}x) - \vartheta_0\right|,$$

it will result that, for $u \in \overline{D}_n$, $\left|n^{1/3}\left(\hat{\lambda}_n(u) - \vartheta_0\right)\right|^3$ is uniformly bounded and $\left|\hat{\lambda}_n^*(u)\right|^3$ is uniformly bounded away from zero. It will result that there exists $M > 0$ such that

$$\begin{aligned}|R_{n,1}| \leqslant & M \int \delta\{x_0 - k_\varepsilon n^{-1/3} \leqslant u \leqslant x_0 + k_\varepsilon n^{-1/3}\} \mathrm{d}\left(P_n - P\right)(u,\delta,z) \\ & + M \int \delta\{x_0 - k_\varepsilon n^{-1/3} \leqslant u \leqslant x_0 + k_\varepsilon n^{-1/3}\} \mathrm{d}P(u,\delta,z) + o_p(1).\end{aligned}$$

Chebyshev's inequality provides that the first term on the right-hand side is $\mathcal{O}_p(n^{-2/3})$. As the function H^{uc} defined above (4.4.1) is assumed to be continuously differentiable in a neighborhood of x_0, the second term on the right-hand side is $\mathcal{O}_p(n^{-1/3})$. We can conclude that $R_{n,1} = o_p(1)$. Completely similar, by using Lemma 4.7 and Lemma 4.8, it can be shown that $R_{n,2} = o_p(1)$. Thus $2\log\xi_n(\vartheta_0) = A_n - B_n + o_p(1)$, where

$$\begin{aligned}A_n = & \frac{2}{\vartheta_0} \sum_{i\in D_n} \Delta_{(i)} \left[\hat{\lambda}_n(T_{(i)}) - \hat{\lambda}_n^0(T_{(i)})\right] \\ & - 2 \sum_{i\in D_n\setminus\{m\}} \left[T_{(i+1)} - T_{(i)}\right]\left[\hat{\lambda}_n(T_{(i)}) - \hat{\lambda}_n^0(T_{(i)})\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \\ & - 2\left[T_{(m+1)} - x_0\right]\left[\hat{\lambda}_n(T_{(m)}) - \vartheta_0\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \\ & - 2\left[x_0 - T_{(m)}\right]\left[\hat{\lambda}_n(T_{(m)}) - \hat{\lambda}_n^0(T_{(m)})\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}},\end{aligned} \tag{4.4.10}$$

and

$$B_n = \frac{1}{\vartheta_0^2} \sum_{i\in D_n} \Delta_{(i)} \left\{\left[\hat{\lambda}_n(T_{(i)}) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(T_{(i)}) - \vartheta_0\right]^2\right\}. \tag{4.4.11}$$

Hence, A_n can be written as $A_n = A_{n1} - A_{n2}$, where

$$A_{n1} = \frac{2}{\vartheta_0} \sum_{i \in D_n} \left[\hat{\lambda}_n(T_{(i)}) - \vartheta_0\right] \left\{ \Delta_{(i)} - \vartheta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\},$$

and

$$\begin{aligned} A_{n2} =& \frac{2}{\vartheta_0} \sum_{i \in D_n \setminus \{m\}} \left[\hat{\lambda}_n^0(T_{(i)}) - \vartheta_0\right] \left\{ \Delta_{(i)} - \vartheta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\ &+ \frac{2}{\vartheta_0} \left[\hat{\lambda}_n^0(T_{(m)}) - \vartheta_0\right] \left\{ \Delta_{(m)} - \vartheta_0 \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\}. \end{aligned}$$

For the term A_{n1}, partition the set of indices D_n into s consecutive blocks of indices $B_1, B_2, \ldots, B_s$, such that $\hat{\lambda}_n$ is constant on each block. Denote by $\hat{v}_{nj}$ the unconstrained estimator $\hat{\lambda}_n(T_{(i)})$, for each $i \in B_j$, with $j = 1, 2, \ldots, s$. By (4.3.6), it follows that

$$\begin{aligned} A_{n1} &= \frac{2}{\vartheta_0} \sum_{j=1}^{s} \sum_{i \in B_j} \left(\hat{v}_{nj} - \vartheta_0\right) \left\{ \Delta_{(i)} - \vartheta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\ &= \frac{2}{\vartheta_0} \sum_{j=1}^{s} \left(\hat{v}_{nj} - \vartheta_0\right) \left\{ \sum_{i \in B_j} \Delta_{(i)} - \vartheta_0 \sum_{i \in B_j} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\ &= \frac{2}{\vartheta_0} \sum_{j=1}^{s} \left(\hat{v}_{nj} - \vartheta_0\right)^2 \sum_{i \in B_j} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \\ &= \frac{2}{\vartheta_0} n \sum_{i \in D_n} \left[\hat{\lambda}_n(T_{(i)}) - \vartheta_0\right]^2 \frac{1}{n} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}. \end{aligned}$$

Define

$$\Phi_n(\beta, x) = \int \{u \geqslant x\} \, \mathrm{e}^{\beta' z} \, \mathrm{d}P_n(u, \delta, z), \tag{4.4.12}$$

and note that

$$\int_{[T_{(i)}, T_{(i+1)})} \Phi_n(\hat{\beta}_n, u) \, \mathrm{d}u = \frac{1}{n} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}},$$

for each $i = 1, 2, \ldots, n-1$. The term A_{n1} can then be written as

$$A_{n1} = \frac{2}{\vartheta_0} n \int \left\{ u \in \overline{D}_n \right\} \left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 \Phi_n(\hat{\beta}_n, u) \, \mathrm{d}u,$$

where $\overline{D}_n$ is the interval on which $\hat{\lambda}_n$ and $\hat{\lambda}_n^0$ differ. Similarly, for the term A_{n2}, partition D_n into q consecutive blocks of indices $B_1^0, B_2^0, \ldots, B_q^0$, such that the constrained estimator $\hat{\lambda}_n^0$ is constant on each block. There is one block, say B_r^0, on which the constrained estimator is ϑ_0, and one block, say B_p^0 that contains m. On all other blocks B_j^0, denote by $\hat{v}_{nj}^0$ the constrained estimator $\hat{\lambda}_n^0(T_{(i)})$, for each $i \in B_j^0$. It will result that,

$$\begin{aligned}
A_{n2} =& \frac{2}{\vartheta_0} \sum_{\substack{j=1\\ j\neq r,p}}^{q} \sum_{i\in B_j^0} \left(\hat{v}_{nj}^0 - \vartheta_0\right) \left\{ \Delta_{(i)} - \vartheta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\
&+ \frac{2}{\vartheta_0} \sum_{i\in B_p^0\setminus\{m\}} \left(\hat{v}_{np}^0 - \vartheta_0\right) \left\{ \Delta_{(i)} - \vartheta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\
&+ \frac{2}{\vartheta_0} \left(\hat{v}_{np}^0 - \vartheta_0\right) \left\{ \Delta_{(m)} - \vartheta_0 \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\
=& \frac{2}{\vartheta_0} \sum_{\substack{j=1\\ j\neq r,p}}^{q} \left(\hat{v}_{nj}^0 - \vartheta_0\right) \left\{ \sum_{i\in B_j^0} \Delta_{(i)} - \vartheta_0 \sum_{i\in B_j^0} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\
&+ \frac{2}{\vartheta_0} \left(\hat{v}_{np}^0 - \vartheta_0\right) \left\{ \sum_{i\in B_p^0} \Delta_{(i)} - \vartheta_0 \left[\sum_{i\in B_p^0\setminus\{m\}} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right.\right. \\
&\left.\left. + \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right] \right\}.
\end{aligned}$$

By (4.3.18) and (4.3.19),

$$\begin{aligned}
A_{n2} =& \frac{2}{\vartheta_0} \sum_{\substack{j=1\\ j\neq r,p}}^{q} \left(\hat{v}_{nj}^0 - \vartheta_0\right)^2 \sum_{i\in B_j^0} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \\
&+ \frac{2}{\vartheta_0} \left(\hat{v}_{np}^0 - \vartheta_0\right)^2 \left\{ \sum_{i\in B_p^0\setminus\{m\}} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right. \\
&\left. + \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\
=& \frac{2}{\vartheta_0} n \sum_{i\in D_n\setminus\{m\}} \left[\hat{\lambda}_n^0(T_{(i)}) - \vartheta_0\right]^2 \frac{1}{n} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \\
&+ \frac{2}{\vartheta_0} n \left[\hat{\lambda}_n^0(T_{(m)}) - \vartheta_0\right]^2 \frac{1}{n} \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}.
\end{aligned}$$

As $\hat{\lambda}_n^0(x) = \hat{\lambda}_n^0(T_{(m)})$ on the interval $[T_{(m)}, x_0)$ and $\hat{\lambda}_n^0(x) = \vartheta_0$ on the interval $[x_0, T_{(m+1)})$, this gives that

$$\begin{aligned}
&\int_{T_{(m)}}^{T_{(m+1)}} \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2 \Phi_n(\hat{\beta}_n, u)\,\mathrm{d}u \\
&\quad = \int_{T_{(m)}}^{x_0} \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2 \Phi_n(\hat{\beta}_n, u)\,\mathrm{d}u + \int_{x_0}^{T_{(m+1)}} \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2 \Phi_n(\hat{\beta}_n, u)\,\mathrm{d}u \\
&\quad = \frac{1}{n}\left[\hat{\lambda}_n^0(T_{(m)}) - \vartheta_0\right]^2 \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}.
\end{aligned}$$

This leads to

$$A_{n2} = \frac{2}{\vartheta_0} n \int \left\{u \in \overline{D}_n\right\} \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2 \Phi_n(\hat{\beta}_n, u)\,\mathrm{d}u,$$

and, thus A_n in (4.4.10) can be written as

$$A_n = \frac{2}{\vartheta_0} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \Phi_n(\hat{\beta}_n, u)\,\mathrm{d}u.$$

In a similar manner, B_n in (4.4.11) can be expressed as

$$B_n = \frac{1}{\vartheta_0^2} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \mathrm{d}V_n(u),$$

by (4.3.4) and by noting that for every $i = 1, 2, \ldots, n-1$,

$$\int_{[T_{(i)}, T_{(i+1)})} \mathrm{d}V_n(u) = V_n(T_{(i+1)}) - V_n(T_{(i)}) = \frac{1}{n}\Delta_{(i)}.$$

Concluding,

$$\begin{aligned}
2\log \xi_n(\vartheta_0) = &\frac{2}{\vartheta_0} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \Phi_n(\hat{\beta}_n, u)\,\mathrm{d}u \\
&- \frac{1}{\vartheta_0^2} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \mathrm{d}V_n(u) + o_p(1).
\end{aligned}$$

Let $V(x) = \int \delta\{u < x\}\,\mathrm{d}P(u, \delta, z)$, and see that, in fact, $V(x) = H^{uc}(x)$, where H^{uc} has been defined above (4.4.1). Thus,

$$\begin{aligned}
2\log \xi_n(\vartheta_0) = &\frac{2}{\vartheta_0} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \Phi(\beta_0, u)\,\mathrm{d}u \\
&- \frac{1}{\vartheta_0^2} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \mathrm{d}V(u) + \overline{R}_n + o_p(1),
\end{aligned}$$

where $\overline{R}_n = \overline{R}_{n1} - \overline{R}_{n2}$, with

$$\overline{R}_{n1} = \frac{2}{\vartheta_0} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \left(\Phi_n(\hat{\beta}_n, u) - \Phi(\beta_0, u)\right) \mathrm{d}u,$$

and

$$\overline{R}_{n2} = \frac{1}{\vartheta_0^2} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \mathrm{d}\left(V_n(u) - V(u)\right).$$

The aim is to show that $\overline{R}_{n1}$ and $\overline{R}_{n2}$, and thus $\overline{R}_n$ is $o_p(1)$. The term $\overline{R}_{n1}$ can be written as

$$\frac{2}{\vartheta_0} n^{1/3} \int \left\{u \in \overline{D}_n\right\} \left\{\left[n^{1/3}\left(\hat{\lambda}_n(u) - \vartheta_0\right)\right]^2 - \left[n^{1/3}\left(\hat{\lambda}_n^0(u) - \vartheta_0\right)\right]^2\right\} \left(\Phi_n(\hat{\beta}_n, u) - \Phi(\beta_0, u)\right) \mathrm{d}u.$$

Lemma 2.8 in Chapter 2 provides that

$$\sup_{x \in \mathbb{R}} \left|\Phi_n(\hat{\beta}_n, x) - \Phi(\beta_0, x)\right| \to 0,$$

with probability one. From Lemma 4.8 and since $\int \{u \in \overline{D}_n\} \,\mathrm{d}u \leqslant 2k_\varepsilon n^{-1/3}$, by Lemma 4.7 and by using similar arguments as for the term $R_{n,1}$, we can conclude that $\overline{R}_{n1}$ is $o_p(1)$. Analogously,

$$\overline{R}_{n2} = \frac{1}{\vartheta_0^2} n^{1/3} \int \left\{u \in \overline{D}_n\right\} \delta \left\{\left[n^{1/3}\left(\hat{\lambda}_n(u) - \vartheta_0\right)\right]^2 - \left[n^{1/3}\left(\hat{\lambda}_n^0(u) - \vartheta_0\right)\right]^2\right\} \mathrm{d}(P_n - P)(u, \delta, z).$$

Once more, by Lemma 4.7 and Lemma 4.8, there exists $M_2 > 0$ such that

$$|\overline{R}_{n2}| \leqslant \frac{M_2^2}{\vartheta_0^2} n^{1/3} \int \delta \left\{u \in \overline{D}_n\right\} \mathrm{d}(P_n - P)(u, \delta, z),$$

with arbitrarily large probability. Chebyshev's inequality along with the same reasoning as for the term $R_{n,1}$ provides that $\overline{R}_{n2} = o_p(1)$. Hence,

$$\begin{aligned} 2\log \xi_n(\vartheta_0) = &\frac{2}{\vartheta_0} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \Phi(\beta_0, u)\,\mathrm{d}(u) \\ &- \frac{1}{\vartheta_0^2} n \int \left\{u \in \overline{D}_n\right\} \left\{\left[\hat{\lambda}_n(u) - \vartheta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \vartheta_0\right]^2\right\} \mathrm{d}V(u) + o_p(1). \end{aligned}$$

Consider the change of variable $x = n^{1/3}(u - x_0)$ and let $\widetilde{D}_n = n^{1/3}\left(\overline{D}_n - x_0\right)$. This yields that

$$
\begin{aligned}
2\log \xi_n(\vartheta_0) =& \frac{2}{\vartheta_0}\int \left\{x \in \widetilde{D}_n\right\}\left[X_n^2(x) - Y_n^2(x)\right]\Phi(\beta_0, x_0 + n^{-1/3}x)\,\mathrm{d}x \\
&- \frac{1}{\vartheta_0^2}\int \left\{x \in \widetilde{D}_n\right\}\left[X_n^2(x) - Y_n^2(x)^2\right]V'(x_0 + n^{-1/3}x)\,\mathrm{d}x + o_p(1) \\
=& \frac{2}{\vartheta_0}\Phi(\beta_0, x_0)\int \left\{x \in \widetilde{D}_n\right\}\left[X_n^2(x) - Y_n^2(x)\right]\,\mathrm{d}x \\
&- \frac{1}{\vartheta_0^2}V'(x_0)\int \left\{x \in \widetilde{D}_n\right\}\left[X_n^2(x) - Y_n^2(x)\right]\,\mathrm{d}x + o_p(1).
\end{aligned}
$$

As inferred in (2.2.9) and (2.4.22) in Chapter 2,

$$
\lambda_0(x) = \frac{\mathrm{d}V(x)/\mathrm{d}x}{\Phi(\beta_0, x)},
$$

which gives that

$$
2\log \xi_n(\vartheta_0) = \frac{1}{\vartheta_0}\Phi(\beta_0, x_0)\int \left\{x \in \widetilde{D}_n\right\}\left[X_n^2(x) - Y_n^2(x)\right]\,\mathrm{d}x + o_p(1).
$$

Thus

$$
2\log \xi_n(\vartheta_0) = \frac{1}{a^2}\int \left\{x \in \widetilde{D}_n\right\}\left[X_n^2(x) - Y_n^2(x)\right]\,\mathrm{d}x + o_p(1),
$$

where a has been defined in (4.4.3). From Lemma 4.7, for every $\varepsilon > 0$, we can find an interval $[-k_\varepsilon, k_\varepsilon]$ such that $\mathbb{P}(\widetilde{D}_n \subset [-k_\varepsilon, k_\varepsilon]) > 1 - \varepsilon$, for n sufficiently large. In order to prove the theorem, we apply Lemma 4.2 in PRAKASA RAO (1969), by taking

$$
\begin{aligned}
Q_n =& \frac{1}{a^2}\int \left\{x \in \widetilde{D}_n\right\}\left[X_n^2(x) - Y_n^2(x)\right]\,\mathrm{d}x, \\
Q_{n\varepsilon} =& \frac{1}{a^2}\int \left\{x \in [-k_\varepsilon, k_\varepsilon]\right\}\left[X_n^2(x) - Y_n^2(x)\right]\,\mathrm{d}x, \\
Q_\varepsilon =& \frac{1}{a^2}\int \left\{x \in [-k_\varepsilon, k_\varepsilon]\right\}\left[(g_{a,b}(x))^2 - \left(g_{a,b}^0(x)\right)^2\right]\,\mathrm{d}x,
\end{aligned}
$$

and

$$
Q = \frac{1}{a^2}\int \left\{x \in D_{a,b}\right\}\left[(g_{a,b}(x))^2 - \left(g_{a,b}^0(x)\right)^2\right]\,\mathrm{d}x,
$$

where $D_{a,b}$ denotes the set on which $g_{a,b}$ and $g_{a,b}^0$ differ. Condition (i) in Lemma 4.2 of Prakasa Rao follows by Lemma 4.7. In addition, Lemma 4.7 and Lemma 4.5 yield condition (ii), since for every $\varepsilon > 0$, we can find $k_\varepsilon > 0$ such that $\mathbb{P}(D_{a,b} \subset [-k_\varepsilon, k_\varepsilon]) >$

$1 - \varepsilon$. The third condition follows, for every fixed ε, by Lemma 4.5 and by continuous mapping theorem. Namely, $(X_n, Y_n) \Rightarrow (g_{a,b}, g^0_{a,b})$ as a process in $\mathscr{L} \times \mathscr{L}$ and $(f,g) \mapsto \int \{x \in [-c,c]\}(f^2(x) - g^2(x))\,\mathrm{d}x$ is a continuous function defined on $\mathscr{L} \times \mathscr{L}$ with values in $\mathbb{R}$. Conclusively,

$$\frac{1}{a^2}\int \left[X_n^2(x) - Y_n^2(x)\right]\left\{x \in \tilde{D}_n\right\}\mathrm{d}x \xrightarrow{d} \frac{1}{a^2}\int\left[(g_{a,b}(x))^2 - \left(g^0_{a,b}(x)\right)^2\right]\{x \in D_{a,b}\}\,\mathrm{d}x,$$
$$\stackrel{d}{=} \int\left[(g_{1,1}(x))^2 - \left(g^0_{1,1}(x)\right)^2\right]\{x \in D_{1,1}\}\,\mathrm{d}x,$$

by continuous mapping theorem and by Brownian scaling, as derived in BANERJEE & WELLNER (2001). This completes the proof. □

The asymptotic distribution of the likelihood ratio statistic in the nonincreasing baseline hazard setting can be derived completely analogous.

THEOREM 4.10. *Suppose (A1) and (A2) hold and let $x_0 \in (0, \tau_H)$. Assume that λ_0 is nonincresing on $[0,\infty)$ and continuously differentiable in a neighborhood of x_0, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) < 0$. Moreover, assume that $H^{uc}(x)$ and $x \to \Phi(\beta_0, x)$, defined in* (4.4.1) *and above* (4.4.1)*, are continuously differentiable in a neighborhood of x_0. Let $2\log\xi_n(\vartheta_0)$ be the likelihood ratio statistic for testing $H_0 : \lambda_0(x_0) = \vartheta_0$, as defined in* (4.3.2)*. Then,*

$$2\log\xi_n(\vartheta_0) \xrightarrow{d} \mathbb{D}.$$

PROOF. Following the same reasoning as in the proof of Theorem 4.9 and by Lemma 4.6, it can be deduced that

$$2\log\xi_n(\vartheta_0) \xrightarrow{d} \frac{1}{a^2}\int\left[(l_{a,b}(x))^2 - \left(l^0_{a,b}(x)\right)^2\right]\left\{x \in \overline{D}_{a,b}\right\}\mathrm{d}x,$$

where $\overline{D}_{a,b}$ is the set on which $l_{a,b}$ and $l^0_{a,b}$ differ. By continuous mapping theorem, it suffices to show that, for t fixed, $l_{a,b}(\overline{X}_{a,b})(t)$ has the same distribution as $g_{a,b}(X_{a,b})(t)$ and $l^0_{a,b}(\overline{X}_{a,b})(t)$ has the same distribution as $g^0_{a,b}(X_{a,b})(t)$. It is noteworthy that

$$\mathrm{slolcm}(\overline{X}_{a,b})(t) = -\mathrm{slogcm}(-\overline{X}_{a,b})(t).$$

Thus, by Brownian motion properties and continuous mapping theorem,

$$P\left(l_{a,b}(t) \leqslant z\right) = P\left(-\mathrm{slogmc}(-a\mathbb{W}(t) + t^2) \leqslant z\right) = P\left(-\mathrm{slogmc}(a\mathbb{W}(t) + t^2) \leqslant z\right)$$
$$= P\left(-g_{a,b}(t) \leqslant z\right).$$

Concluding, $l_{a,b}(\overline{X}_{a,b})(t) \stackrel{d}{=} -g_{a,b}(X_{a,b})(t)$, and a similar reasoning can be applied to show that $l^0_{a,b}(\overline{X}_{a,b})(t) \stackrel{d}{=} -g^0_{a,b}(X_{a,b})(t)$. The proof is then immediate, by continuous mapping theorem. □

REMARK. The same limiting distribution $\mathbb{D}$ is obtained for the loglikleihood ratio statistic in the absence of covariates in BANERJEE (2008), as well as in other censoring frameworks, as derived in BANERJEE & WELLNER (2001). In fact, it has been shown in BANERJEE (2007) that the same holds true for a wide class of monotone response models. This distribution differs from the usual χ_1^2 distribution, that is obtained in the regular parametric setting. It is noteworthy that $\mathbb{D}$ does not depend on any of the parameters of the underlying model, and this property turns out to be particularly useful in constructing confidence intervals for the parameters of interest, as it will be exposed in the subsequent section.

4.5 POINTWISE CONFIDENCE INTERVALS VIA SIMULATIONS

Once having derived the asymptotic distribution of the likelihood ratio statistic, the practical application at hand is to construct, for fixed $x_0 \in (0, \tau_H)$, pointwise confidence intervals. We will derive such intervals, for a nondecreasing baseline hazard function λ_0, evaluated at a fixed point x_0, based on simulated data and compare these intervals with the intervals based on the asymptotic distribution of the nondecreasing NPMLE $\hat{\lambda}_n$. According to Theorem 2.4 in Chapter 2, for fixed x_0,

$$\begin{aligned} n^{1/3}\left(\hat{\lambda}_n(x_0) - \lambda_0(x_0)\right) &\xrightarrow{d} \left(\frac{4\lambda_0(x_0)\lambda_0'(x_0)}{\Phi(\beta_0, x_0)}\right)^{1/3} \underset{x\in\mathbb{R}}{\operatorname{argmin}}\{\mathbb{W}(t) + t^2\} \\ &\equiv C(x_0)\mathbb{Z}, \end{aligned}$$

where $\mathbb{W}$ is standard two-sided Brownian motion starting from zero, and the constant $C(x_0)$ depends on x_0 and on the underlying parameters. An estimator $\hat{C}_n(x_0)$ of $C(x_0)$ will then yield an $1-\alpha$ confidence interval for $\lambda_0(x_0)$

$$C^1_{n,\alpha} \equiv \left[\hat{\lambda}_n(x_0) - n^{-1/3}\hat{C}_n(x_0)q(\mathbb{Z}, 1-\alpha/2), \hat{\lambda}_n(x_0) + n^{-1/3}\hat{C}_n(x_0)q(\mathbb{Z}, 1-\alpha/2)\right],$$

where $q(\mathbb{Z}, 1-\alpha/2)$ is the $(1-\alpha/2)^{th}$ quantile of the distribution $\mathbb{Z}$. These quantiles have been computed in GROENEBOOM & WELLNER (2001), and we will further use $q(\mathbb{Z}, 0.975) = 0.998181$. For simulation purposes, we propose

$$\hat{C}_n(x_0) = \left(\frac{4\hat{\lambda}_n(x_0)\hat{\lambda}_n'(x_0)}{\Phi_n(\hat{\beta}_n, x_0)}\right)^{1/3},$$

where $\Phi_n(\beta, x)$ has been defined in (4.4.12), and $\hat{\beta}_n$ is the maximum partial likelihood estimator. Lemma 2.8 in Chapter 2 ensures that $\Phi_n(\hat{\beta}_n, \cdot)$ is a strong uniform consistent estimator of $\Phi(\beta_0, \cdot)$. Furthermore, as an estimate for $\lambda_0'(x_0)$, we choose the numerical derivative of $\hat{\lambda}_n$ on the interval that contains x_0, that is, the slope of the segment $[\hat{\lambda}_n(T_{(m)}), \hat{\lambda}_n(T_{(m+1)})]$.

Pointwise confidence intervals for $\lambda_0(x_0)$ can also be constructed by making use of Theorem 4.9. Let $2\log\xi_n(\vartheta)$ denote the likelihood ratio for testing $H_0 : \lambda_0(x_0) = \vartheta$ versus $H_1 : \lambda_0(x_0) \neq \vartheta$. A $1-\alpha$ confidence interval is then obtained by inverting the likelihood ratio test $2\log\xi_n(\vartheta)$ for different values of ϑ, namely

$$C^2_{n,\alpha} \equiv \left\{\vartheta : 2\log\xi_n(\vartheta) \leqslant q(\mathbb{D}, 1-\alpha)\right\},$$

where $q(\mathbb{D}, 1-\alpha)$ is the $(1-\alpha)^{th}$ quantile of the distribution $\mathbb{D}$. Quantiles of $\mathbb{D}$, based on discrete approximations of Brownian motion, are provided in BANERJEE & WELLNER (2005), and we will make use of $q(\mathbb{D}, 0.95) = 2.286922$. The parameter ϑ is chosen to take values on a fine grid between 0 and 6. It can be shown immediately that, for large enough n, the coverage probability of $C^2_{n,\alpha}$ is approximately $1-\alpha$.

For the performance analysis, we have constructed and compared, from simulated data, the confidence intervals $C^1_{n,\alpha}$ and $C^2_{n,\alpha}$, for $\alpha = 0.05$ and various n. We will assume a Weibull baseline distribution function for the event times, with shape parameter 2 and scale parameter 1. For simplicity, we will assume that the covariate is single-valued and uniformly $(0,1)$ distributed and take $\beta_0 = 0.5$. Given the covariate, the censoring times are assumed to be uniformly $(0,1)$ distributed. We will choose $x_0 = \sqrt{\log 2}$, the median of the baseline distribution of the event times. For each chosen sample size, we generate 1000 replicates and compute the empirical coverage and the average length of the corresponding confidence intervals. Furthermore, since we are simulating from a Weibull distribution with shape parameter 2 and scale parameter 1, and hence know the true baseline hazard function λ_0 and its derivative, as well as the true underlying regression coefficient, we could also consider a confidence interval $\overline{C}^1_{n,\alpha}$, given by

$$\overline{C}^1_{n,\alpha} \equiv \left[\hat{\lambda}_n(x_0) - n^{-1/3}C_0(x_0)q(\mathbb{Z}, 1-\alpha/2), \hat{\lambda}_n(x_0) + n^{-1/3}C_0(x_0)q(\mathbb{Z}, 1-\alpha/2)\right],$$

where C_0 is a deterministic function given by

$$C_0(x_0) = \left(\frac{4\nu\lambda_0(x_0)\lambda_0'(x_0)}{\Phi(\beta_0, x_0)}\right)^{1/3}.$$

Table 4.1 reveals the performance, for various sample sizes, of the confidence interval $C^2_{n,0.05}$ based on the likelihood ratio method (LR), the confidence interval $C^1_{n,0.05}$, based on the asymptotic distribution (AD) of the scaled differences between the NPMLE $\hat{\lambda}_n$ and the true baseline hazard at a fixed point, as well as the confidence interval $\overline{C}^1_{n,0.05}$ based on the known Weibull distribution (TD).

It is noteworthy that for each sample size, the likelihood ratio method yields, on average, shorter pointwise confidence intervals in comparison with the confidence intervals based on the asymptotic distribution of the NPMLE estimator $\hat{\lambda}_n$. Moreover, the

	LR		AD		TD	
n	AL	CP	AL	CP	AL	CP
50	4.275	0.917	5.203	0.932	1.506	0.964
100	3.837	0.923	4.838	0.941	1.317	0.953
200	3.009	0.931	4.605	0.947	1.247	0.947
500	2.734	0.947	3.372	0.948	0.961	0.964
1000	1.454	0.942	2.259	0.940	0.713	0.957
5000	0.879	0.945	1.768	0.952	0.546	0.953

TABLE 4.1: Simulaton results for constructing 95% pointwise confidence intervals using the likelihood ratio $C^2_{n,0.05}$ (LR) or the asymptotic distribution of the NPMLE estimator $C^1_{n,0.05}$ (AD) and $\overline{C}^1_{n,0.05}$ (TD), in terms of average length (AL) and empirical coverage (CP).

confidence intervals based on the likelihood ratio exhibit comparable coverage probabilities with the confidence intervals $C^2_{n,0.05}$, based on the asymptotic distribution. As expected, the highest coverage rate is attained by the confidence intervals $\overline{C}^1_{n,0.05}$. Furthermore, they also yield confidence intervals with the shortest length, on average.

REFERENCES

BANERJEE, M. (2000). Likelihood Ratio Inference in Regular and Nonregular Problems. *PhD Dissertation.* University of Washington.

BANERJEE, M. (2006). Likelihood based inference for monotone response models. *Preprint*. Available at www.stat.lsa.umich.edu/ moulib/wilks3.pdf.

BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Annals of Statistics*, **35:** 931–956.

BANERJEE, M. (2008). Estimating monotone, unimodal and U-shaped failure rates using asymptotic pivots. *Statistica Sinica*, **18:** 467–492.

BANERJEE, M. & WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Annals of Statistics*, **29:** 1699–1731.

BANERJEE, M. & WELLNER, J. A. (2005). Score statistics for current status data: comparisons with likelihood ratio and Wald statistics. *International Journal of Biostatistics*, **1** Art. 3, 29 pp. (electronic).

COX, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, **34:** 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika*, **62:** 269–276.

DEVROYE, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *Annals of Probability*, **9:** 860–867.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72:** 557–565.

van GELOVEN, N., MARTIN, I., DAMMAN, P., de WINTER, R. J., THIJSSEN, J. G. & LOPUHAÄ, H. P. (2011). Estimation of a decreasing hazard of patients with myocardial infarction. *Statistics in Medicine*, **32:** 1223–1238.

GROENEBOOM, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer II*, 539–555.

GROENEBOOM, P. (1998). Special Topics Course 593C: Nonparametric Estimation for Inverse Problems: Algorithms and Asymptotics. *Technical Report 344, Department of Statistics, University of Washington*.

GROENEBOOM, P. & WELLNER, J. A. (2001). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics*, **10:** 388–400.

HUANG, J. & WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Statistics*, **22:** 3–33.

HUANG, Y. & ZHANG, C. H. (1994). Estimating a monotone density from censored observations. *Annals of Statistics*, **22:** 1256–1274.

KIM, J. & POLLARD, D. (1990). Cube root asymptotics. *Annals of Statistics*, **18:** 191–219.

LOPUHAÄ, H. P. & NANE, G. F. (2013). Shape constrained nonparametric estimators of the baseline distribution in Cox proportional hazards model. *Scandinavian Journal of Statistics*, doi: 10.1002/sjos.12008.

OAKES, D. (1977). The asymptotic information in censored survival data. *Biometrika*, **64:** 441–448.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A*, **31:** 23–36.

SLUD, E. V. (1982). Consistency and efficiency of inferences with the partial likelihood. *Biometrika* **69:** 547–552.

TSIATIS, A. (1981). A large sample study of Cox's regression model. *Annals of Statistics*, **9:** 93–108.

CHAPTER 5

SMOOTH MONOTONE ESTIMATION OF THE BASELINE HAZARD IN THE COX MODEL [1]

We consider estimators of a baseline hazard function under the assumption that the baseline hazard is nondecreasing and smooth. We obtain these estimators by kernel smoothing the shape constrained estimators defined in Chapter 2. Depending on the choice of shape constrained estimators and when the smoothing is performed, three different estimators are studied. Furthermore, we investigate the pointwise consistency of these kernel estimators.

5.1 INTRODUCTION

Nonparametric estimators that account for the assumption of smoothness received considerable attention in the literature. The long and abundant stream of research has shown that a smooth estimator has an interest of its own, along with allowing the estimation of the first or second derivatives of the estimator. For example, the first derivative might be of interest to construct confidence intervals based on the asymptotic distribution, as exposed in Chapter 4. There are various smoothing options for a given estimator, including a local polynomial method, spline methods and kernel methods. Nonparametric estimation of smooth distribution functions based on kernels dates back in 1950's, when ROSENBLATT (1956) was the first to propose a kernel density estimator of a univariate probability distribution.

Smooth estimation under shape constraints received a lot of attention since MAMMEN (1991) addressed this problem for regression functions. Smooth estimation of a monotone density has been considered by van der VAART & van der LAAN (2003), among others, who showed that a monotone kernel estimator with bandwidth $n^{-1/3}$, defined as the slope of the least convex minorant of the convolution of the empirical distribution function with a scaled kernel exhibits the same rate of convergence but different limiting distribution than an unsmoothed monotone estimator. For monotone hazard functions, GROENEBOOM & JONGBLOED (2013) propose a kernel smooth least-squares estimator, along with a smooth estimate based on penalization, and investigate their asymptotic properties. In a paper on testing the equality

[1] By Jongbloed, G., Lopuhaä, H. P. and Nane, G. F. In preparation.

of functions under monotonicity constraints, DUROT *et al.* (2013) consider smooth estimators within a general model that accommodates monotone regression curves, monotone densities and monotone hazards in the random censorship model. Furthermore, GROENEBOOM & JONGBLOED (2010) consider a maximum smoothed likelihood estimator of an nondecreasing hazard. In the current status model, Groeneboom *et al.* (2010) propose various smooth estimators of a nonincreasing density, along with estimators of the cumulative distribution function and plug-in hazard estimators, and examine the asymptotic properties of these estimators.

This chapter focuses on kernel smooth baseline hazard estimators in the Cox model, under the assumption of monotonicity. The Cox model (COX, 1972) is one of the most acknowledged and used in practice semiparametric models. The event of interest for each subject in the study is usually assumed to be subject to right censoring and a number of (time-independent) covariates are typically registered for each subject. The model is expressed through the hazard function, and relates the hazard of each subject with a given covariate vector to the baseline hazard, corresponding to the null covariate vector, and an exponential function of covariates. A key feature of the model is that one is able to efficiently estimate the regression coefficients by a maximum partial likelihood estimator, while leaving the baseline hazard completely unspecified, see, e.g., EFRON (1977), OAKES (1977) and SLUD (1982). Furthermore, Breslow (COX, 1972) focused on hazard estimation and proposed a maximum likelihood estimator of the baseline cumulative hazard function, which is commonly known as the Breslow estimator.

To the authors' best knowledge, ANDERSEN *et al.* (1985) were the first to propose a smooth baseline hazard estimator within the Cox model. Moreover, they claim that this estimator follows an asymptotic normal distribution, for a bandwidth $b = b_n \to 0$, such that $nb_n \to \infty$, by combining the asymptotic results of RAMLAU-HANSEN (1983), that considered smoothing counting processes by means of kernel functions, with the asymptotic properties of the Breslow estimator. DABROWSKA (1997) considered kernel smoothing estimation in a generalized Cox model. The regression parameters estimator is obtained by kernel smoothing the risk and counting processes involved and the baseline cumulative hazard estimator is obtained by further smoothing in the time direction. The asymptotic distribution of the estimators is derived for a bandwidth $b = b_n$ such that $nb_n^4 \to 0$. WELLS (1994) investigated the asymptotic properties of the baseline hazard kernel estimator and shows that the estimator is uniformly consistent and asymptotically normal, when the bandwidth is proportional to $n^{-1/5}$.

Although the baseline hazard can be left completely unspecified, in practice, one might be interested in restricting it qualitatively. LOPUHAÄ & NANE (2013a) proposed a maximum likelihood $\hat{\lambda}_n$ and a Grenander-type estimator $\tilde{\lambda}_n$ of a monotone hazard function. The Grenander-type estimator $\tilde{\lambda}_n$ of a nondecreasing baseline hazard estimator is defined as the left-hand slope of the greatest convex minorant of the Breslow estimator. The two estimators have been proven to be strongly consistent and furthermore, the estimators have been shown to exhibit the same distributional law. The asymptotic distribution is defined in terms of the minimum of a two-sided Brownian motion plus a parabola times a constant depending on the underlying parameters and follows the general shape constrained theory.

In this chapter, we introduce maximum likelihood and Grenander-type smooth estimators of a monotone baseline hazard function. Three different estimators emerge from the interplay of isotonization and smoothing. The first considered estimator is the smoothed maximum likelihood estimator, which is obtained by smoothing the nonparametric maximum likelihood estimator $\hat{\lambda}_n$ of a nondecreasing baseline hazard function. Similarly, the smoothed Grenander-type estimator is obtained by smoothing the Grenander-type estimator $\tilde{\lambda}_n$. Another proposed estimator is the Grenander-type smoothed estimator, which is obtained by first smoothing the Breslow estimator of the baseline cumulative hazard function and then taking slopes of the greatest convex minorant of the smoothed Breslow estimator.

The chapter is organized as follows. Section 5.2 introduces the Cox model, along with necessary definitions and usual assumptions. Section 5.3 introduces the smoothed maximum likelihood estimator and provides its strong uniform consistency, while Section 5.4 and 5.5 focuses on Grenander-type smooth estimators and on proving their strong consistency.

5.2 DEFINITIONS AND ASSUMPTIONS

The observed data consist of the triplets $(T_1, \Delta_1, Z_1), \ldots, (T_n, \Delta_n, Z_n)$, where T_i denotes the i^{th} follow-up time with a corresponding censoring indicator Δ_i and covariate vector $Z_i \in \mathbb{R}^p$. The generic follow-up time is defined as $T = \min(X, C)$, where X denotes the event time and C is the censoring time. The censoring indicator is defined as $\Delta = \{X \leqslant C\}$, where $\{\cdot\}$ denotes the indicator function. Given the covariate Z, the event time X and the censoring time C are assumed to be independent. Furthermore, conditionally on $Z = z$, the event time X is assumed to be a nonnegative random variable with an absolutely continuous distribution function $F(x|z)$ with density $f(x|z)$. Similarly, conditionally on $Z = z$, the censoring time C is assumed to be a nonnegative random variable with an absolutely continuous distribution function $G(c|z)$ and density $g(c|z)$. The distribution functions F and G are assumed to share no parameters, thus the censoring mechanism is assumed to be non-informative. Lastly, the covariate vector $Z \in \mathbb{R}^p$ is assumed to be time invariant.

Within the Cox model, the distribution of the survival time is related to the corresponding covariate by

$$\lambda\left(x \mid z\right) = \lambda_0(x)\, \mathrm{e}^{\beta_0' z}, \quad x \in \mathbb{R}^+,$$

where $\lambda\left(x \mid z\right)$ is the hazard function for a subject with covariate vector $z \in \mathbb{R}^p$, λ_0 represents the underlying baseline hazard function, corresponding to a subject with

$z = 0$, and $\beta_0 \in \mathbb{R}^p$ is the vector of the underlying regression coefficients.

Denote the distribution of the follow-up time T by H. We consider the following assumptions, typically employed when deriving large sample properties of estimators within the Cox model, see, e.g., TSIATIS (1981).

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

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

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

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

where $|\cdot|$ denotes the Euclidean norm.

5.3 Smoothed maximum likelihood estimators

A first natural manner to obtain a smooth estimator of a nondecreasing baseline hazard function is to smooth the nonparametric maximum likelihood estimator (NPMLE) $\hat{\lambda}_n$ defined and characterized in Chapter 2. In this section, we will do so by using kernel smoothing.

As inferred in Section 2.2.1, for fixed β, the NPMLE $\hat{\lambda}_n(x;\beta)$ of a nondecreasing baseline hazard function λ_0 is the maximizer of the (pseudo) loglikelihood function $L_\beta(\lambda_0)$ in (2.2.6) over nondecreasing baseline hazard functions. After obtaining the solution to this maximization problem, we simply replace β by $\hat{\beta}_n$, the maximum partial likelihood proposed by COX (1972,1975). Recall that $\hat{\lambda}_n(x) = \hat{\lambda}_n(x;\hat{\beta}_n)$ is of the form

$$\hat{\lambda}_n(x) = \begin{cases} 0 & x < T_{(1)}, \\ \hat{\lambda}_i & T_{(i)} \leqslant x < T_{(i+1)}, \text{ for } i = 1, 2, \dots, n-1, \\ \infty & x \geqslant T_{(n)}, \end{cases}$$

where $\hat{\lambda}_i$, for $i = 1, 2, \dots, n-1$, are defined as slopes of the greatest convex minorant of a cusum diagram defined in terms of the processes W_n and V_n in (2.2.3) and (2.2.4), but also have the following max-min representation

$$\hat{\lambda}_i = \max_{1\leqslant s\leqslant i} \; \min_{i\leqslant t\leqslant n-1} \frac{\sum_{j=s}^{t} \Delta_{(j)}}{\sum_{j=s}^{t} \left[T_{(j+1)} - T_{(j)}\right] \sum_{l=j+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}},$$

for $i = 1, 2, \dots, n-1$.

Let now k be a kernel density with support $[-1, 1]$. We list below some of the typical properties of the kernel function that will be further employed

1. $k(x) \geqslant 0$, for all $x \in \mathbb{R}$,
2. $\int k(x)\,\mathrm{d}x = 1$,
3. $k(x) = k(-x)$,
4. $\int xk(x)\,\mathrm{d}x = 0$,
5. $\int x^2 k(x)\,\mathrm{d}x = c^2 < \infty$.

Note that property 4 is implied by the symmetry of the kernel density and since the support is $[-1,1]$. Likewise, property 5 follows from previous assumptions. We will consider the scaled version $k_b(u) = b^{-1}k(u/b)$ of the kernel function k, where $0 < b = b_n$ is a bandwidth that depends on the sample size, with $b_n \to 0$, as $n \to \infty$.

We define our kernel smoothed maximum likelihood estimator (SMLE) of the baseline hazard function λ_0 at a point x_0 by

$$\hat{\lambda}_n^{SM}(x_0) = \int k_b(x_0 - u)\hat{\lambda}_n(u)\,\mathrm{d}u. \tag{5.3.1}$$

Since $\hat{\lambda}_n$ is nondecreasing, it can be easily shown that the smoothed estimator $\hat{\lambda}_n^{SM}$ is nondecreasing as well. Moreover, the strong pointwise consistency of the smooth estimator $\hat{\lambda}_n^{SM}$ follows from the strong pointwise consistency of the NPMLE $\hat{\lambda}_n$, as it will be shown with the next theorem.

THEOREM 5.1. *Assume that (A1) and (A2) hold. Suppose that λ_0 is nondecreasing and continuous in a neighborhood of x_0. Then, for any $x_0 \in (0, \tau_H)$,*

$$\hat{\lambda}_n^{SM}(x_0) - \lambda_0(x_0) \to 0,$$

with probability one.

PROOF. For a fixed x_0 in the interior of the support $(0, \tau_H)$, consider

$$\hat{\lambda}_n^{SM}(x_0) - \lambda_0(x_0) = \hat{\lambda}_n^{SM}(x_0) - \lambda_0^s(x_0) + \lambda_0^s(x_0) - \lambda_0(x_0), \tag{5.3.2}$$

where

$$\lambda_0^s(x_0) = \int k_b(x_0 - u)\lambda_0(u)\,\mathrm{d}u \tag{5.3.3}$$

is the convolution of λ_0 with the scaled kernel k_b. For the last two terms on the right-hand side, a change of variable yields

$$\begin{aligned} \lambda_0^s(x_0) - \lambda_0(x_0) &= \int \frac{1}{b} k\left(\frac{x_0 - u}{b}\right)\lambda_0(u)\,\mathrm{d}u - \lambda_0(x_0) \\ &= \int k(y)[\lambda_0(x_0 - b_n y) - \lambda_0(x_0)]\,\mathrm{d}y. \end{aligned} \tag{5.3.4}$$

Since $b_n y \to 0$, as $n \to \infty$, then for every $\varepsilon > 0$, there exists a natural number N such that, for every $n > N$, $|b_n y| < \varepsilon$. Fix $\varepsilon > 0$. As λ_0 is nondecreasing, it follows that

$$\lambda_0(x_0 - \varepsilon) - \lambda_0(x_0) < \lambda_0(x_0 - b_n y) - \lambda_0(x_0) < \lambda_0(x_0 + \varepsilon) - \lambda_0(x_0),$$

and by property 2 of the kernel function,

$$\lambda_0(x_0 - \varepsilon) - \lambda_0(x_0) < \int k(y)[\lambda_0(x_0 - b_n y) - \lambda_0(x_0)]\,\mathrm{d}y < \lambda_0(x_0 + \varepsilon) - \lambda_0(x_0).$$

In addition, by taking $n \to \infty$,

$$\begin{aligned}\lambda_0(x_0 - \varepsilon) - \lambda_0(x_0) &< \limsup_{n\to\infty} \int k(y)[\lambda_0(x_0 - b_n y) - \lambda_0(x_0)]\,\mathrm{d}y \\ &< \lambda_0(x_0 + \varepsilon) - \lambda_0(x_0).\end{aligned}$$

As $\varepsilon > 0$ is chosen arbitrarily, the continuity of λ_0 at x_0 yields that, by (5.3.4),

$$\lambda_0^s(x_0) - \lambda_0(x_0) \to 0.$$

Moreover, by a change of variable, the first two terms on the right-hand side of (5.3.2) can be written as

$$\hat{\lambda}_n^{SM}(x_0) - \lambda_0^s(x_0) = \int k(y)\left[\hat{\lambda}_n(x_0 - b_n y) - \lambda_0(x_0 - b_n y)\right]\mathrm{d}y. \tag{5.3.5}$$

Once more, since $b_n y \to 0$, as $n \to \infty$, then for every $\varepsilon > 0$, there exists a natural number N such that, for every $n > N$, $|b_n y| < \varepsilon$. Fix $\varepsilon > 0$. As both $\hat{\lambda}_n$ and λ_0 are nondecreasing, it follows that

$$\hat{\lambda}_n(x_0 - \varepsilon) - \lambda_0(x_0 + \varepsilon) < \hat{\lambda}_n(x_0 - b_n y) - \lambda_0(x_0 - b_n y) < \hat{\lambda}_n(x_0 + \varepsilon) - \lambda_0(x_0 - \varepsilon),$$

and, once again, by property 2 of the kernel,

$$\begin{aligned}\hat{\lambda}_n(x_0 - \varepsilon) - \lambda_0(x_0 + \varepsilon) &< \int k(y)\left[\hat{\lambda}_n(x_0 - b_n y) - \lambda_0(x_0 - b_n y)\right]\mathrm{d}y \\ &< \hat{\lambda}_n(x_0 + \varepsilon) - \lambda_0(x_0 - \varepsilon),\end{aligned}$$

By taking $n \to \infty$, we get

$$\begin{aligned}\limsup_{n\to\infty}(\hat{\lambda}_n(x_0 - \varepsilon) - \lambda_0(x_0 + \varepsilon)) &< \limsup_{n\to\infty} \int k(y)\left[\hat{\lambda}_n(x_0 - b_n y) - \lambda_0(x_0 - b_n y)\right]\mathrm{d}y \\ &< \limsup_{n\to\infty}(\hat{\lambda}_n(x_0 + \varepsilon) - \lambda_0(x_0 - \varepsilon)).\end{aligned}$$

By Theorem 2.3 in Chapter 2, we obtain

$$\begin{aligned}\lambda_0(x_0 - \varepsilon) - \lambda_0(x_0 + \varepsilon) &< \limsup_{n\to\infty} \int k(y)\left[\hat{\lambda}_n(x_0 - b_n y) - \lambda_0(x_0 - b_n y)\right]\mathrm{d}y \\ &< \lambda_0(x_0 + \varepsilon) - \lambda_0(x_0 - \varepsilon),\end{aligned}$$

with probability one. By the continuity of λ_0 at x_0 and by (5.3.5),

$$\hat{\lambda}_n^{SM}(x_0) - \lambda_0^s(x_0) \to 0,$$

with probability one. This completes the proof. □

5.4 SMOOTHED GRENANDER-TYPE ESTIMATORS

We will consider now smoothing the Grenander-type estimator $\tilde{\lambda}_n$ of a nondecreasing baseline hazard. Recall that $\tilde{\lambda}_n$ is defined as the left-hand slope of the greatest convex minorant $\widetilde{\Lambda}_n$ of the Breslow estimator Λ_n, and its asymptotic properties have been investigated in Chapter 2.

Then, the smoothed Grenander-type (SG) estimator $\tilde{\lambda}_n^{SG}$ at a fixed point x_0 is defined by

$$\tilde{\lambda}_n^{SG}(x_0) = \int k_b(x_0 - u)\tilde{\lambda}_n(u)\,\mathrm{d}u = \int k_b(x_0 - u)\,\mathrm{d}\widetilde{\Lambda}_n(u), \tag{5.4.1}$$

where $b = b_n$ is the bandwidth. Note that the monotonicity of $\tilde{\lambda}_n^{SG}$ follows from the monotonicity of $\tilde{\lambda}_n$. The strong uniform consistency along with rates of convergence for $\tilde{\lambda}_n^{SG}$ are provided in the theorem below.

THEOREM 5.2. *Assume that (A1) and (A2) hold. Suppose that λ_0 is nondecreasing and twice differentiable on $[0,\infty)$, with uniformly bounded first and second derivatives, and that the kernel k is differentiable with a uniformly bounded derivative. Then, for each $\varepsilon > 0$ and for each $M < \tau_H$,*

$$\sup_{x\in[\varepsilon,M]} \left|\tilde{\lambda}_n^{SG}(x) - \lambda_0(x)\right| \to 0,$$

with probability one. Moreover,

$$\sup_{x\in[\varepsilon,M]} \left|\tilde{\lambda}_n^{SG}(x) - \lambda_0(x)\right| = \mathcal{O}_p(b^{-1}n^{-1/2} + b^2),$$

where $b = b_n$ is the bandwidth used in the scaled kernel function k_b.

PROOF. Write

$$\tilde{\lambda}_n^{SG}(x) - \lambda_0(x) = \tilde{\lambda}_n^{SG}(x) - \lambda_0^s(x) + \lambda_0^s(x) - \lambda_0(x),$$

where $\lambda_0^s(x)$ is defined in (5.3.3). A Taylor expansion yields then, for $x \in [\varepsilon, M]$,

$\xi_n \in (x, x - b_n u)$, and n sufficiently large,

$$\begin{aligned}\lambda_0^s(x) - \lambda_0(x) &= \int k(u)\lambda_0(x - b_n u)\,\mathrm{d}u - \lambda_0(x) = \int k(u)[\lambda_0(x - b_n u) - \lambda_0(x)]\,\mathrm{d}u \\ &= \int k(u)\left[-\lambda_0'(x)b_n u + \frac{1}{2}\lambda_0''(\xi_n)b_n^2u^2\right]\mathrm{d}u \\ &= -\lambda_0'(x)b_n\int uk(u)\,\mathrm{d}u + \frac{1}{2}\lambda_0''(\xi_n)b_n^2\int u^2k(u)\,\mathrm{d}u.\end{aligned}$$

The first term on the right-hand side is zero, by property 4 of the kernel density. In addition, by property 5 of the kernel function and as λ_0 has a uniformly bounded second derivative, it can be argued that, for all $x \in [\varepsilon, M]$,

$$\left|\lambda_0^s(x) - \lambda_0(x)\right| \leqslant \frac{1}{2}\sup_{x\in[\varepsilon,M]}|\lambda_0''(x)|b_n^2\int u^2k(u)\,\mathrm{d}u = \mathcal{O}(b_n^2). \tag{5.4.2}$$

Moreover, since $\lambda_0^s(x) = \int k_b(x-u)\mathrm{d}\Lambda_0(u)$, integration by parts yields

$$\begin{aligned}\tilde{\lambda}_n^{SG}(x) - \lambda_0^s(x) &= \int k_b(x-u)\,\mathrm{d}\left(\tilde{\Lambda}_n - \Lambda_0\right)(u) \\ &= \left[k_b(x-u)(\tilde{\Lambda}_n(u) - \Lambda_0(u))\right]\Big|_{x-b}^{x+b} \\ &\quad + \int\frac{\partial}{\partial u}k_b(x-u)\left[\tilde{\Lambda}_n(u) - \Lambda_0(u)\right]\mathrm{d}u \\ &= -\frac{1}{b_n^2}\int k'\left(\frac{x-u}{b_n}\right)\left[\tilde{\Lambda}_n(u) - \Lambda_0(u)\right]\mathrm{d}u.\end{aligned}$$

A change of variable gives that, for all $x \in [\varepsilon, M]$,

$$\begin{aligned}\left|\tilde{\lambda}_n^{SG}(x) - \lambda_0^s(x)\right| &\leqslant \frac{1}{b_n}\int|k'(y)|\left|\tilde{\Lambda}_n(x - b_n y) - \Lambda_0(x - b_n y)\right|\mathrm{d}y \\ &\leqslant \frac{1}{b_n}\sup_{z\in[0,M]}\left|\tilde{\Lambda}_n(z) - \Lambda_0(z)\right|\int|k'(y)|\,\mathrm{d}y.\end{aligned}$$

Marshall's lemma together with Theorem 2.9 in Chapter 2 give that, for all $0 < M < \tau_H$,

$$\sup_{x\in[0,M]}\left|\tilde{\Lambda}_n(x) - \Lambda_0(x)\right| \leqslant \sup_{x\in[0,M]}\left|\Lambda_n(x) - \Lambda_0(x)\right| \to 0,$$

with probability one. Hence, for all $x \in [\varepsilon, M]$,

$$|\tilde{\lambda}_n^{SG}(x) - \lambda_0^s(x)| \to 0,$$

with probability one, as the compactly supported kernel has a uniformly bounded derivative. This together with (5.4.2) proves the first claim of the theorem, since $b = b_n \to 0$, as $n \to \infty$.

Furthermore, Marshall's lemma and Theorem 2.9 in Chapter 2 also provide that, for all $0 < M < \tau_H$,

$$\sup_{x\in[0,M]} \left|\widetilde{\Lambda}_n(x) - \Lambda_0(x)\right| \leqslant \sup_{x\in[0,M]} \left|\Lambda_n(x) - \Lambda_0(x)\right| = \mathcal{O}_p(n^{-1/2}), \tag{5.4.3}$$

which gives that, for all $x \in [\varepsilon, M]$,

$$\left|\tilde{\lambda}_n^{SG}(x) - \lambda_0^s(x)\right| = \mathcal{O}_p(b_n^{-1} n^{-1/2}).$$

This together with (5.4.2) completes the proof of the second claim of the theorem. □

REMARK. It is noteworthy that a KIEFER WOLFOWITZ (1976) type of result for the Breslow estimator, namely a sharper bound in (5.4.3) would be necessary in deriving the asymptotic distribution of the smoothed Grenander-type estimator $\tilde{\lambda}_n^{SG}$, by following the approach used in GROENEBOOM & JONGBLOED (2013). This would also make use of the linearization result of the Breslow estimator derived by LOPUHAÄ & NANE (2013b) and exposed in Chapter 3.

5.5 GRENANDER-TYPE SMOOTHED ESTIMATORS

An alternative method to construct a smooth nondecreasing baseline hazard estimator is to consider a Grenander-type estimator based on the smoothed Breslow estimator. Thus, we first smooth the Breslow estimator Λ_n in the time direction, and consider, for fixed x_0,

$$\Lambda_n^s(x_0) = \int k_b(x_0 - u)\Lambda_n(u)\,\mathrm{d}u = \int K_b(x_0 - u)\,\mathrm{d}\Lambda_n(u), \tag{5.5.1}$$

where $K(x) = \int_{-\infty}^x k(u)\mathrm{d}u$ is the integrated kernel function corresponding to the kernel density k and $K_b(u) = K(u/b)$ its scaled version. The Grenander-type smoothed (GS) baseline hazard estimator $\tilde{\lambda}_n^{GS}$ is defined as the left derivative of the greatest convex minorant of the smoothed Breslow estimator Λ_n^s.

The strong consistency of the estimator $\tilde{\lambda}_n^{GS}$ follows from the uniform strong consistency of the smoothed Breslow estimator, and the proof is in line with the reasoning for the unsmoothed estimator. Hence, we will show first that Λ_n^s is a uniformly strongly consistent estimator of the underlying baseline cumulative hazard function Λ_0.

LEMMA 5.3. *Assume (A1) and (A2) hold that the baseline hazard λ_0 is nondecreasing and continuously differentiable on $[0,\infty)$, with uniformly bounded derivative. Let Λ_n^s be* *the smoothed Breslow estimator defined in* (5.5.1). *Then, for each* $\varepsilon > 0$ *and for each* $M < \tau_H$,

$$\sup_{x\in[\varepsilon,M]} \left|\Lambda_n^s(x) - \Lambda_0(x)\right| \to 0,$$

with probability one. Moreover,

$$\sup_{x\in[\varepsilon,M]} \left|\Lambda_n^s(x) - \Lambda_0(x)\right| = \mathcal{O}_p(n^{-1/2} + b^2),$$

where $b = b_n$ *is the bandwidth used in the scaled kernel function* k_b.

Proof. Write

$$\Lambda_n^s(x) - \Lambda_0(x) = \Lambda_n^s(x) - \Lambda_0^s(x) + \Lambda_0^s(x) - \Lambda_0(x), \tag{5.5.2}$$

where $\Lambda_0^s(x) = \int k_b(x-u)\Lambda_0(u)\mathrm{d}u$ is the convolution of Λ_0 with the scaled kernel function k_b. Similar to the proof of Theorem 5.1, by a change of variable and a Taylor expansion, we can write, for $x \in [\varepsilon, M]$, $\xi_n \in (x, x-b_n u)$, and n sufficiently large,

$$\begin{aligned}
\Lambda_0^s(x) - \Lambda_0(x) &= \int k(y)\Lambda_0(x-b_n y)\,\mathrm{d}y - \Lambda_0(x)\\
&= \int k(y)\left[\Lambda_0(x-b_n y) - \Lambda_0(x)\right]\mathrm{d}y\\
&= \int k(y)\left[-\lambda_0(x)b_n y + \frac{1}{2}\lambda_0'(\xi_n)b_n^2 y^2\right]\mathrm{d}y\\
&= -\lambda_0(x)b_n\int yk(y)\,\mathrm{d}y + \frac{1}{2}\lambda_0'(\xi_n)b_n^2\int y^2k(y)\,\mathrm{d}y.
\end{aligned}$$

By properties 4 and 5 of the kernel function and since λ_0 has a uniformly bounded derivative, it results, for all $x \in [\varepsilon, M]$,

$$\left|\Lambda_0^s(x) - \Lambda_0(x)\right| \leqslant \frac{1}{2}\sup_{x\in[\varepsilon,M]} |\lambda_0'(x)|b_n^2\int y^2k(y)\,\mathrm{d}y = \mathcal{O}_p(b_n^2). \tag{5.5.3}$$

For the first two terms on the right-hand side of (5.5.2), integration by parts yields that, for every $x \in [\varepsilon, M]$,

$$\begin{aligned}
\left|\Lambda_n^s(x) - \Lambda_0^s(x)\right| &\leqslant \int k_b(x-u)\left|\Lambda_n(u) - \Lambda_0(u)\right|\mathrm{d}u\\
&\leqslant \sup_{x\in[0,M]} |\Lambda_n(x) - \Lambda_0(x)|\int k_b(x-u)\,\mathrm{d}u\\
&\to 0,
\end{aligned}$$

with probability one, by Theorem 2.9 in Chapter 2. This together with (5.5.3) proves the first claim of the lemma, since $b = b_n \to 0$, as $n \to \infty$. Furthermore, by (5.4.3) and (5.5.3), the second claim of the lemma is immediate. □

The strong pointwise convergence of $\tilde{\lambda}_n^{GS}$ is then immediate, by using a variation of Marshall's lemma, similar as in the proof of Theorem 2.3 in Chapter 2.

COROLLARY 5.4. *Assume (A1) and (A2)and suppose that λ_0 is nondecreasing on $[0,\infty)$. Let $\tilde{\lambda}_n^{GS}$ be the derivative of the greatest convex minorant of the smoothed Breslow estimator Λ_n^s in* (5.5.1). *Then, for any $x_0 \in (0,\tau_H)$,*

$$\tilde{\lambda}_n^{GS}(x_0) - \lambda_0(x_0) \to 0,$$

with probability one.

REFERENCES

ANDERSEN, P. G., BORGAN, O., HJORT, N. L., ARJAS, E., STENE, J. & AALEN, O. (1985). Counting process models for life history data: a review (With discussion and a reply by the authors). *Scandinavian Journal of Statistics*, **12:** 97–158.

COX, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, **34:**, 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika*, **62:** 269–276.

DABROWSKA, D. M. (1997). Smoothed Cox regression. *Annals of Statistics*, **25:** 1510–1540.

DUROT, C., GROENEBOOM, P. & LOPUHAÄ, H. P. (2013). Testing equality of functions under monotonicity constraints. *Tentatively accepted by Journal of Nonparametric Statistics*.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72:** 557–565.

GROENEBOOM, P. & JONGBLOED, G. (2010). Generalized continuous isotonic regression. *Statistics & Probability Letters*, **80:** 248–253.

GROENEBOOM, P., JONGBLOED, G. & WITTE, B.I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model *Annals of Statistics*, **38:** 352–387.

GROENEBOOM, P. & JONGBLOED, G. (2013). Smooth and non-smooth estimates of a monotone hazard. *IMS Collection. From Probability to Statistics and Back: High-Dimensional Models and Processes*, **9:** 174–196.

KIEFER, J. & WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **34:** 73–85.

LOPUHAÄ, H. P. & NANE, G. F. (2013a). Shape constrained nonparametric estimators of the baseline distribution in Cox proportional hazards model. *Scandinavian Journal of Statistics*, doi: 10.1002/sjos.12008.

LOPUHAÄ, H. P. & NANE, G. F. (2013b). An asymptotic linear representation for the Breslow estimator. *Communications in Statistics - Theory and Methods*, **42:** 1314–1324.

MAMMEN, E. (1991). Estimating a smooth monotone regression function *Annals of Statistics*, **19:** 724–740.

OAKES, D. (1977). The asymptotic information in censored survival data. *Biometrika*, **64:** 441– 448.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Annals of Statistics*, **11:** 453–466.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27:** 832–837.

SLUD, E. V. (1982). Consistency and efficiency of inferences with the partial likelihood. *Biometrika*, **69:** 547–552.

TSIATIS, A. (1981). A large sample study of Cox's regression model. *Annals of Statistics*, **9:** 93–108.

van der VAART, A. W. & van der LAAN, M.J. (2003). Smooth estimation of a monotone density. *Statistics. A Journal of Theoretical and Applied Statistics*, **37:** 189–203.

WELLS, M. T. (1994). Nonparametric kernel estimation in counting processes with explanatory variables. *Biometrika*, **81:** 795–801.

Summary

Shape constrained nonparametric estimation in the Cox model

The events of interest in any survival analysis study are regularly subject to censoring. There are various censoring schemes, including right or left censoring, and interval censoring. The most frequent censoring scheme is the right censoring, where subjects might drop out of the study or simply because not all events of interest occur before the end of the study. Moreover, for each subject, additional information referred to as covariates is registered at the beginning or throughout the study, such as age, sex, undergoing treatment, etc. The classical model to study the distribution of the events of interest, while accounting for additional information, is the Cox model.

The Cox model expresses the hazard function of a subject given a set of covariates in terms of a baseline hazard, for which all covariates are zero, and an exponential function of the covariates and corresponding regression parameters. The baseline hazard can be left completely unspecified while estimating the regression parameters. Nonetheless, in practice, there are numerous studies in which the baseline hazard appears to be monotone. Time to death or to the onset of a disease are observed to have a nondecreasing baseline hazard, while the survival or recovery time after a successful medical treatment usually exhibit a nonincreasing baseline hazard.

The aim of this thesis is to study the behavior of nonparametric baseline hazard and baseline density estimators in the Cox model under monotonicity constraints. The event times are assumed to be right censored and the censoring mechanism is assumed to be independent of the event of interest and non-informative. The covariates are assumed to be time-independent, usually recorded at the beginning of the study. In addition to point estimates, interval estimates of a monotone baseline hazard will be provided, based on a likelihood ratio method, along with testing at a fixed point. Furthermore, kernel smoothed estimates of a monotone baseline hazard will be defined and their behavior will be investigated.

In Chapter 2, we propose several nonparametric monotone estimators of a baseline hazard or a baseline density within the Cox model. We derive the nonparametric maximum likelihood estimator of a nondecreasing baseline hazard and we consider a Grenander-type estimator, defined as the left-hand slope of the greatest convex minorant of the Breslow estimator. The two estimators are then shown to be strongly consistent and asymptotically equivalent. Moreover, we derive their common limit distribution at a fixed point. The two equivalent estimators of a nonincreasing baseline hazard and their asymptotic properties are acquired similarly. Furthermore, we introduce a Grenander-type estimator of a nonincreasing baseline density, defined as the left-hand slope of the least concave majorant of an estimator of the baseline cumulative distribution function derived from the Breslow estimator. This estimator is proven to be strongly consistent and its asymptotic distribution at a fixed point is derived.

Chapter 3 provides an asymptotic linear representation of the Breslow estimator of the baseline cumulative hazard function in the Cox model. This representation can be used to derive the asymptotic distribution of the Grenander type estimator of a monotone baseline hazard estimator. The representation consists of an average of independent random variables and a term involving the difference between the maximum partial likelihood estimator and the underlying regression parameter. The order of the remainder term is arbitrarily close to n^{-1}.

Chapter 4 focuses on interval estimation and on testing whether a monotone baseline hazard function in the Cox model has a particular value at a fixed point, via a likelihood ratio method. Nonparametric maximum likelihood estimators under the null hypothesis are defined for both nondecreasing and nonincreasing baseline hazard functions. These characterizations, along with those of the monotone nonparametric maximum likelihood estimators provide the asymptotic distribution of the likelihood ratio test. This asymptotic distribution enables, via inversion, the construction of pointwise confidence intervals. This method of constructing confidence intervals avoids the issue of estimating the nuisance parameters, as in the case of confidence intervals based on the asymptotic distribution of the estimators. Simulations indicate that the two methods yield confidence intervals with comparable coverage probabilities. Nonetheless, the confidence intervals based on the likelihood ratio are smaller, on average.

Finally, in chapter 5 we consider smooth baseline hazard estimators. The estimators are obtained by kernel smoothing the maximum likelihood and Grenander-type estimators of a monotone baseline hazard function. Three different estimators are proposed for a nondecreasing baseline hazard, which are provided by the interchange of the smoothing and isotonization step. With this respect, we define a smoothed maximum likelihood estimator (SMLE), as well as a smoothed Grenander type (SG) estimator and a Grenander type smoothed (GS) estimator. All estimators are shown to be strongly pointwise or uniformly consistent.

Samenvatting

Niet-parametrische schatting met vormrestricties in het Cox model

In een analyse van levensduren zijn de tijdstippen, waarop de gebeurtenissen plaatsvinden waarin men is geïnteresseerd, regelmatig onderworpen aan censurering. Er zijn verschillende vormen van censurering, waaronder rechts- of links censurering, en interval censurering. De meest voorkomende vorm van censurering is rechts censurering, waarbij deelnemers de studie voortijdig verlaten of waarbij niet alle gebeurtenissen voor het einde van de studie plaatsvinden. Bovendien wordt voor elke deelnemer aan de studie aanvullende informatie geregistreerd, aangeduid als covariaten, ofwel aan het begin dan wel gedurende het onderzoek zoals: leeftijd, geslacht, ondergaande behandeling, etc. Het klassieke model voor de kansverdeling van de gebeurtenissen rekening houdend met de aanvullende informatie is het Cox model.

Het Cox model beschrijft de hazard functie van een deelnemer, gegeven een set covariaten, in termen van een baseline hazard, waarvoor alle covariaten nul zijn, en een exponentiële functie van de covariaten en bijbehorende regressieparameters. Bij het schatten van de regressieparameters hoeft de baseline hazard niet nader gespecificeerd te worden. Toch zijn er in de praktijk talrijke studies waarbij de baseline hazard monotone lijkt. De tijdsduur tot overlijden of tot de aanvang van een ziekte wordt geassocieerd met een niet-dalende baseline hazard, terwijl de levensduur of hersteltijd na een succesvolle medische behandeling gewoonlijk een niet-stijgende baseline hazard vertoont.

Het doel van dit proefschrift is om het gedrag van niet-parametrische schatters voor de baseline hazard en de baseline dichtheid in het Cox model te bestuderen onder de aanname van monotonie. Het wordt verondersteld dat de gebeurtenissen recht gecensureerd zijn en dat het censurerings mechanisme niet-informatief is en onafhankelijk is van de gebeurtenis. De covariaten worden verondersteld als tijdsonafhankelijk, gewoonlijk geregistreerd aan het begin van de studie. Naast punt schatters voor een monotone baseline hazard, worden interval schatters voorgesteld, gebaseerd op een likelihood ratio methode in samenhang met het toetsen op een vast punt. Daarnaast worden kernschatters voor een monotone baseline hazard gedefinieerd en zal hun gedrag worden onderzocht.

In Hoofdstuk 2, introduceren we verschillende niet-parametrische monotone schatters voor de baseline hazard of de baseline kansdichtheid in het Cox model. We leiden de niet-parametrische maximum likelihood schatter af voor een niet-dalende baseline hazard en beschouwen een soort van Grenander-schatter, gedefinieerd als de linker afgeleide van de grootste convexe minorant van de Breslow schatter. De twee schatters blijken sterk consistent te zijn en asymptotisch equivalent. Bovendien leiden we de gemeenschappelijke kansverdeling af in een vast punt. De twee equivalent schatters voor een niet-stijgende baseline hazard en hun asymptotische eigenschappen worden op eenzelfde manier afgeleid. Verder introduceren we een Grenander-schatter voor een niet-stijgende baseline kansdichtheid, gedefinieerd als de linker afgeleide van de kleinste concave majorant van een schatter voor de baseline verdelingsfunctie afgeleid van de Breslow schatter. Deze schatter wordt bewezen sterk consistent te zijn, en de asymptotische kansverdeling in een vast punt wordt afgeleid.

In Hoofdstuk 3 leiden we een asymptotische lineaire representatie af voor de Breslow schatter voor de baseline cumulatieve hazardfunctie in het Cox model. Deze representatie kan gebruikt worden om de asymptotische verdeling van de Grenander-schatter van een monotone baseline hazard te bepalen. De representatie omvat een gemiddelde van onafhankelijke stochastische variabelen en een term die betrekking heeft op het verschil tussen de maximum partial likelihood schatter en de onderliggende regressieparameter. De orde van de resterende term is willekeurig dicht bij n^{-1}.

Hoofdstuk 4 richt zich op interval schatting voor de monotone baseline hazard in het Cox model en toetsen of deze een specifieke waarde heeft in een vast punt, via een likelihood ratio methode. Niet-parametrische maximum likelihood schatters onder de nulhypothese worden gedefinieerd voor zowel niet-dalende en niet-stijgende baseline hazard functies. Deze karakteriseringen, samen met die van de monotone niet-parametrische maximum likelihood schatters, leiden tot de asymptotische kansverdeling van de likelihood ratio toetsingsgrootheid. Deze asymptotische kansverdeling maakt de constructie mogelijk van puntsgewijze betrouwbaarheidsintervallen, via inversie. Deze werkwijze voor het construeren van betrouwbaarheidsintervallen vermijdt het probleem van het schatten van nuisance parameters, zoals bij de betrouwbaarheidsintervallen gebaseerd op de asymptotische kansverdeling van de schatters. Simulaties tonen aan dat beide methoden betrouwbaarheidsintervallen opleveren met een vergelijkbaar overdekkingspercentage. Niettemin, zijn de betrouwbaarheidsintervallen gebaseerd op de likelihood ratio methode gemiddeld kleiner.

Tenslotte worden in Hoofdstuk 5 gladde schatters voor de baseline hazard onderzocht. De schatters worden verkregen door kernel smoothing van de niet-parametrische maximum likelihood schatters en Grenander-schatters voor een monotone baseline hazard functie. Drie verschillende schatters worden voorgesteld voor een niet-dalende baseline hazard, die worden verkregen door verwisseling van de smoothing-stap en de isotonisatie-stap. Op deze manier definiëren we een gladde maximum likelihood schatter (SMLE), een gladde Grenander-schatter (SG) en een Grenander-schatter gebaseerd op een gladde naïeve schatter (GS). Alle schatters blijken sterk consistent te zijn, zowel puntsgewijs als uniform.

Curriculum vitae

Gabriela Florentina Nane, known to almost everyone as Tina, was born on December 16, 1982 in Bucharest, Romania. In 2005, she got her BSc diploma in both Mathematics and Computer Science from the University of Bucharest, with a Bachelor thesis in game theory, graded with 10. From 2005 to 2006, she taught mathematics at Number 41 Gymnasium in Bucharest. In 2006 she joined the master programme "Risk and Environmental modeling" at Delft Institute of Applied Mathematics, Delft University of Technology. Her Master project emerged from a collaboration between Delft University of Technology and NASA Langley Institute and started with a three months internship at National Institute of Aerospace, Hampton, United States. Her Master thesis "System level risk analysis of new merging and spacing protocols" was supervised by prof. dr. R. M. Cooke and was graded with 10, leading to a *cum laude* Master's diploma. Her thesis was also nominated to Risk Management Study Award 2008, for the best master thesis in risk analysis. In December 2008, she commenced her PhD research in Statistics at Delft Institute of Applied Mathematics, Delft University of Technology, under the supervision of dr. H. P. Lopuhaä and prof. dr. ir. G. Jongbloed. The relevant research has materialized in this thesis. During 2007 and 2012, she was also involved in numerous teaching activities as a student assistant for Bachelor and Master courses. In January 2014, she will start working as a scientific researcher at the Center for Science and Technology Studies at Leiden Univeristy.