Faculteit Wetenschappen

Nonparametric Estimation of the Conditional Distribution in Regression with Censored Data

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen, richting Wiskunde
aan het Limburgs Universitair Centrum te verdedigen door

Ingrid Van Keilegom

Promotor : Prof. dr. N. Veraverbeke

1998

Dankwoord

Graag wil ik iedereen bedanken die gedurende de voorbije vijf jaar heeft bijgedragen in het tot stand brengen van deze thesis.

In het bijzonder denk ik hierbij aan mijn promotor Prof. Noël Veraverbeke. Stap voor stap heeft hij mij op zeer deskundige wijze vertrouwd gemaakt met mijn onderzoeksdomein. Steeds was hij bereid om me met veel geduld en bekwaamheid de "knepen van het vak" uit te leggen. Van hem heb ik enorm veel geleerd.

I am also very grateful to Prof. Michael Akritas (Penn State University, U.S.A.), in collaboration with whom part of this work has been carried out. His enormous scientific inspiration and enthusiasm for his work, made me very enthusiastic for doing research in statistics. Also from him I have learned a lot about statistics.

I would like to thank both Noël and Michael for the pleasant working atmosphere and the excellent collaboration.

Ook wens ik alle leden van de onderzoeksgroep Statistiek van het Limburgs Universitair Centrum en in het bijzonder Prof. Paul Janssen te danken voor hun bijdrage in mijn wetenschappelijke opleiding. Tevens ben ik Prof. Paul Janssen dankbaar voor het kritisch nalezen van dit werk.

Mijn collega's assistenten zorgden voor een zeer aangename werksfeer. Zij hebben mij ook op weg gezet bij het opzetten van mijn eerste simulatiestudie.

Ten slotte ben ik mijn ouders en mijn zus Danielle dankbaar voor hun steun tijdens al mijn studiejaren. Hun onmetelijke inzet, geduld en begrip waren voor mij van zeer grote waarde.

Ingrid Van Keilegom
Diepenbeek, 15 mei 1998.

Contents

Chapter 1

Introduction

1.1 Some examples

When analyzing data from economical, social, medical, ... studies, it often happens that data are in some way incomplete or incorrect. There are many sources of incompleteness that can enter the data, one of them being “censoring”. Censoring often occurs in medical studies where one is interested in the survival time of patients receiving e.g. a certain drug or in industrial studies, where the lifetime of a certain machine is of interest. In these situations, it often occurs that another event happens before the event of interest : a patient might die due to another cause, he (or she) might decide to leave the study, ... For these observations, only a lower bound for the survival time is known, and therefore such data are called right censored.

Estimation of the (survival) distribution when data are subject to censoring, has been studied extensively throughout the last decades. For an overview of important results in the literature, we refer to the next section. An extra difficulty, which is often encountered in practice and which will be studied in this thesis, occurs when together with the survival time also another variable is measured for each observation, e.g. the blood pressure in case of a medical study, and one is interested in the distribution of the survival time for a population of individuals having a specific value of the blood pressure. In such situations, the lifetimes of patients with blood pressure close to the value of interest, are informative and therefore should have a large contribution in the estimation procedure. Lifetimes of patients with covariate value (far) away from the value of interest should have (almost) no contribution. This problem fits in a regression context, where the response is the survival time and the predictor (or covariate) represents, in this case, the blood pressure. Estimation of the (conditional) distribution function in a regression model, when the data are not censored, has been studied widely (see Section 1.3 for a formal approach and an overview of existing results). In this thesis, we will focus on the estimation of the distribution function (and its quantile function) in a regression model when the response values are allowed to be subject to censoring.

A well known data set in this context is the one from the Stanford heart transplant program, in which one of the questions of interest was the effect of the age of an individual receiving a heart transplantation, on his survival time after transplantation. 157 patients received a heart transplantation during the study (which lasted from October 1967 till February 1980), and for each of them, the survival time and the age was recorded (among other variables not of interest here). Patients alive beyond February 1980 were considered to be censored (55 in total). The data are presented in Figure 1.1. The methods that will be developed in this thesis, will allow to estimate the distribution of the survival time for a given age value, to estimate the median (or another quantile) of the survival time at a certain age level, ... Many researchers have studied this data set, among them Miller and Halpern (1982), Fan and Gijbels (1994) and Akritas (1996).

Another interesting data set is the one reported in e.g. Sciortino et al. (1990), Akritas, Murphy and LaValley (1995) and Akritas (1996), concerning O-type stars lying 2.5 kiloparsecs from our sun. In these papers, the relation between (the log of) the X-ray luminosity of a star and (the log of) the bolometric luminosity (this is the total emission, principally at visible wavelengths) is studied. Due to the limited spatial resolution of the X-ray detectors used, the X-ray luminosity was not observed for all stars : of the 273 reported data points, only 79 have observed values of X-ray luminosity, for the remaining stars (194), only an upper bound is known. This highly censored data set is shown in Figure 1.2. Note that observations are

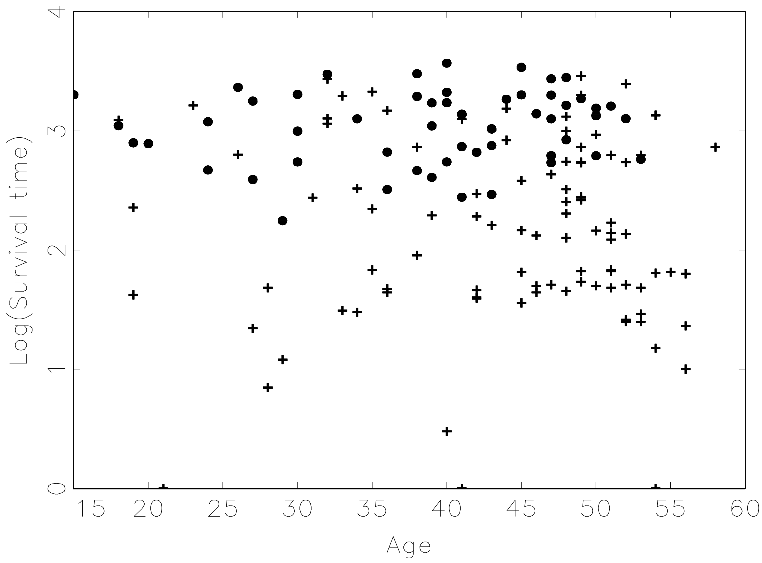

Figure 1.1: *Stanford heart transplant data : uncensored data are indicated by +, censored data by •.*

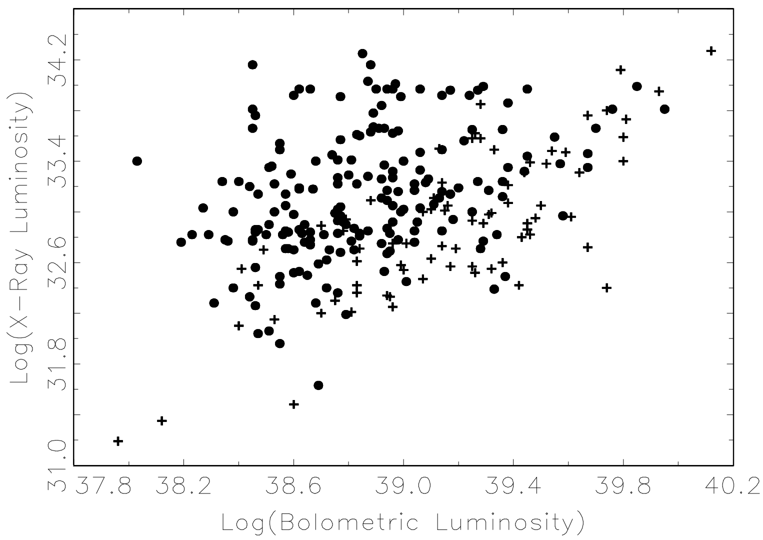

Figure 1.2: *O-type stars data : uncensored data are indicated by +, censored data by •.*

censored to the left here, instead of to the right. However, by inverting the direction of the time axis, results concerning right censored data can be translated into their left censoring-counterparts. In this way, we will be able to apply our results to left censored data as well (which are often encountered in astronomy).

The objective of this thesis is twofold. First, we want to study the asymptotic performance of some estimators of the conditional (survival) distribution and the corresponding quantile function. The second objective is to examine the finite sample performance of these estimators through simulations and to implement the obtained results on a real data set.

There exist many approaches in the statistical literature to analyze censored data when covariates are present. First, a completely nonparametric approach can be followed, in which no assumption on the relation between the covariates and the response is made. The starting point for this research was Beran (1981), who introduced a nonparametric estimator for the conditional distribution of the response given the covariate. Based on this estimator, Doksum and Yandell (1982) studied the conditional mean and median, Cheng (1989) introduced an estimator for the marginal distribution of the response and Akritas (1994) proposed an estimator for the bivariate distribution of the covariate and the response.

Second, a number of models have been proposed, which specify the way on which the survival time depends on the covariate. We willl briefly mention some of these proposals. For a detailed survey, see e.g. Andersen et al. (1993) or Akritas and LaValley (1997). An important and well known model is the proportional hazards model of Cox (1972). It takes the form

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}), \tag{1.1}$$

where $\lambda(t|\boldsymbol{x})$ denotes the hazard function of the response, given the covariate vector $\boldsymbol{x}$ and the unknown parameter vector $\boldsymbol{\beta}$, and $\lambda_0(t)$ is the unknown baseline hazard function. Many researchers have investigated the large sample properties of the estimators in the model or studied the small sample performance through simulation studies. We refer to Akritas and LaValley (1997) for an overview and related research.

Frailty models generalize the usual proportional hazards model by the addition of a random effect. This addition to the model can be motivated by consideration of correlated observations in a study or as a way to incorporate unexplained variation among subjects under study, caused by e.g. unmeasured covariates. In either case, the hazard rate of individual i is modelled as

$$\lambda_i(t|\boldsymbol{x}) = z_i \lambda_0(t) \exp(\boldsymbol{\beta}'\boldsymbol{x}),$$

where z_i is the individual's random effect. The random effects are taken to be generated from a distribution on the real line, often a gamma distribution with mean one and variance θ. Of interest is then the estimation of θ, as well as the baseline hazard function $\lambda_0(t)$ and the vector $\boldsymbol{\beta}$. A number of proposals exist to estimate these parameters, see e.g. Nielsen et al. (1992) and Andersen et al. (1993), where a maximum likelihood approach is used to estimate these parameters (see also Lam and Kuk (1997) where an overview of recent research can be found).

Heteroscedastic regression models are defined as

$$h(Y) = m(\boldsymbol{X}) + \sigma(\boldsymbol{X})\varepsilon, \tag{1.2}$$

where h is a known monotone transformation of the survival time Y, the covariate vector is denoted by $\boldsymbol{X}$, m is the unknown regression function, σ is an unknown scale function, representing possible heteroscedasticity, and the error variable ε is independent of $\boldsymbol{X}$. In case that m is the conditional mean, σ the conditional standard deviation and $\boldsymbol{X}$ is one-dimensional, Fan and Gijbels (1994) studied the estimation of m using local linear regression techniques.

A special case of model (1.2) is the accelerated failure time model, which takes the form

$$\log Y = \boldsymbol{\beta}'\boldsymbol{X} + \varepsilon, \tag{1.3}$$

where the unknown error term ε is independent of $\boldsymbol{X}$. There are two main trends to estimate the parameters in this model : one is to extend the complete data least squares estimator to the censored data case. The other trend is to extend robust estimators to censored data settings. We describe both approaches in more detail in Section 1.5.

Another model is the so-called transformation model, in which, for some unknown monotone function h,

$$h(Y) = \boldsymbol{\beta}'\boldsymbol{X} + \varepsilon,$$

where ε has some known distribution. It can be shown that the proportional hazards model (1.1) is a special case of this model. Doksum (1987), Dabrowska and Doksum (1988) and Cheng and Wu (1994) studied the estimation of the function h and the parameter vector $\boldsymbol{\beta}$.

Recently, two proposals have been worked out to extend the Koziol-Green model to a regression context. The Koziol-Green (1976) model (proposed in the context of censored data in the absence of covariates) specifies that

$$1 - G(t) = (1 - F(t))^{\beta} \tag{1.4}$$

for some $\beta > 0$, where F, respectively G, is the distribution of the survival time Y, respectively censoring time C. In the first extension developed by Veraverbeke and Cadarso Suárez (1997), it is assumed that

$$1 - G(t|\boldsymbol{x}) = (1 - F(t|\boldsymbol{x}))^{\beta_{\boldsymbol{x}}},$$

where $F(\cdot|\boldsymbol{x})$, respectively $G(\cdot|\boldsymbol{x})$, is the conditional distribution of the survival time, respectively censoring time, given the covariate vector $\boldsymbol{x}$ and $\beta_{\boldsymbol{x}} > 0$ is allowed to depend on $\boldsymbol{x}$. The second extension relies on the fact that assumption (1.4) is equivalent to the assumption that the observed survival time $T = Y \wedge C$ is independent of the indicator $\Delta = I(Y \leq C)$. De Uña Álvarez and González Manteiga (1997a,b) extended this assumption to a regression context by assuming that the vector $(\boldsymbol{X}, T)$ is independent of Δ.

Finally, we note that a number of other models, including e.g. the additive risk model and parametric models, have been considered. We refer to Andersen et al. (1993) and Akritas and LaValley (1997) for more details.

In the first part of this thesis, a completely nonparametric approach will be followed. We will study important asymptotic properties of the nonparametric estimator of the conditional distribution of the survival time introduced by Beran

(1981), like its consistency, asymptotic normality, ... A bootstrap procedure will be defined in order to obtain an alternative for the Gaussian approximation. The two approximations will be compared through simulations.

Often, one is interested in the median, or the quartiles, or more generally in any quantile of the distribution function. We will propose an estimator for the quantile function of the conditional distribution function and study its asymptotic and finite sample behavior as well.

In the second part, we assume the heteroscedastic regression model (1.2) with a one-dimensional covariate X. Under the extra assumption that there exists a region of the covariate X where the censoring on Y is "light", we will construct an estimator of the conditional distribution, by transferring tail information from regions of light censoring to regions of heavy censoring.

Apart from the conditional distribution, also the bivariate distribution of X and Y will be studied under model (1.2) and on the basis of its estimator, a least squares estimator for the parameters in a heteroscedastic polynomial regression model will be proposed and its asymptotic features will be examined. This polynomial model reduces to the accelerated failure time model (1.3), when the error terms are homoscedastic and the function h is the log-transformation.

Finally, a simulation study will be carried out in which the completely nonparametric estimator of the conditional distribution introduced by Beran (1981) will be compared with the estimator we will construct under model (1.2). Also, the data from the Stanford heart transplant program will be analyzed and the results will be compared with the results of other approaches found in the literature.

1.2 Nonparametric estimation with censored data

Let $Y_1, \dots, Y_n$ be n independent, identically distributed (i.i.d.), non-negative random variables, such as survival times or failure times, with unknown distribution $F(t) = P(Y_1 \leq t)$. Suppose these survival times are censored by non-negative i.i.d. random variables $C_1, \dots, C_n$, where for each i, C_i is independent of Y_i and suppose their distribution is denoted by $G(t) = P(C_1 \leq t)$. Then, in the right random censorship model, we observe for each i the pair (T_i, Δ_i), where $T_i = Y_i \wedge C_i$ and $\Delta_i = I(Y_i \le C_i)$. Clearly, the distribution $H(t) = P(T_1 \le t)$ satisfies

$$1 - H(t) = (1 - F(t))(1 - G(t)),$$

since Y_1 and C_1 are independent. Further, let $H^u(t) = P(T_1 \le t, \Delta_1 = 1)$ be the subdistribution of the uncensored observations.

The following estimator is widely used for estimating the survival distribution F and is due to Kaplan and Meier (1958) :

$$F_n(t) = 1 - \left\{ \prod_{T_{(i)} \le t} \left(1 - \frac{1}{n - i + 1}\right)^{\Delta_{(i)}} \right\} I(t < T_{(n)}), \tag{1.5}$$

where $T_{(1)}, \ldots, T_{(n)}$ are the ordered T_i, and $\Delta_{(1)}, \ldots, \Delta_{(n)}$ are the corresponding indicators. This estimator is a step function, that jumps at the uncensored observations and has increasing jumps. In the case of no censoring (i.e. $T_i = Y_i$ for all i), the estimator reduces to the empirical distribution function (e.d.f.) $n^{-1} \sum_{i=1}^n I(Y_i \le t)$.

Apart from the distribution function, one is often interested in the estimation of the quantile function. For $0 < p < 1$, we define the p-th quantile $F^{-1}(p)$ as

$$F^{-1}(p) = \inf\{t; F(t) \ge p\}. \tag{1.6}$$

Replacing the distribution function $F(t)$ in (1.6) by its estimator $F_n(t)$, leads to the following estimator for $F^{-1}(p)$:

$$F_n^{-1}(p) = \inf\{t; F_n(t) \ge p\}. \tag{1.7}$$

Földes and Rejtő (1981) showed the strong uniform consistency of the Kaplan-Meier estimator, while the analogue for the quantile estimator was proved by Cheng (1984). A modulus of continuity result for the Kaplan-Meier estimator can be found in Schäfer (1986).

In contrast to the e.d.f., which is an average of i.i.d. terms, the Kaplan-Meier estimator does not have a simple structure. In order to study its asymptotic properties, Lo and Singh (1986) obtained an asymptotic representation, which decomposes $F_n(t) - F(t)$ in an average of i.i.d. terms and a remainder term of lower order.

Theorem 1.1 (Lo and Singh (1986))
Assume F is continuous. Then,

$$F_n(t) - F(t) = n^{-1} \sum_{i=1}^{n} \xi(T_i, \Delta_i, t) + r_n(t),$$

where

$$\xi(z, \delta, t) = (1 - F(t)) \left\{ - \int_0^{z \wedge t} \frac{dH^u(s)}{(1 - H(s))^2} + \frac{I(z \le t, \delta = 1)}{1 - H(z)} \right\},$$

and for $T < \tau_H = \inf\{t; H(t) = 1\}$,

$$\sup_{0 \le t \le T} |r_n(t)| = O(n^{-3/4} (\log n)^{3/4}) \quad a.s.$$

A similar representation can be shown for the quantile estimator (see Lo and Singh (1986)). From this representation, the weak convergence (and in particular the asymptotic normality) of $F_n(t)$ and $F_n^{-1}(p)$ can be easily derived (see Lo and Singh (1986)). However, the expressions of the asymptotic bias and variance of the normal approximation do depend on some unknown quantities, like F, H and H^u. This was one of the motivations for Efron (1981) to propose another approximation method for the distribution of $F_n(t) - F(t)$: the bootstrap. Conditional on the survival times Y_i and the censoring times C_i $(i = 1, \dots, n)$, we define new random variables Y_i^*, C_i^*, T_i^* and Δ_i^* as follows :

$$Y_1^*, \dots, Y_n^* \overset{i.i.d.}{\sim} F_n; C_1^*, \dots, C_n^* \overset{i.i.d.}{\sim} G_n$$
$$Y_i^* \text{ and } C_i^* \text{ are independent (for each } i)$$
$$T_i^* = Y_i^* \wedge C_i^*; \Delta_i^* = I(Y_i^* \le C_i^*),$$

where G_n is the analogue of the Kaplan-Meier estimator for the distribution G (replace Δ in (1.5) with $1 - \Delta$).

We now define the bootstrap analogue of the Kaplan-Meier estimator as

$$F_n^*(t) = 1 - \left\{ \prod_{T_{(i)}^* \le t} \left(1 - \frac{1}{n - i + 1}\right)^{\Delta_{(i)}^*} \right\} I(t < T_{(n)}^*), \tag{1.8}$$

where $T^*_{(1)}, \ldots, T^*_{(n)}$ are the ordered T^*_i, and $\Delta^*_{(1)}, \ldots, \Delta^*_{(n)}$ are the corresponding indicators. Similarly, the bootstrap analogue of the quantile estimator is given by

$$F_n^{*-1}(p) = \inf\{t; F_n^*(t) \geq p\}. \tag{1.9}$$

Via asymptotic representations for $F_n^*(t)$ and $F_n^{*-1}(p)$, the weak convergence of these bootstrapped estimators has been established (Lo and Singh (1986)). As a consequence of this result, a bootstrap approximation of the distribution of $F_n(t)$ and $F_n^{-1}(p)$ can be obtained and confidence bands for the functions F and F^{-1} can be constructed.

1.3 Nonparametric estimation in regression

Suppose that $Y_1, \ldots, Y_n$ are independent random variables, observed at fixed design points $0 \leq x_1 \leq \ldots \leq x_n \leq 1$ and with distribution function $F(t|x_i) = P(Y_i \leq t)$. Let $F(t|x)$ be the conditional distribution of the response at a given design point x. For the (nonparametric) estimation of $F(t|x)$, the following idea is used. If $F(t|x)$ is continuous in x, then observations Y_i with x_i close to x are more informative than observations at design points far away from x. Therefore, observations close to x should have a bigger contribution or weight in the estimator. This leads to an estimator of the following form (Stone (1977)) :

$$F_h(t|x) = \sum_{i=1}^{n} w_{ni}(x; h_n) I(Y_i \leq t), \tag{1.10}$$

where $w_{ni}(x; h_n)$ $(i = 1, \ldots, n)$ are the Gasser-Müller weights

$$w_{ni}(x; h_n) = \frac{1}{c_n(x; h_n)} \int_{x_{i-1}}^{x_i} \frac{1}{h_n} K\left(\frac{x - z}{h_n}\right) dz \qquad i = 1, \ldots, n, \tag{1.11}$$

$$c_n(x; h_n) = \int_0^{x_n} \frac{1}{h_n} K\left(\frac{x - z}{h_n}\right) dz,$$

$x_0 = 0$, K is a known probability density function, called the kernel, and $\{h_n\}$ is a sequence of positive constants, tending to 0 as $n \to \infty$, called the bandwidth sequence.

Note that this estimator is a generalization of the e.d.f., in the sense that the weights n^{-1} are now replaced with more general weights $w_{ni}(x; h_n)$, depending on the relative position of x_i with respect to x.

For $0 < p < 1$, we define the p-th quantile $F^{-1}(p|x)$ and its estimator $F_h^{-1}(p|x)$ as

$$F^{-1}(p|x) = \inf\{t; F(t|x) \geq p\} \tag{1.12}$$

$$F_h^{-1}(p|x) = \inf\{t; F_h(t|x) \geq p\}. \tag{1.13}$$

Stute (1986) showed the consistency of the Stone estimator. The asymptotic normality of $F_h(t|x)$ and $F_h^{-1}(p|x)$ in a general heteroscedastic regression model can be found in Aerts, Janssen and Veraverbeke (1994a). A number of resampling schemes have been proposed in this context. Most schemes resample from the residuals and obtain resampled observations through estimation of the regression function (see Härdle and Mammen (1991) for a bootstrap procedure defined in a homoscedastic model and Härdle and Marron (1990), Härdle and Mammen (1991), Cao Abad (1991) and Härdle, Huet and Jolivet (1995) for the so-called "wild bootstrap" defined in a heteroscedastic regression model). The resampling scheme that will be proposed in this thesis, is based on the one introduced by Aerts, Janssen and Veraverbeke (1994b), which is valid in a general model with no restriction on the relation between the response and the covariate. Let $Y_1, \ldots, Y_n$ be an arbitrary sample. Conditional on this sample, they define the bootstrap sample $Y_1^*, \ldots, Y_n^*$ as independent random variables with

$$Y_j^* \sim F_{x_j g}$$

($j = 1, \ldots, n$). Here, $\{g_n\}$ is a second bandwidth sequence, which is usually chosen to be asymptotically larger than $\{h_n\}$, i.e. $g_n/h_n \to \infty$ in a certain way. (This technique of oversmoothing with the initial bandwidth has also been used by Romano (1988) and Härdle and Mammen (1991).) The bootstrap analogues of (1.10) and (1.13), based on this bootstrap sample, are defined as

$$F_{hg}^*(t|x) = \sum_{i=1}^{n} w_{ni}(x; h_n) I(Y_i^* \leq t) \tag{1.14}$$

$$F_{hg}^{*-1}(p|x) = \inf\{t; F_{hg}^*(t|x) \geq p\}. \tag{1.15}$$

That this resampling method can be used to estimate the distribution of $F_h(t|x)$ and $F_h^{-1}(p|x)$ was shown in Aerts, Janssen and Veraverbeke (1994b).

1.4 Nonparametric estimation in regression with censored data

We will now combine the ideas of Sections 1.2 and 1.3 and present an estimator for the distribution function in a regression model where the data are subject to censoring.

Let us first introduce some notation. Again we let $0 \leq x_1 \leq \ldots \leq x_n \leq 1$ be fixed design points at which we observe independent and non-negative responses $Y_1, \ldots, Y_n$, representing survival or failure times. Denote $F(t|x_i) = P(Y_i \leq t)$ for the distribution of the response Y_i at x_i.

As often occurs in clinical trials or industrial life tests, the responses $Y_1, \ldots, Y_n$ are subject to random right censoring, i.e. the observed random variables at design point x_i are in fact T_i and Δ_i $(i = 1, \ldots, n)$, with

$$T_i = Y_i \wedge C_i \qquad \text{and} \qquad \Delta_i = I(Y_i \leq C_i),$$

where $C_1, \ldots, C_n$ are independent and non-negative censoring variables with distribution functions $G(t|x_i) = P(C_i \leq t)$. We will assume independence of Y_i and C_i for each i. Consequently we have that the distribution function $H(t|x_i) = P(T_i \leq t)$ satisfies the relation

$$1 - H(t|x_i) = (1 - F(t|x_i))(1 - G(t|x_i)). \tag{1.16}$$

At a given fixed design value $x \in]0, 1[$, we write $F(\cdot|x)$, $G(\cdot|x)$, $H(\cdot|x)$ for the distribution function of respectively the response Y_x at x, the censoring variable C_x at x and $T_x = Y_x \wedge C_x$. Also we will write $\Delta_x = I(Y_x \leq C_x)$ (Note that for the design variables x_i we write Y_i, C_i, T_i, Δ_i instead of Y_{x_i}, C_{x_i}, T_{x_i}, Δ_{x_i}).

Beran (1981) was the first one who studied regression problems with censored data in a fully nonparametric way. His estimator is a generalization of the Kaplan-

Meier estimator and is sometimes called the conditional Kaplan-Meier estimator :

$$F_h(t|x) = 1 - \left\{ \prod_{T_{(i)} \leq t} \left(1 - \frac{w_{n(i)}(x; h_n)}{1 - \sum\limits_{j=1}^{i-1} w_{n(j)}(x; h_n)} \right)^{\Delta_{(i)}} \right\} I(t < T_{(n)}). \tag{1.17}$$

Here $T_{(1)} \leq \ldots \leq T_{(n)}$ are the ordered T_i, and $\Delta_{(1)}, \ldots, \Delta_{(n)}$ and $w_{n(1)}(x; h_n), \ldots, w_{n(n)}(x; h_n)$ are the corresponding Δ_i and $w_{ni}(x; h_n)$ (where $w_{ni}(x; h_n)$ is as defined in (1.11)). Note that if we think the weights $w_{ni}(x; h_n)$ all equal to n^{-1}, then $F_h(t|x)$ becomes the classical Kaplan-Meier estimator. Clearly, $F_h(t|x)$ is a step function with jumps only at the uncensored observations. The jump size at $T_{(j)}$ is given by

$$W_{nj} = \Delta_{(j)} \frac{w_{n(j)}(x; h_n)}{1 - \sum_{k=1}^{j-1} w_{n(k)}(x; h_n)} \prod_{i=1}^{j-1} \left(1 - \frac{w_{n(i)}(x; h_n)}{1 - \sum_{k=1}^{i-1} w_{n(k)}(x; h_n)} \right)^{\Delta_{(i)}} \qquad (j < n)$$

$$W_{nn} = \prod_{i=1}^{n-1} \left(1 - \frac{w_{n(i)}(x; h_n)}{1 - \sum_{k=1}^{i-1} w_{n(k)}(x; h_n)} \right)^{\Delta_{(i)}},$$

so that $F_h(t|x)$ can also be expressed as

$$F_h(t|x) = \sum_{j=1}^{n} W_{nj} I(T_{(j)} \leq t).$$

In case of no censoring (i.e. $T_i = Y_i$ for all i), $W_{nj} = w_{n(j)}(x; h_n)$ and $F_h(t|x)$ becomes the kernel estimator of Stone.

For $0 < p < 1$, we denote

$$F^{-1}(p|x) = \inf\{t; F(t|x) \geq p\} \tag{1.18}$$

$$F_h^{-1}(p|x) = \inf\{t; F_h(t|x) \geq p\} \tag{1.19}$$

for the p-th quantile $F^{-1}(p|x)$ and its estimator $F_h^{-1}(p|x)$.

The starting point for the use of the above type of estimators in regression problems with censored data was Beran (1981). His estimator was further studied by Dabrowska (1987, 1989, 1992a,b), McKeague and Utikal (1990) and Akritas (1994) in the random design case, and by González Manteiga and Cadarso Suárez (1994) in the fixed design case. Gentleman and Crowley (1991) proposed the quantile estimator in (1.19) as a graphical tool to present the evolution of the survival time as a function of the covariate. Li and Doss (1995) studied the estimation of the hazard rate, the cumulative hazard function and the survival function in regression with censored data using martingale techniques. In contrast to the Beran estimator, which is a local constant estimator, they constructed a local linear estimator for $F(t|x)$ and performed a similar study as McKeague and Utikal (1990) did for the Beran estimator. Their research was motivated by the knowledge that for the nonparametric estimation of the regression function in the uncensored case, the local constant estimator has larger bias near the endpoints of the covariate region than the local linear estimator.

Beran's (1981) main idea behind the construction of the estimator $F_h(t|x)$ in (1.17) was to use the 1-1 relationship between the distribution function $F(t|x)$ and the cumulative hazard function $\Lambda(t|x)$, defined by

$$\Lambda(t|x) = \int_0^t \frac{dF(s|x)}{1 - F(s-|x)}.$$

Let $H^u(\cdot|x)$ denote the subdistribution function of the uncensored observations, given by

$$H^u(t|x) = P(T_x \leq t, \Delta_x = 1) = \int_0^t (1 - G(s-|x))dF(s|x).$$

Then, using relation (1.16),

$$\Lambda(t|x) = \int_0^t \frac{dH^u(s|x)}{1 - H(s-|x)}. \tag{1.20}$$

We now replace $H(\cdot|x)$ and $H^u(\cdot|x)$ in (1.20) by the following Stone-type estimators :

$$H_h(t|x) = \sum_{i=1}^n w_{ni}(x; h_n) I(T_i \leq t) \tag{1.21}$$

$$H_h^u(t|x) = \sum_{i=1}^n w_{ni}(x; h_n) I(T_i \leq t, \Delta_i = 1), \tag{1.22}$$

which leads to the following Nelson-Aalen type estimator for $\Lambda(t|x)$:

$$\Lambda_h(t|x) = \int_0^t \frac{dF_h(s|x)}{1 - F_h(s-|x)} = \int_0^t \frac{dH_h^u(s|x)}{1 - H_h(s-|x)}. \qquad (1.23)$$

The distribution function corresponding to $\Lambda_h(t|x)$ is given by $1-\prod(1-\Lambda_h(\{a_i\}|x))$, where the product is taken over all atoms a_i of $\Lambda_h(\cdot|x)$ with $a_i \leq t$. Since $\Lambda_h(t|x)$ has only mass at the uncensored T_i, this becomes

$$1 - \prod_{T_i \leq t} \left(1 - \frac{w_{ni}(x;h_n)}{1 - H_h(T_i - |x)}\right)^{\Delta_i},$$

which coincides with the definition of $F_h(t|x)$ in (1.17) for $t < T_{(n)}$.

We will study the Beran estimator $F_h(t|x)$ in more detail in the next chapter. We will start by showing the uniform strong consistency and a modulus of continuity result for $F_h(t|x)$. Next, a representation for $F_h(t|x)$ will be derived, from which the weak convergence (and the asymptotic normality) will follow. In the last section, a number of basic results concerning the Stone estimator, that are needed in the proofs, will be established. All the results of this chapter were developed in Van Keilegom and Veraverbeke (1996, 1997a,b).

The quantile estimator $F_h^{-1}(p|x)$ will be studied in Chapter 4. Also there, the strong consistency, the construction of a representation and the weak convergence are the main results of the study (see Van Keilegom and Veraverbeke (1996, 1997a, 1998)).

It will be shown that, as we have seen for the usual Kaplan-Meier estimator, the asymptotic bias and variance of $F_h(t|x)$ and $F_h^{-1}(p|x)$ depend on some unknown quantities. Therefore, a bootstrap procedure will be very useful, since it will provide an alternative for the normal approximation of the distribution of $F_h(t|x)$ and $F_h^{-1}(p|x)$.

Our procedure combines both the bootstrap ideas of Efron (1981) for censored data and of Aerts, Janssen and Veraverbeke (1994b) for fixed design regression. Given the design points x_i, and conditional on the responses Y_i and censoring times C_i $(i = 1, \ldots, n)$ we define the random variables Y_i^* and C_i^* (independently) as follows :

$$Y_1^*, \ldots, Y_n^* \text{ are independent; } Y_i^* \sim F_{x_i g}$$
$$C_1^*, \ldots, C_n^* \text{ are independent; } C_i^* \sim G_{x_i g}.$$

Here $F_{x_i g}$ is the estimator for F_{x_i} as defined in (1.17), but with a bandwidth sequence $\{g_n\}$, which is different from $\{h_n\}$. The distribution $G_{x_i g}$ is the analogous estimator for G_{x_i}. Then define, for $i = 1, \ldots, n$,

$$T_i^* = Y_i^* \wedge C_i^* \text{ and } \Delta_i^* = I(Y_i^* \leq C_i^*).$$

It is readily verified that the above procedure is equivalent to one where the pairs (T_i^*, Δ_i^*) are drawn (with replacement) from $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$, giving probability $w_{nj}(x_i; g_n)$ to (T_j, Δ_j) for $j = 1, \ldots, n$.

Based on the bootstrap sample $(T_1^*, \Delta_1^*), \ldots, (T_n^*, \Delta_n^*)$, the bootstrap analogues of the Kaplan-Meier type estimator in (1.17) and the quantile estimator in (1.19) are given by

$$F_{hg}^*(t|x) = 1 - \left\{ \prod_{T_{(i)}^* \leq t} \left(1 - \frac{w_{n(i)}(x; h_n)}{1 - \sum_{j=1}^{i-1} w_{n(j)}(x; h_n)} \right)^{\Delta_{(i)}^*} \right\} I(t < T_{(n)}^*) \quad (1.24)$$

$$F_{hg}^{*-1}(p|x) = \inf\{t; F_{hg}^*(t|x) \geq p\}, \quad (1.25)$$

where $T_{(1)}^* \leq \ldots \leq T_{(n)}^*$, and $\Delta_{(i)}^*$ and $w_{n(i)}(x; h_n)$ correspond to $T_{(i)}^*$. In case of ties, we make the usual convention that uncensored observations are considered to occur just before censored observations. It is easy to see that $F_{hg}^*(t|x)$ in (1.24) is well defined in the case that two or more observations occur at the same time and that formula (1.24) can also be written as

$$F_{hg}^*(t|x) = 1 - \left\{ \prod_{T_{(i)} \leq t} \left(1 - \frac{\overline{w}_{n(i)}^u(x; h_n)}{1 - \sum_{j=1}^{i-1} \overline{w}_{n(j)}(x; h_n)} \right) \right\} I(t < T_{(n)}^*),$$

where

$$\overline{w}_{n(i)}(x;h_n) = \sum_{k=1}^{n} w_{nk}(x;h_n) I(T_k^* = T_{(i)})$$

$$\overline{w}^u_{n(i)}(x;h_n) = \sum_{k=1}^{n} w_{nk}(x;h_n) I(T_k^* = T_{(i)}, \Delta_k^* = 1).$$

In a similar way, let

$$H^*_{hg}(t|x) = \sum_{i=1}^{n} w_{ni}(x;h_n) I(T_i^* \leq t) \tag{1.26}$$

$$H^{*u}_{hg}(t|x) = \sum_{i=1}^{n} w_{ni}(x;h_n) I(T_i^* \leq t, \Delta_i^* = 1) \tag{1.27}$$

be the analogues of $H_h(t|x)$ and $H^u_h(t|x)$ in (1.21) and (1.22) for the bootstrapped data.

As was pointed out already in Section 1.3, the parameter g_n which is used to construct the resampled values, is an appropriate pilot bandwidth sequence which is typically asymptotically larger than h_n, i.e. $g_n/h_n \to \infty$ in a certain way. This technique of oversmoothing with the initial bandwidth has been successfully used in other resampling schemes in regression (e.g. Härdle and Mammen (1991), Aerts, Janssen and Veraverbeke (1994b)). It entails that the bootstrap bias and bootstrap variance are asymptotically appropriate estimators for the bias and variance terms.

In Chapter 3, the bootstrapped estimator $F^*_{hg}(t|x)$ will be studied in more detail. First, an asymptotic representation will be obtained, from which the weak convergence will follow. Second, as an application to the weak convergence result, it will be shown that the bootstrap method works (in the sense that it provides us with an alternative to the normal approximation) and bootstrap confidence bands for $F(\cdot|x)$ will be constructed. In the next section, the finite sample performance of the normal and the bootstrap approximation will be compared in a simulation study. Finally, a number of results concerning the bootstrapped Stone estimator, which are needed for showing the main results, will be derived. This chapter contains the results established in Van Keilegom and Veraverbeke (1997a,b).

The bootstrapped quantile estimator $F^{*-1}_{hg}(p|x)$ will be studied in Chapter 4. A similar study as for the bootstrap estimator of the distribution function will be carried out : due to its complicated structure we will start by proving an asymptotic representation for $F_{hg}^{*-1}(p|x)$. As a consequence of this result, the weak convergence of $F_{hg}^{*-1}(p|x)$ will be established, it will be shown that the bootstrap works in this context as well, and confidence bands for $F^{-1}(\cdot|x)$ will be constructed. Also, the normal and the bootstrap approximation of the quantile function $F^{-1}(\cdot|x)$ will be compared by means of some simulations (see Van Keilegom and Veraverbeke (1997a, 1998)).

1.5 Nonparametric estimation in heteroscedastic regression models with censored data

In Section 1.4, we introduced the Beran estimator for estimating the conditional distribution function in a completely general regression model where the data are subject to random right censoring. As with the ordinary Kaplan-Meier estimator, the tails of the Beran estimator may contain, however, little information if the censoring is "heavy". In particular, the Beran estimator cannot estimate the distribution $F(t|x)$ for t greater than the right endpoint of the support of the conditional distribution of the censoring variable at x. This is due to the inherent lack of information and cannot be overcome in a completely general setting. In the second part of this thesis, we will not work in a completely general regression model, but we will assume that a heteroscedastic regression model is valid. Under that model, an estimator for $F(t|x)$ will be constructed, which can have much more information in the upper tail than the Beran estimator.

Consider the heteroscedastic regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.28}$$

where Y is (a possible transformation of) the survival time, X is a random covariate, the error variable ε is independent of X, the function $m(\cdot)$ is the unknown regression curve and $\sigma(\cdot)$ is a conditional scale functional, representing possible heteroscedasticity. We further assume that $m(\cdot)$ and $\sigma(\cdot)$ are, respectively, a location and scale functional. This means that we can write $m(x) = T(F_Y(\cdot|x))$ and $\sigma(x) = S(F_Y(\cdot|x))$ for some functionals T and S, such that

$$T(F_{aY+b}(\cdot|x)) = aT(F_Y(\cdot|x)) + b \tag{1.29}$$

and

$$S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x)) \tag{1.30}$$

for all $a \geq 0$ and $b \in I\!R$, where $F_Y(\cdot|x)$ denotes here the distribution of Y given $X = x$ (see also Huber (1981), p. 59, 202). Under model (1.28), we will study an estimator for the conditional distribution of Y given X, an estimator for the bivariate distribution of X and Y, and a least squares estimator for the regression function in a polynomial regression model.

Model (1.28) has been considered in Fan and Gijbels (1994) who studied estimation of m by local linear regression. When Y is the logarithm of the survival time, model (1.28) can be viewed as a nonparametric version of the accelerated failure time model (1.3).

Let (X_i, T_i, Δ_i) $(i = 1, \ldots, n)$ be independent replications of (X, T, Δ), $T = Y \wedge C$, $\Delta = I(Y \leq C)$ and let C be the censoring random variable which is conditionally independent of Y given X. Model (1.28) implies that

$$F(t|x) = P(Y \leq t|X = x) = F_e\left(\frac{t - m(x)}{\sigma(x)}\right), \tag{1.31}$$

where $F_e(t) = P(\varepsilon \leq t)$ denotes the distribution of the error variable ε. Equation (1.31) suggests that it should be possible to estimate the tail of the conditional distribution $F(\cdot|x)$ for all x, provided that there is a region of x-values where the censoring is "light" (since this implies that the tail of F_e can be estimated). We will estimate $F(t|x)$ by replacing $F_e(\cdot), m(\cdot)$ and $\sigma(\cdot)$ with suitable estimators. First, to estimate m and σ we will work with the particular definitions

$$m(x) = \int_0^1 F^{-1}(s|x)J(s)\,ds, \tag{1.32}$$

$$\sigma^2(x) = \int_0^1 F^{-1}(s|x)^2 J(s)\,ds - m^2(x), \tag{1.33}$$

which are a special case of the general functional for L-statistics (see e.g. Serfling (1980)) and where $F^{-1}(s|x) = \inf\{t; F(t|x) \geq s\}$ is the quantile function of Y given x and $J(s)$ is a given score function satisfying $\int_0^1 J(s)\,ds = 1$. We will show in Proposition 5.1 that if model (1.28) holds for a certain location functional m and scale functional σ, then it holds for all location functionals $\tilde{m}$ and scale functionals $\tilde{\sigma}$, in the sense that the resulting error term $\tilde{\varepsilon}$ is still independent of X. Hence, working with the functionals m and σ defined in (1.32) and (1.33) is no restriction of generality.

To estimate m and σ, replace the distribution $F(s|x)$ in (1.32) and (1.33) with the Beran estimator, introduced in Section 1.4 :

$$\tilde{F}(t|x) = 1 - \prod_{T_{(i)} \leq t} \left\{ 1 - \frac{W_{n(i)}(x; h_n)}{1 - \sum_{j=1}^{i-1} W_{n(j)}(x; h_n)} \right\}^{\Delta_{(i)}}, \tag{1.34}$$

where $W_{ni}(x; h_n)$ $(i = 1, \ldots, n)$ are the Nadaraya-Watson weights

$$W_{ni}(x; h_n) = \frac{K\left(\frac{x - X_i}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)} \tag{1.35}$$

(with K a known probability density function (kernel) and $\{h_n\}$ a bandwidth sequence), where $T_{(i)}$ is the i-th order statistic of $T_1, \ldots, T_n$, and $W_{n(i)}(x; h_n)$ and $\Delta_{(i)}$ are the corresponding $W_{ni}(x; h_n)$ and Δ_i. Note that the Beran estimator introduced in (1.34) is denoted in a different way than in Section 1.4. This is because there is no need here to stress the dependence of the Beran estimator on the particular bandwidth sequence used. We choose to work here with Nadaraya-Watson weights instead of Gasser-Müller weights (which we worked with in Section 1.4), since the former weights are more common in random design regression and are no longer inferior to the latter (as is the case for fixed design) (see also Gasser and Engel (1990) and Chu and Marron (1991)).

Expression (1.34) yields

$$\hat{m}(x) = \int_0^1 \tilde{F}^{-1}(s|x) J(s)\, ds, \tag{1.36}$$

$$\hat{\sigma}^2(x) = \int_0^1 \tilde{F}^{-1}(s|x)^2 J(s)\, ds - \hat{m}^2(x) \tag{1.37}$$

as estimators for $m(x)$ and $\sigma(x)$. Set $\hat{E}_i = (T_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$ for the resulting censored residuals, and let

$$\hat{F}_e(t) = 1 - \prod_{\hat{E}_{(i)} \le t} \left(1 - \frac{1}{n-i+1}\right)^{\Delta_{(i)}} \tag{1.38}$$

denote the Kaplan-Meier estimator of F_e, where $\hat{E}_{(i)}$ is the i-th order statistic of $\hat{E}_1, \ldots, \hat{E}_n$ and $\Delta_{(i)}$ is the corresponding indicator.

Then, relation (1.31) suggests

$$\hat{F}(t|x) = \hat{F}_e\left(\frac{t - \hat{m}(x)}{\hat{\sigma}(x)}\right) \tag{1.39}$$

as an estimator of $F(t|x)$. To see why this has the aforementioned advantage over the Beran estimator, suppose that there exists a subset R of the support of X such that for all $x \in R$,

$$\tau_{F(\cdot|x)} \le \tau_{G(\cdot|x)}, \tag{1.40}$$

where for any distribution F, τ_F denotes the right endpoint of the support of F and where $G(t|x) = P(C \le t|x)$ is the conditional distribution of C given X. Then, $F(t|x)$ can be estimated consistently for all $t < \tau_{F(\cdot|x)}$ and x in R. This implies that the right tail of F_e can be well estimated by the right tail of $\hat{F}_e$. In view of (1.39), it follows that the right tail of $F(t|x)$ can be well estimated for any x. Even if assumption (1.40) would not be satisfied, the estimator (1.39) is still useful, since by transferring tail information, the tails will be estimated at least as well.

The scenario of different degrees of censoring for different regions of the covariates often arises in practice. Consider the typical case in which the covariate is a strong prognostic variable for the event of interest. Then, in regions of the covariate indicating "high risk", the amount of censoring is considerably lower than in regions indicating "low risk", as here extended follow-up is needed to observe the event under study. In this situation, tail information in the "high risk" region can be carried over to the "low risk" region.

Chapter 5 of this thesis deals with the study of the asymptotic behavior of the estimators $\hat{F}_e(t)$ and $\hat{F}(t|x)$ and contains the results established in Van Keilegom and Akritas (1997a). We will prove an asymptotic representation and the weak convergence of both estimators. In order to show these results, a study of the asymptotic properties of the estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$ will be required. Using the uniform (in both x and t) consistency of the Beran estimator and its derivative, the consistency of $\hat{m}(x)$, $\hat{m}'(x)$, $(\hat{m}'(x) - \hat{m}'(y))/(x - y)$ and their analogues for $\hat{\sigma}(x)$ will be established.

In Chapter 7 we will compare the finite sample performance of the Beran estimator $\tilde{F}(t|x)$ (defined in (1.34)) and the estimator $\hat{F}(t|x)$ (defined in (1.39)) through some simulations. An important drawback of the Beran estimator $\tilde{F}(t|x)$ is its inconsistency in the right tail whenever heavy censoring is present. This is due to the lack of information in the right tail and hence this problem cannot be overcome when working in a completely nonparametric way. The estimator $\hat{F}(t|x)$ is not constructed in a completely nonparametric way, but is valid under model (1.28). For many situations (e.g. in the case where condition (1.40) is satisfied) we intuitively expect that the estimator $\hat{F}(t|x)$ does not share this drawback, as is explained above. The simulations which we will perform in Chapter 7, will confirm these intuitive ideas. We will also examine which choice of the location functional $m(x)$ and scale functional $\sigma(x)$ leads (in general) to the best preformance of the estimator $\hat{F}(t|x)$ (recall that model (1.28) holds for all location and scale functionals if it holds for one location and one scale functional (see Proposition 5.1)). Since the estimator $\hat{F}(t|x)$ is constructed under the assumption that model (1.28) is valid, it is interesting to know how $\hat{F}(t|x)$ behaves in situations where model (1.28) is not satisfied. The performance of $\tilde{F}(t|x)$ and $\hat{F}(t|x)$ will be compared in this situation. We will close Chapter 7 by analyzing the data from the Stanford heart transplant program (described in Section 1.1). The conditional distribution of the survival time will be estimated for certain age values, as well as the median regression curve.

The estimator of the conditional distribution of Y given X defined in (1.39), can be used to estimate the bivariate distribution of X and Y, which we denote by $F(x,t) = P(X \leq x, Y \leq t)$. Note that under model (1.28),

$$F(x,t) = \int_{-\infty}^{x} F(t|u)\, dF_X(u) = \int_{-\infty}^{x} F_e\left(\frac{t-m(u)}{\sigma(u)}\right) dF_X(u), \tag{1.41}$$

where $F_X(x) = P(X \le x)$ is the marginal distribution of the covariate X. Hence, we define

$$\hat{F}(x,t) = \int_{-\infty}^{x} \hat{F}(t|u)\, d\hat{F}_X(u) = \int_{-\infty}^{x} \hat{F}_e\left(\frac{t-\hat{m}(u)}{\hat{\sigma}(u)}\right) d\hat{F}_X(u), \tag{1.42}$$

where $\hat{F}_X(x) = n^{-1}\sum_{i=1}^n I(X_i \le x)$ denotes the empirical distribution function of the X_i. This estimator is closely related to the one studied in Akritas (1994) in a general regression model, with no constraint on the relation between X and Y. Instead of replacing $F(t|u)$ in (1.41) with $\hat{F}(t|u)$, he replaced it with the Beran estimator $\tilde{F}(t|u)$. The problem of estimating $F(x,t)$ when Y is subject to censoring, but X is completely observed, has received considerable attention in the literature, but all existing estimators (e.g. Dabrowska (1988), Stute (1993, 1996), Akritas (1994), van der Laan (1996)) have regions of unassigned mass when censoring in the upper tails is heavy (for the estimator in Akritas (1994), this follows from the fact that the tail mass of the Beran estimator will be unassigned when the censoring is heavy). Under assumption (1.40), the estimator defined in (1.42) does not share this feature, because, as we mentioned before, due to the transfer of information in the tails, the estimator $\hat{F}(t|x)$ will be consistent for all $t < \tau_{F(\cdot|x)}$ and all x in R.

In Chapter 6 we will show an asymptotic representation for the estimator $\hat{F}(x,t)$ and show its weak convergence to a Gaussian process. Use will be made of the results obtained in Chapter 5 concerning the estimator $\hat{F}(t|x)$. These results are obtained in Van Keilegom and Akritas (1997b).

Consider now the heteroscedastic polynomial regression model

$$Y = \beta_0 + \beta_1 X + \ldots + \beta_p X^p + \sqrt{\mathrm{Var}(Y|X)}\varepsilon, \tag{1.43}$$

where $E(\varepsilon) = 0, \mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2$ and ε and X are independent. Thus we assume that $E(\boldsymbol{Y}|\boldsymbol{X}) = \boldsymbol{X\beta}$, where $\boldsymbol{Y}$ is the $n \times 1$ response vector, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$ and

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_1 & \ldots & X_1^p \\ \vdots & \vdots & & \vdots \\ 1 & X_n & \ldots & X_n^p \end{pmatrix}$$

denotes the design matrix. We define the least squares estimator as the minimizer w.r.t. $\boldsymbol{\beta}$ of

$$\int (t - \beta_0 - \beta_1 x - \ldots - \beta_p x^p)^2 d\hat{F}(x,t). \tag{1.44}$$

Note that model (1.43) is of the same form as model (1.28), with $m(X) = E(Y|X) = \beta_0 + \ldots + \beta_p X^p$ and $\sigma(X) = \sqrt{\mathrm{Var}(Y|X)}$. Since this m and σ are respectively a location and a scale functional, it can be shown (see Proposition 5.1) that model (1.28) holds for any location functional $\tilde{m}$ and scale functional $\tilde{\sigma}$, in the sense that the resulting error term $\tilde{\varepsilon}$ is still independent of X. Hence, in the construction of the estimator $\hat{F}(x,t)$ in (1.43), one can use any location functional $\tilde{m}$ and scale functional $\tilde{\sigma}$, and not necessarily the functionals from model (1.43). The solution to the minimization problem in (1.44) is

$$\hat{\boldsymbol{\beta}} = n(\boldsymbol{X}'\boldsymbol{X})^{-1} \begin{pmatrix} \int x^0 t \, d\hat{F}(x,t) \\ \vdots \\ \int x^p t \, d\hat{F}(x,t) \end{pmatrix}, \tag{1.45}$$

provided $\hat{F}_e(\infty) = 1$, while the true parameter values are given by

$$\boldsymbol{\beta} = n[E(\boldsymbol{X}'\boldsymbol{X})]^{-1} \begin{pmatrix} \int x^0 t \, dF(x,t) \\ \vdots \\ \int x^p t \, dF(x,t) \end{pmatrix}. \tag{1.46}$$

It follows from (1.44) that $\hat{\boldsymbol{\beta}}$ equals the uncensored data least squares estimator when $\hat{F}(x,t)$ is replaced with the usual bivariate empirical distribution function corresponding to (X_i, Y_i).

Two ideas have been proposed in the literature for extending the least squares estimator to censored data. The first one was to substitute censored data versions in place of corresponding uncensored empirical distributions in expressions yielding the uncensored least squares estimator. The second idea was to use suitably transformed (or "synthetic") data and subsequently apply ordinary least squares. To the first group of extensions of the ordinary least squares, belong the estimators of Miller (1976), Stute (1993, 1996) and Akritas (1994). Also the estimator (1.45) is based on this idea. The second group consists of the estimator of Buckley and James (1979) (which was further studied by Ritov (1990), Lai and Ying (1991a) and Lin and Wei (1992a)), the estimator proposed by Koul, Susarla and Van Ryzin (1981) (Fygenson and Zhou (1994) considered a slight modification of this estimator), the one of Leurgans (1987) (also studied by Zhou (1992a)) and the estimator studied in Akritas (1996). However, all these estimators (except estimator (1.45)) pertain to a homoscedastic regression model.

The estimators studied by Stute (1993, 1996) and Akritas (1994) are somewhat similar to (1.45), but they have the disadvantage that their asymptotic bias increases as the censoring in the upper tails increases. In addition, the estimator in Stute (1993, 1996) uses the assumption that the censoring variable is unconditionally independent of the response variable. The least squares estimator (1.45) minimizes the undesirable effects of heavy censoring in the upper tails, due to the fact that $\hat{F}(x,t)$ minimizes the regions of undefined mass. In particular, if assumption (1.40) holds, then $\hat{F}(x,t)$ has, asymptotically, no regions of undefined mass.

Apart from least squares methods, also a number of other methods have been proposed in the statistical literature for fitting polynomial regression models with censored data. To overcome some of the robustness limitations of the least squares method, Tsiatis (1990), Wei, Ying and Lin (1990), Lai and Ying (1991b, 1994), Zhou (1992b) and Yang (1997) used M-estimators or rank procedures to obtain a more robust estimator for $\boldsymbol{\beta}$. A number of other estimation methods, using other approaches, have been proposed, see LeBlanc and Crowley (1990), Lin and Wei (1992b), Fygenson and Ritov (1994) and Ying, Jung and Wei (1995).

We will study the least squares estimator $\hat{\boldsymbol{\beta}}$ in more detail in Chapter 6. First, an asymptotic representation for each of the $\hat{\beta}_k$ $(k = 0, \ldots, p)$ will be established. As a consequence, the asymptotic normality of $\hat{\beta}_k$ will then follow. These results can be found in Van Keilegom and Akritas (1997b).

Chapter 2

The conditional distribution in regression with censored data

We will focus in this chapter on the nonparametric estimation of the conditional distribution function in (general) regression models with censored data. In particular, the nonparametric estimator introduced by Beran (1981) and defined in (1.17), will be studied. In Section 2.1, the assumptions under which the main results are valid, will be stated. The uniform strong consistency and a modulus of continuity result for the estimator will be derived in Section 2.2. They will be obtained via exponential probability bound results and can also be found in Van Keilegom and Veraverbeke (1996). Section 2.3 contains a representation for the Beran estimator, from which the weak convergence will follow. The former result is part of Van Keilegom and Veraverbeke (1997b), while the latter can be found in Van Keilegom and Veraverbeke (1997a). In the appendix of this chapter a number of useful results concerning the Stone estimator, which is the uncensored version of the Beran estimator, will be established. They are needed in the proofs of the results in Sections 2.2 and 2.3 (see also Van Keilegom and Veraverbeke (1997b)).

2.1 Assumptions

Apart from the notations and definitions introduced in Section 1.4, a number of other notations will be needed throughout Chapters 2 till 4.

For the design points $x_1, \ldots, x_n$ we write $\underline{\Delta}_n = \min_{1\le i\le n}(x_i - x_{i-1})$ and $\overline{\Delta}_n = \max_{1\le i\le n}(x_i - x_{i-1})$ (with $x_0 = 0$). The notations

$$\|K\|_\infty = \sup_{u\in I\!R} K(u), \qquad \|K\|_2^2 = \int_{-\infty}^{\infty} K^2(u)\, du,$$

$$\mu_1^K = \int_{-\infty}^{\infty} uK(u)\, du, \quad \text{and} \quad \mu_2^K = \int_{-\infty}^{\infty} u^2K(u)\, du$$

will be used for the kernel K. The following conditions on the design and on the kernel will be assumed throughout :

$(C1)$ $x_n \to 1$, $\overline{\Delta}_n = O(n^{-1})$, $\overline{\Delta}_n - \underline{\Delta}_n = o(n^{-1})$

$(C2)$ K is a probability density function with finite support $[-L_0, L_0]$ for some $L_0 > 0$, $\mu_1^K = 0$, and K is Lipschitz of order 1.

Note that, for $c_n(x; h_n)$ defined in (1.11), $c_n(x; h_n) = 1$ for n sufficiently large (depending on x) since $x_n \to 1$ and K has finite support. This makes that in all proofs of asymptotic results, we will take $c_n(x; h_n) = 1$.

Concerning the support of distribution functions we will use the following notation : if L is any (sub)distribution function, then τ_L denotes the right endpoint of its support, i.e. $\tau_L = \inf\{t; L(t) = L(\infty)\}$. Clearly, $\tau_{H(\cdot|x)} = \tau_{F(\cdot|x)} \wedge \tau_{G(\cdot|x)}$ by (1.16).

In the formulation of our results, we will need typical types of smoothness conditions on functions like $H(t|x)$ and $H^u(t|x)$. We formulate them here for a general (sub)distribution function $L(t|x)$, $t \in I\!R$, $0 \le x \le 1$, and for a fixed $T > 0$.

$(C3)$ $\dot{L}(t|x) = \dfrac{\partial}{\partial x} L(t|x)$ exists and is continuous in $(x, t) \in [0, 1] \times [0, T]$.

$(C4)$ $L'(t|x) = \dfrac{\partial}{\partial t} L(t|x)$ exists and is continuous in $(x, t) \in [0, 1] \times [0, T]$.

$(C5)$ $\ddot{L}(t|x) = \dfrac{\partial^2}{\partial x^2} L(t|x)$ exists and is continuous in $(x, t) \in [0, 1] \times [0, T]$.

$(C6)$ $L''(t|x) = \dfrac{\partial^2}{\partial t^2} L(t|x)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$.

$(C7)$ $\dot{L}'(t|x) = \dfrac{\partial^2}{\partial x \partial t} L(t|x)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$.

Note that $(C5)$ and $(C7)$ imply $(C3)$ and that $(C6)$ and $(C7)$ imply $(C4)$. Also, $(C3)$ implies that $L(t|x)$ is Lipschitz, in the sense that for all $0 \le x, x' \le 1$,

$$\sup_{0 \le t \le T} |L(t|x) - L(t|x')| \le \|\dot{L}\| \, |x - x'|,$$

where $\|M\| = \sup_{0 \le x \le 1} \sup_{0 \le t \le T} |M(t|x)|$ for any function M on $[0,1] \times [0,T]$. Similarly, $(C4)$ implies that for all $0 \le t, t' \le T$,

$$\sup_{0 \le x \le 1} |L(t|x) - L(t'|x)| \le \|L'\| \, |t - t'|.$$

Also note that imposing conditions $(C3)$ and $(C4)$ on $H(t|x)$ and $H^u(t|x)$ implies that $F(t|x)$ and $G(t|x)$ are continuous in $(x,t) \in [0,1] \times [0,T]$.

Finally, we note that, since most of the results are at a fixed design point x, we could relax the above conditions $(C3) - (C7)$ by requiring the continuity only in $U_x \times [0,T]$ instead of $[0,1] \times [0,T]$, with U_x some open neighborhood of the fixed point x.

2.2 Consistency and modulus of continuity result

A first interesting property of the Beran estimator $F_h(t|x)$ is its uniform strong consistency. This will follow from an exponential bound result for the tail probabilities of the distribution of the Beran estimator. Such type of inequality plays a role similar to the Dvoretzky-Kiefer-Wolfowitz (1956) inequality for the classical empirical distribution function and the inequality established by Főldes and Rejtö (1981) for the Kaplan-Meier estimator. In the appendix of this chapter we will show a similar result for the Stone estimator. They immediately imply the uniform strong consistency of the estimator by making use of the Borel-Cantelli lemma. In a random design regression model with censored data, Dabrowska (1989) obtained an exponential probability bound for $\sup_{0 \le t \le T} |F_h(t|x) - EF_h(t|x)|$, by making use of the theory of copula functions. Under the condition that $nh_n^5 \to 0$, the bias term $\sup_{0\le t\le T} |EF_h(t|x) - F(t|x)|$ is negligible with respect to the stochastic component, and hence the strong consistency of the Beran estimator follows from the Borel-Cantelli lemma. In the next result, we show an exponential probability bound for the complete expression $\sup_{0\le t\le T} |F_h(t|x) - F(t|x)|$ in a fixed design model and derive from this result the uniform strong consistency of $F_h(t|x)$ under weaker conditions on the bandwidth sequence than in Dabrowska (1989). In particular, the next result is valid for the bandwidth $h_n = Cn^{-1/5}$ $(C > 0)$ of optimal rate (see Remark 2.1 later in this chapter).

Theorem 2.1
Assume $(C1)$, $(C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C3)$ *and* $(C5)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $1 - H(T|x) > \delta > 0$, $h_n \to 0$, $nh_n \to \infty$.

(a) *For* $\varepsilon > 0$, n *sufficiently large and*

$$\varepsilon \ge \frac{12}{\delta^2} \max\left(\frac{\sqrt{6}\|K\|_2}{(nh_n)^{1/2}}, 2(\|\dot{H}\| \vee \|\dot{H}^u\|)\overline{\Delta}_n + 2\mu_2^K(\|\ddot{H}\| \vee \|\ddot{H}^u\|)h_n^2\right),$$

we have

$$P\left(\sup_{0\le t\le T} |F_h(t|x) - F(t|x)| > \varepsilon\right) \le C_1\delta^2 nh_n\varepsilon \exp(-C_2\delta^4 nh_n\varepsilon^2),$$

where $C_1 = \dfrac{e^2}{\|K\|_2^2}, C_2 = \dfrac{1}{432\|K\|_2^2}.$

(b) *If* $\dfrac{nh_n^5}{\log n} = O(1)$, *then, as* $n \to \infty$,

$$\sup_{0\le t\le T} |F_h(t|x) - F(t|x)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \qquad \text{a.s.}$$

Proof. (a) Write, using integration by parts,

$$F_h(t|x) - F(t|x) = (1 - F(t|x)) \int_0^t \frac{1 - F_h(s-|x)}{1 - F(s|x)} d(\Lambda_h(s|x) - \Lambda(s|x))$$

(see Shorack and Wellner (1986), p. 305) (where $\Lambda_h(t|x)$ is defined as in (1.23)), from which it follows that

$$\sup_{0\le t\le T} |F_h(t|x) - F(t|x)| \le 3 \sup_{0\le t\le T} |\Lambda_h(t|x) - \Lambda(t|x)|.$$

Straightforward calculations now lead to

$$P\left(\sup_{0\le t\le T} |F_h(t|x) - F(t|x)| > \varepsilon\right) \le 2P\left(\sup_{0\le t\le T} |H_h(t|x) - H(t|x)| > \frac{\varepsilon\delta^2}{12}\right) + P\left(\sup_{0\le t\le T} |H_h^u(t|x) - H^u(t|x)| > \frac{\varepsilon\delta^2}{12}\right). \tag{2.1}$$

The proof is finished after applying Lemma 2.8(b) in the appendix to the right hand side of the above inequality.
(b) Apply part (a) with the choice $\varepsilon = \varepsilon_n = c(nh_n)^{-1/2}(\log n)^{1/2}$ for some appropriately chosen constant $c > 0$. Then, use the Borel-Cantelli lemma.

A somewhat stronger exponential bound result than the one proved in the previous result (leading to the same rate of convergence) can be found in Theorem 1 in Van Keilegom and Veraverbeke (1996). That sharper result requires however that $H(t|x)$ and $H^u(t|x)$ satisfy condition $(C4)$.

The next result deals with the behavior of the modulus of continuity of the Beran estimator $F_h(t|x)$ and will play a key role in deriving an a.s. representation for the quantile estimator $F_h^{-1}(p|x)$, which will be studied in Chapter 4. The result is obtained via an exponential probability bound result for the oscillation of $F_h(t|x) - F(t|x)$. From this result, together with the Borel-Cantelli lemma, a rate for the uniform strong convergence will follow. In the i.i.d. case without censoring, this kind of result was obtained by Stute (1982). A similar result for the Stone estimator will be shown in Lemma 2.10. For the usual Kaplan-Meier estimator, Schäfer (1986) established an exponential bound result and Cheng (1984) and Lo and Singh (1986) proved in this context a result analogous to part (b) of the next theorem for a specific choice of the sequence a_n. Our approach is similar to the one of Schäfer (1986). Using a different proof, Dabrowska (1992b) obtained part (b) of the result below for a specific sequence a_n in a random design regression model. A slightly sharper exponential bound than the one in the next result (leading to the same rate of convergence) can be found in Theorem 3 in Van Keilegom and Veraverbeke (1996). That somewhat stronger result was obtained by making use of the consistency of the Beran estimator established in Theorem 1 of the same paper (for which condition $(C4)$ on $H(t|x)$ and $H^u(t|x)$ was however required), rather than by using the consistency result obtained in the previous result.

Theorem 2.2
Assume $(C1),(C2)$, $H(t|x)$ and $H^u(t|x)$ satisfy $(C3),(C6)$ and $(C7)$ in $[0,T]$ with $T < \tau_{H(\cdot|x)}$, $1-H(T|x) > \delta > 0$, $h_n \to 0$, $nh_n \to \infty$.

(a) *Let $\{a_n\}$ be a sequence of positive constants, tending to 0 as $n \to \infty$ and satisfying $na_n \to \infty$. Then, for n sufficiently large and*

$$\varepsilon \geq \frac{8}{\delta} \max\Big\{ \|\dot{H}^u\|\overline{\Delta}_n + L_0\|\dot{H}^{u'}\|a_n h_n + \|H^{u''}\|a_n^2 \, , \\ \frac{12\|F'\|a_n}{\delta^2} \max\left(\frac{\sqrt{6}\|K\|_2}{(nh_n)^{1/2}}, 4(\|\dot{H}\| \vee \|\dot{H}^u\|)L_0 h_n \right) \Big\},$$

we have

$$P\left(\sup_{0\le s,t\le T; |t-s|\le a_n} |F_h(t|x) - F_h(s|x) - F(t|x) + F(s|x)| > \varepsilon \right)$$

$$\leq C_1 \frac{T\|H^{u'}\|^2 a_n}{\delta^2\varepsilon^2} \exp\left(-C_2 \frac{\delta^2 nh_n \varepsilon^2}{\|H^{u'}\|a_n + \varepsilon} \right) + C_3 \frac{\delta^3 nh_n \varepsilon}{\|F'\|a_n} \exp\left(-C_4 \frac{\delta^6 nh_n \varepsilon^2}{\|F'\|^2 a_n^2} \right)$$

$$+ C_5 \delta^3 nh_n \exp(-C_6 \delta^6 nh_n) + C_7 \frac{T\|H^{u'}\|^2}{\|F'\|^2\delta^2 a_n} \exp\left(-C_8 \frac{\|F'\|^2 \delta^2 nh_n a_n}{\|H^{u'}\| + \|F'\|} \right),$$

where $C_1, \ldots, C_8 > 0$ are absolute constants.

(b) *If $\dfrac{nh_n^5}{\log n} = O(1)$ and $\{a_n\}$ is a sequence of positive constants, tending to 0 as $n \to \infty$ and satisfying $c_1 \left(\dfrac{\log n}{nh_n}\right)^q \leq a_n \leq c_2 \left(\dfrac{\log n}{nh_n}\right)^{1/2}$ for some $c_1, c_2 > 0$*

and $1/2 \le q < 1$, *then, as* $n \to \infty$,

$$\sup_{0 \le s,t \le T; |t-s| \le a_n} |F_h(t|x) - F_h(s|x) - F(t|x) + F(s|x)|$$
$$= O((nh_n)^{-1/2} a_n^{1/2} (\log n)^{1/2}) \qquad \text{a.s.}$$

Proof. (a) Write

$$\begin{aligned} &F_h(t|x) - F_h(s|x) - F(t|x) + F(s|x) \\ &= \left(\tilde{F}_h(t|x) - \tilde{F}_h(s|x) - F(t|x) + F(s|x)\right) \\ &\quad + \left(F_h(t|x) - F_h(s|x) - \tilde{F}_h(t|x) + \tilde{F}_h(s|x)\right), \end{aligned} \tag{2.2}$$

where

$$\tilde{F}_h(t|x) = \int_0^t (1 - G(y|x))^{-1} dH_h^u(y|x).$$

The first term on the right hand side of (2.2) equals

$$\begin{aligned} &\sum_{i=1}^n w_{ni}(x;h_n) \left[\frac{I(s \le T_i \le t, \Delta_i = 1)}{1 - G(T_i|x)} - E\left\{\frac{I(s \le T_i \le t, \Delta_i = 1)}{1 - G(T_i|x)}\right\}\right] \\ &+ \sum_{i=1}^n w_{ni}(x;h_n) \left[E\left\{\frac{I(s \le T_i \le t, \Delta_i = 1)}{1 - G(T_i|x)}\right\} - (F(t|x) - F(s|x))\right] \\ &= T_1 + T_2. \end{aligned}$$

Since $F(t|x) = \int_0^t (1 - G(y|x))^{-1} dH^u(y|x)$, it follows that

$$T_2 = \sum_{i=1}^n w_{ni}(x;h_n) \int_s^t \frac{d(H^u(y|x_i) - H^u(y|x))}{1 - G(y|x)}$$

and hence, using integration by parts, it is easy to show that

$$|T_2| \le \frac{1}{1-G(T|x)} \sup_{0\le s,t\le T;|t-s|\le a_n} |EH_h^u(t|x) - EH_h^u(s|x) - H^u(t|x) + H^u(s|x)|$$

$$+ \frac{2}{(1-G(T|x))^2} \sup_{0\le s,t\le T;|t-s|\le a_n} |G(t|x) - G(s|x)| \sup_{0\le t\le T} |EH_h^u(t|x) - H^u(t|x)|$$

$$\le \frac{1}{1-G(T|x)} (2\|\dot{H}^u\|\overline{\Delta}_n + L_0\|\dot{H}^{u'}\|a_n h_n + \|H^{u''}\|a_n^2) + o(a_n h_n),$$

where the last inequality follows from Lemma 2.6(a) and the proof of Lemma 2.10. Hence,

$$|T_2| \le \frac{2}{\delta}(\|\dot{H}^u\|\overline{\Delta}_n + L_0\|\dot{H}^{u'}\|a_n h_n + \|H^{u''}\|a_n^2) \le \frac{\varepsilon}{4}.$$

For T_1 we have,

$$\mathrm{Var}(T_1) = \sum_{i=1}^n w_{ni}^2(x;h_n) \left[\int_s^t \frac{dH^u(y|x_i)}{(1-G(y|x))^2} - \left(\int_s^t \frac{dH^u(y|x_i)}{1-G(y|x)} \right)^2 \right]$$

$$\le \sum_{i=1}^n w_{ni}^2(x;h_n) \int_s^t \frac{dH^u(y|x_i)}{(1-G(y|x))^2}$$

$$\le \frac{\|H^{u'}\|a_n}{(1-G(T|x))^2} \sum_{i=1}^n w_{ni}^2(x;h_n) \le \frac{\|H^{u'}\|\|K\|_\infty}{\delta^2} \frac{a_n\overline{\Delta}_n}{h_n}.$$

Now, Bernstein's inequality implies that

$$P\left(|\tilde{F}_h(t|x) - \tilde{F}_h(s|x) - F(t|x) + F(s|x)| > \frac{\varepsilon}{2} \right) \le 2\exp\left(-C\frac{\delta^2 n h_n \varepsilon^2}{\|H^{u'}\|a_n + \varepsilon} \right).$$

Partitioning the interval $[0,T]$ into $K = [c/\varepsilon]$ subintervals $[t_i, t_{i+1}]$ of length T/K ($c \ge \varepsilon$ to be chosen later), it follows that

$$P\left(\sup_{0\le s,t\le T;|t-s|\le a_n} \left| \tilde{F}_h(t|x) - \tilde{F}_h(s|x) - F(t|x) + F(s|x) \right| > \frac{\varepsilon}{2} \right) \tag{2.3}$$

$$\le P\left(\max_{0\le j\le K} \sup_{0\le s\le T;|t_j-s|\le a_n} \left| \sum_{i=1}^n w_{ni}(x;h_n) \left[\frac{I(s\le T_i\le t_j, \Delta_i=1)}{1-G(T_i|x)} \right.\right.\right.$$

$$\left.\left.\left. -E\left\{ \frac{I(s\le T_i\le t_j, \Delta_i=1)}{1-G(T_i|x)} \right\} \right] \right| > \frac{\varepsilon}{4} - \max_{1\le i\le n, 1\le j\le K} \left| \int_{t_{j-1}}^{t_j} \frac{dH^u(y|x_i)}{1-G(y|x)} \right| \right).$$

By choosing $c = (32T\|H^{u'}\|)/(1 - G(T|x))$, we have that

$$\left| \int\limits_{t_{j-1}}^{t_j} (1 - G(y|x))^{-1} dH^u(y|x_i) \right| \le \frac{\varepsilon}{16} < \frac{\varepsilon}{8}$$

for every i, j. A similar procedure replaces $\sup_{0 \le s \le T}$ by a maximum and hence the right hand side of (2.3) is smaller than

$$P\left(\max_{0 \le j,k \le K; |t_j - t_k| \le a_n} \left| \sum_{i=1}^n w_{ni}(x; h_n) \left[\frac{I(t_k \le T_i \le t_j, \Delta_i = 1)}{1 - G(T_i|x)} \right.\right.\right.$$

$$\left.\left.\left. - E\left\{ \frac{I(t_k \le T_i \le t_j, \Delta_i = 1}{1 - G(T_i|x)} \right\} \right] \right| > \frac{\varepsilon}{16} \right)$$

$$\le C_1 \frac{T\|H^{u'}\|^2 a_n}{\delta^2 \varepsilon^2} \exp\left(-C_2 \frac{\delta^2 n h_n \varepsilon^2}{\|H^{u'}\| a_n + \varepsilon} \right).$$

Since $F_h(t|x) = \int\limits_0^t (1 - G_h(y - |x))^{-1} dH_h^u(y|x)$, the second term on the right hand side of (2.2) can be written as

$$\int\limits_s^t \frac{dH_h^u(y|x)}{1 - G_h(y - |x)} - \int\limits_s^t \frac{dH_h^u(y|x)}{1 - G(y|x)} = \int\limits_s^t \frac{G_h(y - |x) - G(y|x)}{1 - G_h(y - |x)} d\tilde{F}_h(y|x).$$

Hence,

$$\sup_{0 \le s,t \le T; |t-s| \le a_n} |F_h(t|x) - F_h(s|x) - \tilde{F}_h(t|x) + \tilde{F}_h(s|x)|$$

$$\le \frac{1}{1 - G_h(T|x)} \sup_{0 \le t \le T} |G_h(t|x) - G(t|x)| \sup_{0 \le s,t \le T; |t-s| \le a_n} |\tilde{F}_h(t|x) - \tilde{F}_h(s|x)|.$$

Define

$$A_n = \left\{ 1 - G_h(t|x) \ge \frac{\delta}{2} \text{ for all } 0 \le t \le T \right\}$$

$$B_n = \{ |\tilde{F}_h(t|x) - \tilde{F}_h(s|x)| \le 2\|F'\| a_n \text{ for all } 0 \le s, t \le T \text{ such that } |t - s| \le a_n \}.$$

Because

$$P(A_n^c) \le P\left(\sup_{0 \le t \le T} |G_h(t|x) - G(t|x)| > \frac{\delta}{2} \right)$$

(which follows from the fact that $1 - G(t|x) > 1 - H(t|x) > \delta > 0$ for all $0 \leq t \leq T$) and because

$$P(B_n^c) \leq P\left(\sup_{0 \leq s,t \leq T; |t-s| \leq a_n} |\tilde{F}_h(t|x) - \tilde{F}_h(s|x) - F(t|x) + F(s|x)| > \|F'\| a_n \right),$$

we can write

$$P\left(\sup_{0 \leq s,t \leq T; |t-s| \leq a_n} |F_h(t|x) - F_h(s|x) - \tilde{F}_h(t|x) + \tilde{F}_h(s|x)| > \frac{\varepsilon}{2} \right)$$

$$\leq P\left(\sup_{0 \leq t \leq T} |G_h(t|x) - G(t|x)| > \frac{\delta}{8\|F'\|} \frac{\varepsilon}{a_n} \right) + P\left(\sup_{0 \leq t \leq T} |G_h(t|x) - G(t|x)| > \frac{\delta}{2} \right)$$

$$+ P\left(\sup_{0 \leq s,t \leq T; |t-s| \leq a_n} |\tilde{F}_h(t|x) - \tilde{F}_h(s|x) - F(t|x) + F(s|x)| > \|F'\| a_n \right).$$

Following the proof of Theorem 2.1(a), we have that

$$P\left(\sup_{0 \leq t \leq T} |G_h(t|x) - G(t|x)| > \varepsilon \right) \leq K_1 \delta^2 n h_n \varepsilon \exp(-K_2 \delta^4 n h_n \varepsilon^2),$$

provided $\varepsilon \geq 12\delta^{-2} \max\{\sqrt{6}\|K\|_2 (nh_n)^{-1/2}, 4(\|\dot{H}\| \vee \|\dot{H}^u\|) L_0 h_n\}$ and $H(\cdot|x)$ and $H^u(\cdot|x)$ satisfy $(C3)$. Hence, the result follows using an analogous reasoning as for the first term on the right hand side of (2.2).
(b) This follows from the Borel-Cantelli lemma.

2.3 Asymptotic representation and weak convergence

The estimator $F_h(\cdot|x)$ in (1.17) has a complicated structure, as it is a product of dependent factors. As Lo and Singh (1986) did for the ordinary Kaplan-Meier estimator, we will now prove an a.s. asymptotic representation for $F_h(\cdot|x)$, which decomposes the estimator in a weighted sum of independent terms and a remainder term of order $O((nh_n)^{-3/4}(\log n)^{3/4})$ a.s. Such a representation of the estimator is very useful to show e.g. central limit results.

Theorem 2.3
Assume $(C1), (C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n \to 0$, $\dfrac{\log n}{nh_n} \to 0$, $\dfrac{nh_n^5}{\log n} = O(1)$. *Then, for* $t < \tau_{H(\cdot|x)}$ *:*

$$F_h(t|x) - F(t|x) = \sum_{i=1}^{n} w_{ni}(x;h_n)\xi(T_i,\Delta_i,t|x) + r_n(x,t),$$

where

$$\begin{aligned}&\xi(T_i,\Delta_i,t|x)\\&= (1-F(t|x))\left\{\int_0^t \frac{I(T_i \le s) - H(s|x)}{(1-H(s|x))^2} dH^u(s|x)\right.\\&\quad\left. + \frac{I(T_i \le t, \Delta_i = 1) - H^u(t|x)}{1-H(t|x)} - \int_0^t \frac{I(T_i \le s, \Delta_i = 1) - H^u(s|x)}{(1-H(s|x))^2} dH(s|x)\right\}\end{aligned}$$

and where

$$\sup_{0\le t\le T} |r_n(x,t)| = O((nh_n)^{-3/4}(\log n)^{3/4}) \qquad a.s.$$

as $n \to \infty$.

Proof. Because of continuity, $1-F(t|x) = \exp(-\Lambda(t|x))$. Introducing for $t < \tau_{H_h(\cdot|x)}$,

$$\tilde{\Lambda}_h(t|x) = \int_0^t \frac{dH_h^u(s|x)}{1-H_h(s|x)}, \tag{2.4}$$

we have the following identity :

$$F_h(t|x) - F(t|x) = \left[e^{-\Lambda(t|x)} - e^{-\tilde{\Lambda}_h(t|x)}\right] - \left[1 - F_h(t|x) - e^{-\tilde{\Lambda}_h(t|x)}\right].$$

By a two term Taylor expansion of the first and a one term Taylor expansion of the second term, we obtain

$$F_h(t|x) - F(t|x) = (1-F(t|x))(\tilde{\Lambda}_h(t|x) - \Lambda(t|x)) + R_{n1}(t) + R_{n2}(t), \tag{2.5}$$

where

$$R_{n1}(t) = -\frac{1}{2}\exp(-\Lambda_h^{\circ}(t|x))\left[\tilde{\Lambda}_h(t|x) - \Lambda(t|x)\right]^2$$

$$R_{n2}(t) = \exp(-\Lambda_h^{\circ\circ}(t|x))\left[-\log(1-F_h(t|x)) - \tilde{\Lambda}_h(t|x)\right],$$

with $\Lambda_h^\circ(t|x)$ between $\tilde{\Lambda}_h(t|x)$ and $\Lambda(t|x)$, and $\Lambda_h^{\circ\circ}(t|x)$ between $-\log(1-F_h(t|x))$ and $\tilde{\Lambda}_h(t|x)$. Furthermore, for $t < \tau_{H_h(\cdot|x)}$:

$$\begin{aligned}
&\tilde{\Lambda}_h(t|x) - \Lambda(t|x) \\
&= \int_0^t \frac{dH_h^u(s|x)}{1-H_h(s|x)} - \int_0^t \frac{dH^u(s|x)}{1-H(s|x)} \\
&= \int_0^t \left[\frac{1}{1-H_h(s|x)} - \frac{1}{1-H(s|x)}\right] dH^u(s|x) \\
&\quad + \int_0^t \frac{1}{1-H(s|x)} d\left(H_h^u(s|x) - H^u(s|x)\right) \\
&\quad + \int_0^t \left[\frac{1}{1-H_h(s|x)} - \frac{1}{1-H(s|x)}\right] d\left(H_h^u(s|x) - H^u(s|x)\right).
\end{aligned}$$

Writing for the integrand in the first term

$$\frac{H_h(s|x) - H(s|x)}{(1-H_h(s|x))(1-H(s|x))} = \frac{H_h(s|x)-H(s|x)}{(1-H(s|x))^2} + \frac{(H_h(s|x)-H(s|x))^2}{(1-H(s|x))^2(1-H_h(s|x))}$$

and integrating by parts in the second term, we arrive at

$$\begin{aligned}
&\tilde{\Lambda}_h(t|x) - \Lambda(t|x) \\
&= \int_0^t \frac{H_h(s|x)-H(s|x)}{(1-H(s|x))^2} dH^u(s|x) + \frac{H_h^u(t|x) - H^u(t|x)}{1-H(t|x)} \\
&\quad - \int_0^t \frac{H_h^u(s|x)-H^u(s|x)}{(1-H(s|x))^2} dH(s|x) + R_{n3}(t) + R_{n4}(t),
\end{aligned} \tag{2.6}$$

where

$$\begin{aligned}
R_{n3}(t) &= \int_0^t \frac{(H_h(s|x)-H(s|x))^2}{(1-H(s|x))^2(1-H_h(s|x))} dH^u(s|x) \\
R_{n4}(t) &= \int_0^t \left[\frac{1}{1-H_h(s|x)} - \frac{1}{1-H(s|x)}\right] d\left[H_h^u(s|x) - H^u(s|x)\right].
\end{aligned} \tag{2.7}$$

Because $H(T|x) < 1$ and $H_h(T|x) \to H(T|x)$ a.s. (by Lemma 2.7), we may suppose that $T < \tau_{H_h(\cdot|x)}$. For $R_{n3}(t)$ we have

$$\sup_{0\le t\le T} |R_{n3}(t)| \le \left(\sup_{0\le t\le T} |H_h(t|x) - H(t|x)| \right)^2 \frac{1}{(1 - H_h(T|x))(1 - H(T|x))^2}$$

$$= O((nh_n)^{-1} \log n) \qquad \text{a.s.}$$

by application of Lemma 2.7 and Lemma 2.9. By Lemma 2.1 below,

$$\sup_{0\le t\le T} |R_{n4}(t)| = O((nh_n)^{-3/4} (\log n)^{3/4}) \qquad \text{a.s.}$$

Also from (2.6), Lemma 2.9 and the bounds for $R_{n3}(t)$ and $R_{n4}(t)$:

$$\sup_{0\le t\le T} |\tilde{\Lambda}_h(t|x) - \Lambda(t|x)| = O((nh_n)^{-1/2} (\log n)^{1/2}) \qquad \text{a.s.} \tag{2.8}$$

This gives that

$$\sup_{0\le t\le T} |R_{n1}(t)| = O((nh_n)^{-1} \log n) \qquad \text{a.s.}$$

In Lemma 2.2 below, we will show

$$\sup_{0\le t\le T} |R_{n2}(t)| = O((nh_n)^{-1}) \qquad \text{a.s.}$$

This, together with (2.5) and (2.6), shows that the theorem is proved.

We now prove the two lemmas used above.

Lemma 2.1

Under the conditions of Theorem 2.3, as $n \to \infty$,

$$\sup_{0\le t\le T} |R_{n4}(t)| = O((nh_n)^{-3/4} (\log n)^{3/4}) \qquad \text{a.s.,} \tag{2.9}$$

where $R_{n4}(t)$ is as in (2.7).

Proof. Partitioning the interval $[0,T]$ into $k_n = O((nh_n)^{1/2}(\log n)^{-1/2})$ subintervals $[t_i, t_{i+1}]$ of length $O((nh_n)^{-1/2}(\log n)^{1/2})$, we have, as in the proof of Lemma 2 of Lo and Singh (1986), that the left hand side in (2.9) is bounded above by

$$\begin{aligned} &2 \max_{1\le i\le k_n} \sup_{t_i\le s\le t_{i+1}} \left| \frac{1}{1-H_h(s|x)} - \frac{1}{1-H_h(t_i|x)} - \frac{1}{1-H(s|x)} + \frac{1}{1-H(t_i|x)} \right| \\ &+ k_n \sup_{0\le t\le T} |H_h(t|x) - H(t|x)|(1-H_h(T|x))^{-1}(1-H(T|x))^{-1} \\ &\times \max_{1\le i\le k_n} |H_h^u(t_{i+1}|x) - H_h^u(t_i|x) - H^u(t_{i+1}|x) + H^u(t_i|x)| \,. \end{aligned} \tag{2.10}$$

To estimate the first term in (2.10) we further subdivide each $[t_i, t_{i+1}]$ into $a_n = O((nh_n)^{1/4}(\log n)^{-1/4})$ subintervals $[t_{ij}, t_{i,j+1}]$ of length $O((nh_n)^{-3/4}(\log n)^{3/4})$. Using Lemma 2.9, we obtain, as in Lo and Singh (1986) that the first term in (2.10) is a.s. bounded by

$$C \max_{1\le i\le k_n} \max_{1\le j\le a_n} |H_h(t_{ij}|x) - H_h(t_i|x) - H(t_{ij}|x) + H(t_i|x)| + O((nh_n)^{-3/4}(\log n)^{3/4})$$

for some constant $C > 0$. Applying Corollary 2.2 to the functions $H_h(\cdot|x)$ and $H(\cdot|x)$ gives that this term is $O((nh_n)^{-3/4}(\log n)^{3/4})$ a.s. The second term in (2.10) is treated similarly and leads to the same order.

Lemma 2.2
Assume $(C1), (C2)$, $H(t|x)$ *satisfies* $(C3)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n \to 0$, $\dfrac{\log n}{nh_n} \to 0$. *Then, as* $n \to \infty$,

$$\sup_{0\le t\le T} \left| -\log(1-F_h(t|x)) - \tilde{\Lambda}_h(t|x) \right| = O((nh_n)^{-1}) \qquad \text{a.s.},$$

where $\tilde{\Lambda}_h(t|x)$ *is defined as in (2.4).*

Proof. Because $H(T|x) < 1$ and $H_h(T|x) \to H(T|x)$ a.s., we may suppose that $T < \tau_{H_h(\cdot|x)}$. If $t \le T$, then

$$\begin{aligned} \tilde{\Lambda}_h(t|x) &= \int_0^t \frac{1-G_h(s-|x)}{1-G_h(s|x)} \frac{dF_h(s|x)}{1-F_h(s|x)} \\ &= \int_0^t \frac{G_h(s|x) - G_h(s-|x)}{1-G_h(s|x)} \frac{dF_h(s|x)}{1-F_h(s|x)} - \log(1-F_h(t|x)). \end{aligned}$$

Since $\sup_{0\le t\le T} |G_h(t|x) - G_h(t-|x)| = O((nh_n)^{-1})$ a.s., the result follows.

Asymptotic representations similar to the one in the previous result were studied by Akritas (1994) and González Manteiga and Cadarso Suárez (1994). Akritas (1994) established a representation for $F_h(t|x)$, using nearest neighbors weights instead of the weights in (1.11). González Manteiga and Cadarso Suárez (1994) obtained a representation for (a slightly different version of) $F_h(t|x)$ (but using the same weights as in (1.11)). The remainder term in their representation is $O((nh_n)^{-3/4}(\log n)^{3/4} + h_n^2)$ a.s. and the conditions $\log n/(nh_n) \to 0$ and $nh_n^3 \to \infty$ are needed (where the last condition, however, could be weakened to $nh_n^2 \to \infty$). For situations where the remainder term should be $o((nh_n)^{-1/2})$ a.s., the extra conditions $(\log n)^3/(nh_n) \to 0$ and $nh_n^5 \to 0$ are therefore required. This last condition, however, is not satisfied for the optimal bandwidth sequence $h_n = Cn^{-1/5}$ (which minimizes the asymptotic mean squared error (see Remark 2.1)). The remainder term in the representation of Theorem 2.3 above, does not contain a term of order $O(h_n^2)$ a.s. Hence, for situations where the remainder term should be $o((nh_n)^{-1/2})$ a.s., we do not have the (restrictive) condition $nh_n^5 \to 0$.

The major consequence of Theorem 2.3 is the weak convergence of the Beran estimator $F_h(t|x)$, which will be established in the next theorem. We start with three preliminary lemmas in which the bias and variance of the leading term of the above asymptotic representation will be obtained.

In the sequel, the notation $W(\cdot|x)$ will be used for a Gaussian process with mean function 0 and covariance function $\Gamma(s,t|x)$ given by

$$\Gamma(s,t|x) = \|K\|_2^2(1-F(s|x))(1-F(t|x)) \int_0^{s\wedge t} \frac{dH^u(y|x)}{(1-H(y|x))^2} \tag{2.11}$$

and we will write $\tilde{W}(\cdot|x)$ for a Gaussian process with mean function $b(t|x)$ and covariance function $\Gamma(s,t|x)$, where

$$b(t|x) = \frac{1}{2}\mu_2^K(1-F(t|x)) \int_0^t \left\{ \frac{\ddot{H}(s|x)\, dH^u(s|x)}{(1-H(s|x))^2} + \frac{d\ddot{H}^u(s|x)}{1-H(s|x)} \right\} C^{5/2} \tag{2.12}$$

and $C > 0$.

Lemma 2.3
Assume $(C1),(C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C3)$ *and* $(C5)$ *in* $[0,T]$, $h_n \to 0$. *Then, as* $n \to \infty$,

$$\sup_{0\le t\le T}\left|\sum_{i=1}^n w_{ni}(x;h_n)E\xi(T_i,\Delta_i,t|x)\right.$$
$$\left. -\frac{1}{2}\mu_2^K(1-F(t|x))\int_0^t\left\{\frac{\ddot{H}(s|x)\,dH^u(s|x)}{(1-H(s|x))^2}+\frac{d\ddot{H}^u(s|x)}{1-H(s|x)}\right\}h_n^2\right| = o(h_n^2)+O(n^{-1}).$$

Proof. For fixed $t \le T$,

$$\sum_{i=1}^n w_{ni}(x;h_n)E\xi(T_i,\Delta_i,t|x)$$
$$=(1-F(t|x))\left\{\int_0^t\frac{EH_h(s|x)-H(s|x)}{(1-H(s|x))^2}\,dH^u(s|x)+\int_0^t\frac{d(EH_h^u(s|x)-H^u(s|x))}{1-H(s|x)}\right\}.$$

Now apply Lemma 2.6(b).

Lemma 2.4
Assume $(C1),(C2)$, $h_n \to 0$, $nh_n \to \infty$. *If* $\gamma : [0,1]\times[0,T]^2 \to I\!R$ *is a function for which* $\gamma(\cdot,s,t) : [0,1] \to I\!R$ $(s,t \in [0,T])$ *is Lipschitz with Lipschitz constant uniformly bounded on* $[0,T]^2$ *and for which* $\gamma(x,\cdot,\cdot) : [0,T]^2 \to I\!R$ *is bounded, then, as* $n \to \infty$,

$$\sup_{0\le s,t\le T}\left|\sum_{i=1}^n w_{ni}^2(x;h_n)\gamma(s,t|x_i)-\frac{\|K\|_2^2}{nh_n}\gamma(s,t|x)\right| = o((nh_n)^{-1}). \qquad (2.13)$$

Proof. The expression inside the absolute value on the left hand side of (2.13) equals

$$\frac{1}{h_n}\sum_{i=1}^n w_{ni}(x;h_n)\int_{x_{i-1}}^{x_i}(\gamma(s,t|x_i)-\gamma(s,t|z))K\left(\frac{x-z}{h_n}\right)dz$$

$$+\frac{1}{h_n}\sum_{i=1}^n w_{ni}(x;h_n)\int_{x_{i-1}}^{x_i}(\gamma(s,t|z)-\gamma(s,t|x))K\left(\frac{x-z}{h_n}\right)dz$$

$$+\frac{1}{h_n^2}\sum_{i=1}^{n}\left\{\left(\int_{x_{i-1}}^{x_i} K\left(\frac{x-z}{h_n}\right)dz\right)^2-(x_i-x_{i-1})\int_{x_{i-1}}^{x_i} K^2\left(\frac{x-z}{h_n}\right)dz\right\}\gamma(s,t|x)$$

$$+\left\{\sum_{i=1}^{n}\frac{x_i-x_{i-1}}{h_n^2}\int_{x_{i-1}}^{x_i} K^2\left(\frac{x-z}{h_n}\right)dz-\frac{1}{nh_n}\|K\|_2^2\right\}\gamma(s,t|x).$$

The first term is easily seen to be $O(n^{-2}h_n^{-1})$, while the second term is $O(n^{-1})$. On the fourth term we can apply Lemma A1 in Aerts and Geertsema (1990), which yields the order $o((nh_n)^{-1})$. Finally, the third term equals

$$\left\{\frac{1}{h_n^2}\sum_{i=1}^{n}\int_{x_{i-1}}^{x_i} K\left(\frac{x-z}{h_n}\right)dz\int_{x_{i-1}}^{x_i}\left\{K\left(\frac{x-z}{h_n}\right)-K\left(\frac{x-x_i}{h_n}\right)\right\}dz\right.$$

$$\left.+\frac{1}{h_n^2}\sum_{i=1}^{n}(x_i-x_{i-1})\int_{x_{i-1}}^{x_i} K\left(\frac{x-z}{h_n}\right)\left\{K\left(\frac{x-x_i}{h_n}\right)-K\left(\frac{x-z}{h_n}\right)\right\}dz\right\}\gamma(s,t|x)$$

and this is $O((nh_n)^{-2})$ uniformly in s and t.

This enables us to obtain the following result for the covariance.

Lemma 2.5
Assume $(C1),(C2)$, $H(t|x)$ and $H^u(t|x)$ satisfy $(C3)$ in $[0,T]$ with $T<\tau_{H(\cdot|x)}$, $h_n\to 0$, $nh_n\to\infty$. Then, as $n\to\infty$,

$$\sup_{0\le s,t\le T}\left|\sum_{i=1}^{n} w_{ni}^2(x;h_n)Cov(\xi(T_i,\Delta_i,s|x),\xi(T_i,\Delta_i,t|x))-\frac{1}{nh_n}\Gamma(s,t|x)\right|=o((nh_n)^{-1}).$$

Proof. Using integration by parts, we can write

$$\xi(T_i,\Delta_i,t|x)=(1-F(t|x))\left\{-\int_0^{T_i\wedge t}\frac{dH^u(y|x)}{(1-H(y|x))^2}+\frac{I(T_i\le t,\Delta_i=1)}{1-H(T_i|x)}\right\}.$$

Hence, some straightforward calculations show that

$$
\begin{aligned}
&\operatorname{Cov}(\xi(T_i,\Delta_i,s|x),\xi(T_i,\Delta_i,t|x))\\
&= (1-F(s|x))(1-F(t|x))\left\{\int_0^s \frac{H(y|x)-H(y|x_i)}{(1-H(y|x))^2}\int_0^y \frac{dH^u(z|x)}{(1-H(z|x))^2}dH^u(y|x)\right.\\
&\quad+\int_0^s \frac{1}{1-H(y|x)}\int_0^y \frac{dH^u(z|x)}{(1-H(z|x))^2}d(H^u(y|x)-H^u(y|x_i))\\
&\quad+\int_0^s \frac{1}{(1-H(y|x))^2}\int_y^t \frac{H(z|x)-H(z|x_i)}{(1-H(z|x))^2}dH^u(z|x)dH^u(y|x)\\
&\quad+\int_0^s \frac{1}{(1-H(y|x))^2}\int_y^t \frac{d(H^u(z|x)-H^u(z|x_i))}{1-H(z|x)}dH^u(y|x)+\int_0^{s\wedge t}\frac{dH^u(y|x_i)}{(1-H(y|x))^2}\\
&\quad-\left[-\int_0^s \frac{H(y|x)-H(y|x_i)}{(1-H(y|x))^2}dH^u(y|x)+\int_0^s \frac{d(H^u(y|x_i)-H^u(y|x))}{1-H(y|x)}\right]\\
&\quad\left.\cdot\left[-\int_0^t \frac{H(y|x)-H(y|x_i)}{(1-H(y|x))^2}dH^u(y|x)+\int_0^t \frac{d(H^u(y|x_i)-H^u(y|x))}{1-H(y|x)}\right]\right\}
\end{aligned}
\tag{2.14}
$$

from which the result follows via Lemma 2.4.

Remark 2.1
Lemmas 2.3 and 2.5 allow us to obtain the optimal order of magnitude of the smoothing parameter h_n as a power of n. This is done by minimizing the asymptotic mean squared error (AMSE) :

$$
\begin{aligned}
\operatorname{AMSE}(h_n) &= \operatorname{AsVar}(F_h(t|x)) + (\operatorname{AsBias}(F_h(t|x)))^2\\
&= (nh_n)^{-1}\Gamma(t,t|x) + h_n^4\overline{b}^2(t|x),
\end{aligned}
$$

where

$$
\overline{b}(t|x) = \frac{1}{2}\mu_2^K(1-F(t|x))\int_0^t \left\{\frac{\ddot{H}(s|x)\,dH^u(s|x)}{(1-H(s|x))^2}+\frac{d\ddot{H}^u(s|x)}{1-H(s|x)}\right\}. \tag{2.15}
$$

Hence, the optimal choice is

$$h_{n,\text{opt}} = \left(\frac{\Gamma(t,t|x)}{4\bar{b}^2(t|x)} \right)^{1/5} n^{-1/5}. \tag{2.16}$$

This formula, however, is in practice not directly applicable, since it contains the unknown expressions $H(t|x), H^u(t|x), \ddot{H}(t|x)$ and $\ddot{H}^u(t|x)$ $(1 - F(t|x)$ cancels in numerator and denominator). These quantities can be estimated, respectively, by $H_h(t|x), H_h^u(t|x)$ and their second derivatives with respect to x. However, since both $H_h(t|x)$ and $H_h^u(t|x)$ depend on a smoothing parameter, further bandwidth selection is required. Moreover, the optimal bandwidth for $H_h(t|x)$ will be different from the one for $H_h^u(t|x)$, since the latter uses only part of the observations that is used to estimate the former. Dabrowska (1992a) proposed to use a least squares cross validation procedure to select these two smoothing parameters. However, cross validation procedures are known to be computer time intensive and often give poor estimates of the optimal bandwidth. Further research on this topic is therefore required. One of the possibilities is to estimate the optimal bandwidth for $H_h(t|x)$ and $H_h^u(t|x)$ using ideas in the same spirit as the so-called plug-in rule in e.g. density estimation (see e.g. Park and Marron (1990) and Sheather and Jones (1991)). This method requires, however, the selection of, again, a new smoothing parameter.

We are now ready to state the weak convergence result for the conditional Kaplan-Meier process $(nh_n)^{1/2}(F_h(\cdot|x) - F(\cdot|x))$ in $D[0,T]$, the space of right continuous functions with left hand limits, endowed with the Skorokhod topology.

Weak convergence of a distribution estimator has been the object of study in many papers. In the case of censored data, Breslow and Crowley (1974), Gill (1983) and Lo and Singh (1986) obtained the weak convergence of the Kaplan-Meier process. In the latter paper, the result is obtained by decomposing the Kaplan-Meier estimator in a sum of i.i.d. terms and a remainder term. Dabrowska (1987) established the weak convergence of the Beran estimator in a random design regression model and for bandwidth sequences h_n satisfying $nh_n^5 \to 0$, using a similar approach as Breslow and Crowley (1974). Our approach is similar to the one of Lo and Singh (1986), i.e. we use the a.s. asymptotic representation given in Theorem

2.3. Whereas the result of Dabrowska (1987) does not cover the optimal bandwidth $h_n = Cn^{-1/5}$ $(C > 0)$, we state this case separately in part (b) of the next result.

Theorem 2.4

Assume $(C1), (C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}$.

(a) *If* $nh_n^5 \to 0$ *and* $\dfrac{(\log n)^3}{nh_n} \to 0$, *then, as* $n \to \infty$,

$$(nh_n)^{1/2}(F_h(\cdot|x) - F(\cdot|x)) \to W(\cdot|x) \qquad \text{in } D[0, T]$$

(b) *If* $h_n = Cn^{-1/5}$ *for some* $C > 0$, *then, as* $n \to \infty$,

$$(nh_n)^{1/2}(F_h(\cdot|x) - F(\cdot|x)) \to \tilde{W}(\cdot|x) \qquad \text{in } D[0, T],$$

where $W(\cdot|x)$ *and* $\tilde{W}(\cdot|x)$ *are Gaussian processes with covariance function given by (2.11) and, for* $\tilde{W}(\cdot|x)$, *mean function given by (2.12).*

Proof. From Theorem 2.3 and Lemma 2.3 we have the following asymptotic representation :

$$F_h(t|x) - F(t|x) = \sum_{i=1}^{n} w_{ni}(x; h_n)g(t, T_i, \Delta_i|x) + h_n^2\overline{b}(t|x) + \overline{r}_n(x, t),$$

where $g(t, T_i, \Delta_i|x) = \xi(T_i, \Delta_i, t|x) - E\xi(T_i, \Delta_i, t|x)$, $\overline{b}(t|x)$ is as in (2.15) and $\sup_{0\leq t\leq T} |\overline{r}_n(x, t)| = O((nh_n)^{-3/4}(\log n)^{3/4}) + o(h_n^2)$ a.s. The bias $(nh_n)^{1/2}h_n^2\overline{b}(t|x)$ is $o(1)$ under conditions (a) and equals $b(t|x)$ under conditions (b). Hence, it suffices to prove the weak convergence of $W_h(\cdot|x) = (nh_n)^{1/2}\sum_{i=1}^{n} w_{ni}(x; h_n)g(\cdot, T_i, \Delta_i|x)$ to the process $W(\cdot|x)$ (see e.g. Theorem 4.1 in Billingsley (1968)). For this, we will verify the conditions of Theorem 15.6 in Billingsley (1968). We first show convergence of the finite dimensional distributions, i.e. for any $q = 1, 2, ...$ and any $0 \leq t_1 \leq \ldots \leq t_q \leq T$: $(W_h(t_1|x), \ldots, W_h(t_q|x)) \xrightarrow{d} N(0, (\Gamma(t_i, t_j|x)))$. Since $W_h(t_i|x) = \sum_{k=1}^{n} W_{nki}$ where $W_{nki} = (nh_n)^{1/2}w_{nk}(x; h_n)g(t_i, T_k, \Delta_k|x)$, it suffices to check that (see e.g. Araujo and Giné (1980))

$$\lim_{n\to\infty}\sum_{k=1}^{n} E(W_{nki}W_{nkj}) = \Gamma(t_i,t_j|x) \qquad (1\le i,j\le q)$$
$$\lim_{n\to\infty}\sum_{k=1}^{n} \int\limits_{\{|W_{nk}|>\varepsilon\}} |W_{nk}|^2 dP = 0$$

for every $\varepsilon > 0$, where $|W_{nk}|^2 = \sum_{i=1}^{q} W_{nki}^2$. Now, applying Lemma 2.5,

$$\begin{aligned}\sum_{k=1}^{n} E\left(W_{nki}W_{nkj}\right) &= nh_n \sum_{k=1}^{n} w_{nk}^2(x;h_n)\mathrm{Cov}(g(t_i,T_k,\Delta_k|x), g(t_j,T_k,\Delta_k|x)) \\ &= \Gamma(t_i,t_j|x) + o(1).\end{aligned}$$

Also, since the functions $g(t_i, T_k, \Delta_k|x)$ are uniformly bounded, it follows that $\max_{1\le k\le n}|W_{nk}| = O((nh_n)^{-1/2})$ a.s. and $\sum_{k=1}^{n}|W_{nk}|^2 = O(1)$ a.s., and hence,

$$\sum_{k=1}^{n} \int\limits_{\{|W_{nk}|>\varepsilon\}} |W_{nk}|^2 dP \le O(1)P\left(\max_{1\le k\le n}|W_{nk}| > \varepsilon\right) = o(1).$$

Since the process $W(\cdot|x)$ is equal in law to $B(K(\cdot|x))(1-F(\cdot|x))(1+C(\cdot|x))$ (see Hall and Wellner (1980)), where $\{B(t); 0\le t\le 1\}$ is a Brownian bridge,

$$C(t|x) = \|K\|_2^2 \int_0^t \frac{dH^u(s|x)}{(1-H(s|x))^2} \tag{2.17}$$

and

$$K(t|x) = \frac{C(t|x)}{1+C(t|x)}, \tag{2.18}$$

we have that $W(\cdot|x)$ has continuous sample paths, with probability 1. Finally, we verify the moment condition for tightness : for all $0\le r\le s\le t\le T$:

$$\begin{aligned}&E[(W_h(r|x)-W_h(s|x))^2(W_h(s|x)-W_h(t|x))^2] \\ &\le (V(r)-V(s))(V(s)-V(t)) + a_n\end{aligned} \tag{2.19}$$

for some continuous and non-decreasing function V and some sequence $a_n = o(1)$. Note that the tightness condition in Theorem 15.6 in Billingsley (1968) does not contain the term a_n on the right hand side of (2.19). It is easily verified, however, that this slightly weaker condition is sufficient to guarantee the weak convergence of the process $W_h(\cdot|x)$. Write $g_{ti} = g(t, T_i, \Delta_i|x)$ and $w_{ni} = w_{ni}(x; h_n)$. Then, the left hand side of (2.19) equals

$$(nh_n)^2 \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{\ell=1}^n w_{ni} w_{nj} w_{nk} w_{n\ell} E[(g_{ri} - g_{si})(g_{rj} - g_{sj})(g_{sk} - g_{tk})(g_{s\ell} - g_{t\ell})]$$
$$= (nh_n)^2 \left\{ \sum_{i=1}^n w_{ni}^2 E[(g_{ri} - g_{si})^2] \sum_{j \neq i} w_{nj}^2 E[(g_{sj} - g_{tj})^2] \right.$$
$$+ 2 \sum_{i=1}^n \sum_{j \neq i}^n w_{ni}^2 w_{nj}^2 E[(g_{ri} - g_{si})(g_{si} - g_{ti})] E[(g_{rj} - g_{sj})(g_{sj} - g_{tj})]$$
$$\left. + \sum_{i=1}^n w_{ni}^4 E[(g_{ri} - g_{si})^2 (g_{si} - g_{ti})^2] \right\}$$
$$\leq 3(nh_n)^2 \sum_{i=1}^n w_{ni}^2 E[(g_{ri} - g_{si})^2] \sum_{i=1}^n w_{ni}^2 E[(g_{si} - g_{ti})^2] + O((nh_n)^{-1}),$$

using the inequalities of Cauchy-Schwarz and Hölder in the second term and the fact that the g functions are uniformly bounded in the third term. Now, by Lemma 2.5, we have

$$\sum_{i=1}^n w_{ni}^2 E[(g_{ri} - g_{si})^2]$$
$$= \sum_{i=1}^n w_{ni}^2 [\text{Var}(g_{ri}) + \text{Var}(g_{si}) - 2\text{Cov}(g_{ri}, g_{si})]$$
$$= (nh_n)^{-1} \{(1 - F(r|x))^2 C(r|x) + (1 - F(s|x))^2 C(s|x)$$
$$- 2(1 - F(r|x))(1 - F(s|x)) C(r|x)\} + o((nh_n)^{-1})$$
$$\leq (nh_n)^{-1} \|K\|_2^2 (1 - H(T|x))^{-2} \{(F(s|x) - F(r|x))^2 + H^u(s|x) - H^u(r|x)\}$$
$$+ o((nh_n)^{-1}).$$

Hence, (2.19) is satisfied with $V(s) = \sqrt{3} \|K\|_2^2 (1 - H(T|x))^{-2} (F(s|x) + H^u(s|x))$.

In the following result we state the asymptotic normality of the Beran estimator, which follows readily from the weak convergence, shown above. A direct proof (which does not make use of the weak convergence result) can be found in González Manteiga and Cadarso Suárez (1994) (only case (a)) or in Van Keilegom and Veraverbeke (1997b).

Corollary 2.1

(a) Assume the conditions of Theorem 2.4(a) are satisfied. Then, for $0 \leq t \leq T$, as $n \to \infty$,

$$(nh_n)^{1/2}(F_h(t|x) - F(t|x)) \xrightarrow{d} N(0; \Gamma(t,t|x))$$

(b) Assume the conditions of Theorem 2.4(b) are satisfied. Then, for $0 \leq t \leq T$, as $n \to \infty$,

$$(nh_n)^{1/2}(F_h(t|x) - F(t|x)) \xrightarrow{d} N(b(t|x); \Gamma(t,t|x)),$$

where $\Gamma(t,t|x)$ and $b(t|x)$ are given by (2.11) and (2.12).

Remark 2.2

The above result is not directly applicable for estimating the distribution of $F_h(t|x) - F(t|x)$, since the variance and bias expressions of the above normal approximation contain the unknown quantities $F(t|x), H(t|x), H^u(t|x), \ddot{H}(t|x)$ and $\ddot{H}^u(t|x)$. Two possibilities exist to overcome this difficulty. First, a studentized version of Corollary 2.1 can be developed, in which the variance and the bias terms are replaced by appropriate estimators. Second, the above result can be used to construct a bootstrap estimation procedure for the distribution of $F_h(t|x) - F(t|x)$. In the next chapter, we will follow this second approach.

2.4 Appendix : Some results on conditional empirical distribution functions

In this appendix we prove some basic results for empirical distribution functions of the kernel type (called Stone estimators) which play a major role in fixed design regression models. These results are frequently applied to either $H(t|x)$ or $H^u(t|x)$ in (1.21) and (1.22). We state them for a general empirical

$$L_h(t|x) = \sum_{i=1}^{n} w_{ni}(x; h_n) I(Z_i \leq t),$$

which is an estimator for the (sub)distribution function

$$L(t|x) = P(Z_x \leq t)$$

and where $Z_1, \ldots, Z_n$ are independent random variables with (sub)distributions $L(\cdot|x_1), \ldots, L(\cdot|x_n)$ and Z_x is the response at $x \in]0,1[$. (The proofs are also valid for empiricals of the type $\sum_{i=1}^{n} w_{ni}(x; h_n) I(Z_i \leq t, \Delta_i = 1)$).

We start with a result on bias and variance which can be found in Lemma 1 in Aerts, Janssen and Veraverbeke (1994a).

Lemma 2.6 (Bias and variance)

(a) Assume $(C1), (C2)$, $L(t|x)$ *satisfies* $(C3)$, $h_n \to 0$, $nh_n \to \infty$. *Then, as* $n \to \infty$,

$$\sup_{0 \leq t \leq T} |EL_h(t|x) - L(t|x)| = o(h_n).$$

(b) Assume $(C1), (C2)$, $L(t|x)$ *satisfies* $(C3)$ *and* $(C5)$, $h_n \to 0$. *Then, as* $n \to \infty$,

$$\sup_{0 \leq t \leq T} |EL_h(t|x) - L(t|x)| = O(h_n^2 + n^{-1}).$$

More in particular :

$$\sup_{0 \leq t \leq T} |EL_h(t|x) - L(t|x) - \frac{1}{2}\mu_2^K \ddot{L}(t|x) h_n^2| = o(h_n^2) + O(n^{-1}).$$

(c) Assume $(C1), (C2)$, $L(t|x)$ *satisfies* $(C3)$, $h_n \to 0$, $nh_n \to \infty$. *Then, as* $n \to \infty$,

$$Var\, L_h(t|x) = \frac{1}{nh_n} L(t|x)(1 - L(t|x)) \|K\|_2^2 + o((nh_n)^{-1}).$$

Lemma 2.7 (Pointwise strong consistency)
Assume $(C1), (C2)$, $L(t|x)$ *satisfies* $(C3)$, $h_n \to 0$, $\dfrac{\log n}{nh_n} \to 0$. *Then, as* $n \to \infty$, *for* $t \leq T$,

$$L_h(t|x) - L(t|x) \to 0 \qquad \text{a.s.}$$

Proof. With $X_{in} = w_{ni}(x;h_n)[I(Z_i \leq t) - L(t|x_i)]$ we have $L_h(t|x) - L(t|x) = \sum_{i=1}^n X_{in} + EL_h(t|x) - L(t|x)$, and by Lemma 2.6(a) it suffices to prove strong consistency of $\sum_{i=1}^n X_{in}$. We have : $|X_{in}| \leq w_{ni}(x;h_n) \leq \|K\|_\infty \overline{\Delta}_n/h_n$. Also $EX_{in} = 0$ and

$$\sum_{i=1}^n \mathrm{Var}(X_{in}) = \sum_{i=1}^n w_{ni}^2(x;h_n)L(t|x_i)(1 - L(t|x_i)) \leq \sum_{i=1}^n w_{ni}^2(x;h_n) \leq \|K\|_\infty \overline{\Delta}_n/h_n.$$

Hence, by Bernstein's inequality (see e.g. Serfling (1980)), for all $\varepsilon > 0$,

$$P\left(\left|\sum_{i=1}^n X_{in}\right| > \varepsilon\right) \leq 2\exp(-c\varepsilon^2 h_n/\overline{\Delta}_n)$$

for some constant $c > 0$. By $(C1)$ and the condition $\log n/(nh_n) \to 0$, the right hand side can be made integrable. This shows the complete convergence of X_{in}, and hence also the strong convergence.

Lemma 2.8 (Dvoretzky-Kiefer-Wolfowitz type exponential bound)
Assume (C1), (C2), $h_n \to 0$, $nh_n \to \infty$.

(a) *For* $\varepsilon > 0$ *and* n *sufficiently large such that*

$$\varepsilon^2 \geq \frac{3}{2}\|K\|_2^2 \frac{1}{nh_n}, \tag{2.20}$$

we have for any $T > 0$

$$P\left(\sup_{0\leq t\leq T} |L_h(t|x) - EL_h(t|x)| > \varepsilon\right) \leq d_0 nh_n \varepsilon \exp(-d_1 nh_n \varepsilon^2). \tag{2.21}$$

(b) *If moreover* $L(t|x)$ *satisfies* $(C3)$, *then for* $\varepsilon > 0$ *and* n *sufficiently large such that*

$$\varepsilon \geq \max(\sqrt{6}\|K\|_2(nh_n)^{-1/2}, 2\|\dot{L}\|\overline{\Delta}_n + 2\mu_2^K\|\ddot{L}\|h_n^2), \tag{2.22}$$

we have

$$P\left(\sup_{0\le t\le T}|L_h(t|x)-L(t|x)|>\varepsilon\right)\le\frac{1}{2}d_0nh_n\varepsilon\exp\left(-\frac{1}{4}d_1nh_n\varepsilon^2\right). \quad (2.23)$$

Here $d_0=\dfrac{8e^2}{\|K\|_2^2}$ *and* $d_1=\dfrac{4}{3\|K\|_2^2}$.

Proof. (a) Applying a general exponential bound result of Singh (1975) gives that the left hand side of (2.21) is bounded by

$$\frac{4e^2\varepsilon}{\sum_{i=1}^n w_{ni}^2(x;h_n)}\exp\left\{-2\frac{\varepsilon^2}{\sum_{i=1}^n w_{ni}^2(x;h_n)}\right\},$$

provided $\varepsilon^2\ge\sum_{i=1}^n w_{ni}^2(x;h_n)$. Now,

$$\begin{aligned}&\sum_{i=1}^n w_{ni}^2(x;h_n)-\frac{1}{nh_n}\|K\|_2^2\\&=\sum_{i=1}^n\left(\int_{x_{i-1}}^{x_i}\frac{1}{h_n}K\left(\frac{x-z}{h_n}\right)dz\right)^2-\sum_{i=1}^n\frac{x_i-x_{i-1}}{h_n}\int_{x_{i-1}}^{x_i}\frac{1}{h_n}K^2\left(\frac{x-z}{h_n}\right)dz\\&\quad+\sum_{i=1}^n\frac{x_i-x_{i-1}}{h_n}\int_{x_{i-1}}^{x_i}\frac{1}{h_n}K^2\left(\frac{x-z}{h_n}\right)dz-\frac{1}{nh_n}\|K\|_2^2\\&=O((nh_n)^{-2})+o((nh_n)^{-1})=o((nh_n)^{-1})\end{aligned}$$

(see the proof of Lemma 2.4). Hence, for n sufficiently large,

$$\frac{1}{2}\|K\|_2^2\frac{1}{nh_n}\le\sum_{i=1}^n w_{ni}^2(x;h_n)\le\frac{3}{2}\|K\|_2^2\frac{1}{nh_n}.$$

This, together with condition (2.20) on ε gives the desired bound.
(b) For $\varepsilon>0$, the left hand side in (2.23) is bounded above by

$$P\left(\sup_{0\le t\le T}|L_h(t|x)-EL_h(t|x)|>\varepsilon-\sup_{0\le t\le T}|EL_h(t|x)-L(t|x)|\right). \quad (2.24)$$

Now, from the proof of Lemma 2.6(b) and condition (2.22) on ε,

$$\sup_{0\le t\le T}|EL_h(t|x)-L(t|x)|\le\|\dot L\|\overline{\Delta}_n+\mu_2^K\|\ddot L\|h_n^2\le\frac{\varepsilon}{2}.$$

Again by the condition on ε,

$$\left(\varepsilon - \sup_{0\le t\le T} |EL_h(t|x) - L(t|x)|\right)^2 \ge \frac{\varepsilon^2}{4} \ge \frac{3}{2}\|K\|_2^2 \frac{1}{nh_n}.$$

This allows to apply the (a)-part to (2.24) which leads to the bound in (2.23).

Lemma 2.9 (Rate of uniform strong consistency)
Assume $(C1), (C2)$, $L(t|x)$ *satisfies* $(C3)$ *and* $(C5)$ *and* $\dfrac{nh_n^5}{\log n} = O(1)$. *Then, as* $n \to \infty$,

$$\sup_{0\le t\le T} |L_h(t|x) - L(t|x)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \qquad a.s.$$

Proof. Apply Lemma 2.8(b) with the choice $\varepsilon = \varepsilon_n = c(nh_n)^{-1/2}(\log n)^{1/2}$ for some appropriately chosen constant $c > 0$. Apply Borel-Cantelli.

Lemma 2.10 (Exponential bound for the modulus of continuity)
Assume $(C1), (C2)$, $L(t|x)$ *satisfies* $(C3), (C6)$ *and* $(C7)$, $h_n \to 0$. *Let* $\{a_n\}$ *be a sequence of positive constants, tending to 0 as* $n \to \infty$. *Then, for* n *sufficiently large and*

$$\varepsilon \ge 3(2\|\dot{L}\|\overline{\Delta}_n + L_0\|\dot{L}'\|a_n h_n + \|L''\|a_n^2) \tag{2.25}$$

(where $[-L_0, L_0]$ *is the support of the kernel* K*), we have*

$$P\left(\sup_{0\le s,t\le T; |t-s|\le a_n} |L_h(t|x) - L_h(s|x) - L(t|x) + L(s|x)| \ge \varepsilon\right)$$
$$\le C_1 \frac{T\|L'\|^2 a_n}{\varepsilon^2} \exp\left(-C_2 \frac{nh_n\varepsilon^2}{\|L'\|a_n + \varepsilon}\right),$$

where $C_1, C_2 > 0$ *are absolute constants.*

Proof. Partition the interval $[0, T]$ into $m = [T/a_n]$ subintervals of length $\overline{a}_n = T/m : 0 = t_0 < t_1 < \ldots < t_m = T$ with $t_i = i\overline{a}_n$ for $i = 0, \ldots, m$. Let $I_{ni} = [t_i - \overline{a}_n, t_i + \overline{a}_n]$, $i = 1, \ldots, m-1$. We have : $a_n \le \overline{a}_n < 2a_n$ for n large. Hence, for $s, t \in [0, T]$ with $|t-s| \le a_n$, there exists an interval I_{ni} such that $s, t \in I_{ni}$. Partition each interval I_{ni} by a grid $t_{ij} = t_i + j\overline{a}_n b_n^{-1}$, $j = -b_n, \ldots, b_n$, where $b_n = 12\|L'\|a_n\varepsilon^{-1}$. Using the monotonicity of $L_h(t|x)$ and $EL_h(t|x)$, we have

$$\sup_{0\le s,t\le T;|t-s|\le a_n} |L_h(t|x) - L_h(s|x) - EL_h(t|x) + EL_h(s|x)| \tag{2.26}$$
$$\le \max_{1\le i\le m-1} \max_{-b_n\le j,k\le b_n} |L_h(t_{ik}|x) - L_h(t_{ij}|x) - EL_h(t_{ik}|x) + EL_h(t_{ij}|x)|$$
$$+ 2 \max_{1\le i\le m-1} \max_{-b_n\le j\le b_n-1} |EL_h(t_{i,j+1}|x) - EL_h(t_{ij}|x)|.$$

From the Lipschitz continuity of $L(\cdot|x)$ (implied by condition $(C4)$), it follows that the second term in (2.26) is bounded by $2\|L'\|\overline{a}_n b_n^{-1} \le \varepsilon/3$. The proof will be finished if we can show an exponential inequality for the first term on the right hand side of (2.26) and if the bias term $EL_h(t|x) - EL_h(s|x) - L(t|x) + L(s|x)$ is smaller than $\varepsilon/3$, uniformly over all $|t-s| \le a_n$. As to the first term in (2.26), we have that $L_h(t_{ik}|x) - L_h(t_{ij}|x) - EL_h(t_{ik}|x) + EL_h(t_{ij}|x) = \sum_{r=1}^n X_{rijk}$, where $X_{rijk} = w_{nr}(x;h_n)\{[I(Z_r \le t_{ik}) - I(Z_r \le t_{ij})] - [L(t_{ik}|x_r) - L(t_{ij}|x_r)]\}$. We have : $|X_{rijk}| \le w_{nr}(x;h_n) \le \|K\|_\infty \overline{\Delta}_n/h_n$, $E(X_{rijk}) = 0$ and

$$\begin{aligned}
\mathrm{Var}(X_{rijk}) &= w_{nr}^2(x;h_n)\{L(t_{ik}|x_r)(1 - L(t_{ik}|x_r)) + L(t_{ij}|x_r)(1 - L(t_{ij}|x_r)) \\
&\quad - 2(L(t_{ij} \wedge t_{ik}|x_r) - L(t_{ik}|x_r)L(t_{ij}|x_r))\} \\
&= w_{nr}^2(x;h_n)\{-(L(t_{ik}|x_r) - L(t_{ij}|x_r))^2 \\
&\quad + [L(t_{ik}|x_r) - L(t_{ij} \wedge t_{ik}|x_r)] + [L(t_{ij}|x_r) - L(t_{ij} \wedge t_{ik}|x_r)]\} \\
&\le C\|L'\|w_{nr}^2(x;h_n)a_n
\end{aligned}$$

for some absolute constant $C > 0$, using the Lipschitz continuity of $L(\cdot|x_r)$. It follows that $\sum_{r=1}^n \mathrm{Var}(X_{rijk}) \le C\|K\|_\infty\overline{\Delta}_n\|L'\|a_n h_n^{-1}$. By Bernstein's inequality,

$$P\left(\max_{1\le i\le m-1} \max_{-b_n\le j,k\le b_n} |L_h(t_{ik}|x) - L_h(t_{ij}|x) - EL_h(t_{ik}|x) + EL_h(t_{ij}|x)| > \frac{\varepsilon}{3}\right)$$
$$\le 2(m-1)(2b_n+1)^2 \exp\left\{-\frac{\varepsilon^2}{9} \Big/ \left(2C\frac{\|K\|_\infty\overline{\Delta}_n\|L'\|}{h_n}a_n + \frac{2}{3}\frac{\|K\|_\infty\overline{\Delta}_n}{h_n}\frac{\varepsilon}{3}\right)\right\}$$
$$\le C_1\frac{T\|L'\|^2 a_n}{\varepsilon^2}\exp\left(-C_2\frac{nh_n\varepsilon^2}{\|L'\|a_n + \varepsilon}\right)$$

for some absolute constants $C_1, C_2 > 0$.
Write

$$
\begin{aligned}
&|EL_h(t|x) - EL_h(s|x) - L(t|x) + L(s|x)| \\
&\leq \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) \{|L(t|x_i) - L(t|z)| + |L(s|x_i) - L(s|z)|\}\, dz \\
&\quad + \int_0^{x_n} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) |\, [L(t|z) - L(t|x)] - [L(s|z) - L(s|x)]\,|\, dz.
\end{aligned}
$$

The first term is bounded by $2\|\dot{L}\|\overline{\Delta}_n$. The second term can be written as

$$
\int_{(x-x_n)/h_n}^{x/h_n} K(u)|[L(t|x - h_n u) - L(t|x)] - [L(s|x - h_n u) - L(s|x)]|\, du
$$

$$
= \int_{(x-x_n)/h_n}^{x/h_n} K(u)|(t-s)(L'_{x-h_n u}(s) - L'_x(s)) + \frac{1}{2}(t-s)^2(L''_{x-h_n u}(\theta_0) - L''_x(\theta_0))|\, du,
$$

where θ_0 is an intermediate point between t and s. This can be bounded by $|t-s|\|\dot{L}'\|L_0 h_n + (t-s)^2\|L''\| \leq L_0\|\dot{L}'\|a_n h_n + \|L''\|a_n^2$. Hence,

$$
\sup_{0\leq s,t\leq T; |t-s|\leq a_n} |EL_h(t|x) - EL_h(s|x) - L(t|x) + L(s|x)| \leq \frac{\varepsilon}{3}.
$$

This finishes the proof.

Lemma 2.11 (Almost sure behavior of the modulus of continuity)
Assume $(C1)$, $(C2)$, $L(t|x)$ satisfies $(C3), (C6)$ and $(C7)$, $h_n \to 0$. Let $\{a_n\}$ be a sequence of positive constants, tending to 0 as $n \to \infty$, with

$$
a_n(nh_n)(\log n)^{-1} > \Delta > 0
$$

for all n sufficiently large. Then, as $n \to \infty$,

$$
\begin{aligned}
&\sup_{0\leq s,t\leq T; |t-s|\leq a_n} |L_h(t|x) - L_h(s|x) - L(t|x) + L(s|x)| \\
&= O(a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2} + n^{-1} + a_n h_n + a_n^2) \qquad \text{a.s.}
\end{aligned}
$$

Proof. This follows easily from the proof of Lemma 2.10, combined with the Borel-Cantelli lemma, by choosing $\varepsilon = \varepsilon_n = c a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2}$, for some appropriately chosen constant $c > 0$. Only the derivation of the bias term $|EL_h(t|x) - EL_h(s|x) - L(t|x) + L(s|x)|$ in this proof needs to be slightly modified : we do not assume that condition (2.25) is satisfied, but allow ε_n to be smaller than the right hand side of (2.25), which leads, however, to a more complicated rate of convergence.

Corollary 2.2
If we take $a_n = c_0(nh_n)^{-1/2}(\log n)^{1/2}$ *(*$c_0 > 0$*) in Lemma 2.11, then the order is* $O((nh_n)^{-3/4}(\log n)^{3/4})$*, provided that* $\dfrac{\log n}{nh_n} \to 0$ *and* $\dfrac{nh_n^5}{\log n} = O(1)$.

Chapter 3

Bootstrap estimation of the conditional distribution in regression with censored data

In Chapter 2, we proved that the distribution of $(nh_n)^{1/2}(F_h(t|x) - F(t|x))$ converges to a normal distribution, with variance and bias given by (2.11) and (2.12) (see Corollary 2.1). These variance and bias expressions contain, however, unknown quantities like $H(t|x), \ddot{H}(t|x)$ and $F(t|x)$. To avoid the estimation of these expressions, we will show in Section 3.2 of this chapter that the distribution of $(nh_n)^{1/2}(F_h(t|x) - F(t|x))$ can also be estimated (in a consistent way) by the distribution of $(nh_n)^{1/2}(F^*_{hg}(t|x) - F_g(t|x))$, conditional on the observations (see (1.24) for the definition of the bootstrap estimator of the distribution function $F^*_{hg}(t|x)$). This alternative (bootstrap) approximation has the advantage that no unknown quantities need to be estimated. In Section 3.3 we will compare the Gaussian and the bootstrap approximation in a simulation study.

Another important result of this chapter concerns the construction of a bootstrap confidence band for the distribution function $F(\cdot|x)$. We will construct this band in Section 3.2.

Both results are consequences of the main result in Section 3.1, which is the weak convergence of the bootstrapped process $(nh_n)^{1/2}(F^*_{hg}(t|x) - F_g(t|x))$ to a Gaussian process. This result will be obtained from an a.s. asymptotic representation for $F^*_{hg}(t|x) - F_g(t|x)$. Finally, Section 3.4 contains a number of results on the bootstrapped estimators $H^*_{hg}(t|x)$ and $H^{*u}_{hg}(t|x)$, needed in the proofs of the main results.

The weak convergence result and the construction of confidence bands can be found in Van Keilegom and Veraverbeke (1997a), while the construction of the representation and the validity of the bootstrap procedure are part of Van Keilegom and Veraverbeke (1997b).

The results of this chapter require a slightly stronger version of condition $(C2)$ (defined in Section 2.1). We will denote it by $(C2')$:

$(C2')$ K is a twice differentiable probability density function with finite support $[-L_0, L_0]$ for some $L_0 > 0$, $\mu_1^K = 0$, K'' is continuous and $K(-L_0) = K'(-L_0) = K(L_0) = K'(L_0) = 0$.

We will state the results in this chapter for the bandwidth sequence h_n of optimal rate, i.e. $h_n = Cn^{-1/5}$ for some constant $C > 0$. The notations $P^*, E^*, \text{Var}^*, \ldots$ will be used for probability, expectation, variance, ... conditionally on the original observations. Finally, with ξ as in Theorem 2.3, we write

$$b_n(t|x) = \sum_{i=1}^n w_{ni}(x;h_n) E\xi(T_i, \Delta_i, t|x) \tag{3.1}$$

$$\hat{b}_n(t|x) = \sum_{i=1}^n w_{ni}(x;h_n) E^*\xi(T_i^*, \Delta_i^*, t|x) - \sum_{i=1}^n w_{ni}(x;g_n)\xi(T_i, \Delta_i, t|x). \tag{3.2}$$

3.1 Asymptotic representation and weak convergence

We start by showing an a.s. asymptotic representation, which decomposes $F^*_{hg}(t|x) - F_g(t|x)$ in a leading term and a remainder term of lower order. The leading term is the difference of two weighted sums of independent terms, one depending on the bootstrap sample and one depending on the original sample. In the i.i.d. case with censored data, Lo and Singh (1986) obtained a similar representation for the Kaplan-Meier estimator.

Theorem 3.1
Assume $(C1)$, $(C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Then, for* $t < \tau_{H(\cdot|x)}$,

$$
\begin{aligned}
&F^*_{hg}(t|x) - F_g(t|x)\\
&= \sum_{i=1}^n w_{ni}(x;h_n)\xi(T_i^*, \Delta_i^*, t|x) - \sum_{i=1}^n w_{ni}(x;g_n)\xi(T_i, \Delta_i, t|x) + r_n^*(x,t),
\end{aligned}
$$

where ξ *is as in Theorem 2.3 and where*

$$
\sup_{0\le t\le T} |r_n^*(x,t)| = O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}) \qquad a.s.
$$

as $n \to \infty$.

Proof. We have

$$
\begin{aligned}
F^*_{hg}(t|x) - F_g(t|x) &= (1 - F_g(t|x))[1 - \exp\{\log(1 - F^*_{hg}(t|x)) - \log(1 - F_g(t|x))\}]\\
&= (1 - F_g(t|x))\Big\{-[\log(1 - F^*_{hg}(t|x)) - \log(1 - F_g(t|x))]\\
&\quad - \frac{1}{2}[\log(1 - F^*_{hg}(t|x)) - \log(1 - F_g(t|x))]^2 e^{\theta_n}\Big\} \qquad (3.3)\\
&= (1 - F_g(t|x))\left\{A - \frac{1}{2}B\right\},
\end{aligned}
$$

where θ_n is between 0 and $\log(1 - F^*_{hg}(t|x)) - \log(1 - F_g(t|x))$. It is easy to show that

$$
\begin{aligned}
A = &\int_0^t \frac{H^*_{hg}(s|x) - H_g(s|x)}{(1 - H_g(s|x))^2}\, dH^u_g(s|x) + \frac{H^{*u}_{hg}(t|x) - H^u_g(t|x)}{1 - H_g(t|x)}\\
&- \int_0^t \frac{H^{*u}_{hg}(s|x) - H^u_g(s|x)}{(1 - H_g(s|x))^2}\, dH_g(s|x) + R_{n1}(t) + R_{n2}(t) + R_{n3}(t) + R_{n4}(t),
\end{aligned}
$$

where

$$R_{n1}(t) = -\log(1 - F^*_{hg}(t|x)) - \int_0^t \frac{dH^{*u}_{hg}(s|x)}{1 - H^*_{hg}(s|x)}$$

$$R_{n2}(t) = \int_0^t \frac{(H^*_{hg}(s|x) - H_g(s|x))^2}{(1 - H_g(s|x))^2(1 - H^*_{hg}(s|x))}\, dH^u_g(s|x)$$

$$R_{n3}(t) = \int_0^t \left(\frac{1}{1 - H^*_{hg}(s|x)} - \frac{1}{1 - H_g(s|x)} \right) d(H^{*u}_{hg}(s|x) - H^u_g(s|x))$$

$$R_{n4}(t) = \int_0^t \frac{dH^u_g(s|x)}{1 - H_g(s|x)} + \log(1 - F_g(t|x)).$$

Direct application of Lemma 2.2 gives

$$\sup_{0 \le t \le T} |R_{n4}(t)| = O((ng_n)^{-1}) = O((nh_n)^{-1}) \qquad \text{a.s.}$$

Following the lines of the proof of Lemma 2.2, we obtain

$$\sup_{0 \le t \le T} |R_{n1}(t)| \le \sup_{0 \le t \le T} |G^*_{hg}(t|x) - G^*_{hg}(t-|x)|\, (1 - H^*_{hg}(T|x))^{-1} = O_{P^*}((nh_n)^{-1})$$

a.s., since $H^*_{hg}(T|x) \xrightarrow{P^*} H(T|x)$ by application of Lemmas 2.7 and 3.7. Clearly, $\sup_{0 \le t \le T} |R_{n2}(t)| = O_{P*}((nh_n)^{-1} \log n)$ a.s. by Lemma 3.7. For $R_{n3}(t)$, we use Lemma 3.2. The first term in A equals

$$\int_0^t \frac{H^*_{hg}(s|x) - H_g(s|x)}{(1 - H(s|x))^2}\, dH^u(s|x) + R_{n5}(t) + R_{n6}(t),$$

where

$$R_{n5}(t) = \int_0^t (H^*_{hg}(s|x) - H_g(s|x)) \left(\frac{1}{(1 - H_g(s|x))^2} - \frac{1}{(1 - H(s|x))^2} \right) dH^u_g(s|x)$$

$$R_{n6}(t) = \int_0^t \frac{H^*_{hg}(s|x) - H_g(s|x)}{(1 - H(s|x))^2}\, d(H^u_g(s|x) - H^u(s|x)).$$

$R_{n5}(t)$ is uniformly bounded by

$$\frac{2}{(1-H_g(T|x))^2(1-H(T|x))^2}\sup_{0\le t\le T}|H^*_{hg}(t|x)-H_g(t|x)|\sup_{0\le t\le T}|H_g(t|x)-H(t|x)|$$

and this is $O_{P^*}((nh_n)^{-1}\log n)$ using Lemmas 3.7 and 3.9. Using some analogues of Lemma 2.1, it is easy to show that $\sup_{0\le t\le T}|R_{n6}(t)| = O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4})$ a.s. The second and third term in A can be worked out in a similar way. Hence,

$$\begin{aligned}
A &= \int_0^t \frac{H^*_{hg}(s|x)-H_g(s|x)}{(1-H(s|x))^2}\,dH^u(s|x) + \frac{H^{*u}_{hg}(t|x)-H^u_g(t|x)}{1-H(t|x)} \\
&\quad - \int_0^t \frac{H^{*u}_{hg}(s|x)-H^u_g(s|x)}{(1-H(s|x))^2}\,dH(s|x) + \rho^*_n(t) \\
&= (1-F(t|x))^{-1}\left[\sum_{i=1}^n w_{ni}(x;h_n)\xi(T^*_i,\Delta^*_i,t|x) - \sum_{i=1}^n w_{ni}(x;g_n)\xi(T_i,\Delta_i,t|x)\right] + \rho^*_n(t),
\end{aligned}$$

where $\sup_{0\le t\le T}|\rho^*_n(t)| = O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4})$ a.s.
To deal with the term $-1/2\,B$ in (3.3), we first note that $B = A^2e^{\theta_n}$ and that $e^{\theta_n} \le (1-F_g(T|x))^{-1}$. From Lemma 3.7 it follows that $A = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. Hence, by Lemma 3.1 below, $B = O_{P^*}((nh_n)^{-1}\log n)$ a.s. To complete the proof of Theorem 3.1, we still have to replace the factor $1-F_g(t|x)$ in (3.3) by $1-F(t|x)$. This is allowed by Lemma 3.1 below.

Lemma 3.1
Assume $(C1)$, $(C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C3)$ *and* $(C5)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C>0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Then, as* $n\to\infty$,

$$\sup_{0\le t\le T}|F_g(t|x)-F(t|x)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \qquad \text{a.s.}$$

Proof. In a completely analogous way as in Theorem 2.1, we have that for $1 - H(T|x) > \delta > 0$,

$$P\left(\sup_{0\le t\le T}|F_g(t|x) - F(t|x)| > \varepsilon\right)$$
$$\le 2P\left(\sup_{0\le t\le T}|H_g(t|x) - H(t|x)| > \frac{\varepsilon\delta^2}{12}\right) + P\left(\sup_{0\le t\le T}|H_g^u(t|x) - H^u(t|x)| > \frac{\varepsilon\delta^2}{12}\right)$$

and hence, using the proof of Lemma 3.9, the result follows.

Lemma 3.2
If $(C2')$ holds and $H(t|x)$ and $H^u(t|x)$ satisfy $(C5) - (C7)$ in $[0,T]$ with $T < \tau_{H(\cdot|x)}$, then, as $n \to \infty$,

$$\sup_{0\le t\le T}\left|\int_0^t \left(\frac{1}{1 - H_{hg}^*(s|x)} - \frac{1}{1 - H_g(s|x)}\right) d(H_{hg}^{*u}(s|x) - H_g^u(s|x))\right|$$
$$= O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.}$$

Proof. The proof parallels completely that of Lemma 2.1, i.e. the same partitionings and the same inequalities are used. Also, use is made of Lemmas 3.7, 3.8 and 3.9.

We next prove a lemma on the bootstrap bias which will be needed for showing the weak convergence of the bootstrapped process.

Lemma 3.3
Assume $(C1)$, $(C2')$, $H(t|x)$ and $H^u(t|x)$ satisfy $(C5)-(C7)$ in $[0,T]$ with $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ for some $C > 0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$. Then, as $n \to \infty$,

$$(nh_n)^{1/2}\sup_{0\le t\le T}|\hat{b}_n(t|x) - b_n(t|x)| \to 0 \qquad \text{a.s.},$$

where $b_n(t|x)$ and $\hat{b}_n(t|x)$ are given by (3.1) and (3.2).

Proof. Because

$$
\begin{aligned}
&(nh_n)^{1/2}|\hat{b}_n(t|x) - b_n(t|x)| \\
&= (nh_n)^{1/2}(1 - F(t|x)) \Bigg| \int_0^t \frac{1}{(1 - H(s|x))^2} [E^* H^*_{hg}(s|x) - H_g(s|x) - EH_h(s|x) \\
&\qquad\qquad + H(s|x)]\, dH^u(s|x) \\
&\quad + \frac{1}{1 - H(t|x)} [E^* H^{*u}_{hg}(t|x) - H^u_g(t|x) - EH^u_h(t|x) + H^u(t|x)] \\
&\quad - \int_0^t \frac{1}{(1 - H(s|x))^2} [E^* H^{*u}_{hg}(s|x) - H^u_g(s|x) - EH^u_h(s|x) + H^u(s|x)]\, dH(s|x) \Bigg|,
\end{aligned}
$$

the result follows from Lemma 3.6.

The following lemma calculates the covariance of the first term in the asymptotic representation of Theorem 3.1.

Lemma 3.4
Assume $(C1)$, $(C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C3)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\frac{ng_n^5}{\log n} \to \infty$. *Then, as* $n \to \infty$,

$$
\sup_{0 \le s,t \le T} \left| \sum_{i=1}^n w^2_{ni}(x; h_n) Cov^*(\xi(T^*_i, \Delta^*_i, s|x), \xi(T^*_i, \Delta^*_i, t|x)) - \frac{\Gamma(s, t|x)}{nh_n} \right| = o((nh_n)^{-1})
$$

a.s., where ξ *is as in Theorem 2.3 and* $\Gamma(s, t|x)$ *is as in (2.11).*

Proof. Straightforward calculations show that $\mathrm{Cov}^*(\xi(T^*_i, \Delta^*_i, s|x), \xi(T^*_i, \Delta^*_i, t|x))$ equals expression (2.14), but where every $H(\cdot|x_i)$ is replaced with $H_g(\cdot|x_i)$ and every $H^u(\cdot|x_i)$ with $H^u_g(\cdot|x_i)$. Hence the result follows from Lemma 2.5, since it holds that

$$
\max_{1 \le i \le n} \sup_{0 \le t \le T} |H_g(t|x_i) - H(t|x_i)| \left(\sum_{j=1}^n w^2_{nj}(x; h_n) \right) = o((nh_n)^{-1}) \qquad \text{a.s.},
$$

using a slight modification of Lemma 3.9.

We now show the weak convergence of the bootstrapped process $(nh_n)^{1/2}(F^*_{hg}(t|x) - F_g(t|x))$ for the optimal smoothing parameter $h_n = Cn^{-1/5}$ for some $C > 0$ (see Remark 2.1). The following result shows that the limiting Gaussian process of this bootstrapped process is identical to the limit of the original process $(nh_n)^{1/2}(F_h(t|x) - F(t|x))$, which was obtained in Theorem 2.4.

For censored i.i.d. data, Akritas (1986) and Lo and Singh (1986) proved the weak convergence of the bootstrapped Kaplan-Meier process. The former result makes use of martingale theory, while in the latter paper, the result was established by using a representation for the bootstrapped Kaplan-Meier estimator. Our approach is similar to the one of Lo and Singh (1986), namely, we will use the representation established in Theorem 3.1.

Theorem 3.2
Assume $(C1)$, $(C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)$–$(C7)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\frac{ng_n^5}{\log n} \to \infty$ *and* $\frac{ng_n^5}{\log n}\frac{h_n}{g_n} = O(1)$. *Then, as* $n \to \infty$,

$$(nh_n)^{1/2}(F^*_{hg}(\cdot|x) - F_g(\cdot|x)) \to \tilde{W}(\cdot|x) \qquad \text{in } D[0,T] \text{ a.s.},$$

where $\tilde{W}(\cdot|x)$ *is a Gaussian process with covariance function given by (2.11) and mean function function given by (2.12).*

Proof. In a similar way as in Theorem 2.4 (use Lemma 3.4 instead of Lemma 2.5), one can prove that $W^*_{hg}(\cdot|x) \to W(\cdot|x)$ in $D[0,T]$ a.s., where

$$W^*_{hg}(t|x) = (nh_n)^{1/2} \sum_{i=1}^n w_{ni}(x; h_n) g(t, T_i^*, \Delta_i^*|x),$$

$g(t, T_i^*, \Delta_i^*|x) = \xi(T_i^*, \Delta_i^*, t|x) - E^*\xi(T_i^*, \Delta_i^*, t|x)$ and $W(\cdot|x)$ is given by (2.11). From Theorem 3.1 we know that

$$\begin{aligned}(nh_n)^{1/2}(F^*_{hg}(t|x) - F_g(t|x)) &= W^*_{hg}(t|x) + (nh_n)^{1/2}(\hat{b}_n(t|x) - b_n(t|x)) \\ &\quad + ((nh_n)^{1/2} b_n(t|x) - b(t|x)) + b(t|x) + r_n^*(x,t),\end{aligned}$$

where $\sup_{0\le t\le T} |r_n^*(x,t)| = O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4})$ a.s. Lemma 3.3 entails that $(nh_n)^{1/2} \sup_{0\le t\le T} |\hat{b}_n(t|x) - b_n(t|x)| = o(1)$ a.s., while from Lemma 2.3 it follows that $\sup_{0\leq t\leq T} |(nh_n)^{1/2}b_n(t|x) - b(t|x)| = o(1)$ ($b(t|x)$ being defined as in (2.12)). Hence, the result follows from Theorem 4.1 in Billingsley (1968).

3.2 Applications

In this section we will discuss two applications of the weak convergence result obtained in Theorem 3.2 : the validity of the bootstrap procedure and the construction of confidence bands for $F(\cdot|x)$.

3.2.1 Consistency of the bootstrap approximation

As a first application we have the strong uniform consistency of $P^*((nh_n)^{1/2} (F^*_{hg}(t|x) - F_g(t|x)) \leq y)$ for $P((nh_n)^{1/2}(F_h(t|x) - F(t|x)) \leq y)$. The proof is obvious, since Theorems 2.4 and 3.2 imply that the original and the bootstrapped process have the same limiting process. An alternative proof for this result, which does not make use of the weak convergence of the bootstrapped process, can be found in Theorem 4 in Van Keilegom and Veraverbeke (1997b). The bootstrap approximation is an alternative for the Gaussian approximation obtained in Corollary 2.1. In Section 3.3 we will compare both approximations through some simulations.

Theorem 3.3
Assume $(C1)$, $(C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Then, for* $t \leq T$, *as* $n \to \infty$,

$$\sup_{y\in I\!R} \Big| P^*((nh_n)^{1/2}(F^*_{hg}(t|x) - F_g(t|x)) \leq y) - P((nh_n)^{1/2}(F_h(t|x) - F(t|x)) \leq y)\Big|$$
$$= o(1) \qquad a.s.$$

3.2.2 Confidence bands for the conditional distribution

The construction of confidence bands has been studied in many papers. For i.i.d. observations, not subject to censoring, we refer to Bickel and Freedman (1981) and Shorack (1982). In the case of censored i.i.d. data, Hall and Wellner (1980) constructed a confidence band for the distribution function, starting from the weak convergence of the Kaplan-Meier process, while bootstrapped confidence bands were constructed by Akritas (1986). Below, we will construct bootstrapped confidence bands for the conditional distribution $F(\cdot|x)$, using a similar approach as in Akritas (1986). Recently, Li and Datta (1997) also obtained bootstrapped confidence bands for $F(\cdot|x)$ by using a martingale approach.

Theorem 3.4

Assume $(C1), (C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\frac{ng_n^5}{\log n} \to \infty$ *and* $\frac{ng_n^5}{\log n}\frac{h_n}{g_n} = O(1)$. *Let* $C_h(t|x) = \|K\|_2^2 \int_0^t (1-H_h(s|x))^{-2} dH_h^u(s|x)$, $D_h(t|x) = (nh_n)^{-1/2}(1-F_h(t|x))(1+C_h(t|x))$ *and for each* $0<\alpha<1$ *and for the given sample, let* $c_{\alpha hg}$ *be such that, as* $n \to \infty$,

$$P^*\left(\sup_{0\le t\le T}\left\{|F_{hg}^*(t|x)-F_g(t|x)|D_h^{-1}(t|x)\right\} \le c_{\alpha hg}\right) \to 1-\alpha \qquad \text{a.s.}$$

Then, as $n \to \infty$,

$$P(F_h(t|x) - c_{\alpha hg}D_h(t|x) \le F(t|x) \le F_h(t|x) + c_{\alpha hg}D_h(t|x)$$
$$\text{for all } 0 \le t \le T) \to 1-\alpha.$$

Proof. Since the Gaussian process $\tilde{W}(\cdot|x)$ (see (2.11) and (2.12) for explicit bias and covariance expressions) is equal in law to $B(K(\cdot|x))(1-F(\cdot|x))(1+C(\cdot|x))+b(\cdot|x)$, where $\{B(t); 0\le t\le 1\}$ is a Brownian bridge and $C(t|x)$ and $K(t|x)(0\le t\le T)$ are as in (2.17) and (2.18), it follows from Theorem 3.2 that

$$(nh_n)^{1/2}(F_{hg}^*(\cdot|x)-F_g(\cdot|x))(1-F(\cdot|x))^{-1}(1+C(\cdot|x))^{-1} \to B(K(\cdot|x)) + b'(\cdot|x)$$

in $D[0,T]$ a.s., where

$$b'(t|x) = \frac{1}{2}\int_0^t \left\{\frac{\ddot{H}(s|x)dH^u(s|x)}{(1-H(s|x))^2} + \frac{d\ddot{H}^u(s|x)}{1-H(s|x)}\right\}(1+C(t|x))^{-1}\mu_2^K C^{5/2}. \qquad (3.4)$$

From Lemma 2.7 and Theorem 2.1 together with Slutsky's Theorem, it now follows that

$$(F^*_{hg}(\cdot|x) - F_g(\cdot|x))D_h^{-1}(\cdot|x) \to B(K(\cdot|x)) + b'(\cdot|x) \qquad \text{in } D[0,T] \text{ a.s.}$$

Analogously, $(F_h(\cdot|x) - F(\cdot|x))D_h^{-1}(\cdot|x) \to B(K(\cdot|x)) + b'(\cdot|x)$ in $D[0,T]$ from Theorem 2.4. Let

$$g(c|x) = P\left(\sup_{0\le t\le T} |B(K(t|x)) + b'(t|x)| \le c\right)$$

$$g_h(c|x) = P\left(\sup_{0\le t\le T} \{|F_h(t|x) - F(t|x)|D_h^{-1}(t|x)\} \le c\right)$$

$$g^*_{hg}(c|x) = P^*\left(\sup_{0\le t\le T} \{|F^*_{hg}(t|x) - F_g(t|x)|D_h^{-1}(t|x)\} \le c\right).$$

Since $\sup_{0\le t\le T} |\cdot|$ is a continuous functional, we have that, as $n \to \infty$, $g_h(c|x) \to g(c|x)$ and $g^*_{hg}(c|x) \to g(c|x)$ a.s. for all c. Lemma 3.5 below entails that $g(\cdot|x)$ is continuous, and hence, $\sup_{c\ge 0} |g_h(c|x) - g(c|x)| \to 0$ and $\sup_{c\ge 0} |g^*_{hg}(c|x) - g(c|x)| \to 0$ a.s. by Pólya's Theorem (see e.g. Serfling (1980)). Hence, $g^*_{hg}(c_{\alpha hg}|x) - g(c_{\alpha hg}|x) \to 0$ a.s. From the definition of $c_{\alpha hg}$ we have that $g^*_{hg}(c_{\alpha hg}|x) \to 1-\alpha$ a.s. and hence $g(c_{\alpha hg}|x) \to 1-\alpha$ a.s. Finally, $g_h(c_{\alpha hg}|x) \to 1-\alpha$ a.s., because $g_h(c_{\alpha hg}|x) - g(c_{\alpha hg}|x) \to 0$. This finishes the proof.

We end this section with proving a small technical lemma, which was needed in the proof above.

Lemma 3.5
Let $\{B(t); 0 \le t \le 1\}$ be a Brownian bridge, let $C(t|x)$ and $K(t|x)$ $(0 \le t \le T)$ be as in (2.17) and (2.18) and let $b'(t|x)$ be as in (3.4). Then, $\sup_{0\le t\le T} |B(K(t|x)) + b'(t|x)|$ has a continuous distribution.

Proof. Let $Y(t|x) = B(K(t|x)) + b'(t|x)$. Since $Y(\cdot|x)$ has continuous sample paths, $\sup_{0\le t\le T} |Y(t|x)|$ is actually a maximum. Let $\varepsilon > 0, y_1 \le y_2$ and $|y_2 - y_1| < \delta$ $(\delta > 0$

to be chosen later). Then,

$$
\begin{aligned}
&P\left(y_1 \le \max_{0\le t\le T} |Y(t|x)| \le y_2\right)\\
&= P\left(-y_2 \le Y(t|x) \le y_2 \text{ for all } t \text{ and } Y(t|x) \ge y_1 \text{ or } -Y(t|x) \ge y_1 \text{ for some } t\right)\\
&= P\left(\left\{y_1 \le \max_{0\le t\le T} Y(t|x) \le y_2 \text{ and } \max_{0\le t\le T}(-Y(t|x)) \le y_2\right\}\right.\\
&\qquad\qquad \left.\text{or } \left\{\max_{0\le t\le T} Y(t|x) \le y_2 \text{ and } y_1 \le \max_{0\le t\le T}(-Y(t|x)) \le y_2\right\}\right)\\
&\le P\left(y_1 - m_x \le \max_{0\le t\le T}(B(K(t|x)) + b'(t|x) - m_x) \le y_2 - m_x\right) \qquad (3.5)\\
&\quad + P\left(y_1 + M_x \le \max_{0\le t\le T}(-B(K(t|x)) - b'(t|x) + M_x) \le y_2 + M_x\right),
\end{aligned}
$$

where $\min_{0\le t\le T} b'(t|x) \ge m_x$ and $\max_{0\le t\le T} b'(t|x) \le M_x$. From Theorem 1 in Tsirel'son (1975) it follows that $\max_{0\le t\le T}(B(K(t|x)) + b'(t|x) - m_x)$ and $\max_{0\le t\le T}$ $(-B(K(t|x)) - b'(t|x) + M_x)$ are continuous, and hence (3.5) is smaller than ε for proper choice of $\delta > 0$.

3.3 Simulations

The finite sample performance of the normal approximation (obtained in Corollary 2.1) and the bootstrap approximation (obtained in Theorem 3.4) of the distribution of $(nh_n)^{1/2}(F_h(t|x) - F(t|x))$ will now be compared through some simulations.

Take fixed and equidistant design points $x_i = i/n$ $(i = 1, \ldots, n)$. Assume that the survival times Y_i $(i = 1, \ldots, n)$ are independent random variables and that

$$Y_i \sim \text{Weibull}(a_1 + a_2 x_i, b),$$

i.e., $F(t|x_i) = 1 - \exp(-(a_1 + a_2 x_i)t^b)$ $(t \ge 0)$ for some constants a_1, a_2 and b such that $a_1 > 0 \vee (-a_2)$ and $b > 0$. Similarly, the censoring times C_i $(i = 1, \ldots, n)$ are

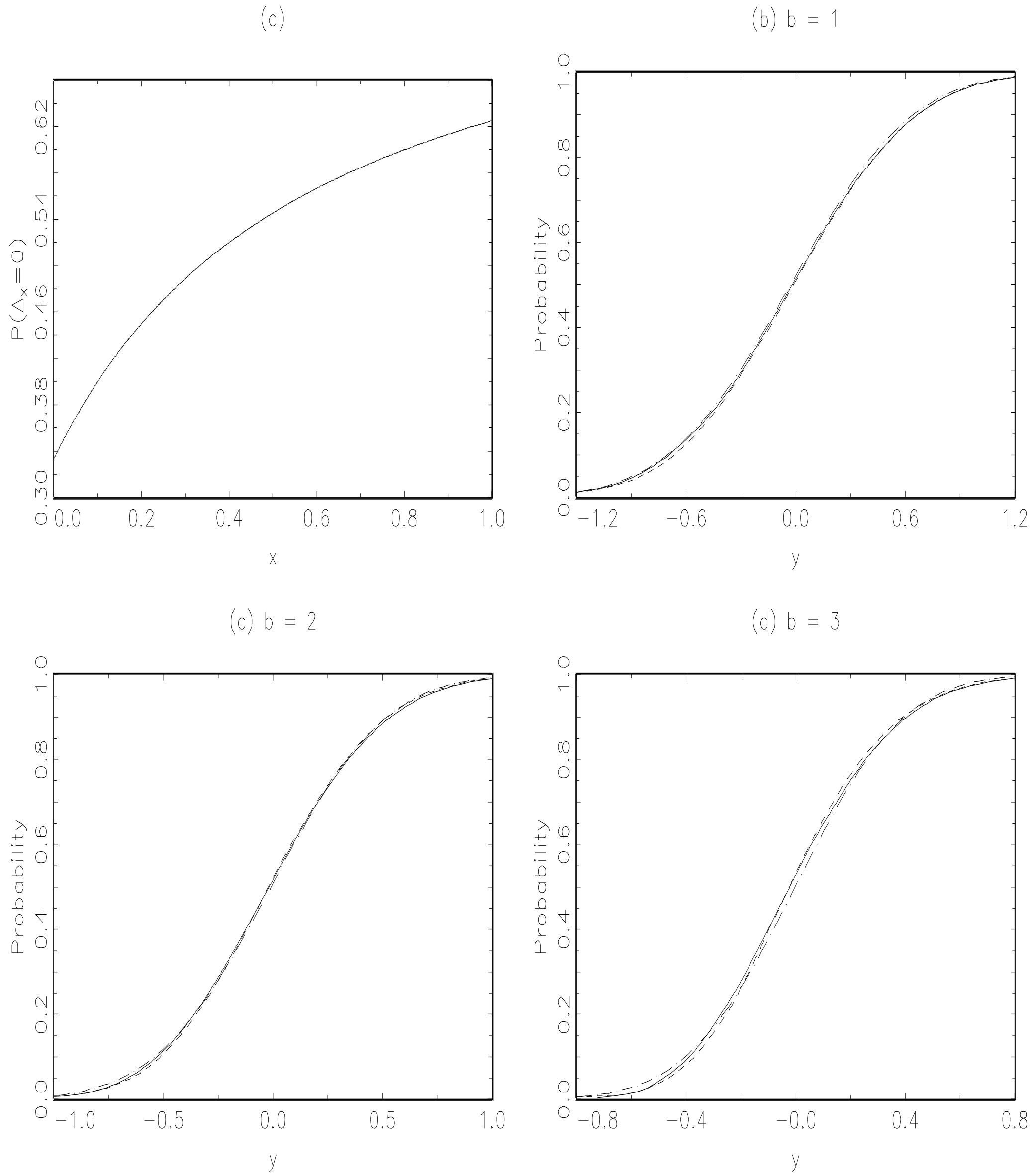

Figure 3.1: *(a) Proportion of censoring when* $a_1 = 1, a_2 = 0.5, c_1 = 0.5$ *and* $c_2 = 2$*. (b)-(d) Normal approximation (curve with dots and dashes) and bootstrap approximation (dashed curve) of* $P((nh_n)^{1/2}(F_h(t|x) - F(t|x)) \le y)$ *(solid curve) for* $a_1 = 1, a_2 = 0.5, c_1 = 0.5, c_2 = 2$ *and* $t = 0.5$*.*

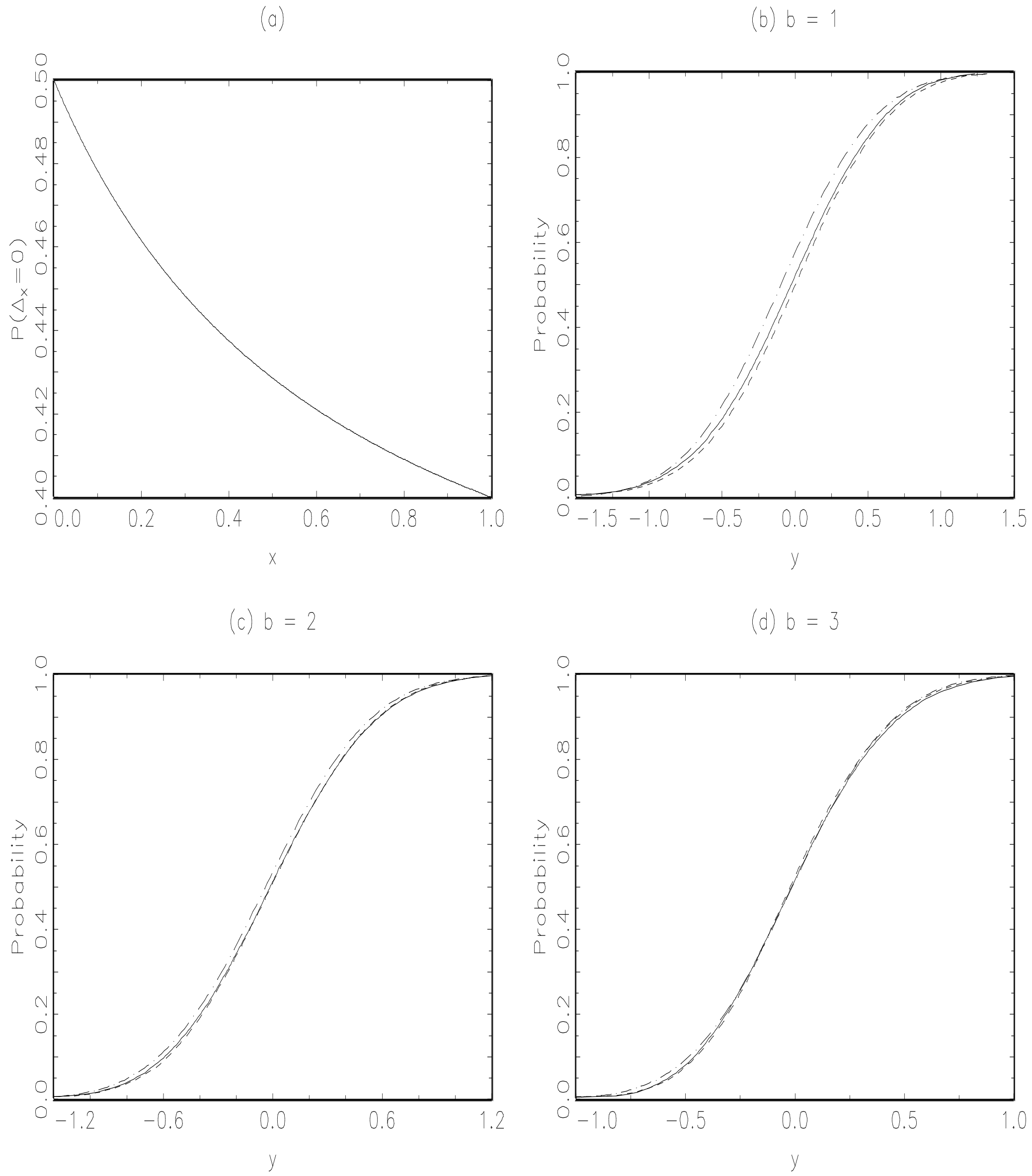

Figure 3.2: *(a) Proportion of censoring when $a_1 = 1, a_2 = 2, c_1 = 1$ and $c_2 = 1$. (b)-(d) Normal approximation (curve with dots and dashes) and bootstrap approximation (dashed curve) of $P((nh_n)^{1/2}(F_h(t|x) - F(t|x)) \leq y)$ (solid curve) for $a_1 = 1, a_2 = 2, c_1 = 1, c_2 = 1$ and $t = 0.5$.*

independent and for each i, $C_i \sim$ Weibull $(c_1 + c_2 x_i, b)$ for some $c_1 > 0 \vee (-c_2)$. As usual, we assume that for each i, the survival time Y_i and the censoring time C_i are independent. Hence, it is readily verified that

$$P(\Delta_x = 0) = \frac{c_1 + c_2 x}{a_1 + a_2 x + c_1 + c_2 x}.$$

We studied two different combinations of the parameters a_1, a_2, c_1 and c_2. Figures 3.1(a) and 3.2(a) show the curve of the function $P(\Delta_x = 0)$ for these two choices of the parameters.

We carried out the simulations for samples of size $n = 50$, for a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$ and for $x = 0.5$ and $t = 0.5$. The normal approximation and the true distribution were obtained using 10 000 samples. For the bootstrap approximation 200 samples were taken and for each sample, 200 resamples were drawn.

In Remark 2.1 we showed that the optimal choice of the bandwidth h_n is $h_{n,\text{opt}} = Cn^{-1/5}$ for some specific $C > 0$, while for calculating the optimal choice of the pilot bandwidth g_n, no methods are available yet in this context. In our simulations, we worked with this optimal bandwidth h_n and with $g_n = 5h_n$. However, in many situations the optimal choice h_n exceeded 0.5. Since we only have observations in a window of width 0.5 around $x = 0.5$, the actual number of observations that is involved in that case is not $2nh_n$, but n. Therefore, we truncated the bandwidth in these situations at $h_n = 0.5$.

Figures 3.1(b)-(d) and 3.2(b)-(d) show that both the normal and the bootstrap approximation are very close to the true distribution, in spite of the fact that the censoring is quite heavy (see Figure 3.1(a) and 3.2(a)). However, when the considered time point t gets much larger (or much smaller), both the quality of the normal and the bootstrap approximation diminish. In Figure 3.3 we have chosen the same set of parameters as in Figure 3.2, except that t is now equal to 1. That this combination of parameters leads to a worse approximation, can be explained in the following way. Since $F(t|x) = 1 - \exp(-(a_1 + a_2 x)t^b) = 0.86$ and the (asymptotic) variance of $F_h(t|x)$ equals 0.24, the estimate $F_h(t|x)$ equals 1 for some of the samples (this can also be seen from the fact that the censoring is quite heavy here

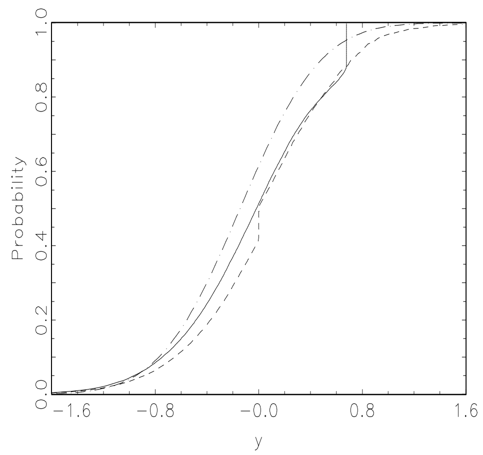

Figure 3.3: *Normal approximation (curve with dots and dashes) and bootstrap approximation (dashed curve) of* $P((nh_n)^{1/2}(F_h(t|x) - F(t|x)) \leq y)$ *(solid curve) for* $a_1 = 1, a_2 = 2, b = 1, c_1 = 1, c_2 = 1$ *and* $t = 1$.

$(P(\Delta_x = 0) = 0.43))$. Because of this, the distribution of $(nh_n)^{1/2}(F_h(t|x) - F(t|x))$ jumps at $(nh_n)^{1/2}(1 - 0.86) = 0.68$. For a similar reason, the bootstrap distribution makes a jump at $y = 0$: indeed, since both $F_g(t|x)$ and $F^*_{hg}(t|x)$ equal 1 for some of the samples, the difference $F^*_{hg}(t|x) - F_g(t|x)$ jumps at zero.

3.4 Appendix : Some results on the bootstrapped Stone estimator

We will now prove a number of technical results concerning the bootstrapped Stone estimator $H^*_{hg}(t|x)$ (defined in (1.26)), which were needed in the proofs of the main results in this chapter. Analogous results hold for the bootstrapped estimator $H^{*u}_{hg}(t|x)$ of the uncensored observations.

Lemma 3.6
Assume $(C1), (C2')$, $H(t|x)$ *satisfies* $(C5) - (C7)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$. *Then, as* $n \to \infty$,

$$\sup_{0\le t\le T} |bias\ H^*_{hg}(t|x) - bias\ H_h(t|x)| = o((nh_n)^{-1/2}) \qquad \text{a.s.},$$

where

$$bias\ H^*_{hg}(t|x) = E^* H^*_{hg}(t|x) - H_g(t|x)$$

$$bias\ H_h(t|x) = EH_h(t|x) - H(t|x).$$

Proof. Partitioning $[0, T]$ into $k_n = O(ng_n(\log n)^{-1})$ subintervals $[t_i, t_{i+1}]$ of length $O((ng_n)^{-1}\log n)$, we have that

$$\begin{aligned}
&\sup_{0\le t\le T} |\text{bias } H^*_{hg}(t|x) - \text{bias } H_h(t|x)| \\
&\le \max_{1\le i\le k_n} |\text{bias } H^*_{hg}(t_i|x) - \text{bias } H_h(t_i|x)| \\
&\quad + \max_{1\le i\le k_n} |H_g(t_i|x) - H_g(t_{i-1}|x) + EH_h(t_i|x) - EH_h(t_{i-1}|x)| \\
&\le \max_{1\le i\le k_n} |\text{bias } H^*_{hg}(t_i|x) - \text{bias } H_h(t_i|x)| \\
&\quad + \max_{1\le i\le k_n} |H_g(t_i|x) - H_g(t_{i-1}|x) - H(t_i|x) + H(t_{i-1}|x)| \\
&\quad + \max_{1\le i\le k_n} |EH_h(t_i|x) - EH_h(t_{i-1}|x) - H(t_i|x) + H(t_{i-1}|x)| \\
&\quad + 2\max_{1\le i\le k_n} |H(t_i|x) - H(t_{i-1}|x)| \\
&= T_1 + T_2 + T_3 + T_4.
\end{aligned}$$

For T_1 we have, as in the proof of Lemma 10 in Aerts, Janssen and Veraverbeke (1994b) that

$$(nh_n)^{1/2} T_1 = \frac{1}{2}\mu_2^K C^{5/2} \max_{1\le i\le k_n} |\ddot{H}_g(t_i|x) - \ddot{H}(t_i|x)| + o(1),$$

where

$$\ddot{H}_g(y|x) = \sum_{j=1}^{n} w^{(2)}_{nj}(x; g_n) I(T_j \le y)$$

$$w^{(2)}_{nj}(x; g_n) = \frac{1}{g_n^3} \int_{x_{j-1}}^{x_j} K''\left(\frac{x-z}{g_n}\right) dz.$$

Now,

$$\max_{1\le i\le k_n} |\ddot{H}_g(t_i|x) - \ddot{H}(t_i|x)|$$

$$\le \max_{1\le i\le k_n} |\ddot{H}_g(t_i|x) - E\ddot{H}_g(t_i|x)| + \max_{1\le i\le k_n} |E\ddot{H}_g(t_i|x) - \ddot{H}(t_i|x)|.$$

The first term is $o(1)$, using Bernstein's theorem. As to the second term, we write

$$\begin{aligned}
&\max_{1\le i\le k_n} |E\ddot{H}_g(t_i|x) - \ddot{H}(t_i|x)| \\
&\le \max_{1\le i\le k_n} \left| \sum_{j=1}^{n} w_{nj}^{(2)}(x;g_n) H(t_i|x_j) - \frac{1}{g_n^3} \int_0^{x_n} K''\left(\frac{x-z}{g_n}\right) H(t_i|z)\,dz \right| \\
&\quad + \max_{1\le i\le k_n} \left| \frac{1}{g_n^3} \int_0^{x_n} K''\left(\frac{x-z}{g_n}\right) H(t_i|z) dz - \ddot{H}(t_i|x) \right| \\
&\le \max_{1\le i\le k_n} \left\{ \sum_{j=1}^{n} \frac{1}{g_n^3} \int_{x_{j-1}}^{x_j} \left| K''\left(\frac{x-z}{g_n}\right) \right| |H(t_i|x_j) - H(t_i|z)|\; dz \right\} \\
&\quad + \max_{1\le i\le k_n} \left| \int_{-L_0}^{L_0} K(u) \left[\ddot{H}(t_i|x-h_n u) - \ddot{H}(t_i|x) \right] du \right| \qquad \text{(for } n \text{ large)},
\end{aligned}$$

which is $o(1)$. Using Lemma 2.11 with $a_n = c(ng_n)^{-1}\log n$ for some $c > 0$, it follows that T_2 and T_3 are $o((nh_n)^{-1/2})$ a.s. Also, $T_4 = o((nh_n)^{-1/2})$, using the Lipschitz continuity of $H(\cdot|x)$.

Lemma 3.7

Assume $(C1), (C2')$, $H(t|x)$ *satisfies* $(C5) - (C7)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$. *Then, as* $n \to \infty$,

$$\sup_{0\le t\le T} |H_{hg}^*(t|x) - H_g(t|x)| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2}) \qquad \text{a.s.}$$

Proof. We write

$$\begin{aligned}
&\left|H_{hg}^*(t|x) - H_g(t|x)\right| \\
&\le \left|H_{hg}^*(t|x) - E^* H_{hg}^*(t|x)\right| + \left|\text{bias } H_{hg}^*(t|x) - \text{bias } H_h(t|x)\right| + \left|\text{bias } H_h(t|x)\right|.
\end{aligned}$$

The second term is $o((nh_n)^{-1/2})$ a.s. by Lemma 3.6. The last term is $O(h_n^2 + n^{-1})$ using Lemma 2.6(b). To obtain the appropriate order bound for the first term, we apply Bernstein's inequality and the usual argument for replacing the supremum by a maximum : partitioning the interval $[0, T]$ in $O((nh_n)^{1/2}(\log n)^{-1/2})$ subintervals $[t_i, t_{i+1}]$ such that $H_g(t_{i+1}|x) - H_g(t_i|x) = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. This is possible since the jump sizes of $H_g(\cdot|x)$ are of order $O((ng_n)^{-1})$ and since $ng_n^5/\log n \to \infty$.

Lemma 3.8
Assume $(C1), (C2)$, $H(t|x)$ *satisfies* $(C3), (C6)$ *and* $(C7)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Then, for any* $c > 0$, *as* $n \to \infty$,

$$\sup_{0 \le s,t \le T; |t-s| \le c(nh_n)^{-1/2}(\log n)^{1/2}} |E^*H_{hg}^*(t|x) - E^*H_{hg}^*(s|x) - H_g(t|x) + H_g(s|x)|$$
$$= O((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.}$$

Proof. We write

$$\begin{aligned}
&E^*H_{hg}^*(t|x) - E^*H_{hg}^*(s|x) - H_g(t|x) + H_g(s|x)\\
&= \sum_{i=1}^n w_{ni}(x; h_n)[H_g(t|x_i) - H_g(s|x_i) - H(t|x_i) + H(s|x_i)]\\
&\quad + [EH_h(t|x) - EH_h(s|x) - H(t|x) + H(s|x)]\\
&\quad + [H(t|x) - H(s|x) - H_g(t|x) + H_g(s|x)].
\end{aligned}$$

The second and the third term have the appropriate order by Lemma 2.11. For the first term the proof is completely analogous to that of Lemma 2.11.

Lemma 3.9
Assume $(C1), (C2)$, $H(t|x)$ *satisfies* $(C3)$ *and* $(C5)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Then, as* $n \to \infty$,

$$\sup_{0 \le t \le T} |H_g(t|x) - H(t|x)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \qquad \text{a.s.}$$

Proof. It follows from Lemma 2.8 that

$$\begin{aligned} & P\left(\sup_{0\le t\le T}|H_g(t|x)-H(t|x)|>c(nh_n)^{-1/2}(\log n)^{1/2}\right) \\ & \le d_0c(nh_n)^{-1/2}(\log n)^{1/2}ng_n\exp\left(-\frac{1}{4}d_1ng_nc^2(nh_n)^{-1}\log n\right) \\ & \le Kng_nn^{-d_1c^2/4} \qquad \text{(for some } K>0 \text{ and for } n \text{ large),} \end{aligned}$$

provided $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n}=O(1)$. Now apply Borel-Cantelli after proper choice of c.

Chapter 4

The conditional quantile function in regression with censored data

The purpose of this chapter is to study the p-th quantile $F_h^{-1}(p|x)$ $(0 < p < 1)$ of the Beran estimator $F_h(\cdot|x)$ (defined in (1.19)), which estimates the p-th quantile of the conditional distribution $F(\cdot|x)$. We start by showing its consistency in Section 4.1, which will be obtained via an exponential probability bound result. Next, an a.s. asymptotic representation will be derived, from which the weak convergence of $F_h^{-1}(\cdot|x)$ will follow (Section 4.2). In Section 4.3 an asymptotic representation and the weak convergence of the bootstrapped quantile estimator $F_{hg}^{*-1}(\cdot|x)$ will be established. Use will be made of the corresponding results on the Beran estimator, studied in Chapters 2 and 3. Section 4.4 deals with the validity of the bootstrap procedure (in the sense that it provides us with an alternative way to approximate $F^{-1}(p|x)$) and the construction of bootstrap confidence bands for the quantile function. The finite sample performance of the Gaussian and the bootstrap approximation of the quantile function will be studied in Section 4.5 by means of a simulation study. Finally, some useful results on the bootstrapped Stone estimator and its quantile estimator, which will be needed in the proofs of the main results, will be proved in Section 4.6.

The consistency of the quantile estimator was shown in Van Keilegom and Veraverbeke (1996), while the weak convergence results and the construction of confidence bands are part of Van Keilegom and Veraverbeke (1997a). The remaining results can be found in Van Keilegom and Veraverbeke (1998).

4.1 Consistency

Our first result is an exponential upper bound for $P(\sup_{0<p\le p_0}|F_h^{-1}(p|x) - F^{-1}(p|x)| > \varepsilon)$. From this result a rate for the uniform strong consistency of the quantile estimator will be easily obtained. Analogous results to this last property were obtained by Aerts, Janssen and Veraverbeke (1994a) in a heteroscedastic regression model with complete observations and by Cheng (1984) and Lo and Singh (1986) for censored data (in the absence of covariates). In a random design model with censored data, Beran (1981) proved the strong consistency of his quantile estimator (without rate of convergence), while Dabrowska (1992b) obtained the a.s. order $O(h_n)$, which is slower than the rate of convergence obtained in the next result, when the bandwidth h_n equals the optimal bandwidth $h_n = Cn^{-1/5}$ for some $C > 0$. An alternative exponential inequality for the one below, which requires however condition $(C4)$ on $H(t|x)$ and $H^u(t|x)$, can be found in Theorem 2 in Van Keilegom and Veraverbeke (1996). The coefficient in front of the exponential factor is (asymptotically) smaller in that result than in the result below, but this has no implication on the rate of convergence of the complete expression.

Theorem 4.1
Assume $(C1), (C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C3)$ *and* $(C5)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $1 - H(T|x) > \delta > 0$, $h_n \to 0$, $nh_n \to \infty$. *Let* $0 < \varepsilon_0 < p_0 < 1$ *be such that* $F^{-1}(p_0|x) < T$ *and* $\inf_{\varepsilon_0 \le p \le p_0} f(F^{-1}(p|x)|x) = \lambda > 0$ *(where* $f(\cdot|x) = F'(\cdot|x)$*).*

(a) *For* $\varepsilon > 0$ *such that* $F^{-1}(\varepsilon_0|x) - \varepsilon \ge 0$, $F^{-1}(p_0|x) + \varepsilon \le T$ *and*

$$\inf_{F^{-1}(\varepsilon_0|x)-\varepsilon \le y \le F^{-1}(p_0|x)+\varepsilon} f(y|x) \ge \frac{\lambda}{2},$$

for n *sufficiently large and*

$$\varepsilon \ge \frac{24}{\delta^2\lambda} \max\left(\frac{\sqrt{6}\|K\|_2}{(nh_n)^{1/2}}, 2(\|\dot{H}\| \vee \|\dot{H}^u\|)\overline{\Delta}_n + 2\mu_2^K(\|\ddot{H}\| \vee \|\ddot{H}^u\|)h_n^2\right)$$

we have

$$P\left(\sup_{\varepsilon_0\le p\le p_0}|F_h^{-1}(p|x)-F^{-1}(p|x)|>\varepsilon\right)\le C_1\delta^2\lambda nh_n\varepsilon\exp\left(-C_2\delta^4\lambda^2nh_n\varepsilon^2\right)$$

where $C_1=\dfrac{e^2\lambda}{\|K\|_2^2}$ *and* $C_2=\dfrac{\lambda^2}{1728\|K\|_2^2}$.

(b) *If* $\dfrac{nh_n^5}{\log n}=O(1)$*, then, as* $n\to\infty$*,*

$$\sup_{\varepsilon_0\le p\le p_0}|F_h^{-1}(p|x)-F^{-1}(p|x)|=O((nh_n)^{-1/2}(\log n)^{1/2})\qquad a.s.$$

Proof. (a) Write

$$\begin{aligned}&P\left(\sup_{\varepsilon_0\le p\le p_0}|F_h^{-1}(p|x)-F^{-1}(p|x)|>\varepsilon\right)\\&\le P\left(F_h^{-1}(p|x)>F^{-1}(p|x)+\varepsilon\text{ for some }\varepsilon_0\le p\le p_0\right)\\&\quad+P\left(F_h^{-1}(p|x)<F^{-1}(p|x)-\varepsilon\text{ for some }\varepsilon_0\le p\le p_0\right).\end{aligned}$$

Since

$$\begin{aligned}&P\left(F_h^{-1}(p|x)>F^{-1}(p|x)+\varepsilon\text{ for some }\varepsilon_0\le p\le p_0\right)\\&\le P\left(p>F_h(F^{-1}(p|x)+\varepsilon|x)\text{ for some }\varepsilon_0\le p\le p_0\right)\\&\le P\left(\sup_{0\le t\le T}|F_h(t|x)-F(t|x)|>\inf_{\varepsilon_0\le p\le p_0}(F(F^{-1}(p|x)+\varepsilon|x)-p)\right)\end{aligned}\tag{4.1}$$

and since

$$\inf_{\varepsilon_0\le p\le p_0}(F(F^{-1}(p|x)+\varepsilon|x)-p)\ge\inf_{F^{-1}(\varepsilon_0|x)\le y\le F^{-1}(p_0|x)+\varepsilon}f(y|x)\varepsilon\ge\frac{1}{2}\lambda\varepsilon>0,$$

we can apply Theorem 2.1(a) to equation (4.1). Analogously, we have that

$$\begin{aligned}&P\left(F_h^{-1}(p|x)<F^{-1}(p|x)-\varepsilon\text{ for some }\varepsilon_0\le p\le p_0\right)\\&\le P\left(p\le F_h(F^{-1}(p|x)-\varepsilon|x)\text{ for some }\varepsilon_0\le p\le p_0\right)\\&\le P\left(\sup_{0\le t\le T}|F_h(t|x)-F(t|x)|\ge\inf_{\varepsilon_0\le p\le p_0}(p-F(F^{-1}(p|x)-\varepsilon|x))\right).\end{aligned}$$

Again, we can apply Theorem 2.1(a), because

$$\inf_{\varepsilon_0 \le p \le p_0} (p - F(F^{-1}(p|x) - \varepsilon|x)) \ge \inf_{F^{-1}(\varepsilon_0|x) - \varepsilon \le y \le F^{-1}(p_0|x)} f(y|x)\varepsilon \ge \frac{1}{2}\lambda\varepsilon > 0.$$

This proves part (a).
(b) This follows from the Borel-Cantelli lemma.

Remark 4.1
The condition

$$\inf_{\varepsilon_0 \le p \le p_0} f(F^{-1}(p|x)|x) > 0 \tag{4.2}$$

needed in Theorem 4.1 also appears in Lo and Singh (1986) and Dabrowska (1992b). One can easily show however that the above exponential bound holds uniformly over all $0 < p \le p_0$ (instead of $\varepsilon_0 \le p \le p_0$) when condition (4.2) is replaced by

$$\inf_{0 < p \le p_0} f(F^{-1}(p|x)|x) > 0, \tag{4.3}$$

which is used by Cheng (1984). This slightly stronger condition is satisfied by distributions like the exponential and Pareto and, for certain values of the parameters, also by the log-logistic, Burr, Weibull and Gamma distribution. Examples of distributions that do not satisfy condition (4.3) include the Rayleigh and the log-normal distribution.

In many of the results in this chapter the condition (4.2) will be assumed. In all of these results (except in Theorems 4.3 and 4.5) the assumption can however be replaced by the somewhat more restrictive condition (4.3), leading to a slightly stronger result.

4.2 Asymptotic representation and weak convergence of the quantile process

The main result of this section concerns the weak convergence of the quantile process $F_h^{-1}(\cdot|x)$ to a Gaussian process. Since $F_h(t|x)$ (and hence also $F_h^{-1}(p|x)$) has a complicated structure, we start with deriving an easier expression for $F_h^{-1}(p|x)$, from which the weak convergence will be readily obtained.

For censored data without covariates, Cheng (1984) proved that $F_n^{-1}(p) - F^{-1}(p) - (p - F_n(F^{-1}(p))/f(F^{-1}(p))$ is uniformly of the a.s. order $O(n^{-3/4}(\log n)^{3/4})$, where F_n is the Kaplan-Meier estimator. In Lo and Singh (1986), it is shown that $F_n^{-1}(p) - F^{-1}(p)$ can be decomposed in an average of i.i.d. terms and a remainder term of order $O(n^{-3/4}(\log n)^{3/4})$. In the next theorem we extend these results to regression quantiles based on censored data. As before, we assume a fixed design. Dabrowska (1992b) generalized the result of Cheng (1984) to a random design regression model with censored data and obtained the a.s. order $O((nh_n)^{-3/4}(\log n)^{3/4} + h_n^2)$. The following representation has the same leading term as in Dabrowska (1992b), but does not contain the remainder term of order $O(h_n^2)$, which dominates the term of order $O((nh_n)^{-3/4}(\log n)^{3/4})$ for a wide range of bandwidth sequences.

Theorem 4.2

Assume $(C1), (C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5) - (C7)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}, h_n \to 0, \dfrac{\log n}{nh_n} \to 0, \dfrac{nh_n^5}{\log n} = O(1)$. *Let* $0 < \varepsilon_0 < p_0 < 1$ *be such that* $F^{-1}(p_0|x) < T$ *and* $\inf_{\varepsilon_0 \le p \le p_0} f(F^{-1}(p|x)|x) = \lambda > 0$, *then, for* $\varepsilon_0 \le p \le p_0$ *:*

$$
\begin{aligned}
&F_h^{-1}(p|x) - F^{-1}(p|x) \\
&= -(F_h(F^{-1}(p|x)|x) - p)/f(F^{-1}(p|x)|x) + r_{n1}(x, p) \\
&= -\sum_{i=1}^n w_{ni}(x; h_n)\xi(T_i, \Delta_i, F^{-1}(p|x)|x)/f(F^{-1}(p|x)|x) + r_{n2}(x, p),
\end{aligned}
$$

where ξ *is as in Theorem 2.3 and where*

$$
\begin{aligned}
\sup_{\varepsilon_0 \le p \le p_0} |r_{n1}(x, p)| &= O((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.} \\
\sup_{\varepsilon_0 \le p \le p_0} |r_{n2}(x, p)| &= O((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.}
\end{aligned}
$$

as $n \to \infty$.

Proof. Using Theorem 2.2 we have that

$$
\begin{aligned}
&\sup_{\varepsilon_0\le p\le p_0}|F_h(F_h^{-1}(p|x)|x)-p|\\
&\le \sup_{\varepsilon_0\le p\le p_0}|F_h(F_h^{-1}(p|x)|x)-F_h(F_h^{-1}(p|x)-|x)|\\
&\le \sup_{0\le s,t\le T;|t-s|\le c(nh_n)^{-1/2}(\log n)^{1/2}}|F_h(t|x)-F_h(s|x)-F(t|x)+F(s|x)|\\
&= O((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\sup_{\varepsilon_0\le p\le p_0}|F_h^{-1}(p|x)-F^{-1}(p|x)-(p-F_h(F^{-1}(p|x)|x))/f(F^{-1}(p|x)|x)|\\
&\le \frac{1}{\lambda}\sup_{\varepsilon_0\le p\le p_0}|f(F^{-1}(p|x)|x)(F_h^{-1}(p|x)-F^{-1}(p|x))\\
&\quad -(F_h(F_h^{-1}(p|x)|x)-F_h(F^{-1}(p|x)|x))|+O((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.}\\
&\le \frac{1}{\lambda}\sup_{\varepsilon_0\le p\le p_0}|f(F^{-1}(p|x)|x)(F_h^{-1}(p|x)-F^{-1}(p|x)) \qquad (4.4)\\
&\quad -(F(F_h^{-1}(p|x)|x)-F(F^{-1}(p|x)|x))|+O((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.,}
\end{aligned}
$$

where the last inequality follows from Theorems 2.2 and 4.1. By means of a Taylor expansion we can write (4.4) as

$$
\frac{1}{2\lambda}\sup_{\varepsilon_0\le p\le p_0}|f'(\theta_h(p|x)|x)|(F_h^{-1}(p|x)-F^{-1}(p|x))^2+O((nh_n)^{-3/4}(\log n)^{3/4}) \ \text{ a.s.,} \quad (4.5)
$$

where $\theta_h(p|x)$ lies between $F^{-1}(p|x)$ and $F_h^{-1}(p|x)$. From condition $(C6)$ it follows that $f'(\cdot|x)$ is bounded on $[0,T]$. Hence, a further application of Theorem 4.1 gives that (4.5) is $O((nh_n)^{-3/4}(\log n)^{3/4})$ a.s. The result now follows from Theorem 2.3.

The next theorem gives the weak convergence result for the quantile process $(nh_n)^{1/2}(F_h^{-1}(\cdot|x)-F^{-1}(\cdot|x))$. Since $F_h^{-1}(\cdot|x)$ is left continuous (in contrast to $F_h(\cdot|x)$, which is right continuous), we consider weak convergence in the space D_L of left continuous functions with right hand limits, endowed with the Skorokhod topology. Since the (a)-part does not cover the optimal bandwidth case $h_n = Cn^{-1/5}$ $(C > 0)$ (see below Corollary 4.1), we state this case separately in part (b). In the context of censored data (in the absence of covariates), a similar result was shown by Lo and Singh (1986). Doss and Gill (1992) showed that the weak convergence of any process $n^{1/2}(\xi_n - \xi)$, satisfying some regularity conditions, entails the weak convergence of the corresponding quantile process $n^{1/2}(\xi_n^{-1} - \xi^{-1})$. This result is however not applicable here, since we have to pre-multiply with $(nh_n)^{1/2}$ instead of $n^{1/2}$.

Theorem 4.3
Assume $(C1), (C2)$, $H(t|x)$ and $H^u(t|x)$ satisfy $(C5)-(C7)$ in $[0,T]$ with $T < \tau_{H(\cdot|x)}$. Let $0 < \varepsilon_0 < p_0 < 1$ be such that $F^{-1}(p_0|x) < T$ and $\inf_{\varepsilon_0 \le p \le p_0} f(F^{-1}(p|x)|x) > 0$.

(a) If $nh_n^5 \to 0$ and $\dfrac{(\log n)^3}{nh_n} \to 0$, then, as $n \to \infty$,

$$(nh_n)^{1/2}(F_h^{-1}(\cdot|x) - F^{-1}(\cdot|x)) \to -\frac{W(F^{-1}(\cdot|x)|x)}{f(F^{-1}(\cdot|x)|x)} \qquad \text{in } D_L[\varepsilon_0, p_0]$$

(b) If $h_n = Cn^{-1/5}$ for some $C > 0$, then, as $n \to \infty$,

$$(nh_n)^{1/2}(F_h^{-1}(\cdot|x) - F^{-1}(\cdot|x)) \to -\frac{\tilde{W}(F^{-1}(\cdot|x)|x)}{f(F^{-1}(\cdot|x)|x)} \qquad \text{in } D_L[\varepsilon_0, p_0],$$

where $W(\cdot|x)$ and $\tilde{W}(\cdot|x)$ are Gaussian processes with covariance function given by (2.11) and, for $\tilde{W}(\cdot|x)$, mean function given by (2.12).

Proof. Theorem 4.2 entails that

$$\begin{aligned}
&\sup_{\varepsilon_0 \le p \le p_0} |(nh_n)^{1/2}(F_h^{-1}(p|x) - F^{-1}(p|x)) \\
&\qquad -(nh_n)^{1/2}(F(F^{-1}(p|x)|x) - F_h(F^{-1}(p|x)|x))/f(F^{-1}(p|x)|x)| \\
&= O((nh_n)^{-1/4}(\log n)^{3/4}) \qquad \text{a.s.}
\end{aligned}$$

Apply Theorem 2.4 together with Theorem 4.1 in Billingsley (1968).

In the following result we state the asymptotic normality of the quantile estimator, which follows readily from the weak convergence, shown above. A direct proof (which does not make use of the weak convergence result) can be found in Van Keilegom and Veraverbeke (1998).

Corollary 4.1
Assume $(C1), (C2)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}$. *Let* $0 < p < 1$ *be such that* $F^{-1}(p|x) < T$ *and* $f(F^{-1}(p|x)|x) > 0$.

(a) If $nh_n^5 \to 0$ *and* $\dfrac{(\log n)^3}{nh_n} \to 0$, *then, as* $n \to \infty$,

$$(nh_n)^{1/2}(F_h^{-1}(p|x) - F^{-1}(p|x)) \xrightarrow{d} N(0; \sigma^2(p|x))$$

(b) If $h_n = Cn^{-1/5}$ *for some* $C > 0$, *then, as* $n \to \infty$,

$$(nh_n)^{1/2}(F_h^{-1}(p|x) - F^{-1}(p|x)) \xrightarrow{d} N(\beta(p|x); \sigma^2(p|x)),$$

where

$$\sigma^2(p|x) = \frac{\Gamma(F^{-1}(p|x), F^{-1}(p|x)|x)}{f^2(F^{-1}(p|x)|x)}, \qquad \beta(p|x) = -\frac{b(F^{-1}(p|x)|x)}{f(F^{-1}(p|x)|x)},$$

and $\Gamma(s, t|x)$ *and* $b(t|x)$ *are given by (2.11) and (2.12).*

From this result, it also follows that the optimal bandwidth for estimating the quantile function $F^{-1}(p|x)$, equals the optimal bandwidth for the Beran estimator given in (2.16), when t is replaced with $F^{-1}(p|x)$ (this is easily seen, since the density $f(F^{-1}(p|x)|x)$ cancels in the numerator and the denominator of formula (2.16)).

4.3 Asymptotic representation and weak convergence of the bootstrapped quantile process

In this section we will first show an a.s. asymptotic representation for the bootstrapped quantile estimator, which will then enable us to prove the weak convergence of the bootstrapped process. In the i.i.d. case with censored data, a similar representation has been derived by Lo and Singh (1986). It represents the estimator $F_{hg}^{*-1}(p|x)$ as a difference of two weighted sums (one depending on the resample and one depending on the original sample) plus a remainder term of lower order.

Theorem 4.4
Assume $(C1), (C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0, g_n \to 0, \dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Let* $0 < \varepsilon_0 < p_0 < 1$ *be such that* $F^{-1}(p_0|x) < T$ *and* $\inf_{\varepsilon_0 \le p \le p_0} f(F^{-1}(p|x)|x) = \lambda > 0$. *Then, for* $\varepsilon_0 \le p \le p_0$ *:*

$$\begin{aligned}
&F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x) \\
&= -\left(F_{hg}^*(F^{-1}(p|x)|x) - F_g(F^{-1}(p|x)|x)\right)/f(F^{-1}(p|x)|x) + r_{n1}^*(x,p) \\
&= -\sum_{i=1}^n w_{ni}(x;h_n)\xi(T_i^*, \Delta_i^*, F^{-1}(p|x)|x)/f(F^{-1}(p|x)|x) \\
&\quad + \sum_{i=1}^n w_{ni}(x;g_n)\xi(T_i, \Delta_i, F^{-1}(p|x)|x)/f(F^{-1}(p|x)|x) + r_{n2}^*(x,p),
\end{aligned}$$

where ξ *is as in Theorem 2.3 and where*

$$\sup_{\varepsilon_0 \le p \le p_0} |r_{n1}^*(x,p)| = O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.}$$

$$\sup_{\varepsilon_0 \le p \le p_0} |r_{n2}^*(x,p)| = O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.}$$

as $n \to \infty$.

Proof. To prove the result, we need two lemmas (Lemma 4.1 and Lemma 4.2), which are given in Section 4.6. First note that since Theorem 4.1 and Lemma 4.1 entail that

$$\sup_{\varepsilon_0 \le p \le p_0} |F_{hg}^{*-1}(p|x) - F^{-1}(p|x)| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2}) \qquad \text{a.s.,}$$

it follows from Lemma 4.2 that

$$\begin{aligned}
&\sup_{\varepsilon_0 \le p \le p_0} |F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x) \\
&\quad - (F_g(F^{-1}(p|x)|x) - F_{hg}^*(F^{-1}(p|x)|x))/f(F^{-1}(p|x)|x)| \\
&\le \frac{1}{\lambda} \sup_{\varepsilon_0 \le p \le p_0} |f(F^{-1}(p|x)|x)(F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x)) \\
&\quad - (F_g(F_{hg}^{*-1}(p|x)|x) - F_{hg}^*(F_{hg}^{*-1}(p|x)|x))| + O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}). \qquad (4.6)
\end{aligned}$$

Next, write

$$\begin{aligned}
&|F_{hg}^*(F_{hg}^{*-1}(p|x)|x) - F_g(F_g^{-1}(p|x)|x)| \\
&\le |F_{hg}^*(F_{hg}^{*-1}(p|x)|x) - p| + |F_g(F_g^{-1}(p|x)|x) - p| \\
&\le |F_{hg}^*(F_{hg}^{*-1}(p|x)|x) - F_{hg}^*(F_{hg}^{*-1}(p|x)-|x)| + |F_g(F_g^{-1}(p|x)|x) - F_g(F_g^{-1}(p|x)-|x)| \\
&\le |F_{hg}^*(F_{hg}^{*-1}(p|x)|x) - F_{hg}^*(F_{hg}^{*-1}(p|x)-|x) - F_g(F_{hg}^{*-1}(p|x)|x) + F_g(F_{hg}^{*-1}(p|x)-|x)| \\
&\quad + |F_g(F_{hg}^{*-1}(p|x)|x) - F_g(F_{hg}^{*-1}(p|x)-|x) - F(F_{hg}^{*-1}(p|x)|x) + F(F_{hg}^{*-1}(p|x)-|x)| \\
&\quad + |F_g(F_g^{-1}(p|x)|x) - F_g(F_g^{-1}(p|x)-|x) - F(F_g^{-1}(p|x)|x) + F(F_g^{-1}(p|x)-|x)|,
\end{aligned}$$

which is $O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4})$ a.s. (uniformly in $[\varepsilon_0, p_0]$) by Theorem 2.2(a) and Lemma 4.2. Hence, (4.6) is not larger than

$$\begin{aligned}
&\frac{1}{\lambda} \sup_{\varepsilon_0 \le p \le p_0} |f(F^{-1}(p|x)|x)(F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x)) \\
&\quad - (F_g(F_{hg}^{*-1}(p|x)|x) - F_g(F_g^{-1}(p|x)|x))| + O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}) \\
&\le \frac{1}{\lambda} \sup_{\varepsilon_0 \le p \le p_0} |f(F^{-1}(p|x)|x)(F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x)) \\
&\quad - (F(F_{hg}^{*-1}(p|x)|x) - F(F_g^{-1}(p|x)|x))| + O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}) \text{ a.s.,} \qquad (4.7)
\end{aligned}$$

where the last inequality follows from Lemmas 2.11 and 4.1. By means of a Taylor expansion we can write (4.7) as

$$\frac{1}{2\lambda} \sup_{\varepsilon_0 \le p \le p_0} |f'(\theta^*_{hg}(p|x)|x)|(F^{*-1}_{hg}(p|x) - F^{-1}_g(p|x))^2 + O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}) \quad (4.8)$$

a.s., where $\theta^*_{hg}(p|x)$ lies between $F^{-1}_g(p|x)$ and $F^{*-1}_{hg}(p|x)$. From condition ($C6$) and Lemma 4.1 we deduce that (4.8) is $O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4})$ a.s. Hence, the result follows from Theorem 3.1.

In a similar way as for the original process studied in Section 4.2, the asymptotic representation established above allows us now to show the weak convergence of the bootstrapped quantile process. In the next section, two important applications of that result will be studied.

Theorem 4.5
Assume $(C1), (C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0, \dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Let* $0 < \varepsilon_0 < p_0 < 1$ *be such that* $F^{-1}(p_0|x) < T$ *and* $\inf_{\varepsilon_0 \le p \le p_0} f(F^{-1}(p|x)|x) > 0$. *Then, as* $n \to \infty$,

$$(nh_n)^{1/2}(F^{*-1}_{hg}(\cdot|x) - F^{-1}_g(\cdot|x)) \to -\frac{\tilde{W}(F^{-1}(\cdot|x)|x)}{f(F^{-1}(\cdot|x)|x)} \qquad \text{in } D_L[\varepsilon_0, p_0] \text{ a.s.}$$

where $\tilde{W}(\cdot|x)$ *is a Gaussian process with covariance function given by (2.11) and mean function given by (2.12).*

Proof. From Theorem 4.4 it follows that

$$\sup_{\varepsilon_0 \le p \le p_0} |(nh_n)^{1/2}(F^{*-1}_{hg}(p|x) - F^{-1}_g(p|x)) - (nh_n)^{1/2}(F_g(F^{-1}(p|x)|x) - F^*_{hg}(F^{-1}(p|x)|x))/f(F^{-1}(p|x)|x)|$$
$$= O_{P^*}((nh_n)^{-1/4}(\log n)^{3/4}) \qquad \text{a.s.}$$

The result now follows from Theorem 3.2 together with Theorem 4.1 in Billingsley (1968).

4.4 Applications

The quantile process $(nh_n)^{1/2}(F_h^{-1}(\cdot|x) - F^{-1}(\cdot|x))$ and the bootstrapped quantile process $(nh_n)^{1/2}(F_{hg}^{*-1}(\cdot|x) - F_g^{-1}(\cdot|x))$ have the same limiting Gaussian process for the optimal bandwidth $h_n = Cn^{-1/5}$ $(C > 0)$ (this follows from Theorems 4.3 and 4.5). This means that the bootstrapped process can be used to approximate the original process. The next two results are direct consequences of this fact : we first show the validity of the bootstrap procedure for the estimation of the quantile function and then construct bootstrap confidence bands for the quantile function. Other applications, like the construction of test statistics for testing hypotheses concerning the quantile function, are beyond the scope of this thesis.

4.4.1 Consistency of the bootstrap approximation

As a first application, we prove that the bootstrap procedure works, in the sense that the bootstrap distribution of $(nh_n)^{1/2}(F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x))$ is a strongly consistent estimator for the distribution of $(nh_n)^{1/2}(F_h^{-1}(p|x) - F^{-1}(p|x))$. This result provides us with an alternative for the normal approximation, obtained in Corollary 4.1. The bootstrap approximation has however the advantage that no additional estimation of unknown parameters is required, while the normal approximation depends on $H(t|x), H^u(t|x)$ and their second derivatives with respect to x.

Theorem 4.6
Assume $(C1), (C2')$, $H(t|x)$ and $H^u(t|x)$ satisfy $(C5)-(C7)$ in $[0,T]$ with $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ for some $C > 0, g_n \to 0, \dfrac{ng_n^5}{\log n} \to \infty$ and $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. Let $0 < p < 1$ be such that $F^{-1}(p|x) < T$ and $f(F^{-1}(p|x)|x) > 0$. Then, as $n \to \infty$,

$$\sup_{y\in\mathbb{R}} |P^*((nh_n)^{1/2}(F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x)) \le y)$$
$$- P((nh_n)^{1/2}(F_h^{-1}(p|x) - F^{-1}(p|x)) \le y)| = o(1) \qquad a.s.$$

The proof follows immediately from Corollary 4.1 and Theorem 4.5. For an alternative proof, which does not make use of the weak convergence of the bootstrapped process, see Theorem 4 in Van Keilegom and Veraverbeke (1998).

4.4.2 Confidence bands for the conditional quantile function

Since the bootstrapped process has the same limiting process as the original quantile process, it can be used to calculate the critical point of a confidence band for the quantile function. This idea is exploited in the next result and has also been used by Akritas (1986) for the construction of confidence bands of the survival distribution in the absence of covariates. Using a different approach, Horváth and Yandell (1987) constructed bootstrap confidence bands for the quantile function when no covariates are present.

Theorem 4.7

Assume $(C1), (C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0$, $g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Let* $0 < \varepsilon_0 < p_0 < 1$ *be such that* $F^{-1}(p_0|x) < T$ *and* $\inf\limits_{\varepsilon_0 \le p \le p_0} f(F^{-1}(p|x)|x) > 0$. *For each* $0 < \alpha < 1$ *and for the given sample, choose* $c_{\alpha hg}$ *such that, as* $n \to \infty$,

$$P^*\left(\sup_{\varepsilon_0 \le p \le p_0} \left\{|F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x)|D_h^{-1}(F^{-1}(p|x)|x)\right\} \le c_{\alpha hg}\right) \to 1-\alpha \qquad \text{a.s.},$$

where $D_h(\cdot|x)$ *is as in Theorem 3.4. Then, as* $n \to \infty$,

$$P(F_h^{-1}(p|x) - c_{\alpha hg}D_h(F^{-1}(p|x)|x) \le F^{-1}(p|x) \le F_h^{-1}(p|x) + c_{\alpha hg}D_h(F^{-1}(p|x)|x)$$
$$\text{for all } \varepsilon_0 \le p \le p_0) \to 1-\alpha.$$

Proof. The proof is completely along the same lines as for the construction of confidence bands for the conditional distribution (see Theorem 3.4). Instead of applying Lemma 3.5, we now need however that $\sup_{0\le t\le T} |(B(K(t|x)) + b'(t|x))/f(t|x)|$ has a continuous distribution (with the same notations as in Lemma 3.5). But, also this result can be shown in nearly the same way as for the distribution function.

4.5 Simulations

The finite sample performance of the asymptotic results obtained in the previous sections will now be examined through some simulations. We will compare the distribution of the quantile estimator $F_h^{-1}(p|x)$ with its normal approximation (Corollary 4.1) and its bootstrap approximation (Theorem 4.6).

Let $x_i = i/n$ $(i = 1, \dots, n)$ be n fixed and equidistant design points in the interval $[0, 1]$. The survival times Y_i $(i = 1, \dots, n)$ are independent, as well as the censoring times C_i $(i = 1, \dots, n)$ and for each i, Y_i and C_i are drawn independently from the exponential distributions

$$Y_i \sim \text{Exp}\left(\frac{a_1}{x_i + b}\right) \qquad \text{and} \qquad C_i \sim \text{Exp}\left(\frac{a_2}{x_i + b}\right)$$

respectively, for some parameters $a_1, a_2, b > 0$. The simulations are performed for samples of size $n = 50$, for $x = 0.5$, for a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$ and for $a_1 = 2$ and 9, $a_2 = 1$ and $b = 0.5$. It is readily verified that $P(\Delta_x = 0) = a_2/(a_1 + a_2)$ for any x and hence these choices of a_1 and a_2 correspond to respectively 33 % and 10 % censoring. We studied the behavior of the median ($p = 0.5$) and also of the 0.2th and 0.8th-quantile. The number of simulations for obtaining the distribution of $(nh_n)^{1/2}(F_h^{-1}(p|x) - F^{-1}(p|x))$ was 10 000. For the bootstrap distribution, 200 samples were drawn and for each sample, 200 bootstrap resamples were considered. The bootstrap approximation shown in Figures 4.1 and 4.2, was obtained by averaging the 200 conditional distributions.

From Corollary 4.1 it follows that the optimal bandwidth (which minimizes the asymptotic mean squared error) for estimating the quantile function $F^{-1}(p|x)$ is given by

$$h_{n,\text{opt}} = \left(\frac{\sigma^2(p|x)}{4\overline{\beta}^2(p|x)}\right)^{1/5} n^{-1/5},$$

where

$$\overline{\beta}(p|x) = -\frac{\overline{b}(F^{-1}(p|x)|x)}{f(F^{-1}(p|x)|x)}$$

and $\overline{b}(\cdot|x)$ is given by (2.15). All the simulations were carried out for this optimal choice. However, in many of our simulations, this optimal bandwidth exceeded 0.5,

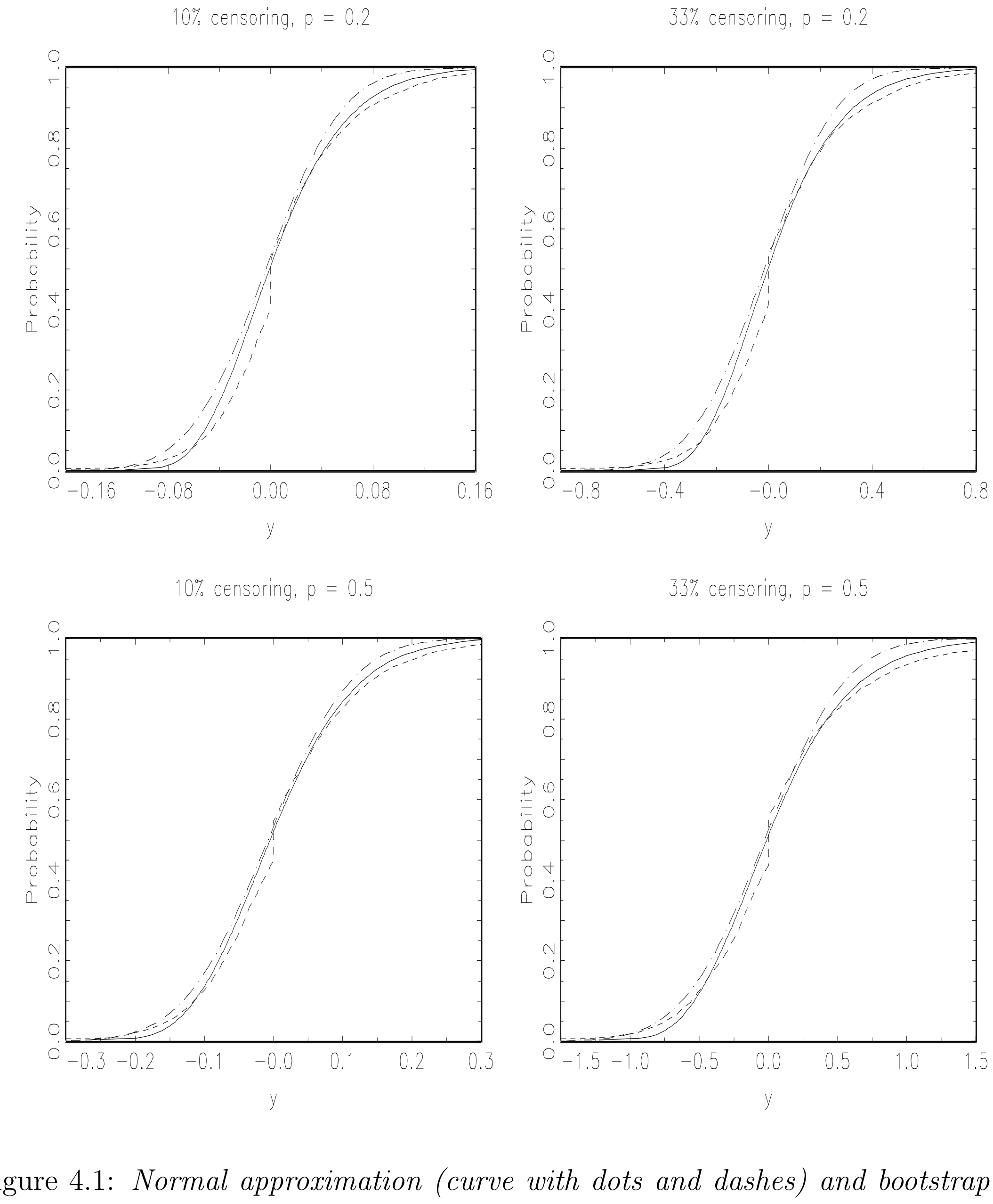

Figure 4.1: *Normal approximation (curve with dots and dashes) and bootstrap approximation (dashed curve) of the distribution of* $(nh_n)^{1/2}(F_h^{-1}(p|x) - F^{-1}(p|x))$ *(solid curve) for exponential survival and censoring times.*

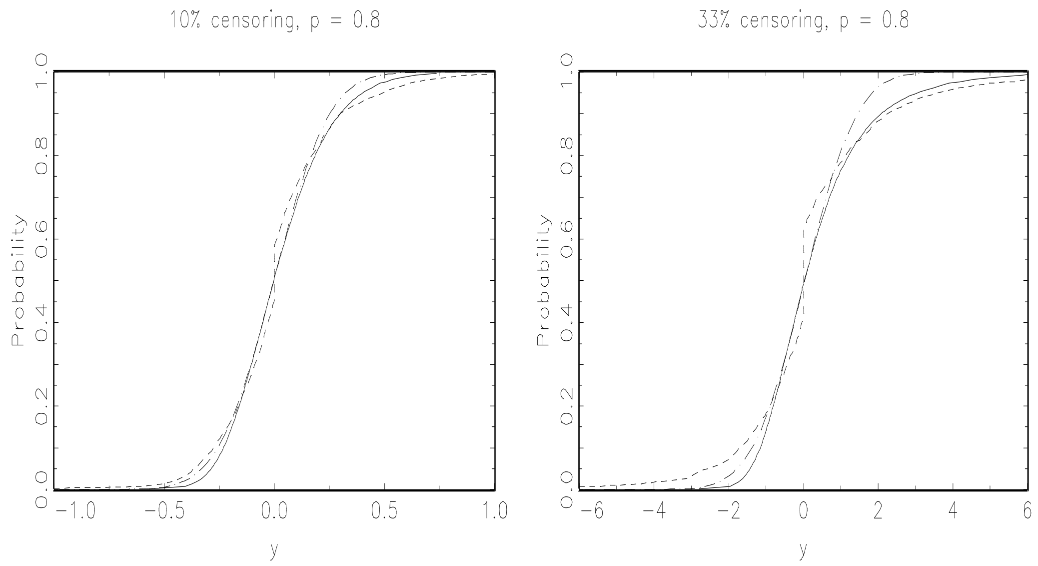

Figure 4.2: *Normal approximation (curve with dots and dashes) and bootstrap approximation (dashed curve) of the distribution of* $(nh_n)^{1/2}(F_h^{-1}(p|x) - F^{-1}(p|x))$ *(solid curve) for exponential survival and censoring times.*

in which case the actual number of observations in the window around x is not $2nh_n$ but n. We therefore truncated the bandwidth h_n at 0.5 in these cases. As to the pilot bandwidth g_n, no methods are available yet for obtaining the optimal value of g_n. We chose $g_n = 5h_n$.

From Figures 4.1 and 4.2 it is apparent that the bootstrap distribution $P^*((nh_n)^{1/2}(F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x)) \leq y)$ jumps at $y = 0$. This is due to the fact that $F_g^{-1}(p|x)$ and $F_{hg}^{*-1}(p|x)$ are equal with nonzero probability, since $F_g(t|x)$ and $F_{hg}^*(t|x)$ have jumps in common. Because of this, the bootstrap approximation does not behave well in a (left) neighborhood of $y = 0$. The redistribution-to-the-right property of the Beran estimator implies that the jumps of this estimator are getting larger when the time points get larger. Because of this, the jump size at 0 of the distribution of $F_{hg}^{*-1}(0.8|x) - F_g^{-1}(0.8|x)$ is much larger than for $p = 0.2$ or 0.5, and makes that this distribution is not approximating the real distribution very well (especially for 33 % censoring). A possible way to obtain a better approximation is to use a smoothed Beran estimator, in analogy with the smoothed Kaplan-Meier estimator introduced by Földes, Rejtő and Winter (1981).

Another feature of these figures is that, as was expected, the distribution of $F_h^{-1}(p|x)$ is much more spread out for 33 % censoring than for 10 % censoring. The figures also show that, in general, the normal approximation is slightly better than the bootstrap approximation. However, in practice, the normal approximation requires the estimation of many unknown quantities, while the bootstrap approximation is applicable without any further estimation.

4.6 Appendix : Some more results on the bootstrapped Stone estimator and its quantile estimator

We end this chapter with two results that were needed in the proof of Theorem 4.4. The first result concerns the uniform strong consistency of the bootstrapped quantile estimator $F_{hg}^{*-1}(p|x)$. The second one is a modulus of continuity result for the bootstrapped estimator of the distribution function.

Lemma 4.1
Assume $(C1), (C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5) - (C7)$ *in* $[0, T]$ *with* $T < \tau_{H(\cdot|x)}, h_n = Cn^{-1/5}$ *for some* $C > 0, g_n \to 0, \dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Let* $0 < \varepsilon_0 < p_0 < 1$ *be such that* $F^{-1}(p_0|x) < T$ *and* $\inf\limits_{\varepsilon_0 \le p \le p_0} f(F^{-1}(p|x)|x) = \lambda > 0$. *Then, as* $n \to \infty$,

$$\sup_{\varepsilon_0 \le p \le p_0} |F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x)| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2}) \qquad \text{a.s.}$$

Proof. Let $C > 0$ and $\varepsilon_n = (nh_n)^{-1/2}(\log n)^{1/2}$. Write

$$P^*\left(\sup_{\varepsilon_0 \le p \le p_0} |F_{hg}^{*-1}(p|x) - F_g^{-1}(p|x)| > C\varepsilon_n\right)$$

$$\le P^*\left(F_{hg}^{*-1}(p|x) > F_g^{-1}(p|x) + C\varepsilon_n \text{ for some } \varepsilon_0 \le p \le p_0\right) \tag{4.9}$$

$$+ P^* \left(F_{hg}^{*-1}(p|x) < F_g^{-1}(p|x) - C\varepsilon_n \text{ for some } \varepsilon_0 \le p \le p_0 \right).$$

We will prove that the first term in (4.9) can be made arbitrarily small for n and C large enough. The proof for the second term will be completely analogous. The first term in (4.9) is not larger than

$$\begin{aligned}
&P^*(p > F_{hg}^*(F_g^{-1}(p|x) + C\varepsilon_n|x) \text{ for some } \varepsilon_0 \le p \le p_0)\\
&\le P^* \left(\sup_{\varepsilon_0 \le p \le p_0} |F(F^{-1}(p|x) + C\varepsilon_n|x) - F_{hg}^*(F_g^{-1}(p|x) + C\varepsilon_n|x)| > A_n \right)\\
&\le P^* \left(\sup_{\varepsilon_0 \le p \le p_0} |F(F^{-1}(p|x) + C\varepsilon_n|x) - F(F_g^{-1}(p|x) + C\varepsilon_n|x)| > \frac{A_n}{3} \right)\\
&\quad + P^* \left(\sup_{\varepsilon_0 \le p \le p_0} |F_g(F_g^{-1}(p|x) + C\varepsilon_n|x) - F(F_g^{-1}(p|x) + C\varepsilon_n|x)| > \frac{A_n}{3} \right)\\
&\quad + P^* \left(\sup_{\varepsilon_0 \le p \le p_0} |F_{hg}^*(F_g^{-1}(p|x) + C\varepsilon_n|x) - F_g(F_g^{-1}(p|x) + C\varepsilon_n|x)| > \frac{A_n}{3} \right)\\
&= T_1 + T_2 + T_3,
\end{aligned}$$

where $A_n = \inf_{\varepsilon_0 \le p \le p_0} (F(F^{-1}(p|x) + C\varepsilon_n|x) - p)$. Choose C such that

$$\sup_{0 \le t \le T} f(t|x) \sup_{\varepsilon_0 \le p \le p_0} |F_g^{-1}(p|x) - F^{-1}(p|x)| \le \frac{1}{6} \lambda C \varepsilon_n.$$

This is possible, because $\sup_{\varepsilon_0 \le p \le p_0} |F_g^{-1}(p|x) - F^{-1}(p|x)| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. by Theorem 4.1(a) and the Borel-Cantelli lemma. Since,

$$\begin{aligned}
&\sup_{\varepsilon_0 \le p \le p_0} |F(F^{-1}(p|x) + C\varepsilon_n|x) - F(F_g^{-1}(p|x) + C\varepsilon_n|x)|\\
&\le \sup_{0 \le t \le T} f(t|x) \sup_{\varepsilon_0 \le p \le p_0} |F_g^{-1}(p|x) - F^{-1}(p|x)|
\end{aligned}$$

and since $A_n \ge \lambda C\varepsilon_n/2$ for n large enough, it follows that $T_1 = 0$. Similarly, if we enlarge C such that $\sup_{0 \le t \le T} |F_g(t|x) - F(t|x)| \le \lambda C\varepsilon_n/6$, which is possible, because $\sup_{0 \le t \le T} |F_g(t|x) - F(t|x)| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. (see Lemma 3.1), we have that $T_2 = 0$. From the proof of Theorem 3.1, it follows that $\sup_{0\le t\le T} |F^*_{hg}(t|x) - F_g(t|x)| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. Hence, T_3 can be made arbitrarily small. This finishes the proof.

Lemma 4.2
Assume $(C1), (C2')$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(C5)-(C7)$ *in* $[0,T]$ *with* $T < \tau_{H(\cdot|x)}$, $h_n = Cn^{-1/5}$ *for some* $C > 0, g_n \to 0$, $\dfrac{ng_n^5}{\log n} \to \infty$ *and* $\dfrac{ng_n^5}{\log n}\dfrac{h_n}{g_n} = O(1)$. *Then, for any* $c > 0$, *as* $n \to \infty$,

$$\sup_{0\le s,t\le T; |t-s|\le c(nh_n)^{-1/2}(\log n)^{1/2}} |F^*_{hg}(t|x) - F^*_{hg}(s|x) - F_g(t|x) + F_g(s|x)|$$

$$= O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4}) \qquad \text{a.s.}$$

Proof. Take $0 \le s \le t \le T$ such that $|t-s| \le a_n$ and let $a_n = c(nh_n)^{-1/2}(\log n)^{1/2}$ and $\varepsilon_n = (nh_n)^{-3/4}(\log n)^{3/4}$. Write

$$\begin{aligned} &F^*_{hg}(t|x) - F^*_{hg}(s|x) - F_g(t|x) + F_g(s|x) \\ &= [\tilde{F}^*_{hg}(t|x) - \tilde{F}^*_{hg}(s|x) - F_g(t|x) + F_g(s|x)] \qquad (4.10) \\ &\quad + [F^*_{hg}(t|x) - F^*_{hg}(s|x) - \tilde{F}^*_{hg}(t|x) + \tilde{F}^*_{hg}(s|x)], \end{aligned}$$

where

$$\tilde{F}^*_{hg}(t|x) = \int_0^t \frac{dH^{*u}_{hg}(y|x)}{1 - G(y|x)}.$$

The first term on the right hand side of (4.10) can be written as

$$\begin{aligned} &\left[\sum_{i=1}^n w_{ni}(x;h_n)\left(\frac{I(s \le T_i^* \le t, \Delta_i^* = 1)}{1 - G(T_i^*|x)} - E^*\left\{\frac{I(s \le T_i^* \le t, \Delta_i^* = 1}{1 - G(T_i^*|x)}\right\}\right)\right] \\ &+ \left[\sum_{i=1}^n w_{ni}(x;h_n)\int_s^t \frac{d(H_g^u(y|x_i) - H_g^u(y|x))}{1 - G(y|x)}\right] \\ &+ \left[\int_s^t \frac{dH_g^u(y|x)}{1 - G(y|x)} - (F_g(t|x) - F_g(s|x))\right] \\ &= T_1 + T_2 + T_3. \end{aligned}$$

Some easy calculations show that

$$T_2 = \frac{E^* H_{hg}^{*u}(t|x) - H_g^u(t|x)}{1 - G(t|x)} - \frac{E^* H_{hg}^{*u}(s|x) - H_g^u(s|x)}{1 - G(s|x)}$$

$$- \int_s^t \frac{E^* H_{hg}^{*u}(y|x) - H_g^u(y|x)}{(1 - G(y|x))^2} dG(y|x)$$

and hence,

$$\sup_{0 \le s,t \le T; |t-s| \le a_n} |T_2|$$

$$\le \frac{1}{1 - G(T|x)} \sup_{0 \le s,t \le T; |t-s| \le a_n} |E^* H_{hg}^{*u}(t|x) - H_g^u(t|x) - E^* H_{hg}^{*u}(s|x) + H_g^u(s|x)|$$

$$+ \frac{2}{(1 - G(T|x))^2} \sup_{0 \le s,t \le T; |t-s| \le a_n} |G(t|x) - G(s|x)| \sup_{0 \le t \le T} |E^* H_{hg}^{*u}(t|x) - H_g^u(t|x)|$$

$$= O((nh_n)^{-3/4} (\log n)^{3/4}) \qquad \text{a.s.}$$

by Lemmas 2.6, 3.6 and 3.8 and by the Lipschitz continuity of $G(\cdot|x)$. Because $T_3 = \tilde{F}_g(t|x) - \tilde{F}_g(s|x) - F_g(t|x) + F_g(s|x)$ where $\tilde{F}_g(t|x) = \int_0^t (1 - G(y|x))^{-1} dH_g^u(y|x)$, it follows from the proof of Theorem 2.2 and from the Borel-Cantelli lemma that $\sup_{0 \le s,t \le T; |t-s| \le a_n} |T_3| = O((nh_n)^{-3/4} (\log n)^{3/4})$ a.s. Next, we will calculate the variance of T_1.

$$\text{Var}^* \left(\frac{I(s \le T_i^* \le t, \Delta_i^* = 1)}{1 - G(T_i^*|x)} \right) = \int_s^t \frac{dH_g^u(y|x_i)}{(1 - G(y|x))^2} - \left(\int_s^t \frac{dH_g^u(y|x_i)}{1 - G(y|x)} \right)^2$$

$$\le \frac{2}{(1 - G(T|x))^2} \sup_{0 \le t \le T} |H_g^u(t|x_i) - H^u(t|x_i)|$$

$$+ \frac{1}{(1 - G(T|x))^2} |H^u(t|x_i) - H^u(s|x_i)|$$

$$= O(a_n) \qquad \text{a.s.}$$

uniformly in s and t ($|t - s| \le a_n$), which follows from the Lipschitz continuity of $H^u(\cdot|x)$ and from Lemma 3.9. Note that this order bound is also uniformly in i

(this can be verified by going through the proof of Lemma 3.9). Hence, $\text{Var}^* T_1 = O(a_n(nh_n)^{-1})$ a.s. Bernstein's inequality (see Serfling (1980)), implies that

$$P^*(|T_1| > M\varepsilon_n) \leq 2\exp\left(-KM\frac{\varepsilon_n^2 nh_n}{a_n}\right) = 2n^{-KM/c}$$

for any $M \geq 1$ and where $K > 0$ is an absolute constant. For dealing with $\sup_{0\leq s,t\leq T;|t-s|\leq a_n} |T_1|$ we partition the interval $[0, T]$ into $k_n = O(\varepsilon_n^{-1})$ subintervals of length $O(\varepsilon_n)$. It follows that (by choosing M large enough)

$$\begin{aligned}
&P^*\left(\sup_{0\leq s,t\leq T;|t-s|\leq a_n} |\tilde{F}^*_{hg}(t|x) - \tilde{F}^*_{hg}(s|x) - F_g(t|x) + F_g(s|x)| > M\varepsilon_n\right) \\
&\leq P^*\left(\max_{0\leq j\leq k_n} \sup_{0\leq s\leq T;|t_j-s|\leq a_n} \left|\sum_{i=1}^n w_{ni}(x;h_n)\left[\frac{I(s\leq T_i^*\leq t_j, \Delta_i^*=1)}{1-G(T_i^*|x)}\right.\right.\right. \qquad (4.11) \\
&\qquad \left.\left. -E^*\left\{\frac{I(s\leq T_i^*\leq t_j, \Delta_i^*=1)}{1-G(T_i^*|x)}\right\}\right]\right| \\
&\qquad \left. > \frac{M\varepsilon_n}{2} - \max_{0\leq j\leq k_n}\left|\sum_{i=1}^n w_{ni}(x;h_n)\int_{t_{j-1}}^{t_j}\frac{dH_g^u(y|x_i)}{1-G(y|x)}\right|\right).
\end{aligned}$$

Lemmas 3.8 and 2.11 together with the Lipschitz continuity of $H^u(\cdot|x)$ imply that

$$\begin{aligned}
&\left|\sum_{i=1}^n w_{ni}(x;h_n)\int_{t_{j-1}}^{t_j}\frac{dH_g^u(y|x_i)}{1-G(y|x)}\right| \\
&\leq \frac{1}{1-G(T|x)}|E^*H_{hg}^{*u}(t_j|x) - E^*H_{hg}^{*u}(t_{j-1}|x)| \\
&\leq \frac{1}{1-G(T|x)}\{|E^*H_{hg}^{*u}(t_j|x) - E^*H_{hg}^{*u}(t_{j-1}|x) - H_g^u(t_j|x) + H_g^u(t_{j-1}|x)| \\
&\quad + |H_g^u(t_j|x) - H_g^u(t_{j-1}|x) - H^u(t_j|x) + H^u(t_{j-1}|x)| + |H^u(t_j|x) - H^u(t_{j-1}|x)|\} \\
&= O(\varepsilon_n) \qquad \text{a.s.}
\end{aligned}$$

Hence, (4.11) is majorized by (choose M large enough and use a similar procedure as above for replacing $\sup_{0\le s\le T}$ by a maximum)

$$P^*\left(\max_{0\le j,l\le k_n;|t_j-t_l|\le a_n}\left|\sum_{i=1}^n w_{ni}(x;h_n)\left[\frac{I(t_l\le T_i^*\le t_j,\Delta_i^*=1)}{1-G(T_i^*|x)}\right.\right.\right.$$
$$\left.\left.\left.-E^*\left\{\frac{I(t_l\le T_i^*\le t_j,\Delta_i^*=1)}{1-G(T_i^*|x)}\right\}\right]\right|>\frac{M\varepsilon_n}{4}\right)\le C\frac{a_n}{\varepsilon_n^2}n^{-KM/c}$$

(for some $C>0$), which can be made arbitrarily small by proper choice of M. This shows that

$$\sup_{0\le s,t\le T;|t-s|\le a_n}|\tilde{F}_{hg}^*(t|x)-\tilde{F}_{hg}^*(s|x)-F_g(t|x)+F_g(s|x)|=O(\varepsilon_n)\qquad \text{a.s.}$$

It remains to prove that the second term on the right hand side of (4.10) is of the given order. It is easy to show that

$$F_{hg}^*(t|x)=\int_0^t\frac{dH_{hg}^{*u}(y|x)}{1-G_{hg}^*(y-|x)}.$$

Hence,

$$|F_{hg}^*(t|x)-F_{hg}^*(s|x)-\tilde{F}_{hg}^*(t|x)+\tilde{F}_{hg}^*(s|x)|$$
$$\le\left|\int_s^t\frac{G_{hg}^*(y-|x)-G(y|x)}{1-G_{hg}^*(y-|x)}d\tilde{F}_{hg}^*(y|x)\right|$$
$$\le\frac{1}{1-G_{hg}^*(T|x)}|\tilde{F}_{hg}^*(t|x)-\tilde{F}_{hg}^*(s|x)|\sup_{0\le t\le T}|G_{hg}^*(t|x)-G(t|x)|.$$

From Lemma 3.1 and the proof of Theorem 3.1 we know that $\sup_{0\le t\le T}|F_{hg}^*(t|x)-F(t|x)|=O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. In a completely similar way (replace any Δ with $1-\Delta$), we can show that $G_{hg}^*(t|x)-G(t|x)$ is of the same order (uniformly in t). Write

$$|\tilde{F}_{hg}^*(t|x)-\tilde{F}_{hg}^*(s|x)|\le|\tilde{F}_{hg}^*(t|x)-\tilde{F}_{hg}^*(s|x)-F_g(t|x)+F_g(s|x)|$$
$$+|F_g(t|x)-F_g(s|x)-F(t|x)+F(s|x)|$$
$$+|F(t|x)-F(s|x)|,$$

which is uniformly (for s, t such that $|t-s| \leq a_n$) of the order $O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. by means of the first part of the present proof, Theorem 2.2 and the Lipschitz continuity of $F(\cdot|x)$. Let

$$A_n^* = \left\{1 - G_{hg}^*(t|x) \geq \frac{\delta}{2} \text{ for all } 0 \leq t \leq T\right\},$$

where $1 - H(T|x) > \delta > 0$. Then,

$$P^* \left(\sup_{0 \leq s,t \leq T; |t-s| \leq a_n} |F_{hg}^*(t|x) - F_{hg}^*(s|x) - \tilde{F}_{hg}^*(t|x) + \tilde{F}_{hg}^*(s|x)| > M\varepsilon_n \right)$$

$$\leq P^* \left(\sup_{0 \leq s,t \leq T; |t-s| \leq a_n} |\tilde{F}_{hg}^*(t|x) - \tilde{F}_{hg}^*(s|x)| \sup_{0 \leq t \leq T} |G_{hg}^*(t|x) - G(t|x)| > \frac{\delta M}{2} \varepsilon_n \right)$$

$$+ P^*(A_n^{*c}),$$

which can be made arbitrarily small by proper choice of M, since

$$P^*(A_n^{*c}) \leq P^* \left(\sup_{0 \leq t \leq T} |G_{hg}^*(t|x) - G(t|x)| > \frac{\delta}{2} \right)$$

(use that $1 - G(t|x) > 1 - H(t|x) > \delta$ for all $0 \leq t \leq T$). This finishes the proof.

Chapter 5

The conditional distribution in heteroscedastic regression models with censored data

In Chapters 2 until 4, we considered a completely nonparametric regression model, in which no assumption is made on the way the response is related to the covariates. The response was allowed to be subject to right censoring. For this situation, an estimator for the conditional distribution of the response given the covariate, called the Beran estimator, was studied (Chapters 2 and 3), as well as the corresponding quantile estimator (Chapter 4).

In the next chapters, we assume a more specific form of relationship between the response and the covariate via the heteroscedastic regression model

$$Y = m(X) + \sigma(X)\varepsilon,$$

where ε is independent of X, and $m(X)$ and $\sigma(X)$ are a location and scale functional. Under this model, we have constructed in the introductory chapter (Chapter 1, equation (1.39)) the estimator $\hat{F}(t|x)$ for the conditional distribution. We discussed in Section 1.5 how this estimator transfers tail information from regions with 'light' censoring to regions with 'heavy' censoring. In this way, the estimator can be more informative in the right tail than the Beran estimator, and we therefore expect that this estimator outperforms the Beran estimator, provided that model (1.28) is valid. In this chapter we will study the asymptotic properties of the estimator $\hat{F}(t|x)$. The finite sample performance of the estimator will be compared with that of the Beran estimator in Chapter 7. In that chapter, the estimator $\hat{F}(t|x)$ will also be used for analyzing a real data set.

We start in Section 5.1 by introducing some notations and definitions and stating the assumptions. That section also contains two basic results, that will be needed throughout the chapter. In Section 5.2, an asymptotic representation will be constructed for the estimator of the residual distribution $\hat{F}_e(t)$, defined in (1.38). As a consequence of that result, the weak convergence of the residual distribution process $\hat{F}_e(\cdot)$ will be established. Section 5.3 deals with the estimator of the conditional distribution $\hat{F}(t|x)$: by making use of the representation for the estimator $\hat{F}_e(t)$, we will develop a representation for $\hat{F}(t|x)$, which will then lead to the weak convergence of the conditional distribution process $\hat{F}(\cdot|x)$. In Section 5.4, we prove a number of results concerning $\hat{m}(x)$ and $\hat{\sigma}(x)$ (defined in (1.36) and (1.37)), which will be needed in Sections 5.2 and 5.3 and which will be obtained by using asymptotic results of the Beran estimator $\tilde{F}(t|x)$ (see (1.34)). Finally, a number of technical results are collected in Section 5.5.

5.1 Definitions and assumptions

Consider the random vector (X, Y) satisfying the nonparametric regression model (1.28), where the smooth functions m and σ are assumed to be, respectively, a location and scale functional (see (1.29) and (1.30) for the definition of a location and scale functional). In Proposition 5.1 later in this section, we will prove that if model (1.28) holds for a location functional m and scale functional σ, then it holds for all location functionals $\tilde{m}$ and scale functionals $\tilde{\sigma}$, in the sense that the resulting error term $\tilde{\varepsilon}$ is still independent of X.

Apart from the notations introduced in Section 1.5, we need some additional notations. Let $G(t|x) = P(C \leq t|x)$, $H(t|x) = P(T \leq t|x), H^u(t|x) = P(T \leq t, \Delta = 1|x)$ and $F_X(x) = P(X \leq x)$. The assumed independence of Y and C

for given X implies that $1 - H(t|x) = (1 - F(t|x))(1 - G(t|x))$. Further, denote $G_e(t) = P((C - m(X))/\sigma(X) \leq t)$ and for $E = T - m(X)/\sigma(X)$ we use the notation $H_e(t) = P(E \leq t)$, $H_e^u(t) = P(E \leq t, \Delta = 1)$, $H_e(t|x) = P(E \leq t|x)$ and $H_e^u(t|x) = P(E \leq t, \Delta = 1|x)$. We will prove in Proposition 5.2 that ε and $(C-m(X))/\sigma(X)$ are independent. This yields that $1-H_e(t) = (1-F_e(t))(1-G_e(t))$. The probability density functions of the distributions defined above will be denoted by the corresponding lower case letters.

The following functions enter in the asymptotic representation for $\hat{F}_e(t)$ and $\hat{F}(t|x)$, which we will establish in Sections 5.2 and 5.3 :

$$\xi_e(z,\delta,t) = (1 - F_e(t)) \left\{ -\int_{-\infty}^{t\wedge z} \frac{dH_e^u(s)}{(1 - H_e(s))^2} + \frac{I(z \leq t, \delta = 1)}{1 - H_e(z)} \right\},$$

$$\xi(z,\delta,t|x) = (1 - F(t|x)) \left\{ -\int_{-\infty}^{t\wedge z} \frac{dH^u(s|x)}{(1 - H(s|x))^2} + \frac{I(z \leq t, \delta = 1)}{1 - H(z|x)} \right\},$$

$$\eta(z,\delta|x) = \sigma^{-1}(x) \int_{-\infty}^{+\infty} \xi(z,\delta,v|x) J(F(v|x))\, dv,$$

$$\zeta(z,\delta|x) = \sigma^{-1}(x) \int_{-\infty}^{+\infty} \xi(z,\delta,v|x) J(F(v|x)) \frac{v - m(x)}{\sigma(x)}\, dv,$$

$$\gamma_1(t|x) = \int_{-\infty}^{t} \frac{h_e(s|x)}{(1 - H_e(s))^2}\, dH_e^u(s) + \int_{-\infty}^{t} \frac{dh_e^u(s|x)}{1 - H_e(s)},$$

$$\gamma_2(t|x) = \int_{-\infty}^{t} \frac{s h_e(s|x)}{(1 - H_e(s))^2}\, dH_e^u(s) + \int_{-\infty}^{t} \frac{d(s h_e^u(s|x))}{1 - H_e(s)}.$$

Easy calculations show that the function ξ defined above equals the function ξ defined in Theorem 2.3. We estimate $H_e(t)$ and $H_e^u(t)$ by the empirical distribution functions (based on the residuals $\hat{E}_i = (T_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$)

$$\hat{H}_e(t) = n^{-1} \sum_{i=1}^{n} I(\hat{E}_i \leq t), \tag{5.1}$$

$$\hat{H}_e^u(t) = n^{-1} \sum_{i=1}^{n} I(\hat{E}_i \leq t, \Delta_i = 1), \tag{5.2}$$

while the conditional (sub)distributions $H(t|x)$ and $H^u(t|x)$ are estimated by the conditional empirical distribution functions (Stone (1977))

$$\hat{H}(t|x) = \sum_{i=1}^{n} W_{ni}(x;h_n) I(T_i \le t), \tag{5.3}$$

$$\hat{H}^u(t|x) = \sum_{i=1}^{n} W_{ni}(x;h_n) I(T_i \le t, \Delta_i = 1), \tag{5.4}$$

where the weights $W_{ni}(x;h_n)$ are the Nadaraya-Watson weights, defined in (1.35). The covariate density $f_X(x)$ is estimated by

$$\hat{f}_X(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right).$$

A few more notations are needed. The support of the covariate X is denoted by R_X. Let $\bar{T}$, respectively $\tilde{T}_x$, be any value less than the upper bound of the support of H_e, respectively $H(\cdot|x)$, such that $\inf_{x \in R_X}(1 - H(\tilde{T}_x|x)) > 0$ and let $\Omega = \{(x,t);\ x \in R_X \text{ and } \frac{t-m(x)}{\sigma(x)} \le \bar{T}\}$. For a (sub)distribution function $L(t|x)$ we will use the notations $l(t|x) = L'(t|x) = \frac{\partial}{\partial t}L(t|x), \dot{L}(t|x) = \frac{\partial}{\partial x}L(t|x)$ and similar notations will be used for higher order derivatives.

The assumptions we need to impose in the proofs of the main results of Chapters 5 and 6 are listed below for convenient reference.

(A1)(i) The sequence h_n satisfies $nh_n^5(\log h_n^{-1})^{-1} = O(1)$ and $nh_n^{3+2\delta}(\log h_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.
(ii) The sequence h_n satisfies $nh_n^4 \to 0$ and $nh_n^{3+2\delta}(\log h_n^{-1})^{-1} \to \infty$ for some $\delta > 0$. (Note that this implies $(A1)(i)$).
(iii) The support R_X of X is bounded, convex and its interior is not empty.
(iv) The probability density function K has compact support, K is twice continuously differentiable and $\int uK(u)\,du = 0$.

(A2)(i) There exist $0 \le s_0 \le s_1 \le 1$ such that $s_1 \le \inf_x F(\tilde{T}_x|x)$, $s_0 \le \inf\{s \in [0,1]; J(s) \ne 0\}$, $s_1 \ge \sup\{s \in [0,1]; J(s) \ne 0\}$ and $\inf_x \inf_{s_0 \le s \le s_1} f(F^{-1}(s|x)|x) > 0$.
(ii) The function J is nonnegative for $0 \le s \le 1$, J is twice continuously differentiable and $\int_0^1 J(s)\,ds = 1$.

(A3)(i) The distribution F_X is thrice continuously differentiable and $\inf_x f_X(x) > 0$.
(ii) The functions m and σ are twice continuously differentiable and $\inf_x \sigma(x) > 0$.

(A4) The functions $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$ are twice continuously differentiable with respect to x and their first and second derivatives (with respect to x) are bounded, uniformly in $x \in R_X, z < T_x$ and δ.

(A5)(i) $L(t|x)$ is continuous.
(ii) $L'(t|x) = l(t|x)$ exists, is continuous in (x, t) and $\sup_{x,t} |tL'(t|x)| < \infty$.
(iii) $L''(t|x)$ exists, is continuous in (x, t) and $\sup_{x,t} |t^2 L''(t|x)| < \infty$.
(iv) $\dot{L}(t|x)$ exists, is continuous in (x, t) and $\sup_{x,t} |t\dot{L}(t|x)| < \infty$.
(v) $\ddot{L}(t|x)$ exists, is continuous in (x, t) and $\sup_{x,t} |t^2 \ddot{L}(t|x)| < \infty$.
(vi) $\dot{L}'(t|x)$ exists, is continuous in (x, t) and $\sup_{x,t} |t\dot{L}'(t|x)| < \infty$.
(vii) $\ddot{L}'(t|x)$ exists, is continuous in (x, t) and $\sup_{x,t} |t\ddot{L}'(t|x)| < \infty$.

We finish this section with two preliminary results. The first one guarantees that model (1.28) is valid for all location and scale functionals provided that it holds for one location and one scale functional. The second result states that under model (1.28), the error term ε is (unconditionally) independent of $(C - m(X))/\sigma(X)$.

Proposition 5.1
Assume that $Y = m(X) + \sigma(X)\varepsilon$, where ε is independent of X, m is a location functional and σ is a scale functional. Let $\tilde{m}(x)$ be an arbitrary location functional and $\tilde{\sigma}(x)$ be an arbitrary scale functional. Then, $\tilde{\varepsilon} = (Y - \tilde{m}(X))/\tilde{\sigma}(X)$ is also independent of X.

Proof. First write

$$\tilde{\varepsilon} = \frac{m(X) - \tilde{m}(X)}{\tilde{\sigma}(X)} + \frac{\sigma(X)}{\tilde{\sigma}(X)}\varepsilon.$$

We will show that $\sigma(X)/\tilde{\sigma}(X)$ is independent of X. In a similar way, one can show that $(m(X) - \tilde{m}(X))/\tilde{\sigma}(X)$ does not depend on X. Let us use the notations $F_Z(\cdot|x)$, respectively $F_Z(\cdot)$, for the conditional (on the covariate X), respectively overall, distribution of any random variable Z (this notation is somewhat different from the notation introduced before, but it is more appropriate here). Since $\tilde{\sigma}(x)$ is a scale functional, there exists a functional $\tilde{S}$, such that $\tilde{\sigma}(x) = \tilde{S}(F_Y(\cdot|x))$ and such that relation (1.30) is satisfied. Hence, for arbitrary x,

$$\tilde{\sigma}(x) = \tilde{S}(F_{m(x)+\sigma(x)\varepsilon}) = \sigma(x)\tilde{S}(F_\varepsilon),$$

i.e. $\sigma(X)/\tilde{\sigma}(X)$ is the same for all values of X. This finishes the proof.

Proposition 5.2
Assume Y is independent of C, conditionally on X, and ε is independent of X. Then, ε and $(C - m(X))/\sigma(X)$ are (unconditionally) independent.

Proof. This follows easily, since

$$P\left(\varepsilon \leq t, \frac{C - m(X)}{\sigma(X)} \leq c\right)$$

$$= \int_{R_X} P(\varepsilon \leq t | X = x) P\left(\frac{C - m(X)}{\sigma(X)} \leq c \middle| X = x\right) dF_X(x)$$

$$= F_e(t) \int_{R_X} P\left(\frac{C - m(X)}{\sigma(X)} \leq c \middle| X = x\right) dF_X(x) = F_e(t) G_e(c).$$

5.2 Asymptotic representation and weak convergence of the residual distribution process

We start by establishing an asymptotic representation for the Kaplan-Meier estimator $\hat{F}_e(t)$ of the error distribution. Below, we give an outline of the proof. A number of technical results needed in the proof of this representation are deferred to Sections 5.4 and 5.5.

Lo and Singh (1986) established an a.s. representation for the Kaplan-Meier estimator of censored and independent data points. Their result can be used on the points $(T_i - m(X_i))/\sigma(X_i)$, but not on $(T_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$ since the latter points are not independent. In the next theorm, we establish a representation that generalizes the Lo and Singh result to the present context. The main term in the representation of Lo and Singh (1986) only contains the first term of the function φ defined below. The two extra terms in our result are caused by the fact that we replaced $(T_i - m(X_i))/\sigma(X_i)$ with $(T_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$.

The remainder term in the next representation is $o_P(n^{-1/2})$. This is precisely the order which is needed to obtain the weak convergence of the process $\hat{F}_e(\cdot) - F_e(\cdot)$.

Theorem 5.1
Assume (A1)-(A4), $H(t|x)$ and $H^u(t|x)$ satisfy (A5) $(i)-(vi)$, and $H_e(t|x)$ and $H_e^u(t|x)$ satisfy (A5) (ii, iii, vi, vii). Then,

$$\hat{F}_e(t) - F_e(t) = n^{-1} \sum_{i=1}^{n} \varphi(X_i, T_i, \Delta_i, t) + R_n(t),$$

where $\sup\{|R_n(t)|; -\infty < t \leq \bar{T}\} = o_P(n^{-1/2})$ *and, with* $S_e(t) = 1 - F_e(t)$,

$$\varphi(x, z, \delta, t) = \xi_e\left(\frac{z - m(x)}{\sigma(x)}, \delta, t\right) - S_e(t)\eta(z, \delta|x)\gamma_1(t|x) - S_e(t)\zeta(z, \delta|x)\gamma_2(t|x).$$

Proof. The scheme of the proof parallels somewhat that of the proof of Theorem 2.3 (representation for the Beran estimator). We start with

$$\int_{-\infty}^{t} \frac{d\hat{H}_e^u(s)}{1 - \hat{H}_e(s)} - \int_{-\infty}^{t} \frac{dH_e^u(s)}{1 - H_e(s)}$$
$$= \int_{-\infty}^{t} \left[\frac{1}{1 - \hat{H}_e(s)} - \frac{1}{1 - H_e(s)}\right] dH_e^u(s) + \int_{-\infty}^{t} \frac{1}{1 - H_e(s)} d(\hat{H}_e^u(s) - H_e^u(s))$$
$$+ \int_{-\infty}^{t} \left[\frac{1}{1 - \hat{H}_e(s)} - \frac{1}{1 - H_e(s)}\right] d(\hat{H}_e^u(s) - H_e^u(s)).$$

The last term on the right hand side is $o_P(n^{-1/2})$ by Proposition 5.17. Using Proposition 5.15, the sum of the first and second term can be written as

$$\int_{-\infty}^{t} \frac{\hat{H}_e(s) - H_e(s)}{(1 - H_e(s))^2} dH_e^u(s) + \int_{-\infty}^{t} \frac{1}{1 - H_e(s)} d(\hat{H}_e^u(s) - H_e^u(s)) + o_P(n^{-1/2})$$
$$= n^{-1} \sum_{i=1}^{n} \varphi(X_i, T_i, \Delta_i, t)(1 - F_e(t))^{-1} + o_P(n^{-1/2}),$$

where the equality follows by Propositions 5.13 and 5.14. Now, write (using a Taylor expansion)

$$\log(1-\hat{F}_e(t)) + \int_{-\infty}^{t} \frac{d\hat{H}_e^u(s)}{1-\hat{H}_e(s-)}$$

$$= -\frac{1}{2}\sum_{i=1}^{n} \frac{I(\hat{E}_{(i)} \leq t, \Delta_{(i)} = 1)}{(n-i+1)^2} \frac{1}{(1-R_i)^2} = O(n^{-1}) \qquad \text{a.s.,}$$

uniformly in t, where R_i is between 0 and $(n-i+1)^{-1}$. The result now follows by noting that $\hat{F}_e(t) - F_e(t) = -(1-F_e(t))(\log(1-\hat{F}_e(t)) - \log(1-F_e(t))) + o_P(n^{-1/2})$, and that $\log(1-F_e(t)) = -\int_{-\infty}^{t}(1-H_e(s))^{-1}dH_e^u(s)$.

We continue with the weak convergence result for the process $n^{1/2}(\hat{F}_e(\cdot) - F_e(\cdot))$. The main novelty here is that the residuals are based on nonparametric regression. In this sense, Theorem 5.2 below (which, clearly, also holds in the uncensored case) extends the classical results of Durbin (1973) and Loynes (1980) concerning the weak convergence of the empirical distribution function when parameters are estimated.

For the proof of the weak convergence result below, we make use of the techniques developed in van der Vaart and Wellner (1996). This method, which requires some knowledge of the theory of bracketing numbers, turned out to be much more powerful than the classical theory for showing weak convergence stated in e.g. Billingsley (1968).

Theorem 5.2
Under the assumptions of Theorem 5.1, the process $n^{1/2}(\hat{F}_e(t) - F_e(t))$, $-\infty < t \leq \bar{T}$ converges weakly to a zero-mean Gaussian process $Z(t)$ with covariance function

$$Cov(Z(t), Z(t')) = E(\varphi(X,T,\Delta,t)\varphi(X,T,\Delta,t')).$$

Proof. We will make use of Theorem 2.5.6 in van der Vaart and Wellner (1996), i.e. we will show that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty, \qquad (5.5)$$

where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of (X, T, Δ), $L_2(P)$ is the L_2-norm, and

$$\mathcal{F} = \{\varphi(X, T, \Delta, t); -\infty < t \leq \bar{T}\}.$$

Proving this entails that the class $\mathcal{F}$ is Donsker and hence the weak convergence of the given process follows from p. 81-82 in van der Vaart and Wellner (1996). Since the functions $x \to S_e(t)\gamma_i(t|x)$ $(i = 1, 2)$ are bounded (uniformly in t), as well as their first derivatives, their bracketing number is $O(\exp(K\varepsilon^{-1}))$ by Corollary 2.7.2 of the aforementioned book. Hence, since $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$ are uniformly bounded, the bracketing number of the second and third term of $\varphi(x, z, \delta, t)$ is $O(\exp(K\varepsilon^{-1}))$. Next, note that the first term of $\xi_e(\frac{z-m(x)}{\sigma(x)}, \delta, t)$ is decreasing in $\frac{z-m(x)}{\sigma(x)}$. Hence, its bracketing number is $m = O(\exp(K\varepsilon^{-1}))$ by Theorem 2.7.5 in van der Vaart and Wellner (1996). Also the class of functions of the form $\frac{z-m(x)}{\sigma(x)} \to (1 - F_e(t))I(\frac{z-m(x)}{\sigma(x)} \leq t)$ needs m brackets by Theorem 2.7.5. Finally, we note that since $I(\delta = 1)(1 - H_e(\frac{z-m(x)}{\sigma(x)}))^{-1}$ is bounded (for $\frac{z-m(x)}{\sigma(x)} \leq \bar{T}$) and independent of t, the second term of $\xi_e(\frac{z-m(x)}{\sigma(x)}, \delta, t)$ has bracketing number m. This concludes the proof, since the integration in (5.5) can be restricted to the interval $[0, 2M]$, if $|\varphi(x, z, \delta, t)| \leq M$ for all x, z, δ and t (for $\varepsilon > 2M$ we take $N_{[]}(\varepsilon, \mathcal{F}, L_2(P)) = 1$).

5.3 Asymptotic representation and weak convergence of the conditional distribution process

The results of the previous section enable us to establish the two main results of this chapter : an asymptotic representation and the weak convergence of the estimator $\hat{F}(t|x)$ of the conditional distribution.

Theorem 5.3

Assume (A1)-(A4), $H(t|x)$ and $H^u(t|x)$ satisfy (A5) $(i) - (vi)$, and $H_e(t|x)$ and

$H_e^u(t|x)$ satisfy (A5) (ii, iii, vi, vii). Then,

$$\hat{F}(t|x) - F(t|x) = \hat{F}_e\left(\frac{t-\hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{t-m(x)}{\sigma(x)}\right)$$

$$= (nh_n)^{-1}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) h_{x,t}(T_i,\Delta_i) + R_n(x,t),$$

where $\sup\{|R_n(x,t)|; (x,t)\in\Omega\} = o_P((nh_n)^{-1/2})$ *and*

$$h_{x,t}(z,\delta) = \left[\eta(z,\delta|x) + \zeta(z,\delta|x)\frac{t-m(x)}{\sigma(x)}\right] f_e\left(\frac{t-m(x)}{\sigma(x)}\right) f_X^{-1}(x).$$

Proof. Write

$$\begin{aligned}
&\hat{F}(t|x) - F(t|x)\\
&= \left[\hat{F}_e\left(\frac{t-\hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{t-\hat{m}(x)}{\hat{\sigma}(x)}\right)\right] + \left[F_e\left(\frac{t-\hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{t-m(x)}{\hat{\sigma}(x)}\right)\right]\\
&\quad + \left[F_e\left(\frac{t-m(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{t-m(x)}{\sigma(x)}\right)\right]\\
&= \alpha_{n1}(x,t) + \alpha_{n2}(x,t) + \alpha_{n3}(x,t).
\end{aligned}$$

We start with $\alpha_{n2}(x,t)$.

$$\alpha_{n2}(x,t) = -\frac{\hat{m}(x)-m(x)}{\hat{\sigma}(x)} f_e\left(\frac{t-m(x)}{\hat{\sigma}(x)}\right) + \frac{1}{2}\left(\frac{\hat{m}(x)-m(x)}{\hat{\sigma}(x)}\right)^2 f_e'(A_x),$$

for some A_x between $\frac{t-m(x)}{\hat{\sigma}(x)}$ and $\frac{t-\hat{m}(x)}{\hat{\sigma}(x)}$. The second term on the right hand side is $O((nh_n)^{-1}\log h_n^{-1})$ a.s. by Proposition 5.7. For the first term, we first replace $\hat{\sigma}(x)$ by $\sigma(x)$ (using Proposition 5.7) and then apply Proposition 5.10. For $\alpha_{n3}(x,t)$ we have,

$$\begin{aligned}
\alpha_{n3}(x,t) = &-\frac{\hat{\sigma}(x)-\sigma(x)}{\hat{\sigma}(x)}\frac{t-m(x)}{\sigma(x)} f_e\left(\frac{t-m(x)}{\sigma(x)}\right)\\
&+\frac{1}{2}\left(\frac{\hat{\sigma}(x)-\sigma(x)}{\hat{\sigma}(x)}\right)^2\left(\frac{t-m(x)}{\sigma(x)}\right)^2 f_e'(B_x),
\end{aligned}$$

where B_x is between $\frac{t-m(x)}{\sigma(x)}$ and $\frac{t-m(x)}{\hat{\sigma}(x)}$. The second term above is $O((nh_n)^{-1}\log h_n^{-1})$ a.s. by Proposition 5.7 and the fact that $\sup_t |t^2 f_e'(t)| < \infty$, and the first term has an asymptotic representation given by Proposition 5.11. The above show that

$$\begin{aligned} &\alpha_{n2}(x,t) + \alpha_{n3}(x,t) \qquad (5.6)\\ &= (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) h_{x,t}(T_i, \Delta_i) + O((nh_n)^{-1}\log h_n^{-1}) \quad \text{a.s.} \end{aligned}$$

Next note that $(nh_n)^{1/2}\alpha_{n1}(x,t) = o_P(1)$, uniformly in $(x,t) \in \Omega$, follows from the order of the remainder term in Theorem 5.1, and the weak convergence established in Theorem 5.2. This and (5.6) complete the proof of the theorem.

Remark 5.1
Condition (A1)(ii) is needed for the proof of Theorem 5.1 which is used in the proof of Theorem 5.3. However, the proof of Theorem 5.3 requires only that the remainder term in Theorem 5.1 satisfies $\sup\{|R_n(t)|; -\infty < t \le \bar{T}\} = o_P((nh_n)^{-1/2})$. This rate of convergence can be obtained by using $nh_n^5 \to 0$ instead of $nh_n^4 \to 0$ of assumption (A1)(ii).

Theorem 5.4
Under the assumptions of Theorem 5.3, the process $(nh_n)^{1/2}(\hat{F}(t|x) - F(t|x))$, $x \in R_X$ *fixed,* $\frac{t-m(x)}{\sigma(x)} \le \bar{T}$ *converges weakly to a zero-mean Gaussian process* $Z(t|x)$ *with covariance function*

$$Cov(Z(t|x), Z(t'|x)) = f_X(x)\, \|K\|_2^2\, Cov(h_{x,t}(T,\Delta), h_{x,t'}(T,\Delta)|X = x). \qquad (5.7)$$

Proof. Let

$$Z_{ni}(t|x) = (nh_n)^{-1/2} K\left(\frac{x-X_i}{h_n}\right) h_{x,t}(T_i, \Delta_i),$$

$i = 1, \ldots, n$, be a triangular array of random processes, with $t \in \mathcal{F} = \{t; \frac{t-m(x)}{\sigma(x)} \le \bar{T}\}$ and x fixed. We will show the validity of the conditions in Theorem 2.11.9 in van der Vaart and Wellner (1996) for $\sum_{i=1}^{n} Z_{ni}$. Endow $\mathcal{F}$ with the semimetric ρ defined by

$$\rho(t,t') = \max\left\{\left|f_e\left(\frac{t'-m(x)}{\sigma(x)}\right) - f_e\left(\frac{t-m(x)}{\sigma(x)}\right)\right|,\right.$$
$$\left.\left|\frac{t'-m(x)}{\sigma(x)} f_e\left(\frac{t'-m(x)}{\sigma(x)}\right) - \frac{t-m(x)}{\sigma(x)} f_e\left(\frac{t-m(x)}{\sigma(x)}\right)\right|\right\}.$$

Since $|f_e(z)|$ and $|zf_e(z)|$ are bounded for $z \in I\!R$, we can divide $\mathcal{F}$ for every $\varepsilon > 0$ into $N_\varepsilon = O(\varepsilon^{-1})$ subintervals $\mathcal{F}_{\varepsilon j}$ $(j = 1, \ldots, N_\varepsilon)$, such that $\rho(t,t') \le C\varepsilon$ $(C > 0)$ for all $t, t' \in \mathcal{F}_{\varepsilon j}$ and hence

$$\sum_{i=1}^{n} \sup_{t,t' \in \mathcal{F}_{\varepsilon j}} |Z_{ni}(t'|x) - Z_{ni}(t|x)|^2 \le \varepsilon^2$$

by proper choice of C, since $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$ are bounded as well. This shows that the bracketing number is $O(\varepsilon^{-1})$ and hence their third condition is satisfied. The other two conditions are easily seen to hold.

It remains to calculate the covariance of the limiting process. Let $\boldsymbol{X} = (X_1, \ldots, X_n)'$. Since $\sup_{x,t} |E(\hat{H}(t|x)|\boldsymbol{X}) - H(t|x)|$ is easily seen to be $O(h_n^2)$ (see e.g. the proof of Proposition 5.3), it follows from (5.20) that $\sum_{i=1}^n W_{ni}(x;h_n) E(\xi(T_i, \Delta_i, t|x)|\boldsymbol{X})$ $= O(h_n^2)$ uniformly in x and t. Noting that $h_{x,t}$ is defined in terms of the functions η and ζ which include the function ξ, it can be seen that the above implies $\sum_{i=1}^n W_{ni}(x;h_n) E(h_{x,t}(T_i,\Delta_i)|\boldsymbol{X}) = O(h_n^2)$. We will show that

$$(nh_n)^{-1} \sum_{i=1}^{n} E\left\{K^2\left(\frac{x - X_i}{h_n}\right)\gamma(X_i, x, t, t')\right\} \tag{5.8}$$

tends to expression (5.7), where

$$\gamma(X_i, x, t, t') = E\{[h_{x,t}(T_i,\Delta_i) - E(h_{x,t}(T_i,\Delta_i)|X_i)]$$
$$\times [h_{x,t'}(T_i,\Delta_i) - E(h_{x,t'}(T_i,\Delta_i)|X_i)]|\boldsymbol{X}\}.$$

Using the fact that for all $s \le t$,

$$E\{[I(T_i \le s) - H(s|X_i)][I(T_i \le t) - H(t|X_i)]|\boldsymbol{X}\}$$
$$= H(s|x) - H(s|x)H(t|x) + O(h_n)$$

holds whenever $K\left(\frac{X_i-x}{h_n}\right) \neq 0$, it follows after some simple algebra that $|\gamma(X_i, x, t, t')$ -$\gamma(x, x, t, t')| = O(h_n)$ for $K\left(\frac{X_i-x}{h_n}\right) \neq 0$, again by using (5.20). Hence, (5.8) equals

$$\begin{aligned} & f_X(x)\, \|K\|_2^2\, \gamma(x, x, t, t') + o(1) \\ &= f_X(x)\, \|K\|_2^2\, \mathrm{Cov}(h_{x,t}(T, \Delta), h_{x,t'}(T, \Delta)|X = x) + o(1). \end{aligned}$$

Remark 5.2

It is easily seen that in the uncensored case and when $J(s) \equiv 1$ (i.e. $m(x) = E(Y|x)$ and $\sigma^2(x) = \mathrm{Var}(Y|x)$), the asymptotic variance of $(nh_n)^{1/2}(\hat{F}(t|x) - F(t|x))$ reduces to

$$f_X^{-1}(x)\, \|K\|_2^2\, f_e^2\left(\frac{t-m(x)}{\sigma(x)}\right)\left[\frac{1}{4}\mathrm{Var}(\varepsilon^2)\left(\frac{t-m(x)}{\sigma(x)}\right)^2 + E(\varepsilon^3)\frac{t-m(x)}{\sigma(x)} + 1\right], \quad (5.9)$$

while the asymptotic variance of $(nh_n)^{1/2}\sum_{i=1}^n W_{ni}(x; h_n)(I(Y_i \leq t) - F(t|x))$, which corresponds to the usual kernel estimator $\tilde{F}(t|x)$, is

$$f_X^{-1}(x)\, \|K\|_2^2\, F(t|x)(1 - F(t|x)). \quad (5.10)$$

In the special case where the distribution of the error variable ε is standard normal, expression (5.9) divided by the common factor $f_X^{-1}(x)\, \|K\|_2^2$ of the two variance functions, reduces to $f_e^2(\frac{t-m(x)}{\sigma(x)})(1 + \frac{1}{2}(\frac{t-m(x)}{\sigma(x)})^2)$, while expression (5.10) divided by this factor equals $F_e(\frac{t-m(x)}{\sigma(x)})(1 - F_e(\frac{t-m(x)}{\sigma(x)}))$. From Figure 5.1, which shows the graph of these two functions, it follows that $\hat{F}(t|x)$ has smaller variance than $\tilde{F}(t|x)$ for all t, and hence the former estimator should be used when model (1.28) can be assumed.

5.4 Appendix A : Some auxiliary results

The asymptotic representation and weak convergence results obtained in Sections 5.2 and 5.3 required some results concerning the location estimator $\hat{m}(x)$ and the scale estimator $\hat{\sigma}(x)$, which we will prove in this section. In particular, the uniform consistency of $\hat{m}(x)$, $\hat{m}'(x)$, $(\hat{m}'(x) - \hat{m}'(y))/(x - y)$ and their analogues for

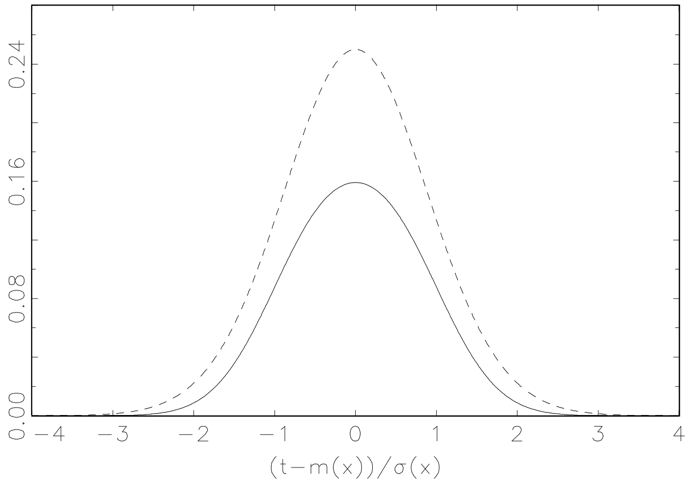

Figure 5.1: *Graph of the variance function of $\hat{F}(t|x)$ (full line) and $\tilde{F}(t|x)$ (dashed line) divided by their common factor $f_X^{-1}(x)\,\|K\|_2^2$.*

$\hat{\sigma}(x)$ will be established, as well as asymptotic representations for $\hat{m}(x)$ and $\hat{\sigma}(x)$. The proofs of these results require the uniform (both in x and t) consistency of the Stone and Beran estimators and their derivatives.

In the uncensored case, the consistency of derivatives of a regression function has been studied by Schuster and Yakowitz (1979), Gasser and Müller (1984), Härdle and Gasser (1985), Mack and Müller (1989), Rychlik (1990), among others. The uniform consistency of a regression function has been considered by Devroye (1978), Schuster and Yakowitz (1979), Müller and Stadtmüller (1987) and Härdle, Janssen and Serfling (1988). In the latter paper strong uniform consistency rates are established for kernel type estimators of functionals of the conditional distribution function, under general conditions.

We start below with showing some results concerning the Stone estimator $\hat{H}(t|x)$ (Propositions 5.3 and 5.4), from which similar results concerning the Beran estimator will follow (Propositions 5.5 and 5.6). From these results, asymptotic properties of $\hat{m}(x)$ and $\hat{\sigma}(x)$ will be derived (Propositions 5.7 - 5.11).

Proposition 5.3
Assume (A1) (i, iii, iv), *the distribution function* F_X *is twice continuously differentiable,* $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(t|x)$ *satisfies (A5)* (i, iv, v). *Then,*

$$\sup_{x\in R_X} \sup_{-\infty<t<\infty} |\hat{H}(t|x) - H(t|x)| = O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2}) \qquad \text{a.s.}$$
$$\sup_{x\in R_X} \sup_{-\infty<t<\infty} |\dot{\hat{H}}(t|x) - \dot{H}(t|x)| = O((nh_n^3)^{-1/2}(\log h_n^{-1})^{1/2}) \qquad \text{a.s.}$$

Proof. We will show the second statement. The proof for the first one is similar. Let $c_n = K_1(nh_n^3)^{-1/2}(\log h_n^{-1})^{1/2}$. Write

$$\hat{H}(t|x) = \frac{(nh_n)^{-1}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) I(T_i \le t)}{(nh_n)^{-1}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} = \frac{\hat{H}_f(t|x)}{\hat{f}_X(x)} \qquad \text{(say)}$$

and let $H(t|x) = \frac{H_f(t|x)}{f_X(x)}$. Then,

$$\begin{aligned}
&\dot{\hat{H}}(t|x) - \dot{H}(t|x)\\
&= \frac{\dot{\hat{H}}_f(t|x) - \dot{H}_f(t|x)}{\hat{f}_X(x)} - \frac{\dot{H}_f(t|x)[\hat{f}_X(x) - f_X(x)]}{f_X(x)\hat{f}_X(x)}\\
&\quad - \frac{\hat{H}_f(t|x)[\hat{f}'_X(x) - f'_X(x)]}{\hat{f}_X^2(x)} + \frac{\hat{H}_f(t|x)f'_X(x)[\hat{f}_X^2(x) - f_X^2(x)]}{f_X^2(x)\hat{f}_X^2(x)}\\
&\quad - \frac{[\hat{H}_f(t|x) - H_f(t|x)]f'_X(x)}{f_X^2(x)}.
\end{aligned}$$

In what follows, we will show that the first term above is $O(c_n)$ a.s. uniformly in x and t. The proofs for the other terms are similar. Let $\boldsymbol{X} = (X_1, \ldots, X_n)'$ and set

$$\begin{aligned}
\dot{\hat{H}}_f(t|x) - \dot{H}_f(t|x) &= \{\dot{\hat{H}}_f(t|x) - E[\dot{\hat{H}}_f(t|x)|\boldsymbol{X}]\}\\
&\quad + \{E[\dot{\hat{H}}_f(t|x)|\boldsymbol{X}] - E[\dot{\hat{H}}_f(t|x)]\} + \{E[\dot{\hat{H}}_f(t|x)] - \dot{H}_f(t|x)\}\\
&= \beta_{n1}(x,t) + \beta_{n2}(x,t) + \beta_{n3}(x,t).
\end{aligned}$$

Partition R_X into $k_n = O(n)$ subintervals $[x_j, x_{j+1}]$ such that $x_{j+1} - x_j \le n^{-1}$. Then,

$$\sup_{x,t} |\beta_{n1}(x,t)|$$

$$\leq (nh_n^2)^{-1} \sup_{t,j} \sup_{x_j \leq x \leq x_{j+1}} \left| \sum_{i=1}^n \left\{ K'\left(\frac{x - X_i}{h_n}\right) - K'\left(\frac{x_j - X_i}{h_n}\right) \right\} \{I(T_i \leq t) - H(t|X_i)\} \right|$$
$$+ (nh_n^2)^{-1} \sup_{t,j} \left| \sum_{i=1}^n K'\left(\frac{x_j - X_i}{h_n}\right) \{I(T_i \leq t) - H(t|X_i)\} \right|.$$

The first term above is easily seen to be $O((nh_n^2)^{-1})$ a.s. Applying the Lemma in Singh (1975) (use $a = c_n$ and $w_i = (nh_n^2)^{-1} K'(\frac{x_j - X_i}{h_n})$), together with the Borel-Cantelli lemma, it can be shown that the second term is $O(c_n)$ a.s. For $\beta_{n2}(x,t)$, we need to partition the line into $m_n = O(n^2)$ subintervals $[t_l, t_{l+1}]$ such that $H(t_{l+1}|x_j) - H(t_l|x_j) \leq n^{-1}$ for all j, l. Then,

$$\sup_{x,t} |\beta_{n2}(x,t)| \leq \max_j \sup_{x_j \leq x \leq x_{j+1}} \sup_t |\beta_{n2}(x,t) - \beta_{n2}(x_j,t)|$$
$$+ \max_{j,l} \sup_{t_l \leq t \leq t_{l+1}} |\beta_{n2}(x_j,t) - \beta_{n2}(x_j,t_l)| + \max_{j,l} |\beta_{n2}(x_j,t_l)| \quad (5.11)$$

The first term above is easily seen to be $O((nh_n^2)^{-1})$ a.s. For the second term write

$$\beta_{n2}(x_j,t) - \beta_{n2}(x_j,t_l) = \frac{1}{nh_n^2} \sum_i K'\left(\frac{x_j - X_i}{h_n}\right) [H(t|X_i) - H(t_l|X_i)]$$
$$- \frac{1}{nh_n^2} \sum_i E\left(K'\left(\frac{x_j - X_i}{h_n}\right) [H(t|X_i) - H(t_l|X_i)]\right). \quad (5.12)$$

Since the two terms on the right hand side of (5.12) are similar, we only consider the first term. Let x_{j_i} denote the grid point in the partitioning of R_X which is closest to X_i. Then,

$$\frac{1}{nh_n^2} \sum_i K'\left(\frac{x_j - X_i}{h_n}\right) [H(t|X_i) - H(t_l|X_i)]$$
$$= \frac{1}{nh_n^2} \sum_i K'\left(\frac{x_j - X_i}{h_n}\right) [H(t|X_i) - H(t|x_{j_i})]$$
$$+ \frac{1}{nh_n^2} \sum_i K'\left(\frac{x_j - X_i}{h_n}\right) [H(t|x_{j_i}) - H(t_l|x_{j_i})]$$
$$+ \frac{1}{nh_n^2} \sum_i K'\left(\frac{x_j - X_i}{h_n}\right) [H(t_l|x_{j_i}) - H(t_l|X_i)].$$

From this and (5.12) it can be seen that the second term in (5.11) is bounded by (where $\|K'\|_\infty = \sup_u |K'(u)|$)

$$\frac{2}{h_n}\|K'\|_\infty \max_{j,l} |H(t_{l+1}|x_j) - H(t_l|x_j)| + \frac{4}{h_n}\|K'\|_\infty \sup_{t,j} \sup_{x_j \le x \le x_{j+1}} |H(t|x) - H(t|x_j)|,$$

which is $O((nh_n)^{-1})$ a.s. On the last term in (5.11) we can apply Bernstein's inequality. Using integration by parts, the term $\beta_{n3}(x,t)$ can be written as

$$\begin{aligned}\beta_{n3}(x,t) &= -\frac{1}{h_n}\int H(t|z) f_X(z)\, dK\left(\frac{x-z}{h_n}\right) - \dot{H}_f(t|x) \\ &= \frac{1}{h_n}\int K\left(\frac{x-z}{h_n}\right)\dot{H}(t|z) f_X(z)\, dz \\ &\quad + \frac{1}{h_n}\int K\left(\frac{x-z}{h_n}\right) H(t|z) f_X'(z)\, dz - \dot{H}_f(t|x) \\ &= \beta_{31}(x,t) + \beta_{32}(x,t) - \dot{H}_f(t|x). \end{aligned} \tag{5.13}$$

We decompose $\beta_{31}(x,t)$ in the following way :

$$\begin{aligned}\beta_{31}(x,t) &= \frac{1}{h_n}\int K\left(\frac{x-z}{h_n}\right)[\dot{H}(t|z) - \dot{H}(t|x)][f_X(z) - f_X(x)]\, dz \\ &\quad + \frac{1}{h_n} f_X(x) \int K\left(\frac{x-z}{h_n}\right)[\dot{H}(t|z) - \dot{H}(t|x)]\, dz \\ &\quad + \frac{1}{h_n}\dot{H}(t|x)\int K\left(\frac{x-z}{h_n}\right)[f_X(z) - f_X(x)]\, dz \\ &\quad + \frac{1}{h_n}\dot{H}(t|x) f_X(x)\int K\left(\frac{x-z}{h_n}\right) dz.\end{aligned}$$

The last term above equals $\dot{H}(t|x)f_X(x)$, while the first one is $O(h_n^2)$. The second and third terms are easily seen to be $O(h_n)$. In a similar way we can show that $\beta_{32}(x,t) = H(t|x)f_X'(x) + O(h_n)$. Noting that $\dot{H}_f(t|x) = \dot{H}(t|x)f_X(x) + H(t|x)f_X'(x)$, (5.13) implies that $\beta_{n3}(x,t)$ is $O(h_n)$ uniformly in x and t. Assumption $(A1)(i)$ implies that this is the same as $O(c_n)$, and this completes the proof.

Proposition 5.4
Assume (A1) (i, iii, iv), the distribution function F_X is twice continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$, and $H(t|x)$ satisfies (A5) (i, iv, v). Then,

$$\sup_{x,y,t} \frac{|\dot{\hat{H}}(t|x) - \dot{H}(t|x) - \dot{\hat{H}}(t|y) + \dot{H}(t|y)|}{|x-y|^\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log h_n^{-1})^{1/2}) \quad \text{a.s.}$$

where $\delta > 0$ is as in assumption (A1).

Proof. As in Proposition 5.3 it suffices to consider (using the same notations)

$$\begin{aligned}
&\dot{\hat{H}}_f(t|x) - \dot{\hat{H}}_f(t|y) - \dot{H}_f(t|x) + \dot{H}_f(t|y) \\
&= \dot{\hat{H}}_f(t|x) - \dot{\hat{H}}_f(t|y) - E[\dot{\hat{H}}_f(t|x) - \dot{\hat{H}}_f(t|y)|\boldsymbol{X}] \\
&\quad + E[\dot{\hat{H}}_f(t|x) - \dot{\hat{H}}_f(t|y)|\boldsymbol{X}] - E[\dot{\hat{H}}_f(t|x) - \dot{\hat{H}}_f(t|y)] \\
&\quad + E[\dot{\hat{H}}_f(t|x) - \dot{\hat{H}}_f(t|y)] - \dot{H}_f(t|x) + \dot{H}_f(t|y) \\
&= (\beta_{n1}(x,y,t) + \beta_{n2}(x,y,t) + \beta_{n3}(x,y,t))(x-y)^\delta,
\end{aligned}$$

where $\boldsymbol{X} = (X_1, \ldots, X_n)'$. Partition R_X into $k_n = O(n)$ subintervals $[x_j, x_{j+1}]$ such that $x_{j+1} - x_j \le n^{-1}$. Then, for $|x-y| \ge n^{-1}$ (for $|x-y| \le n^{-1}$, the proof is similar as for the first term below)

$$\begin{aligned}
&\sup_{x,y,t} |\beta_{n1}(x,y,t)| \\
&\le 2(nh_n^2)^{-1} \sup_{t,j} \sup_{x_j \le x \le x_{j+1}} \Bigg| \sum_{i=1}^n (x-x_j)^{-\delta} \left\{ K'\left(\frac{x-X_i}{h_n}\right) - K'\left(\frac{x_j - X_i}{h_n}\right) \right\} \\
&\qquad \times (I(T_i \le t) - H(t|X_i)) \Bigg| \\
&\quad + (nh_n^2)^{-1} \sup_t \max_{j,k} \Bigg| \sum_{i=1}^n (x_j - x_k)^{-\delta} \left\{ K'\left(\frac{x_j - X_i}{h_n}\right) - K'\left(\frac{x_k - X_i}{h_n}\right) \right\} \\
&\qquad \times (I(T_i \le t) - H(t|X_i)) \Bigg| .
\end{aligned}$$

The first term above is clearly $O((n^{1-\delta}h_n^2)^{-1})$ a.s. which is $o((nh_n^{3+2\delta})^{-1/2}(\log h_n^{-1})^{1/2})$. For the second term note that if $|x_j - x_k| \le h_n$, then $|K'(\frac{x_j - X_i}{h_n}) - K'(\frac{x_k - X_i}{h_n})|/|x_j -$

$x_k|^\delta \leq \|K''\|h_n^{-\delta}$ and if $|x_j - x_k| > h_n$, then $|K'(\frac{x_j - X_i}{h_n}) - K'(\frac{x_k - X_i}{h_n})|/|x_j - x_k|^\delta$ is bounded by $2\|K'\|h_n^{-\delta}$. Using these bounds, it can be seen that the conditions in the Lemma in Singh (1975) hold with $w_i = (nh_n^2)^{-1}[K'(\frac{x_j - X_i}{h_n}) - K'(\frac{x_k - X_i}{h_n})]/(x_j - x_k)^\delta$ and $a = K_1(nh_n^{3+2\delta})^{-1/2}(\log h_n^{-1})^{1/2}(K_1 > 0)$. From this it is easily seen via the Borel-Cantelli lemma that the second term above is of the right order uniformly in j and k. For $\beta_{n2}(x, y, t)$, we divide the line into $m_n = O(n^2)$ subintervals $[t_l, t_{l+1}]$ such that $H(t_{l+1}|x_j) - H(t_l|x_j) \leq n^{-1}$ for all j, l. Then, for $|x - y| \geq n^{-1}$, we have (if $|x - y| \leq n^{-1}$, the derivation is again similar as for the first term below)

$$\sup_{x,y,t} |\beta_{n2}(x, y, t)| \leq 2 \sup_{t,j} \max_{x_j \leq x \leq x_{j+1}} |\beta_{n2}(x, x_j, t)| + 2 \sup_{x,y} \max_{j,l} \sup_{t_l \leq t \leq t_{l+1}} \frac{|\beta_{n2}(x_j, t) - \beta_{n2}(x_j, t_l)|}{|x - y|^\delta} + \max_{j,k,l} |\beta_{n2}(x_j, x_k, t_l)|$$

where $\beta_{n2}(x, t)$ is as in Proposition 5.3. Using what we have shown in Proposition 5.3 for the term $\beta_{n2}(x, t)$, it suffices to show that $\max_{j,k,l} |\beta_{n2}(x_j, x_k, t_l)|$ is of the desired order and this is easily done by Bernstein's inequality, since $|K'(\frac{x_j - X_i}{h_n}) - K'(\frac{x_k - X_i}{h_n})|/|x_j - x_k|^\delta \leq K_1 h_n^{-\delta}$ for some $K_1 > 0$ as before. Finally, for $\beta_{n3}(x, y, t)$, we first consider the case $|x - y| \leq h_n$. Then, $\beta_{n3}(x, y, t)$ equals

$$\frac{1}{h_n(x - y)^\delta} \int \left\{ K\left(\frac{x - z}{h_n}\right) - K\left(\frac{y - z}{h_n}\right) \right\} \dot{H}(t|z) f_X(z)\, dz + \frac{1}{h_n(x - y)^\delta} \int \left\{ K\left(\frac{x - z}{h_n}\right) - K\left(\frac{y - z}{h_n}\right) \right\} H(t|z) f'_X(z)\, dz - \frac{\dot{H}_f(t|x) - \dot{H}_f(t|y)}{(x - y)^\delta}.$$

By decomposing the first two terms above in a similar way as in the proof of Proposition 5.3, we find the order $O(h_n^{1-\delta}) = O((nh_n^{3+2\delta})^{-1/2}(\log h_n^{-1})^{1/2})$. For the case $|x - y| \geq h_n$, we write $\beta_{n3}(x, y, t) = (\beta_{n3}(x, t) - \beta_{n3}(y, t))/(x - y)^\delta$, where $\beta_{n3}(x, t)$ is as in Proposition 5.3. Since we showed there that $\beta_{n3}(x, t)$ is $O(h_n)$ uniformly in x and t, we now have the order $O(h_n^{1-\delta})$. This completes the proof.

Remark 5.3

Replacing all conditions on H by analogous conditions on the subdistribution H^u, it is easily seen that the results of Propositions 5.3 and 5.4 are also valid for H^u.

We are now ready to show some results concerning the Beran estimator $\tilde{F}(t|x)$.

Proposition 5.5
Assume (A1) (i, iii, iv), the distribution function F_X is twice continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$, *and $H(t|x)$ and $H^u(t|x)$ satisfy (A5) (i, iv, v). Then,*

$$\sup_x \sup_{t \leq \tilde{T}_x} |\tilde{F}(t|x) - F(t|x)| = O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2}) \qquad \text{a.s.}$$

$$\sup_x \sup_{t \leq \tilde{T}_x} |\dot{\tilde{F}}(t|x) - \dot{F}(t|x)| = O((nh_n^3)^{-1/2}(\log h_n^{-1})^{1/2}) \qquad \text{a.s.}$$

Proof. As in the proof of Theorem 2.1, one can show that for any $\varepsilon > 0$:

$$P\left(\sup_{x,t} |\tilde{F}(t|x) - F(t|x)| > \varepsilon\right)$$

$$\leq P\left(\sup_{x,t} |\hat{H}(t|x) - H(t|x)| > \frac{\varepsilon\delta_1^2}{12}\right) + P\left(\sup_{x,t} |\hat{H}^u(t|x) - H^u(t|x)| > \frac{\varepsilon\delta_1^2}{12}\right),$$

where $\inf_x(1 - H(\tilde{T}_x|x)) > \delta_1 > 0$. Choosing $\varepsilon = K_1(nh_n)^{-1/2}(\log h_n^{-1})^{1/2}$ and applying Proposition 5.3 and Remark 5.3, entails that $\sup_{x,t} |\tilde{F}(t|x) - F(t|x)|$ is of the desired order. To study $\dot{\tilde{F}}(t|x) - \dot{F}(t|x)$, we first consider

$$\dot{\hat{\Lambda}}(t|x) - \dot{\Lambda}(t|x) = \frac{\partial}{\partial x}\left\{\int_{-\infty}^{t} \frac{d\hat{H}^u(s|x)}{1 - \hat{H}(s-|x)} - \int_{-\infty}^{t} \frac{dH^u(s|x)}{1 - H(s|x)}\right\}. \tag{5.14}$$

This can be decomposed into several terms using

$$\dot{\Lambda}(t|x) = \int_{-\infty}^{t} \frac{d\dot{H}^u(s|x)}{1 - H(s|x)} + \int_{-\infty}^{t} \frac{\dot{H}(s|x)}{(1 - H(s|x))^2}\, dH^u(s|x) \tag{5.15}$$

and similarly for $\dot{\hat{\Lambda}}$. Using Proposition 5.3 and Remark 5.3, each of these terms (and hence (5.14)) is readily seen to be $O(c_n)$ a.s. (here $c_n = K_1(nh_n^3)^{-1/2}(\log h_n^{-1})^{1/2}$ as before). Finally, for $\dot{\tilde{F}}(t|x) - \dot{F}(t|x)$, it suffices to consider $\frac{\partial}{\partial x}\log(1 - \tilde{F}(t|x)) - \frac{\partial}{\partial x}\log(1 - F(t|x))$, because

$$\dot{\tilde{F}}(t|x) - \dot{F}(t|x) = -\frac{\tilde{F}(t|x) - F(t|x)}{1 - F(t|x)}\dot{F}(t|x)$$

$$-(1 - \tilde{F}(t|x))\left(\frac{\partial}{\partial x}\log(1 - \tilde{F}(t|x)) - \frac{\partial}{\partial x}\log(1 - F(t|x))\right).$$

Write

$$V_i(x, h_n) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^n I(T_j \geq T_i) K\left(\frac{x-X_j}{h_n}\right)}, \tag{5.16}$$

and note that

$$\hat{\Lambda}(t|x) = \sum_{i=1}^n V_i(x, h_n) I(T_i \leq t, \Delta_i = 1).$$

Using this and the relation $\Lambda(t|x) = -\log(1 - F(t|x))$ it follows that

$$\begin{aligned}
&\frac{\partial}{\partial x} \log(1 - \tilde{F}(t|x)) - \frac{\partial}{\partial x} \log(1 - F(t|x)) \\
&= \frac{\partial}{\partial x} \sum_{i=1}^n \log(1 - V_i(x, h_n)) I(T_i \leq t, \Delta_i = 1) + \dot{\Lambda}(t|x) \\
&= \left\{ -\sum_{i=1}^n \frac{V_i'(x, h_n)}{1 - V_i(x, h_n)} I(T_i \leq t, \Delta_i = 1) + \dot{\hat{\Lambda}}(t|x) \right\} - \left\{ \dot{\hat{\Lambda}}(t|x) - \dot{\Lambda}(t|x) \right\} \\
&= -\sum_{i=1}^n \frac{V_i(x, h_n) V_i'(x, h_n)}{1 - V_i(x, h_n)} I(T_i \leq t, \Delta_i = 1) + O(c_n) = O(c_n)
\end{aligned}$$

a.s., since $\sum_{j=1}^n I(T_j \geq T_i) K(\frac{x-X_j}{h_n}) \geq K n h_n$ for n large, uniformly in x and $T_i \leq \tilde{T}_x$, because $\inf_x(1 - H(\tilde{T}_x|x)) > 0$. This finishes the proof.

Proposition 5.6

Assume (A1) (i, iii, iv), the distribution function F_X is twice continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$, and $H(t|x)$ and $H^u(t|x)$ satisfy (A5) (i, iv, v). Then,

$$\sup_{x,y} \sup_{t \leq \tilde{T}_x \wedge \tilde{T}_y} \frac{|\dot{\tilde{F}}(t|x) - \dot{F}(t|x) - \dot{\tilde{F}}(t|y) + \dot{F}(t|y)|}{|x - y|^\delta} = O((n h_n^{3+2\delta})^{-1/2} (\log h_n^{-1})^{1/2}) \qquad a.s.,$$

where $\delta > 0$ is as in assumption (A1).

Proof. The proof is along the same lines as the proof of Proposition 5.5. First note that it suffices to consider

$$\frac{\frac{\partial}{\partial x} \log(1 - \tilde{F}(t|x)) - \frac{\partial}{\partial x} \log(1 - F(t|x)) - \frac{\partial}{\partial y} \log(1 - \tilde{F}(t|y)) + \frac{\partial}{\partial y} \log(1 - F(t|y))}{(x - y)^\delta}$$

which equals

$$(x-y)^{-\delta}\left\{-\sum_{i=1}^{n}\left[\frac{V_i(x,h_n)V_i'(x,h_n)}{1-V_i(x,h_n)}-\frac{V_i(y,h_n)V_i'(y,h_n)}{1-V_i(y,h_n)}\right]I(T_i\le t,\Delta_i=1)\right\}$$

$$+(x-y)^{-\delta}\{\dot{\hat{\Lambda}}(t|x)-\dot{\Lambda}(t|x)-\dot{\hat{\Lambda}}(t|y)+\dot{\Lambda}(t|y)\}, \tag{5.17}$$

where $V_i(x,h_n)$ is as in (5.16). Using relation (5.15), the second term of (5.17) can be decomposed into several terms, which are all easily seen to be of the desired order by Propositions 5.3 and 5.4 and Remark A.1. Now, since $T_i\le t\le \tilde{T}_x\wedge\tilde{T}_y$ implies that $\sum_{j=1}^{n}I(T_j\ge T_i)K(\frac{u-X_j}{h_n})\ge Knh_n$ is satisfied for u equal to x and y, and since $|K(\frac{x-z}{h_n})-K(\frac{y-z}{h_n})|/|x-y|^\delta$ is $O(h_n^{-\delta})$ uniformly in x,y and z, it is easily seen that $|V_i(x,h_n)-V_i(y,h_n)|/|x-y|^\delta$ is $O((nh_n^{1+\delta})^{-1})$ a.s. and that $|V_i'(x,h_n)-V_i'(y,h_n)|/|x-y|^\delta$ is $O((nh_n^{2+\delta})^{-1})$ a.s. uniformly in x,y and $T_i\le\tilde{T}_x\wedge\tilde{T}_y$. This shows that the first term of (5.17) is $O((nh_n^{2+\delta})^{-1})$ a.s.

Proposition 5.7

Assume (A1) (i,iii,iv), (A2) (i), the function J is continuous, $\int_0^1 J(s)\,ds=1$, $J(s)\ge 0$ for all $0\le s\le 1$, the distribution function F_X is twice continuously differentiable, $\inf_{x\in R_X}f_X(x)>0$, and $H(t|x)$ and $H^u(t|x)$ satisfy (A5) (i,iv,v). Then,

$$\sup_x|\hat{m}(x)-m(x)|=O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2})\qquad\text{a.s.}$$

If in addition $\inf_x\sigma(x)>0$, then,

$$\sup_x|\hat{\sigma}(x)-\sigma(x)|=O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2})\qquad\text{a.s.}$$

Proof. The proof is along the same lines as the proof of Theorem 3.3 in Härdle, Janssen and Serfling (1988). Let $L(u)=\int_0^u J(s)\,ds$ for $0\le u\le 1$. Then,

$$\begin{aligned}&\sup_x|\hat{m}(x)-m(x)|\\&=\sup_x\left|\int_0^1[\tilde{F}^{-1}(s|x)-F^{-1}(s|x)]J(s)\,ds\right|\\&=\sup_x\left|\int_{-\infty}^{+\infty}[L(F(t|x))-L(\tilde{F}(t|x))]\,dt\right|\end{aligned}$$

$$= \sup_x \left| \int_0^1 \left[L(s) - L(\tilde{F}(F^{-1}(s|x)|x)) \right] (f(F^{-1}(s|x)|x))^{-1} \, ds \right|$$

$$\leq \left(\inf_x \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) \right)^{-1} \sup_{s_0 \leq s \leq s_1} J(s) \sup_x \sup_{t \leq \tilde{T}_x} |\tilde{F}(t|x) - F(t|x)|,$$

which is of the desired order by Proposition 5.5. The derivation for $\hat{\sigma}(x) - \sigma(x)$ is analogous. Since

$$\hat{\sigma}^2(x) - \sigma^2(x) = \int_0^{+\infty} [L(F(\sqrt{t}|x)) - L(\tilde{F}(\sqrt{t}|x))] \, dt$$
$$- \int_0^{+\infty} [L(F(-\sqrt{t}|x)) - L(\tilde{F}(-\sqrt{t}|x))] \, dt - (\hat{m}^2(x) - m^2(x)),$$

a similar procedure as before leads to the given order.

Proposition 5.8
Assume (A1) (i, iii, iv), (A2) (i), the function J is continuously differentiable, $\int_0^1 J(s)ds = 1$, $J(s) \geq 0$ for all $0 \leq s \leq 1$, the distribution function F_X is twice continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$, and $H(t|x)$ and $H^u(t|x)$ satisfy (A5) (i, iv, v). Then,

$$\sup_x |\hat{m}'(x) - m'(x)| = O((nh_n^3)^{-1/2} (\log h_n^{-1})^{1/2}) \qquad \text{a.s.}$$

If in addition the functions m and σ are continuously differentiable and $\inf_{x \in R_X} \sigma(x) > 0$, then,

$$\sup_x |\hat{\sigma}'(x) - \sigma'(x)| = O((nh_n^3)^{-1/2} (\log h_n^{-1})^{1/2}) \qquad \text{a.s.}$$

Proof. We restrict our proof to the first result. The second one can be shown in a completely similar way (as in Proposition 5.7). Write

$$\hat{m}(x) - m(x) = \int_{-\infty}^{+\infty} [L(F(t|x)) - L(\tilde{F}(t|x))] \, dt$$

where $L(u) = \int_0^u J(s)\, ds$ $(0 \le u \le 1)$. Then, $\hat{m}'(x) - m'(x)$ equals

$$\int_{-\infty}^{+\infty} [\{J(F(t|x)) - J(\tilde{F}(t|x))\}\dot{F}(t|x) - J(\tilde{F}(t|x))\{\dot{\tilde{F}}(t|x) - \dot{F}(t|x)\}]\, dt,$$

from which the result follows using Proposition 5.5.

Proposition 5.9

Assume (A1) (i, iii, iv), (A2), the distribution function F_X is twice continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$, and $H(t|x)$ and $H^u(t|x)$ satisfy (A5) (i, iv, v). Then,

$$\sup_{x,y} \frac{|\hat{m}'(x) - m'(x) - \hat{m}'(y) + m'(y)|}{|x-y|^\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log h_n^{-1})^{1/2}) \qquad a.s.$$

If in addition (A3) (ii) holds, then,

$$\sup_{x,y} \frac{|\hat{\sigma}'(x) - \sigma'(x) - \hat{\sigma}'(y) + \sigma'(y)|}{|x-y|^\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log h_n^{-1})^{1/2}) \qquad a.s.,$$

where $\delta > 0$ is as in assumption (A1).

Proof. Assume $\tilde{T}_y \le \tilde{T}_x$. In a similar way as in Proposition 5.8, we can write

$$\begin{aligned}
&\hat{m}'(x) - m'(x) - \hat{m}'(y) + m'(y) \\
&= \int_{-\infty}^{+\infty} [\{J(F(t|x)) - J(\tilde{F}(t|x)) - J(F(t|y)) + J(\tilde{F}(t|y))\}\dot{F}(t|x) \\
&\quad + \{J(F(t|y)) - J(\tilde{F}(t|y))\}\{\dot{F}(t|x) - \dot{F}(t|y)\} \\
&\quad - \{J(\tilde{F}(t|x)) - J(\tilde{F}(t|y))\}\{\dot{\tilde{F}}(t|x) - \dot{F}(t|x)\} \\
&\quad - J(\tilde{F}(t|y))\{\dot{\tilde{F}}(t|x) - \dot{F}(t|x) - \dot{\tilde{F}}(t|y) + \dot{F}(t|y)\}]\, dt.
\end{aligned}$$

Dividing the above terms by $(x-y)^\delta$ and applying Propositions 5.5 and 5.6, yields the desired order. A similar procedure can be followed for the second result.

Proposition 5.10
Assume (A1) (i, iii, iv), (A2) (i), the function J is continuously differentiable, $\int_0^1 J(s)ds = 1$, $J(s) \geq 0$ for all $0 \leq s \leq 1$, the distribution function F_X is twice continuously differentiable, $\inf_{x\in R_X} f_X(x) > 0$, and $H(t|x)$ and $H^u(t|x)$ satisfy (A5) $(i) - (vi)$. Then,

$$\hat{m}(x) - m(x) = -(nh_n)^{-1} f_X^{-1}(x)\sigma(x) \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \eta(T_i, \Delta_i|x) + R_n(x),$$

where $\sup\{|R_n(x)|; x \in R_X\} = O((nh_n)^{-3/4}(\log h_n^{-1})^{3/4})$ *a.s.*

Proof. Write, as in Proposition 5.7,

$$\hat{m}(x) - m(x) = \int_{-\infty}^{+\infty} [L(F(t|x)) - L(\tilde{F}(t|x))]\, dt$$

$$= \int_{-\infty}^{+\infty} J(F(t|x))(F(t|x) - \tilde{F}(t|x))\, dt - \frac{1}{2}\int_{-\infty}^{+\infty} J'(\xi_{tx})(F(t|x) - \tilde{F}(t|x))^2\, dt, \quad (5.18)$$

where ξ_{tx} is between $F(t|x)$ and $\tilde{F}(t|x)$. The second term is $O((nh_n)^{-1}\log h_n^{-1})$ a.s. by Proposition 5.5. The first one equals

$$-\sum_{i=1}^n W_{ni}(x; h_n) \int_{-\infty}^{+\infty} \xi(T_i, \Delta_i, t|x) J(F(t|x))\, dt + \int_{-\infty}^{+\infty} J(F(t|x)) r_{x,n}(t)\, dt, \quad (5.19)$$

where the remainder term $r_{x,n}(t)$ can be shown to satisfy $\sup_x \sup_{t\leq \tilde{T}_x} |r_{x,n}(t)| = O((nh_n)^{-3/4}\, (\log h_n^{-1})^{3/4})$ a.s. in a completely similar way as is done in Theorem 2.3 for the fixed design case. Finally, that the first term of (5.19) equals

$$-(nh_n)^{-1} f_X^{-1}(x)\sigma(x) \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \eta(T_i, \Delta_i|x) + O((nh_n)^{-1}\log h_n^{-1}),$$

follows from $\sup_x |\hat{f}_X(x) - f_X(x)| = O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2})$ a.s. (see proof of Proposition 5.3) and from the fact that

$$\sum_{i=1}^n W_{ni}(x; h_n)\xi(T_i, \Delta_i, t|x) = \int_{-\infty}^{t} \frac{\hat{H}(s|x) - H(s|x)}{(1 - H(s|x))^2} dH^u(s|x) \quad (5.20)$$

$$+ \frac{\hat{H}^u(t|x) - H^u(t|x)}{1 - H(t|x)} - \int_{-\infty}^{t} \frac{\hat{H}^u(s|x) - H^u(s|x)}{(1 - H(s|x))^2} dH(s|x)$$

is $O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2})$ a.s. uniformly in x and $t \leq \tilde{T}_x$ by Proposition 5.3 and Remark 5.3.

Proposition 5.11
Assume (A1) (i, iii, iv), (A2) (i), the function J is continuously differentiable, $\int_0^1 J(s)ds = 1$, $J(s) \geq 0$ for all $0 \leq s \leq 1$, the distribution function F_X is twice continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$, $\inf_{x \in R_X} \sigma(x) > 0$, *and $H(t|x)$ and $H^u(t|x)$ satisfy (A5) $(i) - (vi)$. Then,*

$$\hat{\sigma}(x) - \sigma(x) = -(nh_n)^{-1} f_X^{-1}(x)\sigma(x) \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \zeta(T_i, \Delta_i|x) + \tilde{R}_n(x),$$

where $\sup\{|\tilde{R}_n(x)|; x \in R_X\} = O((nh_n)^{-3/4}(\log h_n^{-1})^{3/4})$ *a.s.*

Proof. Write

$$\hat{\sigma}(x) - \sigma(x) = \frac{\hat{\sigma}^2(x) - \sigma^2(x)}{2\sigma(x)} - \frac{[\hat{\sigma}(x) - \sigma(x)]^2}{2\sigma(x)},$$

and note that the second term is $O((nh_n)^{-1}\log h_n^{-1})$ a.s. by Proposition 5.7. As in the proof of Proposition 5.7 we write

$$\begin{aligned}
\hat{\sigma}^2(x) - \sigma^2(x) &= \int_0^{+\infty} [L(F(\sqrt{t}|x)) - L(\tilde{F}(\sqrt{t}|x))]\, dt \\
&\quad - \int_0^{+\infty} [L(F(-\sqrt{t}|x)) - L(\tilde{F}(-\sqrt{t}|x))]\, dt - (\hat{m}^2(x) - m^2(x)) \\
&= 2\int_{-\infty}^{+\infty} [L(F(t|x)) - L(\tilde{F}(t|x))]t\, dt - (\hat{m}^2(x) - m^2(x)). \qquad (5.21)
\end{aligned}$$

For the second term on the right hand side of (5.21) write

$$\begin{aligned}
&\hat{m}^2(x) - m^2(x) = 2m(x)(\hat{m}(x) - m(x)) + (\hat{m}(x) - m(x))^2 \\
&= -2m(x)(nh_n)^{-1} f_X^{-1}(x)\sigma(x) \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \eta(T_i, \Delta_i|x) \\
&\quad + O((nh_n)^{-3/4}(\log h_n^{-1})^{3/4}),
\end{aligned}$$

by Propositions 5.10 and 5.7. For the first term of (5.21) use an expansion similar to the one in (5.18), and write the first term in this expansion in a way similar to (5.19). Combining these expressions one gets the desired representation with remainder term of the stated order.

5.5 Appendix B : Some technical results

We end this chapter with a number of technical results that are needed in the proof of Theorem 5.1.

Proposition 5.12
Assume $(A1)(i, iii, iv), (A2), (A3)(ii)$, *the distribution* F_X *is twice continuously differentiable,* $\inf_{x \in R_X} f_X(x) > 0$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(A5)(i, iv, v)$, *and* $H_e(t|x)$ *satisfies* $(A5)(ii)$. *Then,*

$$\sup_{-\infty<t<+\infty} |n^{-1} \sum_{i=1}^{n} \{I(\hat{E}_i \le t) - I(E_i \le t) - P(\hat{E} \le t|\mathcal{X}_n) + P(E \le t)\}| = o_P(n^{-1/2}),$$

where $P(\hat{E} \le t|\mathcal{X}_n)$ *is the distribution of* $\hat{E} = \frac{Z-\hat{m}(X)}{\hat{\sigma}(X)}$ *conditioning on* (X_j, T_j, Δ_j), $j = 1, \ldots, n$.

Proof. The expression between braces equals $I(E_i \le td_{n2}(X_i) + d_{n1}(X_i)) - I(E_i \le t) - P(E \le td_{n2}(X) + d_{n1}(X)) + P(E \le t)$, where $d_{n1}(x) = (\hat{m}(x) - m(x))/\sigma(x)$ and $d_{n2}(x) = \hat{\sigma}(x)/\sigma(x)$. The proof is based on results in van der Vaart and Wellner (1996). Let

$$\begin{aligned}\mathcal{F} = \{&I(E \le td_2(X) + d_1(X)) - I(E \le t) - P(E \le td_2(X) + d_1(X)) + P(E \le t);\\ &-\infty < t < +\infty, d_1 \in C_1^{1+\delta}(R_X) \text{ and } d_2 \in \tilde{C}_2^{1+\delta}(R_X)\},\end{aligned}$$

where $C_1^{1+\delta}(R_X)$ is the class of all differentiable functions d defined on the domain R_X of X such that $\|d\|_{1+\delta} \le 1$, $\tilde{C}_2^{1+\delta}(R_X)$ is the class of all differentiable functions d defined on R_X such that $\|d\|_{1+\delta} \le 2$ and $\inf_x\{d(x)\} \ge \frac{1}{2}$ and

$$\|d\|_{1+\delta} = \max\{\sup_x |d(x)|, \sup_x |d'(x)|\} + \sup_{x,y} \frac{|d'(x) - d'(y)|}{|x - y|^\delta}.$$

Note that by Propositions 5.7, 5.8 and 5.9, we have that $P(d_{n1} \in C_1^{1+\delta}(R_X)$ and $d_{n2} \in \tilde{C}_2^{1+\delta}(R_X)) \to 1$ as $n \to \infty$. In a first step we will show that the class $\mathcal{F}$ is Donsker. From Theorem 2.5.6 in van der Vaart and Wellner (1996), it follows that it suffices to show that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty, \tag{5.22}$$

where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of (E, X), and $L_2(P)$ is the L_2-norm. We will restrict ourselves to showing (5.22) for the class

$$\mathcal{F}_1 = \{I(E \le td_2(X)+d_1(X)); -\infty < t < +\infty,\ d_1 \in C_1^{1+\delta}(R_X) \text{ and } d_2 \in \tilde{C}_2^{1+\delta}(R_X)\},$$

since the other terms are similar, but much easier. In Corollary 2.7.2 of the aforementioned book it is stated that

$$m_1 = N_{[]}(\varepsilon^2, C_1^{1+\delta}(R_X), L_2(P)) \le \exp(K\varepsilon^{-\frac{2}{1+\delta}})$$
$$m_2 = N_{[]}(\varepsilon^2, \tilde{C}_2^{1+\delta}(R_X), L_2(P)) \le \exp(K\varepsilon^{-\frac{2}{1+\delta}}).$$

Let $d_1^L \le d_1^U, \ldots, d_{m_1}^L \le d_{m_1}^U$ be the functions defining the m_1 brackets for $C_1^{1+\delta}(R_X)$ and let $\tilde{d}_1^L \le \tilde{d}_1^U, \ldots, \tilde{d}_{m_2}^L \le \tilde{d}_{m_2}^U$ be the functions defining the m_2 brackets for $\tilde{C}_2^{1+\delta}(R_X)$. Thus, for each d_1 and d_2 and each fixed t :

$$I(E \le t\tilde{d}_j^L(X) + d_i^L(X)) \le I(E \le td_2(X) + d_1(X)) \le I(E \le t\tilde{d}_j^U(X) + d_i^U(X)).$$

Define $F_{ij}^L(t) = P(E \le t\tilde{d}_j^L(X) + d_i^L(X))$ and let t_{ijk}^L, $k = 1, \ldots, O(\varepsilon^{-2})$, partition the line in segments having F_{ij}^L-probability less than or equal to a fraction of ε^2. Similarly, define $F_{ij}^U(t) = P(E \le t\tilde{d}_j^U(X) + d_i^U(X))$ and let t_{ijk}^U, $k = 1, \ldots, O(\varepsilon^{-2})$, partition the line in segments having F_{ij}^U-probability less than or equal to a fraction of ε^2. Now, define the following bracket for t :

$$t_{ijk_1}^L \le t \le t_{ijk_2}^U,$$

where $t_{ijk_1}^L$ is the largest of the t_{ijk}^L with the property of being less than or equal to t and $t_{ijk_2}^U$ is the smallest of the t_{ijk}^U with the property of being greater than or equal to t. We will now show that the brackets for our function are given by

$$I(E \leq t^L_{ijk_1}\tilde{d}^L_j(X)+d^L_i(X)) \leq I(E \leq td_2(X)+d_1(X)) \leq I(E \leq t^U_{ijk_2}\tilde{d}^U_j(X)+d^U_i(X)).$$

Let's calculate

$$\begin{aligned} &\|I(E \leq t^U_{ijk_2}\tilde{d}^U_j(X) + d^U_i(X)) - I(E \leq t^L_{ijk_1}\tilde{d}^L_j(X) + d^L_i(X))\|_2^2 \\ &= F^U_{ij}(t^U_{ijk_2}) - F^L_{ij}(t^L_{ijk_1}) = F^U_{ij}(t) - F^L_{ij}(t) + K\varepsilon^2. \end{aligned}$$

Applying a Taylor expansion to the function H_e, yields

$$\begin{aligned} &F^U_{ij}(t) - F^L_{ij}(t) \\ &= \int [H_e(t\tilde{d}^U_j(x) + d^U_i(x)|x) - H_e(t\tilde{d}^L_j(x) + d^L_i(x)|x)]\, dF_X(x) \\ &= \int h_e(t\tilde{\xi}_j(x) + \xi_i(x)|x)[t(\tilde{d}^U_j(x) - \tilde{d}^L_j(x)) + (d^U_i(x) - d^L_i(x))]\, dF_X(x) \\ &= \int h_e(t\tilde{\xi}_j(x) + \xi_i(x)|x)(t\tilde{\xi}_j(x) + \xi_i(x))\tilde{\xi}_j(x)^{-1}(\tilde{d}^U_j(x) - \tilde{d}^L_j(x))\, dF_X(x) \\ &\quad - \int h_e(t\tilde{\xi}_j(x) + \xi_i(x)|x)\xi_i(x)\tilde{\xi}_j(x)^{-1}(\tilde{d}^U_j(x) - \tilde{d}^L_j(x))\, dF_X(x) \\ &\quad + \int h_e(t\tilde{\xi}_j(x) + \xi_i(x)|x)(d^U_i(x) - d^L_i(x))\, dF_X(x). \end{aligned} \tag{5.23}$$

Here, $\xi_i(x)$ is between $d^L_i(x)$ and $d^U_i(x)$ and $\tilde{\xi}_j(x)$ is between $\tilde{d}^L_j(x)$ and $\tilde{d}^U_j(x)$. Since we can choose the brackets d_i and $\tilde{d}_j$ such that $\sup_x |d^U_i(x)| \leq 1$ and $\inf_x |\tilde{d}^L_j(x)| \geq 1/2$ (for all i and j) and since $\sup_{x,t} |th_e(t|x)| < \infty$, (5.23) is bounded in absolute value by

$$K_1\|\tilde{d}^U_j - \tilde{d}^L_j\|_{P,1} + K_2\|d^U_i - d^L_i\|_{P,1} \leq (K_1 + K_2)\varepsilon^2$$

(since $\|d\|_{P,1} \leq \|d\|_{P,2}$ for any function d, where $\|d\|_{P,1}$ respectively $\|d\|_{P,2}$ is the $L_1(P)$-norm respectively $L_2(P)$-norm of d). Hence, for the class $\mathcal{F}_1$ and for each $\varepsilon > 0$, we have at most $O(\varepsilon^{-2}\exp(K\varepsilon^{-\frac{2}{1+\delta}}))$ brackets in total. However, for $\varepsilon > 1$, one bracket suffices. So we have,

$$\int_0^\infty \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{F}_1, L_2(P))}\, d\varepsilon < \infty.$$

This shows that the class $\mathcal{F}_1$ (and hence $\mathcal{F}$) is Donsker.

Next, let's calculate

$$\begin{aligned}
&\mathrm{Var}(I(E \le td_{n2}(X) + d_{n1}(X)) - I(E \le t) \\
&\qquad -P(E \le td_{n2}(X) + d_{n1}(X)) + P(E \le t)) \\
&= \mathrm{Var}(I(E \le td_{n2}(X) + d_{n1}(X)) - I(E \le t)) \\
&\le E[E(\{I(E \le td_{n2}(X) + d_{n1}(X)) - I(E \le t)\}^2|X)] \\
&= E[H_e(td_{n2}(X) + d_{n1}(X)|X) - H_e(\min(t, td_{n2}(X) + d_{n1}(X))|X)] \\
&\quad + E[H_e(t|X) - H_e(\min(t, td_{n2}(X) + d_{n1}(X))|X)] \\
&= E[h_e(ta_{n2}(X) + a_{n1}(X)|X)\,|t(d_{n2}(X) - 1) + d_{n1}(X)|], \qquad (5.24)
\end{aligned}$$

for some $a_{n1}(X)$ between 0 and $d_{n1}(X)$ and some $a_{n2}(X)$ between 1 and $d_{n2}(X)$. Since

$$\begin{aligned}
&\sup_x |th_e(ta_{n2}(x) + a_{n1}(x)|x)| \\
&\le \sup_x \{|ta_{n2}(x) + a_{n1}(x)|a_{n2}(x)^{-1} + |a_{n1}(x)|a_{n2}(x)^{-1}\}h_e(ta_{n2}(x) + a_{n1}(x)|x) \\
&\le K_1 \quad \text{(say)}
\end{aligned}$$

(by Proposition 5.7, because $\sup_{x,t} h_e(t|x) < \infty$, $\sup_{x,t} |th_e(t|x)| < \infty$ and $\inf_x \sigma(x) > 0$), (5.24) is bounded by

$$\begin{aligned}
&K_1 E|d_{n2}(X) - 1| + \sup_{x,t} h_e(t|x) E|d_{n1}(X)| \\
&\le K_1 \sup_x \left|\frac{\hat{\sigma}(x)}{\sigma(x)} - 1\right| + K_2 \sup_x |\hat{m}(x) - m(x)| \to 0 \qquad \text{a.s.},
\end{aligned}$$

again by Proposition 5.7. Since the class $\mathcal{F}$ is Donsker, it follows from Corollary 2.3.12 in van der Vaart and Wellner (1996) that

$$\lim_{\alpha \downarrow 0} \limsup_{n \to \infty} P\left(\sup_{f \in \mathcal{F}, \mathrm{Var}(f) < \alpha} n^{-1/2} \left| \sum_{i=1}^{n} f(X_i) \right| > \varepsilon \right) = 0,$$

for each $\varepsilon > 0$. By restricting the supremum inside this probability to the elements in $\mathcal{F}$ corresponding to $d_1(X) = d_{n1}(X)$ and $d_2(X) = d_{n2}(X)$ as defined above, the result follows.

We now use the above result to establish an asymptotic representation for the estimator $\hat{H}_e(t)$. Note that the main term in the decomposition below, contains the empirical distribution based on the $E_i = (T_i - m(X_i))/\sigma(X_i)$ and an extra term caused by replacing $m(\cdot)$ and $\sigma(\cdot)$ by $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$.

Proposition 5.13
Assume $(A1)-(A4)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(A5)(i)-(vi)$, *and* $H_e(t|x)$ *satisfies* $(A5)(ii, iii, vi, vii)$. *Then,*

$$\begin{aligned}&\hat{H}_e(t) - H_e(t)\\&= n^{-1}\sum_{i=1}^n[-\{\eta(T_i,\Delta_i|X_i) + \zeta(T_i,\Delta_i|X_i)t\}h_e(t|X_i) + I(E_i \le t) - H_e(t)] + R_n(t),\end{aligned}$$

where $\sup\{|R_n(t)|; -\infty < t < +\infty\} = o_P(n^{-1/2})$.

Proof. Using Proposition 5.12 and the notation in the statement of that proposition,

$$\begin{aligned}&\hat{H}_e(t) - H_e(t)\\&= \int\left\{H_e\left(\frac{t\hat{\sigma}(x) + \hat{m}(x) - m(x)}{\sigma(x)}\middle| x\right) - H_e(t|x)\right\} dF_X(x)\\&\quad + n^{-1}\sum_{i=1}^n\{I(E_i \le t) - H_e(t)\}\\&\quad + n^{-1}\sum_{i=1}^n\{I(\hat{E}_i \le t) - I(E_i \le t) - P(\hat{E} \le t|\mathcal{X}_n) + P(E \le t)\}\\&= \int h_e(t|x)\frac{t(\hat{\sigma}(x) - \sigma(x)) + (\hat{m}(x) - m(x))}{\sigma(x)}\, dF_X(x)\\&\quad + \frac{1}{2}\int h_e'(t_i|x)\left(\frac{t(\hat{\sigma}(x) - \sigma(x)) + (\hat{m}(x) - m(x))}{\sigma(x)}\right)^2 dF_X(x)\\&\quad + n^{-1}\sum_{i=1}^n\{I(E_i \le t) - H_e(t)\} + o_P(n^{-1/2}),\end{aligned} \tag{5.25}$$

where t_i is between t and $(t\hat{\sigma}(X_i) + \hat{m}(X_i) - m(X_i))/\sigma(X_i)$. The second term above is $o(n^{-1/2})$ a.s. by Proposition 5.7 and since $\sup_{x,t}|t^2h_e'(t|x)| < \infty$. The first term on the right hand side of (5.25) splits naturally in two parts. We will deal with the second one only. Using Proposition 5.10 it follows that

$$\begin{aligned}
&\int h_e(t|x) \frac{\hat{m}(x) - m(x)}{\sigma(x)} \, dF_X(x) \\
&= -(nh_n)^{-1} \sum_{i=1}^n \int f_X^{-1}(x) h_e(t|x) K\left(\frac{x - X_i}{h_n}\right) \eta(T_i, \Delta_i|x) \, dF_X(x) + o(n^{-1/2}) \\
&= -(nh_n)^{-1} \sum_{i=1}^n \int K\left(\frac{x - X_i}{h_n}\right) \beta(x, T_i, \Delta_i, t) \, dF_X(x) + o(n^{-1/2}), \qquad (5.26)
\end{aligned}$$

a.s., uniformly in t, where $\beta(x, z, \delta, t) = f_X^{-1}(x) h_e(t|x) \eta(z, \delta|x)$. Using a two term Taylor expansion of $\beta(x, T_i, \Delta_i, t)$ around X_i, (5.26) can be written as

$$\begin{aligned}
&-(nh_n)^{-1} \sum_{i=1}^n \beta(X_i, T_i, \Delta_i, t) \int K\left(\frac{x - X_i}{h_n}\right) \, dF_X(x) \\
&-(nh_n)^{-1} \sum_{i=1}^n \beta'(X_i, T_i, \Delta_i, t) \int K\left(\frac{x - X_i}{h_n}\right) (x - X_i) \, dF_X(x) \\
&-\frac{1}{2}(nh_n)^{-1} \sum_{i=1}^n \int K\left(\frac{x - X_i}{h_n}\right) (x - X_i)^2 \beta''(\xi_i, T_i, \Delta_i, t) \, dF_X(x) + o(n^{-1/2}) \\
&= -n^{-1} \sum_{i=1}^n \beta(X_i, T_i, \Delta_i, t) f_X(X_i) + O(h_n^2) + o(n^{-1/2}) \\
&= -n^{-1} \sum_{i=1}^n h_e(t|X_i) \eta(T_i, \Delta_i|X_i) + o(n^{-1/2}),
\end{aligned}$$

where ξ_i is between x and X_i. This finishes the proof.

Using arguments which are entirely analogous to the proof of Proposition 5.13 we obtain

Proposition 5.14
Assume $(A1)-(A4)$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(A5)(i)-(vi)$, *and* $H_e^u(t|x)$ *satisfies* $(A5)(ii, iii, vi, vii)$. *Then,*

$$\begin{aligned}
\hat{H}_e^u(t) - H_e^u(t) = n^{-1} \sum_{i=1}^n [&-\{\eta(T_i, \Delta_i|X_i) + \zeta(T_i, \Delta_i|X_i)t\} H_e^u(t|X_i) \\
&+ I(E_i \le t, \Delta_i = 1) - H_e^u(t)] + \tilde{R}_n(t),
\end{aligned}$$

where $\sup\{|\tilde{R}_n(t)|; -\infty < t < +\infty\} = o_P(n^{-1/2})$.

Proposition 5.15

Assume $(A1)(i, iii, iv), (A2), (A3)(ii)$, *the distribution* F_X *is twice continuously differentiable,* $\inf_{x \in R_X} f_X(x) > 0$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(A5)(i, iv, v)$, *and* $H_e(t|x)$ *satisfies* $(A5)(ii)$. *Then,*

$$\sup_{-\infty<t<+\infty} |\hat{H}_e(t) - H_e(t)| = O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2}) \quad a.s.$$

Proof. Similarly as in the proof of Proposition 5.13, we can write

$$\hat{H}_e(t) - H_e(t) = n^{-1}\sum_{i=1}^{n} h_e(t_i|X_i)\frac{t(\hat{\sigma}(X_i) - \sigma(X_i)) + (\hat{m}(X_i) - m(X_i))}{\sigma(X_i)} + n^{-1}\sum_{i=1}^{n}\{I(E_i \le t) - H_e(t)\} + o_P(n^{-1/2}),$$

where t_i is between t and $(t\hat{\sigma}(X_i) + \hat{m}(X_i) - m(X_i))/\sigma(X_i)$. The first term is $O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2})$ a.s. by Proposition 5.7. For the second one, the Dvoretzky-Kiefer-Wolfowitz (1956) inequality yields the order $O(n^{-1/2}(\log n)^{1/2})$ a.s.

Remark 5.4

The representation in Proposition 5.13 can also be used to obtain a bound of the order $O_P(n^{-1/2})$, instead of $O((nh_n)^{-1/2}(\log h_n^{-1})^{1/2})$ a.s. as established above. Note that the nature of such results is entirely different from the case where Euclidean parameters for location and scale are estimated. The rate of convergence of the present nonparametric estimators is slower than $n^{-1/2}$ and recovery of the root n convergence is due to the averaging over the covariate values. Similar results can be found in Cristóbal Cristóbal, Faraldo Roca and González Manteiga (1987), and in Akritas (1994, 1996).

Proposition 5.16

Assume $(A1)(i, iii, iv), (A2), (A3)(ii)$, *the distribution* F_X *is twice continuously differentiable,* $\inf_{x \in R_X} f_X(x) > 0$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(A5)(i, iv, v)$, *and* $H_e(t|x)$ *satisfies* $(A5)(ii, iii, vi)$. *Let* $J_c = \{(s,t); |H_e(t) - H_e(s)| \le c\}$. *Then,*

$$\sup\{|\hat{H}_e(t) - \hat{H}_e(s) - H_e(t) + H_e(s)|; (s,t) \in J_{\bar{h}_n}\} = o_P(n^{-1/2}),$$

where $\bar{h}_n$ is any sequence of positive numbers tending to zero as n tends to infinity that satisfies $\bar{h}_n h_n^{-1} \log h_n^{-1} \to 0$.

Proof. Consider the decomposition

$$\begin{aligned}
&\hat{H}_e(t) - \hat{H}_e(s) - H_e(t) + H_e(s) \\
&= \int \left\{ H_e\left(\frac{t\hat{\sigma}(x) + \hat{m}(x) - m(x)}{\sigma(x)} \middle| x \right) - H_e(t|x) \right. \qquad (5.27) \\
&\qquad \left. - H_e\left(\frac{s\hat{\sigma}(x) + \hat{m}(x) - m(x)}{\sigma(x)} \middle| x \right) + H_e(s|x) \right\} dF_X(x) \\
&\quad + n^{-1} \sum_{i=1}^{n} \{ I(E_i \le t) - H_e(t) - I(E_i \le s) + H_e(s) \} + o_P(n^{-1/2}),
\end{aligned}$$

uniformly in s and t by Proposition 5.12. The second term on the right hand side of (5.27) is $O(\bar{h}_n^{1/2} n^{-1/2} (\log n)^{1/2})$ a.s. by Lemma 2.4 in Stute (1982). The first term equals

$$\begin{aligned}
&\int \left\{ \frac{\hat{m}(x) - m(x)}{\sigma(x)} (h_e(t|x) - h_e(s|x)) \right. \\
&+ \left. \frac{\hat{\sigma}(x) - \sigma(x)}{\sigma(x)} (th_e(t|x) - sh_e(s|x)) \right\} dF_X(x) + O((nh_n)^{-1} \log h_n^{-1}) \\
&\le \sup_x |\hat{m}(x) - m(x)| (\inf_x \sigma(x))^{-1} \int |h_e(t|x) - h_e(s|x)| \, dF_X(x) \qquad (5.28) \\
&\quad + \sup_x |\hat{\sigma}(x) - \sigma(x)| (\inf_x \sigma(x))^{-1} \int |th_e(t|x) - sh_e(s|x)| \, dF_X(x) \\
&\quad + O((nh_n)^{-1} \log h_n^{-1}).
\end{aligned}$$

We will show that for all $x \in R_X$,

$$|th_e(t|x) - sh_e(s|x)| \le \frac{K}{\bar{h}_n^{1/2}} |H_e(t|x) - H_e(s|x)| + 2\bar{h}_n^{1/2}. \qquad (5.29)$$

Showing this implies that for $(s,t) \in J_{\bar{h}_n}$,

$$\begin{aligned}
&\int |th_e(t|x) - sh_e(s|x)| \, dF_X(x) \\
&\le \frac{K}{\bar{h}_n^{1/2}} \int |H_e(t|x) - H_e(s|x)| \, dF_X(x) + 2\bar{h}_n^{1/2} \\
&= \frac{K}{\bar{h}_n^{1/2}} |H_e(t) - H_e(s)| + 2\bar{h}_n^{1/2} \le (K+2)\bar{h}_n^{1/2}.
\end{aligned}$$

This together with Proposition 5.7 and the assumption on $\bar{h}_n$ implies the order $o_P(n^{-1/2})$ for the second term of (5.28). The derivation for the first term is similar, but easier. For the proof of (5.29) define $A_x = \{t; |th_e(t|x)| \ge \bar{h}_n^{1/2}\}$. Clearly, in the case where $s \notin A_x$ and $t \notin A_x$ there is nothing to show. Consider next the case where $s \notin A_x$ and $t \in A_x$. Then there exists an r between s and t such that $|rh_e(r|x)| = \bar{h}_n^{1/2}$ and $u \in A_x$ for all u between r and t. Using this r write

$$|th_e(t|x) - sh_e(s|x)| \le |th_e(t|x) - rh_e(r|x)| + |rh_e(r|x) - sh_e(s|x)|. \quad (5.30)$$

The second term on the right hand side of (5.30) is clearly less than $2\bar{h}_n^{1/2}$. To deal with the first term define $v_e(t|x) = th_e(t|x)$ and write

$$\begin{aligned}
&|v_e(t|x) - v_e(r|x)| \\
&= |(v_e(\cdot|x) \circ H_e^{-1}(\cdot|x))'(H_e(u|x))|(H_e(t|x) - H_e(r|x)) \\
&= \frac{|v_e'(u|x)|}{h_e(u|x)}(H_e(t|x) - H_e(r|x)) \\
&= \frac{|h_e(u|x) + uh_e'(u|x)|}{h_e(u|x)}(H_e(t|x) - H_e(r|x)) \\
&\le \left(1 + \frac{\sup_u |u^2 h_e'(u|x)|}{\inf_{u \in A_x} |uh_e(u|x)|}\right)(H_e(t|x) - H_e(r|x)) \\
&\le K\bar{h}_n^{-1/2}(H_e(t|x) - H_e(r|x)) \qquad (5.31)
\end{aligned}$$

where the u in the first equality is between r and t. Finally consider the case where both $s, t \in A_x$. If we have that $r \in A_x$ for all r between s and t then (5.31) shows that $|th_e(t|x) - sh_e(s|x)| \leq K\bar{h}_n^{-1/2}(H_e(t|x) - H_e(s|x))$. It remains to consider the possibility that there exists an r between s and t with $r \notin A_x$. Let r_1 (r_2) be the smallest (largest) number between s and t such that $r_1 h_e(r_1|x) = \bar{h}_n^{1/2}$ (and similarly for r_2). Also assume without loss of generality that $s < t$. Then $|th_e(t|x) - sh_e(s|x)| \leq K\bar{h}_n^{1/2}$ follows from the decomposition $|th_e(t|x) - sh_e(s|x)| \leq |th_e(t|x) - r_2 h_e(r_2|x)| + |r_2 h_e(r_2|x) - r_1 h_e(r_1|x)| + |r_1 h_e(r_1|x) - sh_e(s|x)|$ and (5.31).

Proposition 5.17
Assume $(A1)(i, iii, iv), (A2), (A3)(ii)$, *the distribution* F_X *is twice continuously differentiable,* $\inf_{x \in R_X} f_X(x) > 0$, $H(t|x)$ *and* $H^u(t|x)$ *satisfy* $(A5)(i, iv, v)$, *and* $H_e(t|x)$ *and* $H_e^u(t|x)$ *satisfy* $(A5)(ii, iii, vi)$. *Then,*

$$\sup_{-\infty < t \leq \bar{T}} \left| \int_{-\infty}^{t} \left[\frac{1}{1 - \hat{H}_e(y)} - \frac{1}{1 - H_e(y)} \right] d(\hat{H}_e^u(y) - H_e^u(y)) \right| = o_P(n^{-1/2}). \quad (5.32)$$

Proof. Partitioning the interval $(-\infty, \bar{T}]$ into k_n subintervals $[t_i, t_{i+1}]$ such that

$$H_e(t_{i+1}) - H_e(t_i) \leq (nh_n)^{-1/2}(\log h_n^{-1})^{1/2} = \bar{h}_n,$$

where $k_n = O((nh_n)^{1/2}(\log h_n^{-1})^{-1/2})$, we have that the integral in (5.32) is bounded by

$$\begin{aligned} & k_n \sup\left\{\left|\frac{1}{1 - \hat{H}_e(t)} - \frac{1}{1 - H_e(t)}\right|; -\infty < t \leq \bar{T}\right\} \\ & \quad \times \sup\{|\hat{H}_e^u(t) - \hat{H}_e^u(s) - H_e^u(t) + H_e^u(s)|; (s,t) \in J_{\bar{h}_n}\} \\ & + 2\sup\left\{\left|\frac{1}{1 - \hat{H}_e(t)} - \frac{1}{1 - \hat{H}_e(s)} - \frac{1}{1 - H_e(t)} + \frac{1}{1 - H_e(s)}\right|; \right. \\ & \quad \left. (s,t) \in J_{\bar{h}_n} \cap (-\infty, \bar{T}]^2 \right\}, \end{aligned}$$

where $J_{\bar{h}_n} = \{(s,t); |H_e(t) - H_e(s)| \leq \bar{h}_n\}$. Applying Proposition 5.15 and an analogue of Proposition 5.16 for the distribution H_e^u, the first term above is easily seen to be $o_P(n^{-1/2})$. Again using Proposition 5.15, the expression inside the supremum of the second term can be written as

$$
\begin{aligned}
&|(1-H_e(s))^{-2}(\hat{H}_e(s)-H_e(s))-(1-H_e(t))^{-2}(\hat{H}_e(t)-H_e(t))| \\
&+O((nh_n)^{-1}\log h_n^{-1}) \\
&=(1-H_e(s))^{-2}|\hat{H}_e(t)-\hat{H}_e(s)-H_e(t)+H_e(s)|+O((nh_n)^{-1}\log h_n^{-1}) \qquad \text{a.s.,}
\end{aligned}
$$

uniformly on $(-\infty,\bar{T}]$. This is $o_P(n^{-1/2})$ by Proposition 5.16.

Chapter 6

The bivariate distribution in heteroscedastic regression models with censored data

In the previous chapter, we constructed an estimator for the conditional distribution of the response Y (subject to random right censoring) given the covariate X (completely observed) when X and Y are related through a heteroscedastic regression model. We will now consider two functionals of this conditional distribution. First, we will estimate the bivariate distribution of X and Y by averaging the conditional distribution estimator over all values of X. To show asymptotic properties of this estimator, the results of Chapter 5 will be used. Second, using that estimator of the bivariate distribution, we will construct new least squares estimators for the regression coefficients in a polynomial regression model and study their asymptotic properties.

This chapter is organized as follows. In the first section, we will collect the assumptions under which the main results are valid and we will introduce some additional notation. In Section 6.2, an asymptotic representation and the weak convergence of the bivariate distribution estimator will be established. Section 6.3 deals with the least squares estimators : by making use of the representation for the bivariate distribution estimator we will construct a representation for the least squares estimators and obtain a central limit result. A number of technical derivations are given in Section 6.4.

6.1 Definitions and assumptions

Because the asymptotic theory is based on an i.i.d. representation for $\hat{F}_e(t)$ which is valid up to any point smaller than τ_{H_e} (see Theorem 5.1), we need to work with a slightly modified version of the estimator $\hat{F}(x,t)$ of the bivariate distribution, defined in (1.42). Namely the asymptotic theory which we will develop in Section 6.2, pertains to

$$\hat{F}_T(x,t) = \int_{-\infty}^{x} \hat{F}(t \wedge \bar{T}_u|u) d\hat{F}_X(u) = \int_{-\infty}^{x} \hat{F}_e\left(\frac{t \wedge \bar{T}_u - \hat{m}(u)}{\hat{\sigma}(u)}\right) d\hat{F}_X(u), \qquad (6.1)$$

where $\bar{T}_u \leq \bar{T}\sigma(u) + m(u)$ for all $u \in R_X$ and where $\bar{T} < \tau_{H_e}$. This is actually an estimator of

$$F_T(x,t) = \int_{-\infty}^{x} F(t \wedge \bar{T}_u|u) dF_X(u) = \int_{-\infty}^{x} F_e\left(\frac{t \wedge \bar{T}_u - m(u)}{\sigma(u)}\right) dF_X(u),$$

which can become arbitrarily close to $F(x,t)$ if $\tau_{F_e} \leq \tau_{G_e}$ and $\bar{T}_u$, respectively $\bar{T}$, is chosen sufficiently close to $\bar{T}\sigma(u) + m(u)$, respectively τ_{H_e}, for all u.

Because the least squares estimator will be studied as a functional of the estimator of the bivariate distribution, its asymptotic theory pertains to

$$\hat{\boldsymbol{\beta}}_T = n(\boldsymbol{X}'\boldsymbol{X})^{-1} \begin{pmatrix} \int x^0 t \, d\hat{F}_T(x,t) \\ \vdots \\ \int x^p t \, d\hat{F}_T(x,t) \end{pmatrix}, \qquad (6.2)$$

which estimates $\boldsymbol{\beta}_T$ defined as in (1.46) with $F_T(x,t)$ replacing $F(x,t)$. By suitable choice of $\bar{T}$ and $\bar{T}_u$ ($u \in R_X$), $\boldsymbol{\beta}_T$ becomes arbitrarily close to $\boldsymbol{\beta}$ if $\tau_{F_e} \leq \tau_{G_e}$.

The following result reveals that in the homoscedastic linear regression model with uncensored data, the proposed slope estimator $\hat{\beta}_1$ is the same as that obtained by least squares on the pairs $(X_i, \hat{m}(X_i))$, $i = 1, \ldots, n$. In particular, if $J(s) \equiv 1$, the present estimator is a special case of the estimator in Cristóbal Cristóbal, Faraldo Roca and González Manteiga (1987). In this particular case, they showed (see their Theorem 1.5) that the estimator coincides with the ridge regression estimator, where the ridge factor depends on the kernel function and the bandwidth sequence.

Example 6.1
In the homoscedastic linear regression model with uncensored data, the slope estimator $\hat{\beta}_1$ equals the least squares estimator obtained from the pairs $(X_i, \hat{m}(X_i))$, $i = 1, \ldots, n$.

Proof. For $p = 1$, we have

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{1}{n\widehat{\text{Var}}(X)} \begin{pmatrix} n^{-1}\sum_{i=1}^n X_i^2 & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix},$$

where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and $\widehat{\text{Var}}(X) = n^{-1}\sum_{i=1}^n X_i^2 - \bar{X}^2$. Hence,

$$\hat{\beta}_1 = \frac{1}{\widehat{\text{Var}}(X)} \left(\int xt\, d\hat{F}(x,t) - \bar{X} \int t\, d\hat{F}(x,t) \right).$$

Since for homoscedastic uncensored data,

$$\hat{F}(x,t) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n I(Y_j - \hat{m}(X_j) \le t - \hat{m}(X_i)) I(X_i \le x),$$

it easily follows that

$$\int xt\, d\hat{F}(x,t) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n X_i(Y_j - \hat{m}(X_j) + \hat{m}(X_i))$$
$$\int t\, d\hat{F}(x,t) = n^{-1} \sum_{j=1}^n Y_j$$

and hence

$$\hat{\beta}_1 = \frac{1}{\widehat{\text{Var}}(X)} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) \hat{m}(X_i),$$

i.e. $\hat{\beta}_1$ equals the usual least squares estimator of the slope based on $(X_i, \hat{m}(X_i))$ $(i = 1, \ldots, n)$.

Throughout this chapter we continue working with assumptions $(A1) - (A5)$ made in Chapter 5 (see Section 5.1). In addition, we will need the following extra conditions (the function $L(t|x)$ in $(A5)(viii)$ below is an arbitrary (sub)distribution function) :

(A3)(iii) The error variable ε has finite expectation.

(A5)(viii) $L'''(t|x)$ exists, is continuous in (x, t) and $\sup_{x,t} |t^3 L'''(t|x)| < \infty$.

(A6) The function $t \to P(m(X) + e\sigma(X) \le t)$ $(t \in I\!R)$ is differentiable for all $e \in I\!R$ and the derivative is uniformly bounded over all $e \in I\!R$.

Let $\tilde{T}_x$ $(x \in R_X)$ be such that $\tilde{T}_x < \tau_{H(\cdot|x)}$ and $\inf_{x \in R_X}(1 - H(\tilde{T}_x|x)) > 0$. The results in this chapter require a slightly stronger version of assumptions $(A2)(i)$ and $(A4)$:

(A2)(i′) The function $x \to \bar{T}_x$ $(x \in R_X)$ is twice continuously differentiable and there exist $0 \le s_0 \le s_1 \le 1$ such that $s_1 \le \inf_x F(\tilde{T}_x|x)$, $s_0 \le \inf\{s \in [0,1]; J(s) \ne 0\}$, $s_1 \ge \sup\{s \in [0,1]; J(s) \ne 0\}$ and $\inf_{x \in R_X} \inf_{s_0 \le s \le s_1} f(F^{-1}(s|x)|x) > 0$.

(A4′) The functions $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$ are twice continuously differentiable with respect to x, their first and second derivatives (with respect to x) are bounded, uniformly in $x \in R_X, z < \tilde{T}_x$ and δ, and for any $\delta = 0, 1$, the first derivatives (considered as functions in z) are of bounded variation and the variation norms are uniformly bounded over all x.

6.2 Asymptotic representation and weak convergence of the bivariate distribution process

This section presents the asymptotic theory of the bivariate distribution estimator. The estimator in (6.1) will be studied by means of the decomposition

$$\hat{F}_T(x,t) - F_T(x,t) = \int_{-\infty}^{x} \left[\hat{F}_e\left(\frac{t \wedge \bar{T}_u - \hat{m}(u)}{\hat{\sigma}(u)}\right) - F_e\left(\frac{t \wedge \bar{T}_u - m(u)}{\sigma(u)}\right)\right] d\hat{F}_X(u)$$
$$+ \int_{-\infty}^{x} F_e\left(\frac{t \wedge \bar{T}_u - m(u)}{\sigma(u)}\right) d(\hat{F}_X(u) - F_X(u)). \quad (6.3)$$

For the first term on the right hand side of (6.3), use will be made of some of the results in Chapter 5.

Other nonparametric estimators for the bivariate distribution of X (completely observed) and Y (subject to right censoring) have been studied by Dabrowska (1988), Stute (1993,1996), Akritas (1994) and van der Laan (1996). The main advantage of the present estimator lies in its good behavior for large values of the variable Y : if the censoring is "light" in a region R of the covariate space (in the sense that $\tau_{F(\cdot|x)} \leq \tau_{G(\cdot|x)}$ for all x in R), then the estimator $\hat{F}(t|x)$ of the conditional distribution of Y given $X = x$ behaves well in the right tail for all x. From equation (1.42) it follows that the same is true for the estimator $\hat{F}(x,t)$ of the bivariate distribution.

Theorem 6.1
Assume (A1),(A2)(i', ii),(A3)(i, ii, iii),(A4'),(A6), $H(t|x)$ and $H^u(t|x)$ satisfy (A5) $(i) - (vi)$, and $H_e(t|x)$ and $H^u_e(t|x)$ satisfy (A5)($ii, iii, vi, vii, viii$). Then,

$$\hat{F}_T(x,t) - F_T(x,t) = n^{-1}\sum_{i=1}^{n} g_{x,t}(X_i, T_i, \Delta_i) + R_n(x,t),$$

where $\sup\{|R_n(x,t)|; x \in R_X, t \in I\!R\} = o_P(n^{-1/2})$, *and* $g_{x,t}(u,z,\delta) = \sum_{i=1}^{3} g_{x,t,i}(u,z,\delta)$, *where*

$$g_{x,t,1}(u,z,\delta) = E\left\{\varphi\left(u,z,\delta,\frac{t \wedge \bar{T}_X - m(X)}{\sigma(X)}\right) I(X \leq x)\right\},$$

$$g_{x,t,2}(u,z,\delta) = \left[\eta(z,\delta|u) + \zeta(z,\delta|u)\frac{t \wedge \bar{T}_u - m(u)}{\sigma(u)}\right] f_e\left(\frac{t \wedge \bar{T}_u - m(u)}{\sigma(u)}\right) I(u \leq x),$$

$$g_{x,t,3}(u,z,\delta) = F_e\left(\frac{t \wedge \bar{T}_u - m(u)}{\sigma(u)}\right) I(u \leq x) - E\left[F_e\left(\frac{t \wedge \bar{T}_X - m(X)}{\sigma(X)}\right) I(X \leq x)\right].$$

Proof. For the first term on the right hand side of (6.3) we will make use of Theorem 5.3. However, the remainder term in that representation is not $o_P(n^{-1/2})$, as is required. Let $\alpha_{ni}(x,t)$, $i = 1,2,3$, be as in the proof of that representation. Then, the first term on the right hand side of (6.3) equals

$$\int_{-\infty}^{x} \left[\alpha_{n1}(u, t \wedge \bar{T}_u) + \alpha_{n2}(u, t \wedge \bar{T}_u) + \alpha_{n3}(u, t \wedge \bar{T}_u)\right] d\hat{F}_X(u). \tag{6.4}$$

Using equation (5.6) in the proof of Theorem 5.3, $\alpha_{n2}(u, t \wedge \bar{T}_u) + \alpha_{n3}(u, t \wedge \bar{T}_u)$ can be expressed as the sum of a leading term and a remainder term which is $o_P(n^{-1/2})$ uniformly in (u,t). Thus only the term $\alpha_{n1}(u, t \wedge \bar{T}_u)$ needs further attention. But, using Theorem 5.1 together with Proposition 6.1 (see appendix), it is easily seen that this term can also be written as a leading term and a remainder of order $o_P(n^{-1/2})$. Combining the two leading terms, it follows that (6.4) equals

$$n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi\left(X_i, T_i, \Delta_i, \frac{t \wedge \bar{T}_{X_j} - m(X_j)}{\sigma(X_j)}\right) I(X_j \le x) \tag{6.5}$$

$$+ n^{-2} h_n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{X_j - X_i}{h_n}\right) f_X^{-1}(X_j)$$

$$\times \left[\eta(T_i, \Delta_i | X_j) + \zeta(T_i, \Delta_i | X_j) \frac{t \wedge \bar{T}_{X_j} - m(X_j)}{\sigma(X_j)}\right] f_e\left(\frac{t \wedge \bar{T}_{X_j} - m(X_j)}{\sigma(X_j)}\right) I(X_j \le x)$$

$$+ o_P(n^{-1/2}).$$

The result now follows from Propositions 6.4, 6.5 and 6.7 in the appendix of this chapter.

Theorem 6.2

Under the assumptions of Theorem 6.1, the process $n^{1/2}(\hat{F}_T(x,t) - F_T(x,t))$, $x \in R_X, t \in I\!R$, converges weakly to a zero-mean Gaussian process $Z(x,t)$, whose covariance function $Cov(Z(x,t), Z(x',t'))$ equals

$$E\{[g_{x,t,1}(X,T,\Delta) + g_{x,t,2}(X,T,\Delta)][g_{x',t',1}(X,T,\Delta) + g_{x',t',2}(X,T,\Delta)]\}$$

$$+ Cov\left[F_e\left(\frac{t \wedge \bar{T}_X - m(X)}{\sigma(X)}\right) I(X \le x), F_e\left(\frac{t' \wedge \bar{T}_X - m(X)}{\sigma(X)}\right) I(X \le x')\right].$$

Proof. To prove the weak convergence of the given process, we will make use of results in van der Vaart and Wellner (1996). Let $\mathcal{F} = \{(u, z, \delta) \to g_{x,t}(u, z, \delta); x \in R_X, t \in I\!R\}$. We will prove that the class $\mathcal{F}$ is Donsker by showing that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty, \tag{6.6}$$

(see p. 130 in van der Vaart and Wellner (1996)) where $N_{[]}$ is the bracketing number. Set $e_T(x) = \frac{t \wedge \bar{T}_x - m(x)}{\sigma(x)}$ and write

$$\begin{aligned} g_{x,t,1}(u, z, \delta) = &- \int_{-\infty}^{x} \int_{-\infty}^{\frac{z-m(u)}{\sigma(u)}} \frac{1 - F_e(e_T(v))}{(1 - H_e(s))^2} I\left(s \le e_T(v)\right)\, dH_e^u(s)\, dF_X(v) \\ &+ \frac{I(\delta = 1)}{1 - H_e\left(\frac{z-m(u)}{\sigma(u)}\right)} \int_{-\infty}^{x} (1 - F_e(e_T(v))) I\left(\frac{z - m(u)}{\sigma(u)} \le e_T(v)\right) dF_X(v) \\ &- \eta(z, \delta|u) \int_{-\infty}^{x} S_e(e_T(v)) \gamma_1(e_T(v)|u)\, dF_X(v) \\ &- \zeta(z, \delta|u) \int_{-\infty}^{x} S_e(e_T(v)) \gamma_2(e_T(v)|u)\, dF_X(v) \\ = &\sum_{i=1}^{4} g_{x,t,1,i}(u, z, \delta). \end{aligned}$$

The following argumentation uses Theorem 2.7.5 and Corollary 2.7.2 in van der Vaart and Wellner (1996). The term $g_{x,t,1,1}(u, z, \delta)$ is decreasing as a function of $\frac{z-m(u)}{\sigma(u)}$ and bounded and hence $O(\exp(K\varepsilon^{-1}))$ brackets are required. Similarly, the bracketing number for the integral in $g_{x,t,1,2}(u, z, \delta)$ is at most $O(\exp(K\varepsilon^{-1}))$ and hence the same holds for $g_{x,t,1,2}(u, z, \delta)$ itself since the expression in front of the integral is bounded and independent of x and t. Since the integrals in $g_{x,t,1,3}(u, z, \delta)$ and $g_{x,t,1,4}(u, z, \delta)$ are bounded (uniformly in u), as well as their first derivatives (with respect to u), their bracketing number is $O(\exp(K\varepsilon^{-1}))$. Therefore, the terms $g_{x,t,1,3}(u, z, \delta)$ and $g_{x,t,1,4}(u, z, \delta)$ themselves also need $O(\exp(K\varepsilon^{-1}))$ brackets, because $\eta(z, \delta|u)$ and $\zeta(z, \delta|u)$ are bounded and independent of x and t. All together,

we have $O(\exp(K\varepsilon^{-1}))$ brackets for $g_{x,t,1}(u,z,\delta)$.
For $g_{x,t,2}(u,z,\delta)$ we first note that the bracketing number for the class of functions of the form $u \to I(u \le x)$ is $O(\exp(K\varepsilon^{-1}))$. For the functions $u \to f_e(\frac{t\wedge \bar{T}_u - m(u)}{\sigma(u)})$, we consider three cases. If $t \le \min_u \bar{T}_u$, then $O(\exp(K\varepsilon^{-1}))$ brackets are needed. If $t \ge \max_u \bar{T}_u$, then $f_e(\frac{t\wedge \bar{T}_u - m(u)}{\sigma(u)})$ does not depend on t and hence one bracket suffices. For the intermediate case, i.e. for $\min_u \bar{T}_u \le t \le \max_u \bar{T}_u$, we apply Theorem 2.11.9 in van der Vaart and Wellner (1996) on

$$\sum_{i=1}^{n} Z_{ni}(t) = n^{-1/2} \sum_{i=1}^{n} f_e \left(\frac{t \wedge \bar{T}_{X_i} - m(X_i)}{\sigma(X_i)} \right).$$

For each $\varepsilon > 0$, divide the interval $[\min_u \bar{T}_u, \max_u \bar{T}_u]$ into $N_\varepsilon = O(\varepsilon^{-1})$ subintervals $\mathcal{F}^n_{\varepsilon j}$ of length not more than $K\varepsilon$ for some $K > 0$. Then, for each $j = 1, \ldots, N_\varepsilon$:

$$\sum_{i=1}^{n} \sup_{t,t' \in \mathcal{F}^n_{\varepsilon j}} |Z_{ni}(t) - Z_{ni}(t')|^2 \le \varepsilon^2,$$

by proper choice of $K > 0$. This shows that the bracketing number is $O(\varepsilon^{-1})$ in this case. In total, we have $O(\exp(K\varepsilon^{-1}))$ brackets for the class $\{u \to f_e(\frac{t\wedge \bar{T}_u - m(u)}{\sigma(u)}); t \in I\!R\}$. Similarly, the class $\{u \to \frac{t\wedge \bar{T}_u - m(u)}{\sigma(u)} f_e(\frac{t\wedge \bar{T}_u - m(u)}{\sigma(u)})\}$ also requires $O(\exp(K\varepsilon^{-1}))$ brackets. This shows that the bracketing number for $g_{x,t,2}(u,z,\delta)$ is $O(\exp(K\varepsilon^{-1}))$. Finally, for the term $g_{x,t,3}(u,z,\delta)$, analogous arguments as before show that also $O(\exp(K\varepsilon^{-1}))$ brackets are needed for this term.
This shows that the integral in (6.6) is bounded by

$$K \int_0^{2M} \frac{d\varepsilon}{\varepsilon^{1/2}} < \infty$$

where M is an upper bound for $|g_{x,t}(u,z,\delta)|$ (because for $\varepsilon > 2M$, one bracket suffices to cover $\mathcal{F}$). This shows that the class $\mathcal{F}$ is Donsker. The result now follows from p. 81-82 in van der Vaart and Wellner (1996).

6.3 Asymptotic representation and asymptotic normality of the least squares estimator

We proceed with presenting an asymptotic representation for the proposed least squares estimator $\hat{\beta}_T$ defined in (6.2). The representation of $\hat{\beta}_T$ will be obtained via the representation for $\hat{F}_T(x,t)$ shown in Theorem 6.1. Recall that if there is a region R of X with "light" censoring (in the sense that $\tau_{F(\cdot|x)} \le \tau_{G(\cdot|x)}$ for all x in R), then the vectors β_T and $\hat{\beta}_T$ are arbitrarily close to, respectively, β and $\hat{\beta}$ by choosing $\bar{T}$ and $\bar{T}_u$ large enough for all u in R_X. This is because $F_T(x,t)$ and $\hat{F}_T(x,t)$ are in that case arbitrarily close to $F(x,t)$ and $\hat{F}(x,t)$.

A large number of other generalizations of the (uncensored) least squares estimator to the censored data case exist. We refer to Section 1.5 for an overview. All of these methods pertain however to homoscedastic regression models.

As notation we use $n^{-1}(\boldsymbol{X}'\boldsymbol{X}) = A_n = (\hat{a}_{rs})$ and $n^{-1}E(\boldsymbol{X}'\boldsymbol{X}) = A = (a_{rs})$. The $(k+1, l+1)$-th element of A^{-1} and A_n^{-1} will be denoted by, respectively, $g_{kl}(A)$ and $g_{kl}(A_n)$. Also, let $\tilde{G}_{rs}(A) = (\tilde{g}_{kl}^{rs}(A))$ be the matrix of the partial derivative of the elements $g_{kl}(A)$, $k, l = 0, \ldots, p$, with respect to a_{rs}.

Theorem 6.3
Assume (A1),(A2)(i', ii),(A3)(i, ii, iii),(A4'),(A6), $H(t|x)$ and $H^u(t|x)$ satisfy (A5) $(i) - (vi)$, $H_e(t|x)$ and $H_e^u(t|x)$ satisfy (A5)($ii, iii, vi, vii, viii$) and the p-variate distribution of $(X, X^2, \ldots, X^p)$ is nonsingular. Then, the vector

$$\begin{pmatrix} \hat{\beta}_{T,0} - \beta_{T,0} \\ \vdots \\ \hat{\beta}_{T,p} - \beta_{T,p} \end{pmatrix} - n[E(\boldsymbol{X}'\boldsymbol{X})]^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n \Psi_0(X_i, T_i, \Delta_i) \\ \vdots \\ n^{-1}\sum_{i=1}^n \Psi_p(X_i, T_i, \Delta_i) \end{pmatrix} \tag{6.7}$$

$$- n[E(\boldsymbol{X}'\boldsymbol{X})]^{-1} \begin{pmatrix} \int x^0 \int t\, dF_e(\frac{t\wedge \bar{T}_x - m(x)}{\sigma(x)})\, d(\hat{F}_X(x) - F_X(x)) \\ \vdots \\ \int x^p \int t\, dF_e(\frac{t\wedge \bar{T}_x - m(x)}{\sigma(x)})\, d(\hat{F}_X(x) - F_X(x)) \end{pmatrix}$$

$$- \sum_{r,s=0}^{p} \tilde{G}_{rs}(A) \begin{pmatrix} \int x^0 t\, dF_T(x,t) \\ \vdots \\ \int x^p t\, dF_T(x,t) \end{pmatrix} \int x^{r+s}\, d(\hat{F}_X(x) - F_X(x)),$$

is $o_P(n^{-1/2})$ *(in* L_2*-norm), where for* $k = 0, \ldots, p$*,*

$$\Psi_k(u,z,\delta) = \int_{R_X} x^k \int_{-\infty}^{\bar{T}_x} t\, d\varphi\left(u, z, \delta, \frac{t - m(x)}{\sigma(x)}\right) dF_X(x)$$
$$+ u^k \int_{-\infty}^{\bar{T}_u} t\, d\left[\left(\eta(z,\delta|u) + \zeta(z,\delta|u)\frac{t - m(u)}{\sigma(u)}\right) f_e\left(\frac{t - m(u)}{\sigma(u)}\right)\right].$$

Proof. Write

$$\begin{pmatrix} \hat{\beta}_{T,0} - \beta_{T,0} \\ \vdots \\ \hat{\beta}_{T,p} - \beta_{T,p} \end{pmatrix} = n(\boldsymbol{X}'\boldsymbol{X})^{-1} \begin{pmatrix} \int x^0 t\, d(\hat{F}_T(x,t) - F_T(x,t)) \\ \vdots \\ \int x^p t\, d(\hat{F}_T(x,t) - F_T(x,t)) \end{pmatrix} \tag{6.8}$$
$$+ n((\boldsymbol{X}'\boldsymbol{X})^{-1} - [E(\boldsymbol{X}'\boldsymbol{X})]^{-1}) \begin{pmatrix} \int x^0 t\, dF_T(x,t) \\ \vdots \\ \int x^p t\, dF_T(x,t) \end{pmatrix}.$$

We start with the first term on the right hand side of (6.8).

$$\int x^k t\, d(\hat{F}_T(x,t) - F_T(x,t)) \tag{6.9}$$
$$= \int x^k \int t\, d\left[\hat{F}_e\left(\frac{t \wedge \bar{T}_x - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{t \wedge \bar{T}_x - m(x)}{\sigma(x)}\right)\right] d\hat{F}_X(x)$$
$$+ \int x^k \int t\, dF_e\left(\frac{t \wedge \bar{T}_x - m(x)}{\sigma(x)}\right) d(\hat{F}_X(x) - F_X(x)).$$

Using integration by parts, the first term on the right hand side of (6.9) can be written as

$$-\int\int x^k I(t \le \bar{T}_x)\left[\hat{F}_e\left(\frac{t-\hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{t-m(x)}{\sigma(x)}\right)\right] d\hat{F}_X(x)\,dt \qquad (6.10)$$

$$+\int x^k \bar{T}_x\left[\hat{F}_e\left(\frac{\bar{T}_x-\hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{\bar{T}_x-m(x)}{\sigma(x)}\right)\right] d\hat{F}_X(x).$$

This is somewhat similar to the first term of (6.3) for which an asymptotic representation was shown in Theorem 6.1. A completely similar derivation as for Theorem 6.1 shows that (6.10) equals

$$-n^{-1}\sum_{i=1}^n \int\int x^k I(t \le \bar{T}_x)\,\varphi\left(X_i, T_i, \Delta_i, \frac{t-m(x)}{\sigma(x)}\right) dF_X(x)\,dt$$

$$-n^{-1}\sum_{i=1}^n \int_{-\infty}^{\bar{T}_{X_i}} X_i^k\left[\eta(T_i,\Delta_i|X_i) + \zeta(T_i,\Delta_i|X_i)\frac{t-m(X_i)}{\sigma(X_i)}\right] f_e\left(\frac{t-m(X_i)}{\sigma(X_i)}\right) dt$$

$$+n^{-1}\sum_{i=1}^n \int x^k \bar{T}_x\,\varphi\left(X_i, T_i, \Delta_i, \frac{\bar{T}_x-m(x)}{\sigma(x)}\right) dF_X(x)$$

$$+n^{-1}\sum_{i=1}^n X_i^k \bar{T}_{X_i}\left[\eta(T_i,\Delta_i|X_i) + \zeta(T_i,\Delta_i|X_i)\frac{\bar{T}_{X_i}-m(X_i)}{\sigma(X_i)}\right] f_e\left(\frac{\bar{T}_{X_i}-m(X_i)}{\sigma(X_i)}\right)$$

$$+o_P(n^{-1/2})$$

and this equals $n^{-1}\sum_{i=1}^n \Psi_k(X_i,T_i,\Delta_i) + o_P(n^{-1/2})$. For the second term on the right hand side of (6.8) we have that

$$n((\boldsymbol{X}'\boldsymbol{X})^{-1} - [E(\boldsymbol{X}'\boldsymbol{X})]^{-1}) = A_n^{-1} - A^{-1} = (g_{kl}(A_n) - g_{kl}(A)).$$

Since the distribution of $(X, X^2, \ldots, X^p)$ is nonsingular, it follows that A has full rank and using similar techniques as in Theorem 17.8 in Arnold (1981), we also have that A_n has full rank with probability 1. Hence, a first order Taylor expansion applied to each $g_{kl}(A_n) - g_{kl}(A)$ yields that $A_n^{-1} - A^{-1}$ is asymptotically equivalent to

$$\sum_{r=0}^p\sum_{s=0}^p \tilde{G}_{rs}(A)(\hat{a}_{rs} - a_{rs}) = \sum_{r=0}^p\sum_{s=0}^p \tilde{G}_{rs}(A)\int x^{r+s}\,d(\hat{F}_X(x) - F_X(x)). \qquad (6.11)$$

Finally, by the use of Bernstein's inequality, it is seen that $n^{-1}\sum_{i=1}^{n}\Psi_k(X_i,T_i,\Delta_i)$ is $O(n^{-1/2}(\log n)^{1/2})$ a.s. Therefore, since the second term on the right hand side of (6.9) and the integral in (6.11) are $O(n^{-1/2}(\log n)^{1/2})$ a.s., we can replace the factor $n(\boldsymbol{X}'\boldsymbol{X})^{-1}$ in the first term on the right hand side of (6.8) by $n[E(\boldsymbol{X}'\boldsymbol{X})]^{-1}$. This finishes the proof.

Theorem 6.4

Under the assumptions of Theorem 6.3, we have

$$n^{1/2}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_T) \xrightarrow{d} N(0;\boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = (\sigma^2_{kl})$ *and*

$$\begin{aligned}\sigma^2_{k\ell} = \mathit{Cov}&\left\{\sum_{i=0}^{p} g_{ki}(A)\Psi_i(X,T,\Delta), \sum_{i=0}^{p} g_{\ell i}(A)\Psi_i(X,T,\Delta)\right\}\\ &+ \mathit{Cov}\left\{\sum_{i=0}^{p} g_{ki}(A)X^i\int t dF_e\left(\frac{t\wedge \bar{T}_X - m(X)}{\sigma(X)}\right) + \sum_{i,r,s=0}^{p}\tilde{g}^{rs}_{ki}(A)\int x^i t dF_T(x,t)X^{r+s},\right.\\ &\qquad\left.\sum_{i=0}^{p} g_{\ell i}(A)X^i\int t dF_e\left(\frac{t\wedge \bar{T}_X - m(X)}{\sigma(X)}\right) + \sum_{i,r,s=0}^{p}\tilde{g}^{rs}_{\ell i}(A)\int x^i t dF_T(x,t)X^{r+s}\right\}.\end{aligned}$$

The proof of this result is straightforward.

6.4 Appendix : Some technical results

In this appendix we show a number of technical results, needed in the proofs of the main results.

Proposition 6.1

Let the assumptions imposed in Theorem 6.1 hold. Then,

$$\begin{aligned}n^{-1/2}\sup_{\substack{x\in R_X\\ t\in I\!R}}\Bigg|\sum_{i=1}^{n}\Bigg\{\varphi&\left(X_i,T_i,\Delta_i,\frac{t\wedge \bar{T}_x - \hat{m}(x)}{\hat{\sigma}(x)}\right)\\ &- \varphi\left(X_i,T_i,\Delta_i,\frac{t\wedge \bar{T}_x - m(x)}{\sigma(x)}\right)\Bigg\}\Bigg| \xrightarrow{P} 0.\end{aligned}$$

Proof. Let $\mathcal{F}_1 = \{\varphi_{1t}(X,T,\Delta); -\infty < t \leq \bar{\bar{T}}\}$ and $\mathcal{F}_2 = \{\varphi_{2t}(X,T,\Delta); -\infty < t \leq \bar{\bar{T}}\}$, where

$$\varphi_{1t}(x,z,\delta) = \xi_e\left(\frac{z-m(x)}{\sigma(x)}, \delta, t\right),$$

$$\varphi_{2t}(x,z,\delta) = -S_e(t)\eta(z,\delta|x)\gamma_1(t|x) - S_e(t)\zeta(z,\delta|x)\gamma_2(t|x),$$

and $\bar{T} < \bar{\bar{T}} < \tau_{H_e}$. (Note that $\varphi(x,z,\delta,t) = \varphi_{1t}(x,z,\delta) + \varphi_{2t}(x,z,\delta)$). In the proof of Theorem 5.2 it was shown that the classes $\mathcal{F}_1$ and $\mathcal{F}_2$ are Donsker. Hence, it follows from Theorem 2.3.12 in van der Vaart and Wellner (1996) that for every decreasing sequence $\delta_n \downarrow 0$,

$$n^{-1/2}\text{sup}^\star\Big|\sum_{i=1}^n\{\varphi_{kt_2}(X_i,T_i,\Delta_i) - \varphi_{kt_1}(X_i,T_i,\Delta_i)\}\Big| \xrightarrow{P} 0 \quad (k=1,2), \tag{6.12}$$

where $\sup^\star$ is the supremum over all (t_1,t_2) such that $\text{Var}(\varphi_{kt_2} - \varphi_{kt_1}) \leq \delta_n$. We will show that for n large, this condition is satisfied for all pairs of the form $(e,\hat{e}) = (\frac{t\wedge\bar{T}_x - m(x)}{\sigma(x)}, \frac{t\wedge\bar{T}_x - \hat{m}(x)}{\hat{\sigma}(x)})$. Let $\delta_n = (nh_n)^{-1/2}(\log h_n^{-1})^{1/2}$. First note that, by Proposition 5.7, for n large, $|F_e(\hat{e}) - F_e(e)| \leq \delta_n(K_1 \sup_t f_e(t) + K_2 \sup_t |tf_e(t)|)$ and similarly for H_e^u, where $K_1, K_2 > 0$ do not depend on x or t. Using this, and Proposition 5.7, it is easy to see that $\sup_{u,z,\delta}|\varphi_{2\hat{e}}(u,z,\delta) - \varphi_{2e}(u,z,\delta)| \leq C_2\delta_n$ for some $C_2 > 0$, which implies that $\text{Var}(\varphi_{2\hat{e}} - \varphi_{2e}) \leq C_2^2\delta_n^2$. This argument cannot be used for $(\varphi_{1\hat{e}} - \varphi_{1e})(u,z,\delta)$, because φ_{1t} is not continuous in t. Note, however, that since the function ξ_e is the function ξ in the notation of Lo and Singh (1986), it follows that for any t_1, t_2,

$$\text{Cov}(\varphi_{1t_1}, \varphi_{1t_2}) = (1-F_e(t_1))(1-F_e(t_2))\int_{-\infty}^{t_1\wedge t_2}\frac{d\,H_e^u(s)}{(1-H_e(s))^2},$$

and hence $\text{Var}(\varphi_{1\hat{e}} - \varphi_{1e}) = \text{Var}(\varphi_{1e}) + \text{Var}(\varphi_{1\hat{e}}) - 2\text{Cov}(\varphi_{1e}, \varphi_{1\hat{e}}) \leq C_1\delta_n$ for some $C_1 > 0$. The result now follows from (6.12).

Proposition 6.2

Let the assumptions imposed in Theorem 6.1 hold. Then,

$$n^{-3/2}h_n^{-1}\sup_{\substack{x\in R_X\\ t\in I\!R}}\left|\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{X_j - X_i}{h_n}\right)\left[f_X^{-1}(X_j)\eta(T_i,\Delta_i|X_j)f_e\left(\frac{t\wedge\bar{T}_{X_j} - m(X_j)}{\sigma(X_j)}\right)\right.\right.$$

$$-f_X^{-1}(X_i)\eta(T_i,\Delta_i|X_i)\ f_e\left(\frac{t\wedge \bar T_{X_i}-m(X_i)}{\sigma(X_i)}\right)\Bigg]\ I(X_j\le x)\Bigg| \to 0 \quad \text{a.s.}$$

Proof. Define $h(x,z,\delta,t)=f_X^{-1}(x)\eta(z,\delta|x)f_e\left(\frac{t-m(x)}{\sigma(x)}\right)$. Then,

$$\begin{aligned}
&h(X_j,T_i,\Delta_i,t\wedge \bar T_{X_j})\\
&=h(X_j,T_i,\Delta_i,t)I(t\le \bar T_{X_j})+h(X_j,T_i,\Delta_i,\bar T_{X_j})I(t>\bar T_{X_j})\\
&=h(X_j,T_i,\Delta_i,t)I(t\le \bar T_{X_i})+h(X_j,T_i,\Delta_i,\bar T_{X_j})I(t>\bar T_{X_i})\\
&\quad+[h(X_j,T_i,\Delta_i,t)-h(X_j,T_i,\Delta_i,\bar T_{X_j})][I(t\le \bar T_{X_j})-I(t\le \bar T_{X_i})]\\
&=T_{ij}^{(1)}(t)+T_{ij}^{(2)}(t).
\end{aligned}$$

The term $T_{ij}^{(2)}(t)$ differs from zero only if t is between $\bar T_{X_i}$ and $\bar T_{X_j}$. Since $|\bar T_{X_j}-\bar T_{X_i}|=O(h_n)$ if $K(\frac{X_j-X_i}{h_n})\ne 0$, it follows that $|\bar T_{X_j}-t|=O(h_n)$ in that case. Hence, using a one-term Taylor expansion,

$$n^{-2}h_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{X_j-X_i}{h_n}\right)T_{ij}^{(2)}(t)I(X_j\le x)=O(h_n^2)$$

uniformly in x and t. It now follows that

$$\begin{aligned}
&n^{-2}h_n^{-1}\sum_{i,j=1}^n K\left(\frac{X_j-X_i}{h_n}\right)[h(X_j,T_i,\Delta_i,t\wedge \bar T_{X_j})-h(X_i,T_i,\Delta_i,t\wedge \bar T_{X_i})]I(X_j\le x)\\
&=n^{-2}h_n^{-1}\sum_{i,j=1}^n K\left(\frac{X_j-X_i}{h_n}\right)[h(X_j,T_i,\Delta_i,t)-h(X_i,T_i,\Delta_i,t)]I(t\le \bar T_{X_i})I(X_j\le x)\\
&\quad+n^{-2}h_n^{-1}\sum_{i,j=1}^n K\left(\frac{X_j-X_i}{h_n}\right)[h(X_j,T_i,\Delta_i,\bar T_{X_j})-h(X_i,T_i,\Delta_i,\bar T_{X_i})]\\
&\qquad\times I(t>\bar T_{X_i})I(X_j\le x).
\end{aligned}$$

We will show that the second term above is $o(n^{-1/2})$ a.s. The proof for the first term is completely analogous. Let $g(x,z,\delta)=h(x,z,\delta,T_x)$. Writing $g(X_j,T_i,\Delta_i)-g(X_i,T_i,\Delta_i)=(X_j-X_i)g'(X_i,T_i,\Delta_i)+\frac{1}{2}(X_j-X_i)^2g''(\xi_{ij},T_i,\Delta_i)$ for some ξ_{ij} between X_i and X_j (and where g' and g'' denote respectively the first and second derivative of $g(x,z,\delta)$ with respect to x), it is clear that it suffices to consider

$$n^{-2}h_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{X_j-X_i}{h_n}\right)(X_j-X_i)g'(X_i,T_i,\Delta_i)I(t>\bar{T}_{X_i})I(X_j\le x)$$
$$=n^{-2}h_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{X_j-X_i}{h_n}\right)(X_j-X_i)g'(X_i,T_i,\Delta_i)I(t>\bar{T}_{X_i})I(X_i\le x)$$
$$+n^{-2}h_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{X_j-X_i}{h_n}\right)(X_j-X_i)g'(X_i,T_i,\Delta_i)I(t>\bar{T}_{X_i})$$
$$\times[I(X_j\le x)-I(X_i\le x)].$$

The second term above is $O(h_n^2)$, since there are only $O(nh_n)$ i's and $O(nh_n)$ j's involved. The first one equals

$$(nh_n)^{-1}\sum_{i=1}^n\int K\left(\frac{u-X_i}{h_n}\right)(u-X_i)\,d(\hat{F}_X(u)-F_X(u))$$
$$\times g'(X_i,T_i,\Delta_i)I(t>\bar{T}_{X_i})I(X_i\le x)+O(h_n^2)$$
$$=n^{-3/2}h_n^{-1}\sum_{i=1}^n\int K\left(\frac{u-X_i}{h_n}\right)(u-X_i)\,d(\alpha_n(u)-\alpha_n(x_i^\star)) \tag{6.13}$$
$$\times g'(X_i,T_i,\Delta_i)I(t>\bar{T}_{X_i})I(X_i\le x)+O(h_n^2),$$

where $\alpha_n(u)=n^{1/2}(\hat{F}_X(u)-F_X(u))$ and $x_i^\star\in R_X$ is such that $K\left(\frac{x_i^\star-X_i}{h_n}\right)\neq 0$. Making the substitution $v=\frac{u-X_i}{h_n}$, using integration by parts and the fact that $\sup_{|x_2-x_1|\le h_n}|\alpha_n(x_2)-\alpha_n(x_1)|=O(h_n^{1/2}(\log h_n^{-1})^{1/2})$ a.s. (see Theorem 0.2 in Stute (1982)), (6.13) is easily seen to be $o(n^{-1/2})$ a.s.

Proposition 6.3
Let the assumptions imposed in Theorem 6.1 hold. Then,

$$\sup\left|\int_{-\infty}^{z}\int_{-\infty}^{x}(\hat{F}_X(x_1-vh_n)-F_X(x_1-vh_n))d\left(\hat{F}_D(x_1,z_1,\delta)-F_D(x_1,z_1,\delta)\right)\right|$$
$$=O_P(n^{-1}), \tag{6.14}$$

where the supremum is taken over all v belonging to the support of the kernel K, all $x \in R_X$ and all $-\infty < z < +\infty$, where

$$\hat{F}_D(x,z,\delta) - F_D(x,z,\delta) \qquad (6.15)$$
$$= n^{-1} \sum_{i=1}^{n} I(X_i \le x, T_i \le z, \Delta_i = \delta) - P(X \le x, T \le z, \Delta = \delta)$$

is the empirical process of the data (X_i, T_i, Δ_i) $(i = 1, \ldots, n)$ and where $\delta = 0$ or 1 is fixed.

Proof. We will apply Theorem 7 in Nolan and Pollard (1988) on the degenerate class of functions

$$\mathcal{F} = \Big\{(v,x,z) \to I(X_1 \le X_2 - vh_n, X_2 \le x, T_2 \le z, \Delta_2 = \delta)$$
$$- F_X(X_2 - vh_n) I(X_2 \le x, T_2 \le z, \Delta_2 = \delta)$$
$$- \int_{-\infty}^{x} \int_{-\infty}^{z} I(X_1 \le x_1 - vh_n)\, dF_D(x_1, z_1, \delta) + \int_{-\infty}^{x} \int_{-\infty}^{z} F_X(x_1 - vh_n)\, dF_D(x_1, z_1, \delta) \Big\}.$$

Showing that the three displayed conditions in the aforementioned theorem are satisfied will imply the weak convergence of the process stated between absolute values in (6.14) and hence the result will follow. In what follows we will show that these conditions are satisfied for the class

$$\mathcal{F}_1 = \{(v,x,z) \to I(X_1 \le X_2 - vh_n, X_2 \le x, T_2 \le z, \Delta_2 = \delta)\}.$$

The other terms are dealt with in an easier way. First note that the envelope function F_1 equals 1. Let $\varepsilon > 0$ be fixed. For the first condition, we need to find a subset $\mathcal{F}_1^\star$ of $\mathcal{F}_1$ (allowed to depend on the given sample), such that for each f in $\mathcal{F}_1$, there exists a $f^\star$ in $\mathcal{F}_1^\star$, such that $E_{T_n}|f - f^\star|^2 \le \varepsilon^2 n^2$, where the measure T_n is as defined on p. 1293 in Nolan and Pollard (1988). If f (respectively $f^\star$) corresponds to the triplet (v,x,z) (respectively $(v^\star, x^\star, z^\star)$), this means that

$$E_{T_n}|I(X_1 \le X_2 - vh_n, X_2 \le x, T_2 \le z, \Delta_2 = \delta)$$
$$-I(X_1 \le X_2 - v^\star h_n, X_2 \le x^\star, T_2 \le z^\star, \Delta_2 = \delta)| \le \varepsilon^2 n^2. \qquad (6.16)$$

Partition the domain of $X_2 - X_1$ into $O(\varepsilon^{-2})$ subintervals $[v_j, v_{j+1}]$ such that the number of $X_{i_2} - X_{i_1}$'s between each $v_j h_n$ and $v_{j+1} h_n$ is less than $K_1 \varepsilon^2 n^2$ for some constant $K_1 > 0$ to be specified later. Divide in a similar way R_X into $O(\varepsilon^{-2})$ intervals $[x_k, x_{k+1}]$ such that $\hat{F}_X(x_{k+1}) - \hat{F}_X(x_k) \leq K_2 \varepsilon^2$. Finally, divide the real line into $O(\varepsilon^{-2})$ subintervals $[z_l, z_{l+1}]$, such that the number of T_i's between z_l and z_{l+1} is never more than $K_3 \varepsilon^2 n^2$. It is easily seen that, by proper choice of K_1, K_2 and K_3, for any (v, x, z) there exist j, k and l such that (6.16) is satisfied for $(v^\star, x^\star, z^\star) = (v_j, x_k, z_l)$. This means that the covering number for the class $\mathcal{F}_1$ is $O(\varepsilon^{-6})$ and hence the covering integral $J(1, T_n, \mathcal{F}_1, F_1)$ is 6, uniformly over all samples and over all n. The second condition in Theorem 7 in Nolan and Pollard (1988) is satisfied by noting that for any $\gamma > 0$, $J(\gamma, T_n, \mathcal{F}_1, F_1) = 6\gamma(1 - \log \gamma)$, which equals 0 for $\gamma = e$. Finally, for the third condition, we have to calculate both $N(\varepsilon, P \times P_n, \mathcal{F}_1, F_1)$ and $N(\varepsilon, P_n \times P, \mathcal{F}_1, F_1)$, since the functions of the class $\mathcal{F}_1$ are not symmetric in (X_1, T_1) and (X_2, T_2). For $N(\varepsilon, P \times P_n, \mathcal{F}_1, F_1)$, we can use the same partitions for x and z as before and for v we choose $O(\varepsilon^{-2})$ points v_j such that $|v_{j+1} - v_j| \leq K_1 \varepsilon^2$ for some $K_1 > 0$. In a similar way as before, the class of functions corresponding to all triplets of the form (v_j, x_k, z_l) can be used to show that the covering number $N(\varepsilon, P \times P_n, \mathcal{F}_1, F_1)$ is $O(\varepsilon^{-6})$ uniformly over all samples and all n. Selecting the points v_j, x_k and z_l $(j, k, l = 1, \ldots, O(\varepsilon^{-2}))$ satisfying $|v_{j+1} - v_j| \leq K_1 \varepsilon^2, |x_{k+1} - x_k| \leq K_2 \varepsilon^2$ and $|H(z_{l+1}) - H(z_l)| \leq K_3 \varepsilon^2$ (where H is the distribution of T), it is easy to show that also $N(\varepsilon, P_n \times P, \mathcal{F}_1, F_1)$ is $O(\varepsilon^{-6})$. This shows that also the third condition of Theorem 7 in Nolan and Pollard (1988) is satisfied and hence the result follows.

Proposition 6.4

Let the assumptions imposed in Theorem 6.1 hold. Then,

$$n^{-2} h_n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{X_j - X_i}{h_n}\right) f_X^{-1}(X_j) \eta(T_i, \Delta_i | X_j) f_e\left(\frac{t \wedge \bar{T}_{X_j} - m(X_j)}{\sigma(X_j)}\right) I(X_j \leq x)$$
$$= n^{-1} \sum_{i=1}^{n} \eta(T_i, \Delta_i | X_i) f_e\left(\frac{t \wedge \bar{T}_{X_i} - m(X_i)}{\sigma(X_i)}\right) I(X_i \leq x) + R_{n1}(x, t),$$

where $\sup\{|R_{n1}(x, t)|; x \in R_X, t \in I\!R\} = o_P(n^{-1/2})$.

Proof. Using Proposition 6.2, it is clear that it suffices to show that

$$n^{-2}h_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{X_j-X_i}{h_n}\right) f_X^{-1}(X_i)\eta(T_i,\Delta_i|X_i)f_e\left(\frac{t\wedge \bar{T}_{X_i}-m(X_i)}{\sigma(X_i)}\right) I(X_j\le x)$$

$$-\,n^{-1}\sum_{i=1}^n \eta(T_i,\Delta_i|X_i)f_e\left(\frac{t\wedge \bar{T}_{X_i}-m(X_i)}{\sigma(X_i)}\right) I(X_i\le x) \qquad (6.17)$$

$$=n^{-1}\sum_{i=1}^n[\hat{f}_X(X_i)-f_X(X_i)]f_X^{-1}(X_i)\eta(T_i,\Delta_i|X_i)f_e\left(\frac{t\wedge \bar{T}_{X_i}-m(X_i)}{\sigma(X_i)}\right) I(X_i\le x)$$

$$+\,n^{-2}h_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{X_j-X_i}{h_n}\right) f_X^{-1}(X_i)\eta(T_i,\Delta_i|X_i)f_e\left(\frac{t\wedge \bar{T}_{X_i}-m(X_i)}{\sigma(X_i)}\right)$$

$$\times[I(X_j\le x)-I(X_i\le x)]$$

is $o_P(n^{-1/2})$ uniformly in (x,t), where $\hat{f}_X(x)=(nh_n)^{-1}\sum_{i=1}^n K(\frac{x-X_i}{h_n})$ is an estimator for the density $f_X(x)$. We will prove that the first term on the right hand side of (6.17) is $o_P(n^{-1/2})$. The proof for the second term is similar and will be left to the reader. Writing

$$\begin{aligned}\hat{f}_X(x)-f_X(x)&=h_n^{-1}\int K\left(\frac{x-u}{h_n}\right) d(\hat{F}_X(u)-F_X(u))+O(h_n^2)\\ &=h_n^{-1}\int(\hat{F}_X(x-vh_n)-F_X(x-vh_n))K'(v)\,dv+O(h_n^2),\end{aligned}$$

and introducing the notation $\hat{F}_D(D)-F_D(D)$ for the empirical process that corresponds to the data $D_i=(X_i,T_i,\Delta_i)$, $i=1,\dots,n$, and the notation

$$h_t(x_1,z_1,\delta_1)=f_X^{-1}(x_1)\eta(z_1,\delta_1|x_1)f_e\left(\frac{t\wedge \bar{T}_{x_1}-m(x_1)}{\sigma(x_1)}\right),$$

the first term on the right hand side of (6.17) can be written as

$$\begin{aligned}&h_n^{-1}\int(\hat{F}_X(x_1-vh_n)-F_X(x_1-vh_n))h_t(x_1,z_1,\delta_1)\\ &\times I(x_1\le x)\,d(\hat{F}_D(D_1)-F_D(D_1))\,K'(v)\,dv.\end{aligned}$$

Further, let

$$A_v(x,z)=\int_{-\infty}^{z}\int_{-\infty}^{x} n^{1/2}(\hat{F}_X(x_1-vh_n)-F_X(x_1-vh_n))\,dn^{1/2}(\hat{F}_D(x_1,z_1,\delta)-F_D(x_1,z_1,\delta))$$

for fixed $\delta = 0$ or 1. Then, it can be easily seen, that it suffices to consider

$$(nh_n)^{-1} \int h_t(x_1, z_1, \delta) I(x_1 \le x)\, dA_v(x_1, z_1) K'(v)\, dv \tag{6.18}$$

$$= (nh_n)^{-1} \int \int_{-\infty}^{z_1} \int_{x_L}^{x_1} dh_t(x_2, z_2, \delta) I(x_1 \le x)\, dA_v(x_1, z_1)\, K'(v)\, dv$$

$$+ (nh_n)^{-1} \int h_t(x_L, z_1, \delta) I(x_1 \le x)\, dA_v(x_1, z_1)\, K'(v)\, dv,$$

$$+ (nh_n)^{-1} \int \int_{x_L}^{x_1} dh_t(x_2, -\infty, \delta) I(x_1 \le x)\, dA_v(x_1, z_1)\, K'(v)\, dv,$$

where x_L is the left endpoint of the support of R_X. The first term on the right hand side of (6.18) can be written as

$$(nh_n)^{-1} \int [A_v(x, +\infty) - A_v(x, z_2) - A_v(x_2, +\infty) + A_v(x_2, z_2)]$$
$$\times I(x_2 > x_L)\, dh'_t(x_2, z_2, \delta)\, dx_2\, K'(v)\, dv,$$

(where $h'_t(x_2, z_2, \delta)$ denotes the partial derivative of $h_t(x_2, z_2, \delta)$ with respect to x_2) and this is bounded by (see e.g. Lemma B p. 254 in Serfling (1980))

$$K(nh_n)^{-1} \sup_{x,z,v} |A_v(x, z)| \sup_{t,x_2} \|h'_t(x_2, z_2, \delta)\|_V,$$

for some $K > 0$ ($\|h'_t(x_2, z_2, \delta)\|_V$ is the variation norm of the function $h'_t(x_2, z_2, \delta)$ considered as a function in z_2), which is $O_P((nh_n)^{-1})$ by Proposition 6.3 and assumption ($A4$). The derivation for the second and third term of (6.18) is similar, but easier. This finishes the proof.

Using arguments which are entirely analogous to the proof of Proposition 6.4 we obtain

Proposition 6.5
Let the assumptions imposed in Theorem 6.1 hold. Then,

$$n^{-2}h_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{X_j-X_i}{h_n}\right)f_X^{-1}(X_j)\zeta(T_i,\Delta_i|X_j)\frac{t\wedge\bar{T}_{X_j}-m(X_j)}{\sigma(X_j)}$$
$$\times f_e\left(\frac{t\wedge\bar{T}_{X_j}-m(X_j)}{\sigma(X_j)}\right)I(X_j\leq x)$$
$$=n^{-1}\sum_{i=1}^{n}\zeta(T_i,\Delta_i|X_i)\frac{t\wedge\bar{T}_{X_i}-m(X_i)}{\sigma(X_i)}f_e\left(\frac{t\wedge\bar{T}_{X_i}-m(X_i)}{\sigma(X_i)}\right)I(X_i\leq x)$$
$$+R_{n2}(x,t),$$

where $\sup\{|R_{n2}(x,t)|; x\in R_X, t\in I\!R\}=o_P(n^{-1/2})$.

The next result will be used in Proposition 6.7 for establishing uniform convergence to zero of a U-process.

Proposition 6.6
Let D, $\tilde{D}$ *be random vectors of dimension* r, $\tilde{r}$, *respectively, and let* $(D_i,\tilde{D}_i)$, $i=1,\ldots,n$, *be a random sample drawn from the joint distribution of* $(D,\tilde{D})$. *Let* $V(d_1,\tilde{d}_1,d_2,v)$ *be a real-valued function defined on* $I\!R^r\times I\!R^{\tilde{r}}\times I\!R^r\times T$, *where* $T=I\!R^r\times I\!R^s$ *(where* s *can be zero) and where* $v=(v_1,v_2)$ *with* $v_1\in I\!R^r$ *and* $v_2\in I\!R^s$. *Assume that the function* V *satisfies* $\sup_{d_1,\tilde{d}_1,d_2,v}|V(d_1,\tilde{d}_1,d_2,v)|<\infty$, $E_{D_1,\tilde{D}_1}(V(D_1,\tilde{D}_1,d_2,v))=0$ *for all* $d_2\in I\!R^r$ *and all* $v\in T$, *that as a process in* v_1, $\mathcal{B}_n(d,\tilde{d},v_1,v_2)=n^{-1/2}\sum_{j=1}^n V(d,\tilde{d},D_j,v_1,v_2)$ *is a pure jump process with jumps at* $v_1=D_j$, *and that* $\mathcal{B}_n(d,\tilde{d},v)$ *converges weakly to a zero-mean Gaussian process* $\mathcal{B}(d,\tilde{d},v)$, *which is continuous in all the indices* $d,\tilde{d}$ *and* v. *Then,*

$$n^{-3/2}\sup_{v\in T}|\sum_{i=1}^{n}\sum_{j=1}^{n}V(D_i,\tilde{D}_i,D_j,v)|\xrightarrow{P}0.$$

Proof. The continuity of the limiting Gaussian process together with the Skorohod-Dudley-Wichura theorem (cf. Shorack and Wellner (1986), p. 47) implies the existence of versions of the processes $\mathcal{B}_n(d,\tilde{d},v)$ $(n \geq 1)$, and $\mathcal{B}(d,\tilde{d},v)$ such that

$$\sup_{d,\tilde{d},v} |\mathcal{B}_n(d,\tilde{d},v) - \mathcal{B}(d,\tilde{d},v)| \to 0,$$

almost surely. (For simplicity, we denote the a.s. convergent versions of the processes by the same symbols as the original processes. We will do the same for the versions of the random vectors $(D_j,\tilde{D}_j)$, $j = 1,\ldots,n$ to be defined in the new space.) Note that in this new space $\mathcal{B}_n(d,\tilde{d},v_1,v_2)$ is also a pure jump process in v_1 and the points where it jumps define the realization of the random vectors D_j. Thus the representation $\mathcal{B}_n(d,\tilde{d},v_1,v_2) = n^{-1/2}\sum_{j=1}^n V(d,\tilde{d},D_j,v_1,v_2)$ holds also in the new space. Let $\tilde{D}_j$ be a random variable generated according to the conditional distribution of $\tilde{D}$ given $D = D_j$. Assumption $E_{D_1,\tilde{D}_1}(V(D_1,\tilde{D}_1,d_2,v)) = 0$ implies

$$E_{D_1,\tilde{D}_1}\left[\mathcal{B}_n(D_1,\tilde{D}_1,v)|\omega\right] = \int \mathcal{B}_n(\omega,d_1,\tilde{d}_1,v)dF_{D_1,\tilde{D}_1}(d_1,\tilde{d}_1) = 0, \tag{6.19}$$

where conditioning on ω means conditioning on each sample path of the process $\mathcal{B}_n(d,\tilde{d},v)$, and $\mathcal{B}_n(\omega,d,\tilde{d},v)$ denotes the sample path. Consider now the process $n^{-1}\sum_{i=1}^n \mathcal{B}(D_i,\tilde{D}_i,v)$ as a process in v and write it as $n^{-1}\sum_{i=1}^n \mathcal{B}_v(\omega,D_i,\tilde{D}_i)$ to stress the fact that D_i and $\tilde{D}_i$, $i = 1,\ldots,n$, are not the only sources of randomness in this process. We will show the weak convergence of this process to the zero process. To show the convergence of the finite dimensional distributions, it suffices to show that for a fixed v, $n^{-1}\sum_{i=1}^n \mathcal{B}_v(\omega,D_i,\tilde{D}_i) \xrightarrow{P} 0$. Write

$$\begin{aligned} &E\left[n^{-1}\sum_{i=1}^n \mathcal{B}_v(\omega,D_i,\tilde{D}_i)\right]^2 \qquad (6.20)\\ &= n^{-2}\sum_{i_1\neq i_2} E[E\{\mathcal{B}_v(\omega,D_{i_1},\tilde{D}_{i_1})|\omega\}E\{\mathcal{B}_v(\omega,D_{i_2},\tilde{D}_{i_2})|\omega\}]\\ &\quad + n^{-2}\sum_{i=1}^n E[\mathcal{B}_v^2(\omega,D_i,\tilde{D}_i)]. \end{aligned}$$

By adding and subtracting the zero term in (6.19), it can be seen that

$$|E(\mathcal{B}_v(\omega,D_i,\tilde{D}_i)|\omega)| \leq \sup_{d,\tilde{d},v} |\mathcal{B}_n(\omega,d,\tilde{d},v) - \mathcal{B}(\omega,d,\tilde{d},v)| \to 0,$$

for almost all ω, which implies that $E(\mathcal{B}_v(\omega, D_i, \tilde{D}_i)|\omega) = 0$ a.s. Thus, the first term on the right hand side of (6.20) equals zero. The second term of (6.20) is $O(n^{-1})$, since V is uniformly bounded over all variables. Hence, by Chebyshev's inequality, $n^{-1}\sum_{i=1}^n \mathcal{B}_v(\omega, D_i, \tilde{D}_i) \xrightarrow{P} 0$. Finally, we consider

$$P\left(\sup_{\rho(v,v')<\delta}\left|n^{-1}\sum_{i=1}^n[\mathcal{B}_{v'}(\omega, D_i, \tilde{D}_i) - \mathcal{B}_v(\omega, D_i, \tilde{D}_i)]\right| > \varepsilon\right)$$

$$\leq P\left(\sup_{\rho(v,v')<\delta}\sup_{d,\tilde{d}}|\mathcal{B}(d,\tilde{d},v') - \mathcal{B}(d,\tilde{d},v)| > \varepsilon\right).$$

Since $\mathcal{B}$ is a tight process, this shows the tightness of the process $n^{-1}\sum_{i=1}^n \mathcal{B}_v(\omega, D_i, \tilde{D}_i)$ (see e.g. Theorem 10.2 in Pollard (1990)) and finishes the proof.

Proposition 6.7

Let the assumptions imposed in Theorem 6.1 hold. Then,

$$n^{-2}\sum_{i=1}^n\sum_{j=1}^n \varphi\left(X_i, T_i, \Delta_i, \frac{t\wedge\bar{T}_{X_j} - m(X_j)}{\sigma(X_j)}\right) I(X_j \leq x)$$

$$= n^{-1}\sum_{i=1}^n E\left\{\varphi\left(X_i, T_i, \Delta_i, \frac{t\wedge\bar{T}_X - m(X)}{\sigma(X)}\right) I(X\leq x)\middle| X_i, T_i, \Delta_i\right\} + R_{n3}(x,t),$$

where $\sup\{|R_{n3}(x,t)|; x \in R_X, t \in I\!R\} = o_P(n^{-1/2})$.

Proof. We start with the second term of φ to which we will apply Proposition 6.6 with $D = X$, $\tilde{D} = (T, \delta)$, $v = (x, t)$ and

$$V(d_1, \tilde{d}_1, d_2, u) = V(x_1, z_1, \delta_1, x_2, x, t)$$

$$= S_e\left(\frac{t\wedge\bar{T}_{x_2} - m(x_2)}{\sigma(x_2)}\right)\eta(z_1, \delta_1|x_1)\,\gamma_1\left(\frac{t\wedge\bar{T}_{x_2} - m(x_2)}{\sigma(x_2)}\middle| x_1\right) I(x_2\leq x)$$

$$-E\left[S_e\left(\frac{t\wedge\bar{T}_{X_2} - m(X_2)}{\sigma(X_2)}\right)\eta(z_1, \delta_1|x_1)\,\gamma_1\left(\frac{t\wedge\bar{T}_{X_2} - m(X_2)}{\sigma(X_2)}\middle| x_1\right) I(X_2\leq x)\right].$$

Since $E[\xi(T,\Delta,t|X)|X] = 0$, we also have that $E[\eta(T,\Delta|X)|X] = 0$ and hence it suffices for the second term of φ to show the weak convergence of the process

$n^{-1/2}\sum_{j=1}^n V(x_1, z_1, \delta_1, X_j, x, t)$ to a zero-mean Gaussian process. This will be done by showing that the class $\mathcal{F} = \{x_2 \to V^\star(x_1, z_1, \delta_1, x_2, x, t)\}$ is Donsker, where $V^\star$ equals the first term of V (see van der Vaart and Wellner (1996), p. 81). For this we need to show that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty,$$

where $N_{[]}$ denotes the bracketing number (see Theorem 2.5.6 in the same book). Since the function $x_2 \to S_e(\frac{t-m(x_2)}{\sigma(x_2)})$ is bounded and continuously differentiable, it follows from Corollary 2.7.2 in the aforementioned book that $m = O(\exp(K\varepsilon^{-1}))$ brackets are required for the class $\{x_2 \to S_e(\frac{t-m(x_2)}{\sigma(x_2)})\}$. Hence, by truncating these brackets at $S_e(\frac{\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)})$, the same number is needed for the class $\{x_2 \to S_e(\frac{t\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)})\}$, since $S_e(\frac{t\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}) = S_e(\frac{t-m(x_2)}{\sigma(x_2)}) \vee S_e(\frac{\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)})$. For γ_1, we first divide R_X into $O(\varepsilon^{-1})$ subintervals $[t_i, t_{i+1}]$ such that $|t_{i+1}-t_i| \le K\varepsilon$ for some $K > 0$. For each fixed i, there exist m brackets that cover the class $\{x_2 \to \gamma_1(\frac{t-m(x_2)}{\sigma(x_2)}|t_i); t \in I\!R\}$. Truncating these brackets at $\gamma_1(\frac{\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}|t_i)$ shows that the bracketing number for the truncated class $\{x_2 \to \gamma_1(\frac{t\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}|t_i); t \in I\!R\}$ is also m. Using the (more general) definition of bracketing number stated on p. 211 in van der Vaart and Wellner (1996), it is clear that the class $\{x_2 \to \gamma_1(\frac{t\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}|x_1); x_1 \in R_X, t \in I\!R\}$ can be covered by using all the $O(\varepsilon^{-1}\exp(K\varepsilon^{-1}))$ brackets corresponding to all the points t_i $(i = 1, \ldots, O(\varepsilon^{-1}))$, since for any $t_i \le x_1, x_1' \le t_{i+1}$ and for any $t, t' \in I\!R$,

$$\begin{aligned}
&\gamma_1\left(\frac{t'\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}\middle| x_1'\right) - \gamma_1\left(\frac{t\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}\middle| x_1\right) \\
&\le \gamma_1\left(\frac{t'\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}\middle| x_1'\right) - \gamma_1\left(\frac{t'\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}\middle| t_i\right) \\
&\quad + \gamma_1\left(\frac{t'\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}\middle| t_i\right) - \gamma_1\left(\frac{t\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}\middle| t_i\right) \\
&\quad + \gamma_1\left(\frac{t\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}\middle| t_i\right) - \gamma_1\left(\frac{t\wedge\bar{T}_{x_2}-m(x_2)}{\sigma(x_2)}\middle| x_1\right).
\end{aligned}$$

The L_2-norm of each of the three terms above is less than (a constant times) ε, provided t and t' belong to the same bracket from the class of m brackets determined by t_i. Finally, since $\eta(z_1, \delta_1|x_1)I(x_2 \leq x)$ is increasing in x_2 and bounded, it requires $O(\exp(K\varepsilon^{-1}))$ brackets by Theorem 2.7.5 in van der Vaart and Wellner (1996). This shows the weak convergence of the process $n^{-1/2}\sum_j V(x_1, z_1, \delta_1, X_j, x, t)$. The third term of φ is dealt with in a similar way. For the first term, we use the following decomposition of the function ξ_e :

$$\begin{aligned}\xi_e(E,\Delta,t) &= (1-F_e(t))\left\{-\int_{-\infty}^{E\wedge t}\frac{dH_e^u(s)}{(1-H_e(s))^2}+\int_{-\infty}^{t}\frac{dH_e^u(s)}{1-H_e(s)}\right\}\\ &\quad+(1-F_e(t))\left\{\frac{I(E\leq t,\Delta=1)}{1-H_e(E)}-\int_{-\infty}^{t}\frac{dH_e^u(s)}{1-H_e(s)}\right\}\\ &=\xi_{e1}(E,\Delta,t)+\xi_{e2}(E,\Delta,t).\end{aligned}$$

For $n^{-2}\sum_{i,j}\xi_{e1}(E_i, \Delta_i, \frac{t\wedge\bar{T}_{X_j}-m(X_j)}{\sigma(X_j)})I(X_j \leq x)$, the same arguments as for the second term of φ show that this term is asymptotically equivalent to $n^{-1}\sum_{i=1}^n E\{\xi_{e1}(E_i, \Delta_i, \frac{t\wedge\bar{T}_X-m(X)}{\sigma(X)})I(X \leq x)|E_i, \Delta_i\}$. (To show that the class

$$\left\{u\rightarrow\left(1-F_e\left(\frac{t\wedge\bar{T}_u-m(u)}{\sigma(u)}\right)\right)\int_{-\infty}^{\frac{t\wedge\bar{T}_u-m(u)}{\sigma(u)}}\frac{I(s\leq E)}{(1-H_e(s))^2}\,dH_e^u(s)\,I(u\leq x)\right\}$$

is Donsker, use similar arguments as above for $F_e(\frac{t\wedge\bar{T}_u-m(u)}{\sigma(u)})$ and for $I(u \leq x)$, and apply Corollary 2.7.2 in van der Vaart and Wellner (1996) on the integral). On the term ξ_{e2} we will apply Theorem 7 in Nolan and Pollard (1988). Let

$$\begin{aligned}\mathcal{F}=\Bigg\{&(x,t)\rightarrow\xi_{e2}\left(\frac{T_1-m(X_1)}{\sigma(X_1)},\Delta_1,\frac{t\wedge\bar{T}_{X_2}-(X_2)}{\sigma(X_2)}\right)I(X_2\leq x)\\ &-E\left[\xi_{e2}\left(\frac{T_1-(X_1)}{\sigma(X_1)},\Delta_1,\frac{t\wedge\bar{T}_{X_2}-m(X_2)}{\sigma(X_2)}\right)I(X_2\leq x)\middle|X_1,T_1,\Delta_1\right];\\ &x\in R_X,t\in I\!R\Bigg\}.\end{aligned}$$

Note that this is a degenerate class of functions. We will restrict ourselves to verifying the conditions in that theorem for the class :

$$\mathcal{F}_1 = \left\{(x,t) \to \left(1 - F_e\left(\frac{t \wedge \bar{T}_{X_2} - m(X_2)}{\sigma(X_2)}\right)\right)\left(1 - H_e\left(\frac{T_1 - m(X_1)}{\sigma(X_1)}\right)\right)^{-1}\right.$$
$$\left.\times I\left(\frac{T_1 - m(X_1)}{\sigma(X_1)} \leq \frac{t \wedge \bar{T}_{X_2} - m(X_2)}{\sigma(X_2)}, \Delta_1 = 1\right) I(X_2 \leq x)\right\}.$$

The verification of the conditions in Theorem 7 in Nolan and Pollard (1988) for the other terms is similar, but easier. Let $F_1 \equiv (1 - H_e(\bar{T}))^{-1}$ denote the envelope function for this class and let $\varepsilon > 0$. We start with showing that there exists a partition $[t_i, t_{i+1}]$ of the real line such that for all i, all $t_i \leq t \leq t_{i+1}$ and all $x \in R_X$,

$$F_e\left(\frac{t \wedge \bar{T}_x - m(x)}{\sigma(x)}\right) - F_e\left(\frac{t_i \wedge \bar{T}_x - m(x)}{\sigma(x)}\right) \leq \varepsilon. \tag{6.21}$$

To see this, first note that for ε small enough $F_e^{-1}(\varepsilon) \geq -\varepsilon^{-1}$ and $F_e^{-1}(1-\varepsilon) \leq \varepsilon^{-1}$. Hence, all $t \leq -\varepsilon^{-1}\sup_x \sigma(x) + \inf_x m(x) = a_\varepsilon$ satisfy $F_e\left(\frac{t-m(x)}{\sigma(x)}\right) \leq \varepsilon$ (for all x) and all $t \geq \varepsilon^{-1}\sup_x \sigma(x) + \sup_x m(x) = b_\varepsilon$ satisfy $F_e\left(\frac{t-m(x)}{\sigma(x)}\right) \geq 1 - \varepsilon$ (for all x). Now divide the interval $[a_\varepsilon, b_\varepsilon]$ into $K = O(\varepsilon^{-2})$ subintervals $[t_i, t_{i+1}]$ ($t_1 = a_\varepsilon, t_K = b_\varepsilon$) such that $|t_{i+1} - t_i| \leq \varepsilon(\sup_z f_e(z))^{-1}\inf_x \sigma(x)$. Then, if $t_0 = -\infty$ and $t_{K+1} = +\infty$, it is readily shown that (6.21) is satisfied. Next, let us divide the line into $O(\varepsilon^{-2})$ subintervals $[\tilde{t}_k, \tilde{t}_{k+1}]$ such that the number of couples (i,j) for which $\tilde{t}_k \leq \frac{T_i - m(X_i)}{\sigma(X_i)}\sigma(X_j) + m(X_j) \leq \tilde{t}_{k+1}$ is less than $\varepsilon^2 n^2$. Finally, partition R_X into $O(\varepsilon^{-2})$ subintervals $[x_i, x_{i+1}]$ such that $\hat{F}_X(x_{i+1}) - \hat{F}_X(x_i) \leq \varepsilon^2$. Now let $\mathcal{F}_1^\star$ be the subclass of $\mathcal{F}_1$ consisting of all functions for which the corresponding couple (x,t) equals (x_i, t_j) or $(x_i, \tilde{t}_k)$ $(i,j,k = 1, \ldots, O(\varepsilon^{-2}))$. It is readily seen that for any f in $\mathcal{F}_1$, there is a $f^\star$ in $\mathcal{F}_1^\star$ for which $E_{T_n}|f - f^\star|^2 \leq 3(1 - H_e(\bar{T}))^{-2}\varepsilon^2 n^2$, where T_n is the measure defined on p. 1293 in Nolan and Pollard (1988). This shows that the covering number $N(\varepsilon, T_n, \mathcal{F}_1, F_1)$ is $O(\varepsilon^{-4})$ uniformly over all samples and over all n and hence the covering integral $J(1, T_n, \mathcal{F}_1, F_1)$ is 4. Hence, the first condition in Theorem 7 in Nolan and Pollard (1988) is satisfied. The second one is easily verified by an application of Chebyshev's inequality and by noting that the covering integral $J(\gamma, T_n, \mathcal{F}_1, F_1) = 4\gamma(1-\log\gamma)$ equals zero for $\gamma = e$. Finally, for the third condition, we calculate first the covering number $N(\varepsilon, P\times P_n, \mathcal{F}_1, F_1)$ (note that since the functions in the class $\mathcal{F}_1$ are not symmetric, we have to consider both $P_n \times P$ and $P \times P_n$). For this, the partitions x_i ($i = 1, \ldots, O(\varepsilon^{-2})$) and t_j ($j = 1, \ldots, O(\varepsilon^{-2})$) constructed above can be used here to take care of respectively the first and the last factor of the functions in $\mathcal{F}_1$. For the indicator $I(\frac{T_1-m(X_1)}{\sigma(X_1)} \leq \frac{t\wedge \bar{T}_{X_2}-m(X_2)}{\sigma(X_2)}, \Delta_1 = 1)$, we partition the real line into $O(\varepsilon^{-2})$ subintervals $[\bar{t}_k, \bar{t}_{k+1}]$ such that

$$H_e^u\left(\frac{t\wedge \bar{T}_x - m(x)}{\sigma(x)}\right) - H_e^u\left(\frac{\bar{t}_k\wedge \bar{T}_x - m(x)}{\sigma(x)}\right) \leq \varepsilon$$

for all $\bar{t}_k \leq t \leq \bar{t}_{k+1}$ and for all $x \in R_X$ (this can be done in a similar way as for equation (6.21)). The subclass of $\mathcal{F}_1$ corresponding to all couples (x_i, t_j) and $(x_i, \bar{t}_k)$ can be used to show that the covering number is not more than $O(\varepsilon^{-4})$. For the covering number with respect to the measure $P_n \times P$, we again use the partition t_j ($j = 1, \ldots, O(\varepsilon^{-2})$) for the factors $1 - F_e(\frac{t\wedge \bar{T}_{X_2}-m(X_2)}{\sigma(X_2)})$. For $I(X_2 \leq x)$ we divide R_X into $O(\varepsilon^{-2})$ intervals $[\tilde{x}_i, \tilde{x}_{i+1}]$ such that $F_X(\tilde{x}_{i+1}) - F_X(\tilde{x}_i) \leq \varepsilon^2$. A more complicated partition is needed for the remaining factor, for which subintervals (say $[t_k^\star, t_{k+1}^\star]$) need to be constructed satisfying

$$\sum_{i=1}^n \left\{ P\left(\frac{T_i - m(X_i)}{\sigma(X_i)}\sigma(X) + m(X) \leq t_{k+1}^\star \,\middle|\, X_i, T_i\right) - P\left(\frac{T_i - m(X_i)}{\sigma(X_i)}\sigma(X) + m(X) \leq t_k^\star \,\middle|\, X_i, T_i\right)\right\} \leq \varepsilon^2 n. \tag{6.22}$$

Let V be the number of $\frac{T_i-m(X_i)}{\sigma(X_i)}$ that are less than $H_e^{-1}(\varepsilon^2/2)$ or greater than $H_e^{-1}(1-\varepsilon^2/2)$. Then, $V \sim \text{Bin}(n, \varepsilon^2)$ and hence, for n large, $V - n\varepsilon^2 \leq 2(\varepsilon^2(1-\varepsilon^2)n\log\log n)^{1/2}$ a.s. (see e.g. Serfling (1980) p. 35). So, we only need to consider the terms in (6.22) for which $H_e^{-1}(\varepsilon^2/2) \leq \frac{T_i-m(X_i)}{\sigma(X_i)} \leq H_e^{-1}(1-\varepsilon^2/2)$. As before, we have that $[H_e^{-1}(\varepsilon^2/2), H_e^{-1}(1-\varepsilon^2/2)] \subseteq [-\frac{\varepsilon^{-2}}{2}, \frac{\varepsilon^{-2}}{2}]$ for ε small enough. Hence, it suffices to consider values of t for which $t_L = -\frac{\varepsilon^{-2}}{2}\sup_x \sigma(x) + \inf_x m(x) \leq t \leq \frac{\varepsilon^{-2}}{2}\sup_x \sigma(x) + \sup_x m(x) = t_U$. Using assumption $(A5)$, (6.22) can be achieved for these values of t by using at most $O(\varepsilon^{-4})$ subintervals $[t_k^\star, t_{k+1}^\star]$. This shows that also the last condition of Theorem 7 in Nolan and Pollard (1988) is satisfied and hence

$$n^{-1} \sup_{\substack{x \in R_X \\ t \in I\!R}} \left| \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \xi_{e2} \left(\frac{T_i - m(X_i)}{\sigma(X_i)}, \Delta_i, \frac{t \wedge \bar{T}_{X_j} - m(X_j)}{\sigma(X_j)} \right) I(X_j \leq x) \right. \right.$$

$$\left. \left. -E \left[\xi_{e2} \left(\frac{T_i - (X_i)}{\sigma(X_i)}, \Delta_i, \frac{t \wedge \bar{T}_{X_j} - m(X_j)}{\sigma(X_j)} \right) I(X_j \leq x) \middle| X_i, T_i, \Delta_i \right] \right\} \right| = O_P(1).$$

This completes the proof.

Chapter 7

Comparison of two estimators of the conditional distribution in regression with censored data

In Chapters 2 and 3 we studied the Beran estimator $\tilde{F}(t|x)$, which estimates (in a completely nonparametric way) the conditional distribution of the response Y (subject to censoring) given the covariate x, in a completely nonparametric way. The main disadvantage of this estimator is its inconsistency in the right tail whenever heavy censoring is present. In Chapter 5 we constructed an alternative for the Beran estimator, denoted by $\hat{F}(t|x)$, which is valid for heteroscedastic regression models. We explained, in an intuitive way, why the estimator $\hat{F}(t|x)$ behaves well in the right tail, while the Beran estimator $\tilde{F}(t|x)$ is inconsistent in the tail. In this chapter we will verify our intuitive expectations of $\hat{F}(t|x)$ in a simulation study. In Section 7.1 we start by comparing the finite sample performance of the estimators $\tilde{F}(t|x)$ and $\hat{F}(t|x)$ through some simulations. Next, we will compare the performance of $\hat{F}(t|x)$ for different choices of the location functional $m(\cdot)$ and the scale functional $\sigma(\cdot)$ (recall that under model (1.28), the choice of the location and scale functional has no influence on the validity of the model (see Proposition 5.1)). The sensitivity of the estimator $\hat{F}(t|x)$ to the model assumption (1.28) will also be investigated, more precisely we will examine the performance of $\hat{F}(t|x)$ when model (1.28) is not satisfied and we will compare the behavior of $\tilde{F}(t|x)$ and $\hat{F}(t|x)$ in this situation. In Section 7.2, the data from the Stanford heart transplant program (see Figure 1.1) will be analyzed. In this study, the effect of the age of a patient receiving a heart transplantation, on his (or her) survival time after transplantation was one of the points of interest. We will estimate the median regression curve, as well as the conditional distribution of the survival time for several ages and we will compare our analysis with that of other approaches found in the literature.

The results of this chapter can be found in Van Keilegom, Akritas and Veraverbeke (1998).

7.1 Simulations

7.1.1 Comparison of the two estimators

Throughout this chapter, we will work with a random design. (Note that we studied the Beran estimator $\tilde{F}(t|x)$ under a fixed design, while for the estimator $\hat{F}(t|x)$ we worked with a random design). Assume that the covariate X is uniformly distributed on the interval $[0, 1]$. The conditional distribution of the response Y given the covariate $X = x$ is given by

$$(Y|X = x) \sim \text{Exp}\left(\frac{1}{a_0 + a_1 x + a_2 x^2}\right) \tag{7.1}$$

(the regression function is a quadratic function in x), while the censoring time C has the conditional distribution

$$(C|X = x) \sim \text{Exp}\left(\frac{1}{b_0 + b_1 x + b_2 x^2}\right). \tag{7.2}$$

Here, the constants a_0, a_1, a_2, b_0, b_1 and b_2 are chosen in such a way that $a_0 + a_1 x + a_2 x^2 > 0$ for all $0 \leq x \leq 1$ and similarly for the b's. As usual, we assume that the response Y and the censoring time C are independent, conditionally on X. Hence, it is readily verified that $(T|X = x) \sim \text{Exp}(1/(a_0 + a_1 x + a_2 x^2) + 1/(b_0 + b_1 x + b_2 x^2))$ and that

$$P(\Delta = 0|x) = \frac{a_0 + a_1 x + a_2 x^2}{a_0 + b_0 + (a_1 + b_1)x + (a_2 + b_2)x^2}.$$

To show that model (1.28) holds for these distributions, it suffices to check assumption (1.28) when $m(x)$ is the conditional mean function and $\sigma(x)$ the conditional standard deviation (since model (1.28) holds for any location and scale functional if it is valid for one location and one scale functional (see Proposition 5.1)). This follows easily by showing that $P(\varepsilon \leq t|x) = 1 - \exp(-(t+1))$, which is independent of x.

We carried out the simulations for samples of size $n = 150$ and for the covariate $x = 0.5$. The results are obtained by using 1000 simulations, except for Figure 7.2 where 2500 simulations are used. For the weights that appear in the Beran estimator $\tilde{F}(t|x)$, we chose a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$, while for the bandwidth sequence h_n, we selected the minimizer of the asymptotic mean squared error (AMSE) :

$$\begin{aligned}\mathrm{AMSE}(h_n) &= \mathrm{AsVar}(\tilde{F}(t|x)) + (\mathrm{AsBias}(\tilde{F}(t|x)))^2 \\ &= (nh_n)^{-1}\Gamma(t,t|x) + h_n^4 \bar{b}^2(t|x),\end{aligned}$$

where

$$\begin{aligned}\bar{b}(t|x) = \frac{1}{2}\mu_2^K (1 - F(t|x)) &\left[\int_{-\infty}^{t} \left\{ \frac{\ddot{H}(s|x)\, dH^u(s|x)}{(1 - H(s|x))^2} + \frac{d\ddot{H}^u(s|x)}{1 - H(s|x)} \right\} \right. \\ &\left. + 2 f_X'(x) f_X^{-1}(x) \int_{-\infty}^{t} \left\{ \frac{\dot{H}(s|x)\, dH^u(s|x)}{(1 - H(s|x))^2} + \frac{d\dot{H}^u(s|x)}{1 - H(s|x)} \right\} \right],\end{aligned}$$

$$\Gamma(s,t|x) = f_X^{-1}(x)\, \|K\|_2^2\, (1 - F(s|x))(1 - F(t|x)) \int_{-\infty}^{s\wedge t} \frac{dH^u(y|x)}{(1 - H(y|x))^2}$$

(see Dabrowska (1992a)). Hence, the optimal choice is

$$h_n = h_n(x,t) = \left(\frac{\Gamma(t,t|x)}{4\bar{b}^2(t|x)} \right)^{1/5} n^{-1/5}. \tag{7.3}$$

In Remark 2.1, we calculated the optimal bandwidth of the Beran estimator for fixed, asymptotically equidistant design points. The present formula, which is valid for any random design, reduces to expression (2.16) obtained in Remark 2.1 by choosing $f_X(x) = I(-1 \leq x \leq 1)$, i.e. by choosing a uniform design. Note that formula (7.3) depends on both x and t. Since we need to know the optimal bandwidth at all positive values of t, an appropriate approximation rule for formula (7.3) is therefore required. We worked with the following procedure : we divided the region $[0, F^{-1}(0.99|x)]$ into 20 intervals and in each interval, we worked with the optimal bandwidth for its midpoint. In Hall, Marron and Titterington (1995) another approximation for full local smoothing is introduced in the context of density estimation. Instead of considering the midpoint of the interval, their 'partial local bandwidth' is obtained by averaging the different pieces in the formula of the local bandwidth over the considered interval. However, we chose to work with the former method because of its simplicity. As a consequence of the fact that different bandwidths are used in the calculation of the Beran estimator, it might happen that the Beran estimator is not increasing everywhere. In these cases, the estimator is kept constant until it starts increasing again. In a number of situations, the optimal bandwidth h_n in a point x was larger than the distance from x to both the left and right endpoint of the interval. In such cases, the bandwidth was redefined as the maximum of these two distances.

For the location and scale functionals $m(x)$ and $\sigma(x)$, defined in (1.32) and (1.33), we need to choose a proper weight function J. Since $\tilde{F}^{-1}(s|x)$ is only defined for s less than $\tilde{F}(+\infty|x)$, the function $J(s)$ must be 0 from that point on (otherwise $\hat{m}(x)$ and $\hat{\sigma}(x)$ are not defined). We worked with $J(s) = I(a \leq s \leq b)/(b-a)$ (which leads to trimmed means and variances), where $a = 0$ and $b = \min_{1 \leq i \leq n} \tilde{F}(+\infty|X_i)$ (since we need to calculate $\hat{m}(X_i)$ and $\hat{\sigma}(X_i)$ for all $i = 1, \ldots, n$).

Figure 7.1 shows the function $P(\Delta = 0|x)$ when $a_0 = 1, a_1 = 5, a_2 = 100, b_0 = 20, b_1 = -10$ and $b_2 = 5$. At $x = 0.5$, we have approximately 65% censoring. In Figure 7.2, the exponential distribution $F(t|0.5)$ is plotted for this set of parameters. The median and the 2.5 and 97.5 percentiles of the estimates of $\tilde{F}(t|0.5)$ and $\hat{F}(t|0.5)$ from the 2500 runs, are also plotted. We chose to work with percentiles as measures of location and scale, rather than with the mean and the variance, because it is clear from Figure 7.2 that the distribution of $\tilde{F}(t|0.5)$ and $\hat{F}(t|0.5)$ are skewed to the left for large values of t, and hence the mean and the variance do not seem very

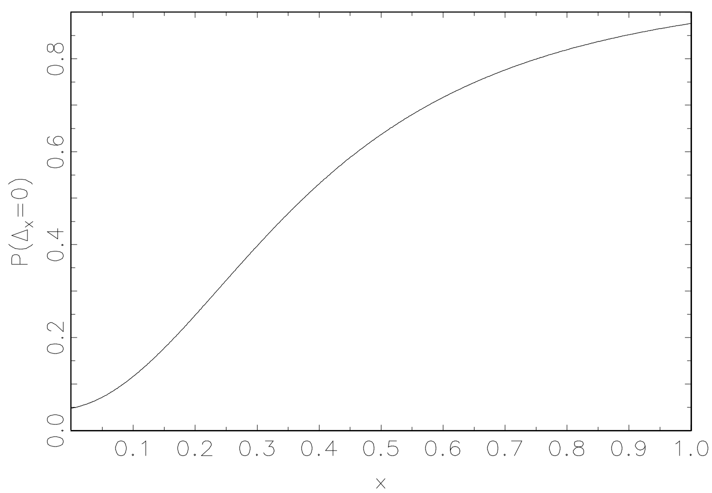

Figure 7.1: *Probability of censoring for exponential survival and censoring distributions when* $a_0 = 1, a_1 = 5, a_2 = 100, b_0 = 20, b_1 = -10$ *and* $b_2 = 5$.

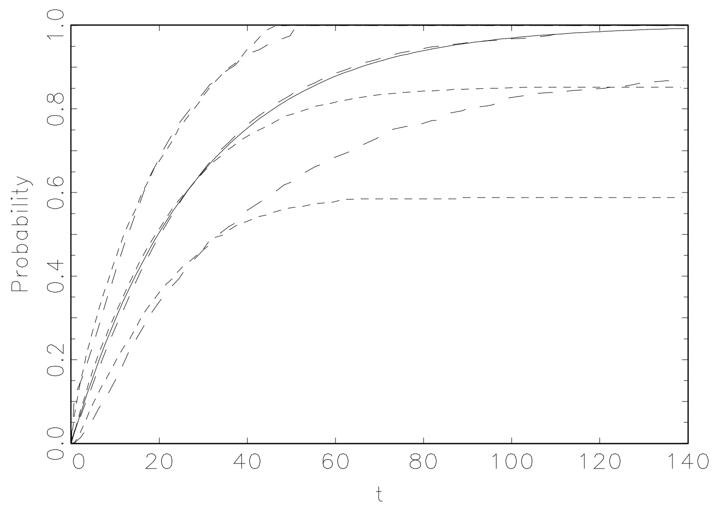

Figure 7.2: *Curve of the exponential distribution* $F(t|0.5)$ *(full line) when* $a_0 = 1, a_1 = 5, a_2 = 100, b_0 = 20, b_1 = -10$ *and* $b_2 = 5$. *The three curves with short dashes (respectively long dashes) represent the median and the 2.5 and 97.5 percentiles of the estimates of* $\tilde{F}(t|0.5)$ *(respectively* $\hat{F}(t|0.5)$*) from the 2500 runs.*

Table 7.1: *Simulation results for exponential survival and censoring distributions.* $\xi_{0.025}$*,* $\xi_{0.5}$ *and* $\xi_{0.975}$ *refer to the 2.5 percentile, the median and the 97.5 percentile of the estimates of* $\tilde{F}(t|x)$ *and* $\hat{F}(t|x)$ *from the 1000 runs.*

Para-meters	$P(\Delta = 0\|x)$	$F(t\|x)$	$\tilde{F}(t\|x)$			$\hat{F}(t\|x)$		
			$\xi_{0.025}$	$\xi_{0.5}$	$\xi_{0.975}$	$\xi_{0.025}$	$\xi_{0.5}$	$\xi_{0.975}$
a_0=5	0.25	0.80	0.6115	0.7748	0.8623	0.5792	0.7762	0.9403
a_1=-30		0.90	0.6912	0.8616	0.9859	0.6966	0.8903	1.0000
b_0=5		0.99	0.8013	0.9919	1.0000	0.8935	0.9803	1.0000
b_1=-10	0.50	0.80	0.5760	0.7507	0.9222	0.5486	0.7625	0.9238
b_2=20		0.90	0.6253	0.8433	1.0000	0.6559	0.8677	1.0000
		0.99	0.6640	0.9593	1.0000	0.8190	0.9620	1.0000
	0.75	0.80	0.4107	0.6718	1.0000	0.4995	0.7626	1.0000
		0.90	0.4191	0.7279	1.0000	0.5879	0.8460	1.0000
		0.99	0.4244	0.7812	1.0000	0.6601	0.9286	1.0000
a_0=1	0.25	0.80	0.6556	0.7674	0.8622	0.6362	0.7679	0.8759
a_1=-5		0.90	0.7418	0.8771	0.9740	0.7509	0.8805	0.9634
b_0=5		0.99	0.8816	0.9933	1.0000	0.9197	0.9861	1.0000
b_1=2	0.50	0.80	0.6120	0.7594	0.8949	0.5991	0.7820	0.9152
b_2=8		0.90	0.6694	0.8561	1.0000	0.7297	0.8877	1.0000
		0.99	0.7238	0.9656	1.0000	0.8816	1.0000	1.0000
	0.75	0.80	0.4186	0.6920	1.0000	0.5247	0.8043	1.0000
		0.90	0.4302	0.7226	1.0000	0.6261	0.8900	1.0000
		0.99	0.4315	0.7321	1.0000	0.7823	1.0000	1.0000
a_0=1	0.25	0.80	0.6586	0.7733	0.8660	0.6639	0.7812	0.8807
a_1=2		0.90	0.7688	0.8799	0.9596	0.7839	0.8864	0.9591
b_0=4		0.99	0.9005	0.9923	1.0000	0.9360	0.9871	1.0000
b_1=5	0.50	0.80	0.5924	0.7608	0.9146	0.6059	0.7812	0.9246
b_2=50		0.90	0.6690	0.8615	1.0000	0.7143	0.8913	1.0000
		0.99	0.7434	0.9897	1.0000	0.8513	0.9736	1.0000
	0.75	0.80	0.3762	0.6733	1.0000	0.4762	0.7521	1.0000
		0.90	0.3969	0.7405	1.0000	0.5735	0.8447	1.0000
		0.99	0.4098	0.8202	1.0000	0.6714	0.9584	1.0000

appropriate to work with here. As expected, the median of $\hat{F}(t|0.5)$ is closer to the true value $F(t|0.5)$ than the median of $\tilde{F}(t|0.5)$ and the variability of $\hat{F}(t|0.5)$ in the right tail is much smaller than that of the Beran estimator.

In Table 7.1, we collected the results of the simulations performed for other combinations of parameter values. For given values of a_0, a_1, b_0, b_1 and b_2, the constant a_2 was determined in such a way that the probability of censoring at $x = 0.5$ achieved the given value (0.25, 0.50 or 0.75). Since the right tails of the distributions are of primary interest, we only compared the results at the 0.8th, 0.9th and 0.99th quantile of the true distribution. The table shows that for all levels of censoring and all choices of the parameters, the estimator $\hat{F}(t|x)$ performs better than the Beran estimator $\tilde{F}(t|x)$. Both the median and the quantiles $\xi_{0.025}$ and $\xi_{0.975}$ of $\hat{F}(t|x)$ (which measure the variability of the estimator) behave almost everywhere better than those of $\tilde{F}(t|x)$. For the heavy censoring cases (75 % censoring), the difference between the two estimators is even very pronounced. The better performance of $\hat{F}(t|x)$ (especially in heavy censoring regions) is due to the fact that $\hat{F}(t|x)$ transfers tail information from light censoring regions to regions with heavy censoring, as explained in Chapter 5. In this way, the right tail of the distribution can be well estimated, in spite of the heavy censoring. The Beran estimator does not have this feature.

Next we consider Weibull survival and censoring distributions with the following parameters

$$(Y|X = x) \sim \text{Weibull}(c_0 + c_1 x + c_2 x^2, d) \quad (7.4)$$

$$(C|X = x) \sim \text{Weibull}(e_0 + e_1 x + e_2 x^2, d), \quad (7.5)$$

i.e. $F(t|x) = 1 - \exp(-(c_0 + c_1 x + c_2 x^2)t^d)$ and similarly for $G(t|x)$. Here, $d > 0$ and c_0, c_1, c_2, e_0, e_1 and e_2 are such that $c_0 + c_1 x + c_2 x^2 > 0$ and $e_0 + e_1 x + e_2 x^2 > 0$ for all $0 \le x \le 1$. From the conditional independence of Y and C for given X, it follows that $(T|X = x) \sim \text{Weibull}(c_0 + e_0 + (c_1 + e_1)x + (c_2 + e_2)x^2, d)$ and that

$$P(\Delta = 0|x) = \frac{e_0 + e_1 x + e_2 x^2}{c_0 + e_0 + (c_1 + e_1)x + (c_2 + e_2)x^2}. \quad (7.6)$$

It is easily shown that if $m(x)$ equals the conditional mean and $\sigma(x)$ the conditional

Table 7.2: *Simulation results for Weibull survival and censoring distributions.* $\xi_{0.025}$*,* $\xi_{0.5}$ *and* $\xi_{0.975}$ *refer to the 2.5 percentile, the median and the 97.5 percentile of the estimates of* $\tilde{F}(t|x)$ *and* $\hat{F}(t|x)$ *from the 1000 runs.*

Para-meters	$P(\Delta = 0\|x)$	$F(t\|x)$	$\tilde{F}(t\|x)$			$\hat{F}(t\|x)$		
			$\xi_{0.025}$	$\xi_{0.5}$	$\xi_{0.975}$	$\xi_{0.025}$	$\xi_{0.5}$	$\xi_{0.975}$
c_0=1	0.25	0.80	0.6570	0.7726	0.8817	0.6943	0.7975	0.8926
c_2=20		0.90	0.7653	0.8782	0.9701	0.8120	0.9038	0.9729
d=3		0.99	0.9067	0.9884	1.0000	0.9487	1.0000	1.0000
e_0=0.1	0.50	0.80	0.2976	0.7140	0.9776	0.4474	0.7672	1.0000
e_1=0		0.90	0.3152	0.8050	1.0000	0.5116	0.8658	1.0000
e_2=25		0.99	0.4088	0.9309	1.0000	0.6472	0.9729	1.0000
	0.75	0.80	0.2703	0.6128	1.0000	0.4341	0.7383	1.0000
		0.90	0.2899	0.6772	1.0000	0.4961	0.8235	1.0000
		0.99	0.3370	0.7958	1.0000	0.6308	0.9182	1.0000
c_0=1	0.25	0.80	0.6725	0.7761	0.8752	0.6799	0.7854	0.8761
c_2=10		0.90	0.7799	0.8766	0.9677	0.8000	0.8928	0.9693
d=2		0.99	0.8982	0.9858	1.0000	0.9217	1.0000	1.0000
e_0=2	0.50	0.80	0.6117	0.7679	0.9263	0.6625	0.8126	0.9453
e_1=-1		0.90	0.6739	0.8630	1.0000	0.7629	0.9076	1.0000
e_2=3		0.99	0.7238	0.9563	1.0000	0.8274	1.0000	1.0000
	0.75	0.80	0.3755	0.6296	1.0000	0.5687	0.8438	1.0000
		0.90	0.3755	0.6380	1.0000	0.6062	0.8774	1.0000
		0.99	0.3755	0.6380	1.0000	0.6154	0.8927	1.0000
c_0=5	0.25	0.80	0.7031	0.7996	0.8868	0.7167	0.8165	0.8920
c_2=20		0.90	0.8050	0.8933	0.9669	0.8235	0.9083	0.9761
d=2		0.99	0.9249	0.9950	1.0000	0.9422	1.0000	1.0000
e_0=1	0.50	0.80	0.6908	0.8335	0.9763	0.7448	0.8624	0.9652
e_1=-1		0.90	0.7769	0.9268	1.0000	0.8533	0.9443	1.0000
e_2=5		0.99	0.8065	1.0000	1.0000	0.9031	1.0000	1.0000
	0.75	0.80	0.3574	0.8083	1.0000	0.6725	0.9079	1.0000
		0.90	0.3574	0.8265	1.0000	0.7692	0.9559	1.0000
		0.99	0.3574	0.8289	1.0000	0.8330	1.0000	1.0000

standard deviation, then

$$P(\varepsilon \leq t|x) = 1 - \exp(-\{t[\Gamma(1+2d^{-1}) - \Gamma^2(1+d^{-1})]^{1/2} + \Gamma(1+d^{-1})\}^d), \quad (7.7)$$

which is independent of x. Hence, model (1.28) is satisfied for these distributions.

Three levels of censoring are considered at $x = 0.5$, namely $P(\Delta = 0|x) = 0.25$, 0.50 and 0.75. For given values of c_0, c_2, d, e_0, e_1 and e_2, the constant c_1 is determined such that these levels are achieved. The results are shown in Table 7.2. Note that in the last (of the three) blocks in Table 7.2, the median of $\hat{F}(t|x)$ is quite a bit larger than the true value $F(t|x)$ for $p = 0.5$ and 0.75. This is because, for these situations, the functions $m(x)$ and $\sigma(x)$ peak at a point very close to $x = 0.5$, where they attain a maximum. Therefore, the estimators $\hat{m}(0.5)$ and $\hat{\sigma}(0.5)$, which are obtained by using observations in a window around $x = 0.5$, seriously underestimate the true $m(0.5)$ and $\sigma(0.5)$ and hence the median of $\hat{F}(t|0.5) = \hat{F}_e(\frac{t-\hat{m}(0.5)}{\hat{\sigma}(0.5)})$ differs somewhat from the true $F(t|0.5)$. We would like to note, however, that this situation is quite exceptional and that in general, as can be seen from the first two blocks in Table 7.2, both the median and the quantiles $\xi_{0.025}$ and $\xi_{0.975}$ of $\hat{F}(t|x)$ behave much better than those of $\tilde{F}(t|x)$.

7.1.2 Choice of the score function

Until now, we worked with the score function $J(s) = I(a \leq s \leq b)/(b-a)$ for $a = 0$ and $b = \min_i \tilde{F}(+\infty|X_i)$. We will now study the performance of $\hat{F}(t|x)$ for other choices of a and b. From the definition of $\hat{m}(x)$ and $\hat{\sigma}(x)$ (see (1.36) and (1.37)) it follows however that the largest possible value for b is $\min_i \tilde{F}(+\infty|X_i)$. Let us therefore consider $b = \bar{b} \wedge \min_i \tilde{F}(+\infty|X_i)$ for different values of $0 \leq \bar{b} \leq 1$. The expression $\min_i \tilde{F}(+\infty|X_i)$ differs among the samples in the simulation and in exceptional cases this minimum could be very small. To avoid that b would be smaller than a in such situations, we put $a = 0$ everywhere.

We first consider the exponential distributions defined in (7.1) and (7.2) for the set of parameters $a_0 = 1, a_1 = 5, a_2 = 100, b_0 = 5, b_1 = 10$ and $b_2 = 75$. At the covariate value of interest, $x = 0.5$, we have 50 % censoring. The estimator $\hat{F}(t|x)$ is calculated for $\bar{b} = 0.4, 0.6, 0.8$ and 1. From the results shown in Table 7.3, it

Table 7.3: *Comparison of the estimator $\hat{F}(t|x)$ for different score functions $J(s) = I(0 \leq s \leq b)/b$, where $b = \bar{b} \wedge \min_i \tilde{F}(+\infty|X_i)$. The parameters of the exponential distribution are $a_0 = 1, a_1 = 5, a_2 = 100, b_0 = 5, b_1 = 10$ and $b_2 = 75$. For the extreme value distribution we selected $u_0 = 1, u_1 = -10, v_0 = 3, v_1 = -12$ and $r = 2$. $\xi_{0.025}$, $\xi_{0.5}$ and $\xi_{0.975}$ refer to the 2.5 percentile, the median and the 97.5 percentile of the estimates of $\tilde{F}(t|x)$ and $\hat{F}(t|x)$ from the 1000 runs.*

$F(t\|x)$	Estimator	Exponential distribution			Extreme value distribution		
		$\xi_{0.025}$	$\xi_{0.5}$	$\xi_{0.975}$	$\xi_{0.025}$	$\xi_{0.5}$	$\xi_{0.975}$
0.80	$\tilde{F}(t\|x)$	0.5919	0.7616	0.9160	0.5527	0.6662	0.7740
0.90		0.6752	0.8613	1.0000	0.6409	0.7464	0.8390
0.99		0.7456	0.9889	1.0000	0.7889	0.8752	0.9438
0.80	$\hat{F}(t\|x)$	0.6070	0.7796	0.9295	0.6247	0.7269	0.8259
0.90	$\bar{b} = 1$	0.7154	0.8911	1.0000	0.7103	0.8158	0.9102
0.99		0.8515	0.9738	1.0000	0.8468	0.9343	1.0000
0.80	$\hat{F}(t\|x)$	0.5867	0.7865	0.9194	0.6230	0.7348	0.8606
0.90	$\bar{b} = 0.8$	0.6948	0.8964	1.0000	0.7157	0.8250	0.9427
0.99		0.8553	0.9650	1.0000	0.8532	0.9443	1.0000
0.80	$\hat{F}(t\|x)$	0.6310	0.7797	0.9229	0.6156	0.7421	0.8909
0.90	$\bar{b} = 0.6$	0.7363	0.8823	1.0000	0.7068	0.8317	0.9644
0.99		0.8546	0.9673	1.0000	0.8124	0.9440	1.0000
0.80	$\hat{F}(t\|x)$	0.6378	0.7868	0.9298	0.5822	0.7442	0.9114
0.90	$\bar{b} = 0.4$	0.7359	0.8844	1.0000	0.6575	0.8342	0.9732
0.99		0.8500	0.9705	1.0000	0.7747	0.9433	1.0000

follows that there are only small differences among the considered choices of $\bar{b}$. In particular, for any of the values of $\bar{b}$, the Beran estimator $\tilde{F}(t|x)$ behaves worse than $\hat{F}(t|x)$.

Next, we compare $\hat{F}(t|x)$ for different score functions when Y and C follow an extreme value distribution :

$$(Y|X=x) \sim F(t|x) = 1 - \exp\left(-\exp\left(\frac{t - u_0 - u_1 x}{r}\right)\right) \tag{7.8}$$

$$(C|X=x) \sim G(t|x) = 1 - \exp\left(-\exp\left(\frac{t - v_0 - v_1 x}{r}\right)\right) \tag{7.9}$$

for some $u_0, u_1, v_0, v_1 \in I\!R$ and $r > 0$. Note that Y may be any monotone transformation of the survival time. In particular, we can choose a logarithmic transformation for Y. Since the logarithm of a random variable following a Weibull distribution, has an extreme value distribution, this motivates why we consider this class of distributions here. Since $E(Y|x) = -rC + u_0 + u_1 x$ ($C = -0.5772$ being the Euler constant) and $\mathrm{Var}(Y|x) = r^2\pi^2/6$, it easily follows that if $m(x) = E(Y|x)$ and $\sigma^2(x) = \mathrm{Var}(Y|x)$, then $P(\varepsilon \leq t|x) = 1 - \exp(-\exp(t\pi/\sqrt{6} - C))$. Since this is independent of x, model (1.28) holds. Again we consider the values 0.4, 0.6, 0.8 and 1 for $\bar{b}$ and compare the results. We selected $u_0 = 1, u_1 = -10, v_0 = 3, v_1 = -12$ and $r = 2$. This leads to 38 % censoring at $x = 0.5$. For this situation, we find (see Table 7.3) that the larger the value of $\bar{b}$, the better the result (when $\bar{b}$ becomes larger, the median behaves slightly worse, but the variability decreases a lot). From other simulations we did, we can recommend that for any distributions F and G, the support of J should be chosen as large as possible, i.e. $a = 0$ and $b = \min_i \tilde{F}(+\infty|X_i)$.

7.1.3 Sensitivity to model assumption

Let us now examine the performance of the estimator $\hat{F}(t|x)$ when model (1.28) is not satisfied. We do not aim, however, at performing a detailed sensitivity analysis of the estimator, but will only consider a few situations. In particular, we wish to know whether the estimator $\hat{F}(t|x)$ performs still better than $\tilde{F}(t|x)$ under departures from model (1.28). We consider the case where the conditional distribution of Y, respectively C, is given by the Weibull distributions displayed in (7.4), respectively (7.5), with $d = d_0 + d_1 x$. As was shown in (7.7), the conditional distribution of

Table 7.4: *Simulation results for Weibull survival and censoring distributions, which do not satisfy model (1.28).* $\xi_{0.025}$, $\xi_{0.5}$ *and* $\xi_{0.975}$ *refer to the 2.5 percentile, the median and the 97.5 percentile of the estimates of* $\tilde{F}(t|x)$ *and* $\hat{F}(t|x)$ *from the 1000 runs.*

Para-meters	$P(\Delta = 0\|x)$	$F(t\|x)$	$\tilde{F}(t\|x)$			$\hat{F}(t\|x)$		
			$\xi_{0.025}$	$\xi_{0.5}$	$\xi_{0.975}$	$\xi_{0.025}$	$\xi_{0.5}$	$\xi_{0.975}$
c_0=10	0.25	0.80	0.7233	0.8137	0.9015	0.7347	0.8223	0.9021
c_2=50		0.90	0.8256	0.9040	0.9753	0.8373	0.9131	0.9728
d_0=1		0.99	0.9234	0.9944	1.0000	0.9507	1.0000	1.0000
d_1=1	0.50	0.80	0.7178	0.8372	0.9525	0.7333	0.8580	0.9601
e_0=1		0.90	0.7634	0.9065	1.0000	0.8322	0.9365	1.0000
e_1=-5		0.99	0.8198	1.0000	1.0000	0.9297	1.0000	1.0000
e_2=20	0.75	0.80	0.5154	0.8198	1.0000	0.6672	0.8928	1.0000
		0.90	0.5469	0.8695	1.0000	0.7340	0.9509	1.0000
		0.99	0.5561	0.9063	1.0000	0.8741	1.0000	1.0000
c_0=5	0.25	0.80	0.6765	0.7773	0.8656	0.6735	0.7744	0.8597
c_2=50		0.90	0.8005	0.8817	0.9508	0.7956	0.8787	0.9471
d_0=1		0.99	0.9316	0.9850	1.0000	0.9330	0.9825	1.0000
d_1=2	0.50	0.80	0.6387	0.7669	0.8833	0.6589	0.7734	0.8877
e_0=1		0.90	0.7273	0.8673	0.9779	0.7487	0.8640	0.9615
e_1=-20		0.99	0.8389	0.9703	1.0000	0.8771	0.9696	1.0000
e_2=100	0.75	0.80	0.4022	0.6828	0.9986	0.5208	0.7156	0.9190
		0.90	0.4457	0.7903	1.0000	0.6079	0.8032	0.9703
		0.99	0.6506	0.9181	1.0000	0.7541	0.9365	1.0000
c_0=20	0.25	0.80	0.7217	0.8362	0.9342	0.7631	0.8489	0.9262
c_2=100		0.90	0.8437	0.9285	1.0000	0.8586	0.9304	0.9870
d_0=1		0.99	0.9261	1.0000	1.0000	0.9565	1.0000	1.0000
d_1=3	0.50	0.80	0.7026	0.8387	0.9890	0.7387	0.8686	0.9654
e_0=5		0.90	0.7323	0.8987	1.0000	0.8337	0.9385	1.0000
e_1=-30		0.99	0.8092	1.0000	1.0000	0.9247	1.0000	1.0000
e_2=50	0.75	0.80	0.3941	0.6937	1.0000	0.5832	0.8435	1.0000
		0.90	0.4046	0.7659	1.0000	0.6563	0.9050	1.0000
		0.99	0.4498	0.8246	1.0000	0.7468	0.9738	1.0000

ε given $X = x$ depends on d and hence, in this situation, also on x. This shows that model (1.28) is not satisfied here. Table 7.4 contains the output for three different choices of the parameters $c_0, c_2, d_0, d_1, e_0, e_1$ and e_2. The remaining parameter c_1 is determined such that the probability of censoring at $x = 0.5$ (given by formula (7.6)) equals 0.25, 0.50 or 0.75. The table shows that the median of $\tilde{F}(t|x)$ is often closer to $F(t|x)$ than the median of $\hat{F}(t|x)$, but its variability is in almost all cases much larger. Other simulations for other choices of the parameters gave similar results.

7.2 Data analysis

To illustrate our method on a real data set, the data from the Stanford heart transplant program will be used. This program started in October 1967 and by February 1980, 184 patients had received a heart transplantation. For reasons of comparison with other work (Miller and Halpern (1982), Doksum and Yandell (1982)), we restrict our attention to the 157 out of 184 individuals who had complete tissue typing. Patients alive beyond February 1980 were considered censored (55 in total). We focus on two variables here, with the (base 10 logarithm of the) survival time (in days) as the response and the age at transplantation as covariate. The conditional distribution of the response given the covariate, as well as the median survival time given the covariate, will be examined.

We selected the bandwidth $h_n = 7$ to calculate the Beran estimator $\tilde{F}(t|x)$ (for simplicity we selected a global bandwidth). Other reasonable values for the bandwidth were found not to give significantly different results. Following the recommendation in Section 7.1.2 , we took $a = 0$ and $b = \min_i \tilde{F}(+\infty|X_i) = 0.587$ as left and right endpoints of the support of the score function $J(s) = I(a \leq s \leq b)/(b - a)$. In order to overcome the boundary problems near the endpoints of the support of the covariate space, a boundary corrected kernel function was chosen, as developed by Müller and Wang (1994). In particular, we selected the boundary corrected

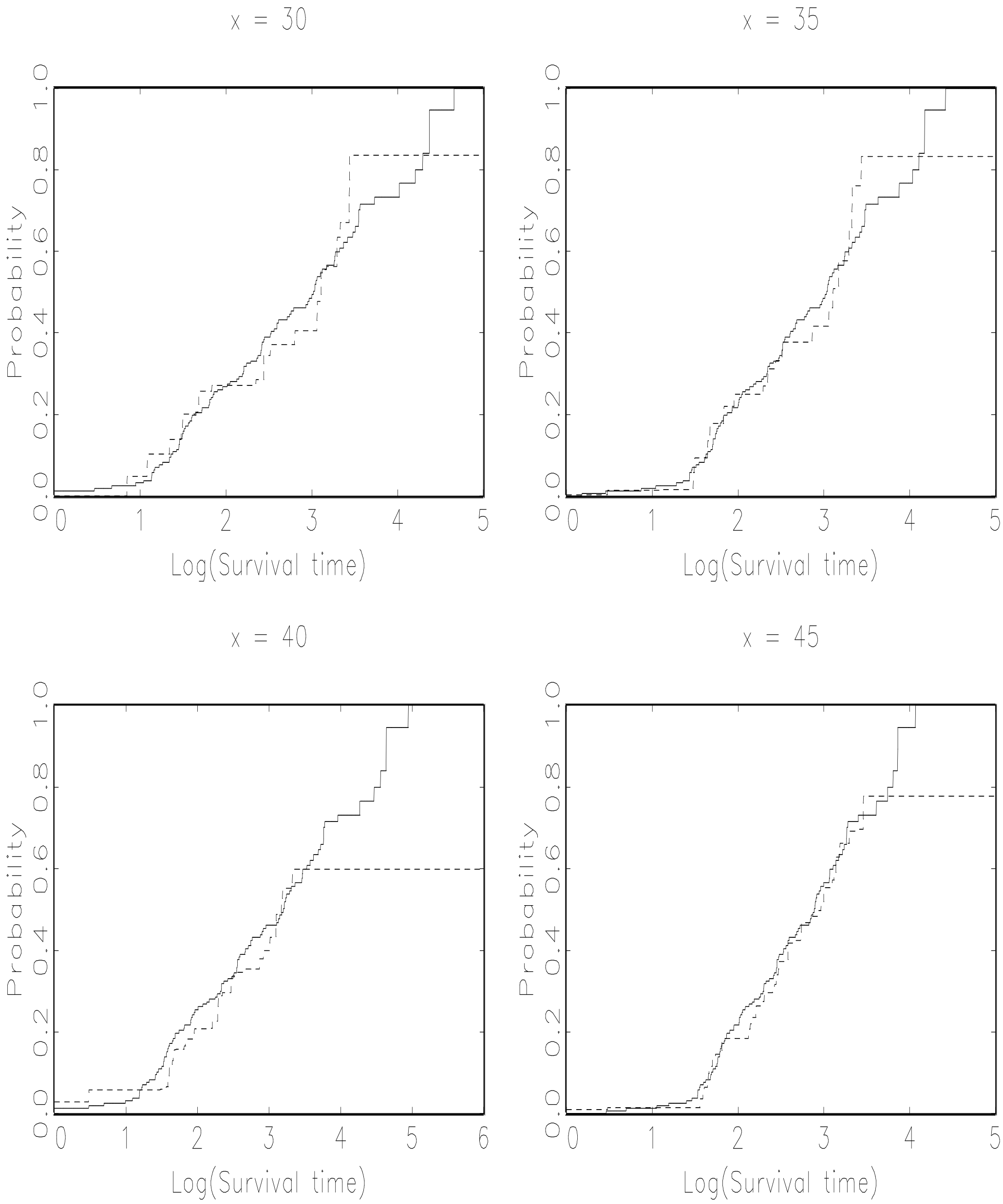

Figure 7.3: *Conditional distribution of the logarithm of the survival time at age 30, 35, 40 and 45 for the Stanford heart transplant data. The solid curve represents $\hat{F}(t|x)$, the dashed curve is $\tilde{F}(t|x)$.*

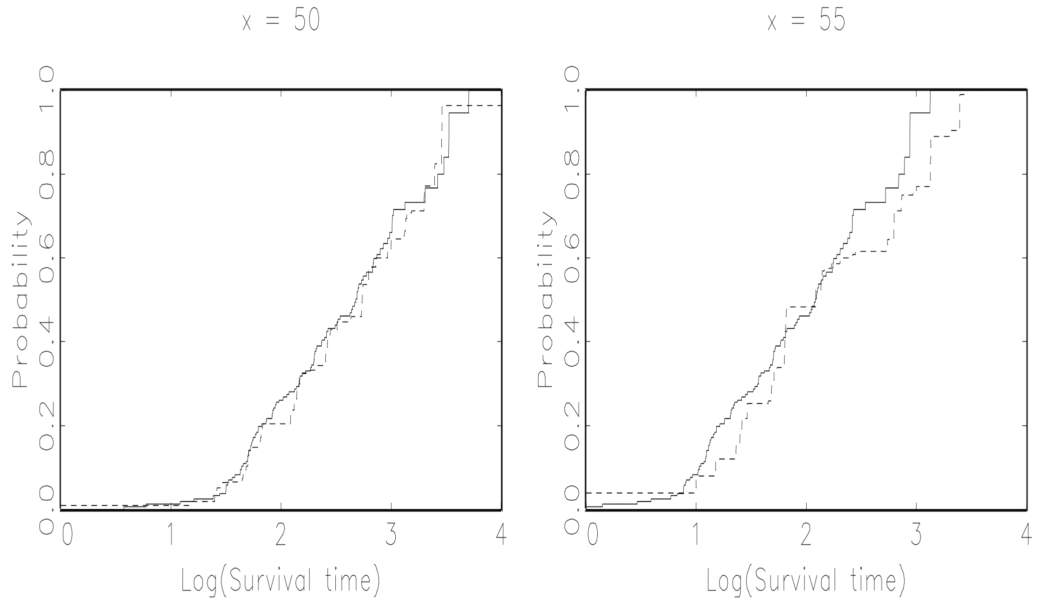

Figure 7.4: *Conditional distribution of the logarithm of the survival time at age 50 and 55 for the Stanford heart transplant data. The solid curve represents* $\hat{F}(t|x)$*, the dashed curve is* $\tilde{F}(t|x)$*.*

biquadratic kernel

$$K_x(z) = \frac{15}{(1+q(x))^5}(z+1)^2(q(x)-z) \times \left\{2z\left(5\frac{1-q(x)}{1+q(x)}-1\right)+3q(x)-1+5\frac{(1-q(x))^2}{1+q(x)}\right\} I(|z|\le 1),$$

where

$$q(x) = \begin{cases} \dfrac{x-m}{h_n} & m \le x \le m+h_n \\ 1 & m+h_n \le x \le M-h_n \\ \dfrac{M-x}{h_n} & M-h_n \le x \le M \end{cases}$$

and m and M are, respectively, the left and right endpoints of the support of the covariate space. We chose $m = 9$ and $M = 67$. (In case that $q(x) \equiv 1$ (i.e. when no boundary correction is used), this kernel reduces to the usual biquadratic kernel $K(z) = (15/16)(1-z^2)^2 I(|z| \le 1)$).

Figures 7.3 and 7.4 show the graph of both the Beran estimator $\tilde{F}(\cdot|x)$ and the estimator $\hat{F}(\cdot|x)$ for $x = 30, 35, 40, 45, 50$ and 55. The two estimators are in good agreement (although the Beran estimator is much rougher) up to the point in time where the Beran estimator $\tilde{F}(t|x)$ becomes flat. Since the latter estimator is consistent (for t smaller than the largest uncensored observation) (see Proposition 5.5) and the estimator $\hat{F}(\cdot|x)$ is consistent if model (1.28) holds (see Theorem 5.4), this is an indication that model (1.28) is indeed valid here. Note that the jumps of $\tilde{F}(\cdot|x)$ are much larger than those of $\hat{F}(\cdot|x)$. This is because $\tilde{F}(\cdot|x)$ is using only part of the data that is utilized by $\hat{F}(\cdot|x)$.

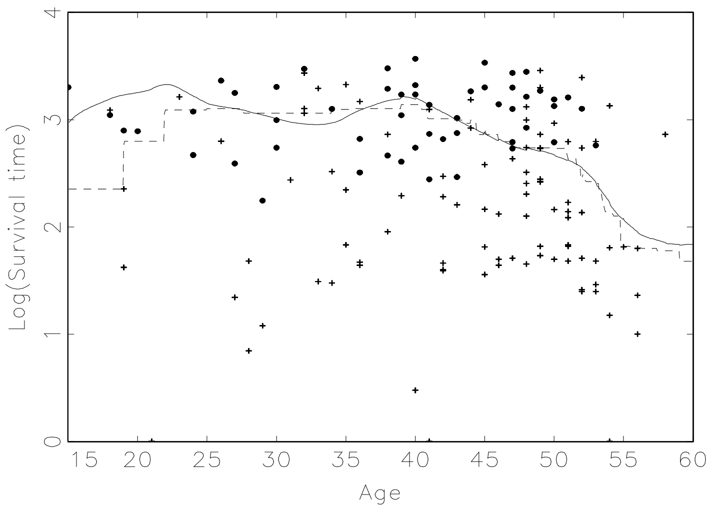

Figure 7.5: *Median regression curve for the Stanford heart transplant data by using* $\hat{F}^{-1}(0.5|x)$ *(solid curve) and* $\tilde{F}^{-1}(0.5|x)$ *(dashed curve). Uncensored data are indicated by +, censored data by* $\bullet$.

The estimation of the conditional mean of the present data set has been studied in many papers (Miller (1976), Doksum and Yandell (1982), Miller and Halpern (1982), Fan and Gijbels (1994), Akritas (1996) among others), while Doksum and Yandell (1982) and Miller and Halpern (1982) studied in addition the estimation of the conditional median. Since distributions of lifetimes tend to be skewed (to the right), we prefer to work with the median, defined by

$$F^{-1}(0.5|x) = \inf\{t; F(t|x) \geq 0.5\},$$

as measure of location. Figure 7.5 contains the graph of the Beran median curve $\tilde{F}^{-1}(0.5|\cdot)$ and the Van Keilegom-Akritas median $\hat{F}^{-1}(0.5|\cdot)$. These curves are obtained by connecting the medians $\tilde{F}^{-1}(0.5|\cdot)$ and $\hat{F}^{-1}(0.5|\cdot)$ for 720 equally spaced points in the interval [15,60]. The latter curve is smooth, while the Beran median $\tilde{F}^{-1}(0.5|x)$ makes jumps and is somewhat rougher. The two estimators are however in good accordance, except for ages less than 25, probably due to the sparseness of the data. Note that the median survival time starts decreasing quite rapidly from the age of about 40. The Beran running median has also been considered by Doksum and Yandell (1982) using nearest neighbor kernels, while Miller and Halpern (1982) and again Doksum and Yandell (1982) studied the median under the Cox proportional hazards model. In comparison with these two estimators, $\hat{F}^{-1}(0.5|x)$ shows a more detailed estimation of the median.

References

Aerts, M. and Geertsema, J. (1990). Bounded length confidence intervals in nonparametric regression. *Sequential Analysis*, **9**, 171-192.

Aerts, M., Janssen, P. and Veraverbeke, N. (1994a). Asymptotic theory for regression quantile estimators in the heteroscedastic regression model. In : *Asymptotic Statistics* (eds. P. Mandl and M. Hušková). Physica-Verlag, Heidelberg, 151-161.

Aerts, M., Janssen, P. and Veraverbeke, N. (1994b). Bootstrapping regression quantiles. *J. Nonparametric Statistics*, **4**, 1-20.

Akritas, M. G. (1986). Bootstrapping the Kaplan-Meier estimator. *J. Amer. Statist. Assoc.*, **81**, 1032-1038.

Akritas, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann. Statist.*, **22**, 1299-1327.

Akritas, M. G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*, **52**, 1342-1362.

Akritas, M. G. and LaValley, M. (1997). Statistical analysis with incomplete data : a selective review. In : *Handbook of statistics* (eds. G. S. Maddala and C.R. Rao). Elsevier Science, **15**, 551-632.

Akritas, M. G., Murphy, S. and LaValley, M. (1995). The Theil-Sen estimator with doubly censored data and applications to astronomy. *J. Amer. Statist. Assoc.*, **90**, 170-177.

Andersen, P. K., Borgan, O., Gill, R. and Keiding, N. (1993). *Statistical models based on counting processes.* Springer-Verlag, New York.

Araujo, A. and Giné, E. (1980). *The central limit theorem for real and Banach valued random variables.* Wiley, New York.

Arnold, S. F. (1981). *The theory of linear models and multivariate analysis.* Wiley, New York.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196 - 1217.

Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.*, **2**, 437-453.

Buckley, J. and James, I. R. (1979). Linear regression with censored data. *Biometrika*, **66**, 429-436.

Cao Abad, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Ann. Statist.*, **19**, 2226-2231.

Cheng, K.-F. (1984). On almost sure representation for quantiles of the product-limit estimator with applications. *Sankhyā Ser. A*, **46**, 426-443.

Cheng, K.-F. and Wu, J. W. (1994). Adjusted least squares estimates for the scaled regression coefficients with censored data. *J. Amer. Statist. Assoc.*, **89**, 1483-1491.

Cheng, P. E. (1989). Nonparametric estimation of survival curve under dependent censoring. *J. Statist. Planning Inf.*, **23**, 181-192.

Chu, C.-K. and Marron, J. S. (1991). Choosing a kernel regression estimator. *Statist. Science*, **6**, 404-436.

Cox, D. R. (1972). Regression models and life-tables. *J. Royal Statist. Soc. Ser. B*, **34**, 187-202.

Cristóbal Cristóbal, J. A., Faraldo Roca, P. and González Manteiga, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Ann. Statist.*, **15**, 603-609.

Dabrowska, D. M. (1987). Non-parametric regression with censored survival time data. *Scand. J. Statist.*, **14**, 181-197.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.*, **16**, 1475-1489.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.*, **17**, 1157-1167.

Dabrowska, D. M. (1992a). Variable bandwidth conditional Kaplan-Meier estimate. *Scand. J. Statist.*, **19**, 351-361.

Dabrowska, D. M. (1992b). Nonparametric quantile regression with censored data. *Sankhyā Ser. A*, **54**, 252-259.

Dabrowska, D. M. and Doksum, K. (1988). Partial likelihood in transformation models with censored data. *Scand. J. Statist.*, **15**, 1-24.

Devroye, L. P. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canad. J. Statist.*, **6**, 179-191.

de Uña Álvarez, J. and González Manteiga, W. (1997a). Strong consistency under proportional censorship when covariables are present. *Statist. Probab. Letters* (in press).

de Uña Álvarez, J. and González Manteiga, W. (1997b). Distributional convergence under proportional censorship when covariables are present. *Statist. Probab. Letters* (in press).

Doksum, K. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.*, **15**, 325-345.

Doksum, K. and Yandell, B. S. (1982). Properties of regression estimates based on censored survival data. In : *Festschrift for E. L. Lehmann* (eds. P. J. Bickel, K. A. Doksum and J. L. Hodges). Wadsworth, 140-156.

Doss, H. and Gill, R. (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *J. Amer. Statist. Assoc.*, **87**, 869-877.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.*, **1**, 279-290.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, **27**, 642-669.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.*, **76**, 312-319.

Fan, J. and Gijbels, I. (1994). Censored regression : local linear approximations and their applications. *J. Amer. Statist. Assoc.*, **89**, 560-570.

Földes, A. and Rejtő, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete*, **56**, 75-86.

Földes, A., Rejtő, L. and Winter, B. B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data, II : Estimation of density and failure rate. *Periodica Math. Hungarica*, **12**, 15-29.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.*, **22**, 732-746.

Fygenson, M. and Zhou, M. (1994). On using stratification in the analysis of linear regression models with right censoring. *Ann. Statist.*, **22**, 747-762.

Gasser, T. and Engel, J. (1990). The choice of weights in kernel regression estimation. *Biometrika*, **77**, 377-381.

Gasser, T. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171-185.

Gentleman, R. and Crowley, J. (1991). Graphical methods for censored data. *J. Amer. Statist. Assoc.*, **86**, 678-683.

Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.*, **11**, 49-58.

González Manteiga, W. and Cadarso Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *J. Nonparametric Statistics*, **4**, 65-78.

Hall, P., Marron, J. S. and Titterington, D. M. (1995). On partial local smoothing rules for curve estimation. *Biometrika*, **82**, 575-587.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika*, **67**, 113-143.

Härdle, W. and Gasser, T. (1985). On robust kernel estimation of derivatives of regression functions. *Scand. J. Statist.*, **12**, 233-240.

Härdle, W., Huet, S. and Jolivet, E. (1995). Better bootstrap confidence intervals for regression curve estimation. *Statistics*, **26**, 287-306.

Härdle, W., Janssen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.*, **16**, 1428-1449.

Härdle, W. and Mammen, E. (1991). Bootstrap methods in nonparametric regression. In : *Nonparametric functional estimation and related topics* (ed. G. Roussas). Kluwer, 111-123.

Härdle, W. and Marron, J. S. (1990). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.*, **19**, 778-796.

Huber, P. J. (1981). *Robust statistics.* Wiley, New York.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457-481.

Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.*, **9**, 1276-1288.

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

Lai, T. and Ying, Z. (1991a). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.*, **19**, 1370-1402.

Lai, T. and Ying, Z. (1991b). Rank regression methods for left truncated and right censored data. *Ann. Statist.*, **19**, 531-554.

Lai, T. and Ying, Z. (1994). A missing information principle and M-estimators in regression analysis with censored or truncated data. *Ann. Statist.*, **22**, 1222-1255.

Lam, K. F. and Kuk, A. Y. C. (1997). A marginal likelihood approach to estimation in frailty models. *J. Amer. Statist. Assoc.*, **92**, 985-990.

LeBlanc, M. and Crowley, J. (1995). Semiparametric regression functionals. *J. Amer. Statist. Assoc.*, **90**, 95-105.

Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika*, **74**, 301-309.

Li, G. and Datta, S. (1997). Bootstrapped confidence bands for conditional survival functions from right censored data (submitted).

Li, G. and Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.*, **23**, 787-823.

Lin, D. Y. and Wei, L. J. (1992a). Linear regression analysis based on Buckley-James estimating equations. *Biometrics*, **48**, 679-681.

Lin, D. Y. and Wei, L. J. (1992b). Linear regression analysis of multivariate failure time observations. *J. Amer. Statist. Assoc.*, **87**, 1091-1097.

Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap : some asymptotic representations. *Probab. Th. Rel. Fields*, **71**, 455-465.

Loynes, R. M. (1980). The empirical sample distribution function of residuals from generalized regression. *Ann. Statist.*, **8**, 285-298.

Mack, Y. P. and Müller, H.-G. (1989). Derivative estimation in nonparametric regression with random predictor variable. *Sankhyā Ser. A*, **51**, 59-72.

McKeague, I. W. and Utikal, K. J. (1990). Inference for a nonlinear counting process regression model. *Ann. Statist.*, **18**, 1172-1187.

Miller, R. (1976). Least squares regression with censored data. *Biometrika*, **63**, 449-464.

Miller, R. and Halpern, J. (1982). Regression with censored data. *Biometrika*, **69**, 521-531.

Müller, H. G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.*, **15**, 610-625.

Müller, H. G. and Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, **50**, 61-76.

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sørensen, T. I. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.*, **19**, 25-43.

Nolan, D. and Pollard, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.*, **16**, 1291-1298.

Park, B. U. and Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.*, **85**, 66-72.

Pollard, D. (1990). *Empirical processes : theory and applications.* IMS, Hayward, California.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.*, **18**, 303-328.

Romano, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.*, **16**, 629-647.

Rychlik, T. (1990). Unbiased nonparametric estimation of the derivative of the mean. *Statist. Probab. Letters*, **10**, 329-333.

Schäfer, H. (1986). Local convergence of empirical measures in the random censorship situation with application to density and rate estimators. *Ann. Statist.*, **14**, 1240-1245.

Schuster, E. and Yakowitz, S. (1979). Contributions to the theory of nonparametric regression, with application to system identification. *Ann. Statist.*, **7**, 139-149.

Sciortino, S., Vaiana, G. S., Harnden, Jr., F. R., Ramella, M., Morossi, C., Rosner, R. and Schmitt, J. H. H. M. (1990). Relationship between optical and X-ray properties of O-type stars surveyed with the Einstein observatory. *The Astrophysical Journal*, **361**, 621-643.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B*, **53**, 683-690.

Shorack, G. R. (1982). Bootstrapping robust regression. *Commun. Statist., - Theory Meth.*, **11**, 961-972.

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

Singh, R. S. (1975). On the Glivenko-Cantelli theorem for weighted empiricals based on independent random variables. *Ann. Probab.*, **3**, 371-374.

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595-645.

Stute, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.*, **10**, 86-107.

Stute, W. (1986). Conditional empirical processes. *Ann. Statist.*, **14**, 638-647.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multiv. Analysis*, **45**, 89-103.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.*, **23**, 461-471.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.*, **18**, 354-372.

Tsirel'son, V. S. (1975). The density of the distribution of the maximum of a Gaussian process. *Theory Probab. Appl.*, **20**, 847-856.

van der Laan, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.*, **24**, 596-627.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer-Verlag, New York.

Van Keilegom, I. and Akritas, M. G. (1997a). Estimation of the conditional distribution in censored regression models via transfer of tail information (conditionally accepted by *Ann. Statist.*).

Van Keilegom, I. and Akritas, M. G. (1997b). The least squares method in heteroscedastic censored regression models (conditionally accepted by *Ann. Statist.*).

Van Keilegom, I., Akritas, M. G. and Veraverbeke, N. (1998). Estimation of the conditional distribution in regression with censored data : a comparative study (submitted).

Van Keilegom, I. and Veraverbeke, N. (1996). Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles. *Commun. Statist., - Theory Meth.*, **25**, 2251-2265.

Van Keilegom, I. and Veraverbeke, N. (1997a). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. *Commun. Statist., - Theory Meth.*, **26**, 853-869.

Van Keilegom, I. and Veraverbeke, N. (1997b). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, **49**, 467-491.

Van Keilegom, I. and Veraverbeke, N. (1998). Bootstrapping quantiles in a fixed design regression model with censored data. *J. Statist. Planning Inf.*, **69**, 115-131.

Veraverbeke, N. and Cadarso Suárez, C. (1997). Regression with censored data under the Koziol-Green model (submitted).

Wei, L. J., Ying, Z. and Lin, D. (1990). Linear regression analysis of censored survival data based on linear rank tests. *Biometrika*, **77**, 845-851.

Yang, S. (1997). Extended weighted log-rank estimating functions in censored regression. *J. Amer. Statist. Assoc.*, **92**, 977-984.

Ying, Z., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.*, **90**, 178-184.

Zhou, M. (1992a). Asymptotic normality of the "synthetic data" regression estimator for censored survival data. *Ann. Statist.*, **20**, 1002-1021.

Zhou, M. (1992b). M-estimation in censored linear models. *Biometrika*, **79**, 837-841.

Samenvatting

In dit proefschrift worden regressiemodellen bestudeerd waarbij de respons data onderworpen zijn aan (rechtse) censurering. We behandelen twee verschillende regressiemodellen. Eerst onderstellen we een volledig niet-parametrisch model, waarin de relatie tussen de respons en de covariaat niet gespecifieerd is (dit model zal ondersteld worden in Hoofdstukken 2 tot 4). In het tweede model, dat we onderstellen in Hoofdstukken 5 tot 7, wordt het verband tussen de respons Y en de covariaat X gegeven door

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1}$$

voor zekere functies m en σ en voor ε onafhankelijk van X. We bespreken deze twee modellen uitvoerig in Hoofdstuk 1.

In Hoofdstuk 2 bestuderen we een kernschatter voor de conditionele verdeling $F(t|x) = P(Y \leq t|X = x)$ van de respons Y, gegeven een waarde x van de covariaat X onder een algemeen regressiemodel. Deze schatter wordt de Beran schatter genoemd en wordt genoteerd door $F_h(t|x)$. We stellen eerst een asymptotische representatie voor de Beran schatter op, die de schatter ontbindt in een som van onafhankelijke termen en een restterm van lagere orde. Dankzij deze representatie kunnen we asymptotische eigenschappen van de Beran schatter bestuderen, zoals de zwakke convergentie naar een Gauss proces.

De uitdrukkingen voor de vertekening en de variantie van dit Gauss proces bevatten ongekende, ingewikkelde functies. Voor het construeren van bv. betrouwbaarheidsbanden voor $F(\cdot|x)$, zouden we kunnen gebruik maken van de verdeling van dit limiet proces, maar dit vereist dan wel dat de ongekende functies worden geschat. Dit probleem kan omzeild worden door gebruik te maken van een zogenaamde “bootstrap” procedure, die de verdeling van de Beran schatter op een alternatieve manier benadert. Het bestuderen van de bootstrap schatter $F^*_{hg}(t|x)$ is het onderwerp van Hoofdstuk 3. Aan de hand van een asymptotische representatie voor $F^*_{hg}(t|x)$ verkrijgen we de zwakke convergentie van het proces $F^*_{hg}(\cdot|x)$. Uit dit resultaat leiden we een alternatieve benadering voor de verdeling van de Beran schatter af en verkrijgen we ook bootstrap betrouwbaarheidsbanden voor $F(\cdot|x)$. In simulaties worden de normale en de bootstrap benadering van de verdeling van de Beran schatter met elkaar vergeleken.

Naast het schatten van de verdelingsfunctie, is het ook van belang om de conditionele kwantielfunctie van Y gegeven X te kennen. Deze functie is gedefinieerd als

$$F^{-1}(p|x) = \inf\{t; F(t|x) \geq p\}$$

$(0 < p < 1)$ en wordt geschat door de verdeling $F(t|x)$ te vervangen door de Beran schatter $F_h(t|x)$:

$$F_h^{-1}(p|x) = \inf\{t; F_h(t|x) \geq p\}.$$

In Hoofdstuk 4 voeren we een asymptotische studie uit van deze schatter. Naast nog andere resultaten verkrijgen we ook hier twee benaderingen voor de verdeling van de schatter $F_h^{-1}(p|x)$ (een normale en een bootstrap approximatie) die we met elkaar vergelijken in een simulatiestudie.

In Hoofdstuk 5 (en ook in Hoofdstukken 6 en 7) onderstellen we dat het heteroscedastische model (1) voldaan is. Dit houdt in dat we de verdeling $F(t|x)$ kunnen schrijven als

$$F(t|x) = F_e\left(\frac{t - m(x)}{\sigma(x)}\right),$$

waar F_e de verdelingsfunctie van de variabele ε voorstelt. Vervangen we in deze uitdrukking de ongekende functies F_e, m en σ door geschikte schatters, dan verkrijgen we een alternatief voor de Beran schatter, dat we noteren door $\hat{F}(t|x)$. Een belangrijk voordeel van $\hat{F}(t|x)$ is dat de staart van deze schatter zich beter gedraagt dan die van $F_h(t|x)$, die inconsistent is in vele situaties. Bovendien kunnen we voor deze schatter alle gegeven data gebruiken, terwijl dit niet het geval is voor de Beran schatter. We stellen een asymptotische representatie op voor $\hat{F}(t|x)$ en verkrijgen de zwakke convergentie naar een Gauss proces. Als toepassing vergelijken we de asymptotische variantie van deze schatter met die van de Beran schatter, in het geval dat alle data niet gecensureerd zijn.

Een ander interessant probleem is het schatten van de gezamenlijke verdeling van de respons Y en de covariaat X, gegeven door :

$$F(x,t) = P(X \leq x, Y \leq t) = \int_{-\infty}^{x} F(t|u)\, dF_X(u).$$

Een voor de hand liggende manier om $F(x,t)$ te schatten bestaat in het vervangen van $F(t|u)$ door de hierboven gedefinieerde schatter $\hat{F}(t|u)$ (voor alle u) en het vervangen van de ongekende verdeling $F_X(u)$ van de covariaat X door de empirische verdelingsfunctie. Voor de schatter die we op die manier verkrijgen (en die we noteren door $\hat{F}(x,t)$), bewijzen we in Hoofdstuk 6 een representatie en de zwakke convergentie. Onderstel nu dat we de functie $m(x)$ in (1) kunnen schrijven als

$$m(x) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p$$

voor zekere ongekende parameters $\beta_0, \ldots, \beta_p$ en onderstel dat we deze parameters wensen te schatten. We definiëren de schatter $(\hat{\beta}_0, \ldots, \hat{\beta}_p)$ als de waarde van de vector $(\beta_0, \ldots, \beta_p)$ waarvoor de uitdrukking

$$\int (t - \beta_0 - \beta_1 x - \ldots - \beta_p x^p)^2 \, d\hat{F}(x,t)$$

minimaal is. Een expliciete uitdrukking voor deze schatter wordt gegeven. In Hoofdstuk 6 stellen we een asymptotische representatie op voor elk van de $\hat{\beta}_k$ $(k = 0, \ldots, p)$ waaruit de asymptotische normaliteit van de vector $(\hat{\beta}_0, \ldots, \hat{\beta}_p)$ volgt.

Ten slotte voeren we in Hoofdstuk 7 een simulatiestudie uit omtrent het gedrag voor eindige steekproeven van de Beran schatter $F_h(t|x)$ (geconstrueerd onder een algemeen regressie model) en de schatter $\hat{F}(t|x)$ (geconstrueerd onder het heteroscedastische model (1)). We vergelijken eerst het gedrag van beide schatters met elkaar. Vervolgens onderzoeken we voor welke keuze van de functie $m(x)$ de schatter $\hat{F}(t|x)$ zich het best gedraagt en we gaan ook na hoe gevoelig de schatter $\hat{F}(t|x)$ is aan het voldaan zijn van model (1). Als toepassing op de ontwikkelde methoden, analyzeren we een data set (omtrent de overlevingstijd van patiënten die een hart transplantatie hebben ondergaan) met behulp van de schatters $F_h(t|x)$ en $\hat{F}(t|x)$.