

Bootstrapping the conditional survival function estimator in the partial
Koziol-Green model

Non Peer-reviewed author version

BRAEKERS, Roel & VERAVERBEKE, Noel (2005) Bootstrapping the conditional survival function estimator in the partial Koziol-Green model. In: JOURNAL OF NONPARAMETRIC STATISTICS, 17(3). p. 299-318.

DOI: 10.1080/10485250500038330

Handle: <http://hdl.handle.net/1942/2071>

Bootstrapping the conditional survival function estimator in the partial Koziol-Green model

Roel Braekers and Noël Veraverbeke

Limburgs Universitair Centrum, B-3590 Diepenbeek, Belgium

Keywords: Bootstrap; Fixed design; Informative censoring; Nonparametric regression; Proportional hazards; Right censoring; Weak convergence

Running head: bootstrap partial Koziol-Green model

AMS Subject classifications: Primary 62G09; secondary 62G05, 62G20

ABSTRACT

In the partial Koziol-Green regression model, the lifetime variable may be censored by two types of censoring variables. One is called informative because it satisfies the Koziol-Green assumption on proportionality of hazards, and the other one is general. [3] proposed a nonparametric estimator for the conditional lifetime distribution and obtained a Gaussian approximation for the corresponding process. In the present paper we propose an appropriate resampling scheme and show that this leads to a valid bootstrap approximation for the process.

1 Introduction

Consider a regression model with independent and non-negative responses Y_1, \dots, Y_n at fixed design points $0 \leq x_1 \leq \dots \leq x_n \leq 1$. These responses are subject to random right censoring in the following way: each Y_i can be censored by the minimum of two independent and non-negative variables C_i and D_i . The C_i are informative censoring times, satisfying the proportional hazards condition, while the D_i are arbitrary non informative censoring times. The observed random variables at design point x_i are (Z_i, δ_i) where $Z_i = Y_i \wedge C_i \wedge D_i$ and $\delta_i = 1$ if $Y_i \leq C_i \wedge D_i$, $\delta_i = 0$ if $C_i \leq Y_i \wedge D_i$ and $\delta_i = -1$ if $D_i \leq Y_i \wedge C_i$. For a given design value $x \in]0, 1[$, we write F_x, G_{1x}, G_{2x}, H_x for the distribution function of respectively, the response Y_x at x , the informative censoring variable C_x , the non informative censoring variable D_x and $Z_x = Y_x \wedge C_x \wedge D_x$. Also

we will write δ_x for the indicator variable that takes the values 1, 0, -1 according as $Z_x = Y_x, C_x, D_x$. (Note that for the design values x_i we write $Y_i, C_i, D_i, Z_i, \delta_i$ instead of $Y_{x_i}, C_{x_i}, D_{x_i}, Z_{x_i}, \delta_{x_i}$). We will assume the usual conditions of independence of the Y_i , the C_i and the D_i , and also between Y_i , C_i and D_i . The proportional hazards assumption on the censoring variable C_x can be expressed as follows: for $t \geq 0$,

$$1 - G_{1x}(t) = (1 - F_x(t))^{\beta_x} \quad (1.1)$$

for some $\beta_x > 0$.

We will assume that the distribution functions F_x and G_{2x} are continuous. It then follows by (1.1) that G_{1x} is also continuous. We also assume that the probabilities $p_{xk} = P(\delta_x = k)$ ($k = -1, 0, 1$) are strictly between 0 and 1 for all x . An easy calculation in [3] shows that $\beta_x = \frac{p_{x0}}{p_{x1}}$.

In that same paper we introduced the estimator $\hat{F}_{xh}(t)$ for $F_x(t)$ in the following way:

$$1 - \hat{F}_{xh}(t) = (1 - L_{xh}(t))^{\gamma_{xh}} \quad (1.2)$$

where

$$L_{xh}(t) = 1 - \left\{ \prod_{Z_{(i)} \leq t} \left(1 - \frac{w_{n(i)}(x; h_n)}{1 - \sum_{j=1}^{i-1} w_{n(j)}(x; h_n)} \right)^{\varepsilon_{(i)}} \right\} I(t < Z_{(n)}). \quad (1.3)$$

and

$$\gamma_{xh} = \frac{p_{x1h}}{p_{x0h} + p_{x1h}}. \quad (1.4)$$

with $p_{xkh} = \sum_{i=1}^n w_{ni}(x; h_n) I(\delta_i = k)$ ($k = -1, 0, 1$) and $\varepsilon_i = I(\delta_i \neq -1)$. Here $Z_{(1)} \leq \dots \leq Z_{(n)}$ are the ordered Z_1, \dots, Z_n and the $\varepsilon_{(i)}$ and $w_{n(i)}(x; h_n)$ are the corresponding ε_i and $w_{ni}(x; h_n)$ of the ordered Z_i .

These estimators involve a sequence of smoothing weights $\{w_{ni}(x; h_n)\}$, depending on a positive bandwidth sequence $\{h_n\}$, tending to zero as $n \rightarrow \infty$. In our present case of fixed design points it is customary to take the Gasser-Müller type weights, given by

$$\begin{aligned} w_{ni}(x; h_n) &= \frac{1}{c_n(x; h_n)} \int_{x_{i-1}}^{x_i} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) dz & (i = 1, \dots, n) \\ c_n(x; h_n) &= \int_0^{x_n} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) dz. \end{aligned}$$

Here $x_0 = 0$ and K is a known probability density function (kernel).

For simplicity we have used the same kernel K for estimation of $L_{xh}(t)$ and p_{xkh} . The extension to different kernels is of course straightforward.

2 The bootstrap procedure

The purpose of this paper is to propose a bootstrap method for approximating the distribution function of $(nh_n)^{1/2}(\hat{F}_{xh}(t) - F_x(t))$. This will provide an alternative for the normal approximation derived in Theorem 3 of [3]. The normal approximation suffers from complicated expressions for the asymptotic mean and variance (see (5.1) and (5.2) below) which cannot be calculated in a real data analysis. In this situation, the bootstrap approximation allows us to find some bands for the conditional distribution estimator $\hat{F}_{xh}(t)$, as will be shown in Section 7. For the Kaplan Meier estimator the bootstrap has been introduced by [4] and for the bootstrapped Beran estimator we refer to [10].

Given the design points x_i , the minima Z_i and indicators δ_i ($i = 1, \dots, n$), we propose the following

Resampling procedure

Resample (Z_i^*, δ_i^*) ($i = 1, \dots, n$) with replacement from $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$, giving probability $w_{nj}(x_i; g_n)$ to (Z_j, δ_j) , $j = 1, \dots, n$. The sequence $\{g_n\}$ is a pilot bandwidth sequence which is different from $\{h_n\}$ in the sense that $g_n/h_n \rightarrow \infty$ in a certain way.

It is easily verified that the above procedure is equivalent to two alternative procedures.

Alternative procedure 1

Define the random variables Y_i^* , C_i^* , D_i^* (independently) as follows:

$$\begin{array}{lll} Y_1^*, \dots, Y_n^* & \text{are independent;} & Y_i^* \sim \hat{F}_{x_i g} \\ C_1^*, \dots, C_n^* & \text{are independent;} & C_i^* \sim \hat{G}_{1x_i g} \\ D_1^*, \dots, D_n^* & \text{are independent;} & D_i^* \sim G_{2x_i g} \end{array}$$

Here $\hat{F}_{x_i g}$ is an estimator for F_{x_i} as defined in (1.2). The distribution $\hat{G}_{1x_i g}$ is defined

as $1 - (1 - \hat{F}_{x_i g})^{\beta_{x_i g}}$ where $\beta_{x_i g} = (1 - \gamma_{x_i g})/\gamma_{x_i g}$. Finally $G_{2x_i g}$ is a Beran type estimator for the distribution G_{2x_i} . It is given by

$$G_{2x_i g}(t) = 1 - \left\{ \prod_{Z_{(j)} \leq t} \left(1 - \frac{w_{n(j)}(x_i; g_n)}{1 - \sum_{k=1}^{j-1} w_{n(k)}(x_i; g_n)} \right)^{I(\delta_{(j)} = -1)} \right\} I(t < Z_{(n)}).$$

Then define, for $i = 1, \dots, n$, $Z_i^* = Y_i^* \wedge C_i^* \wedge D_i^*$ and $\delta_i^* = 1$ if $Y_i^* \leq C_i^* \wedge D_i^*$, $\delta_i^* = 0$ if $C_i^* \leq Y_i^* \wedge D_i^*$, $\delta_i^* = -1$ if $D_i^* \leq Y_i^* \wedge C_i^*$.

Alternative procedure 2

Another equivalent resampling procedure can be set up by exploiting the following characterization of the model: assumption (1.1) holds for some $\beta_x > 0$ if and only if Z_x and δ_x are independent, conditionally on the event $\{\delta_x \neq -1\}$. (see [3]).

Now define the random variables U_i^* , ε_i^* , D_i^* (independently) as follows:

$$\begin{array}{lll} U_1^*, \dots, U_n^* & \text{are independent;} & U_i^* \sim L_{x_i g} \\ \varepsilon_1^*, \dots, \varepsilon_n^* & \text{are independent;} & \varepsilon_i^* \sim \text{Bin}(1, \gamma_{x_i g}) \\ D_1^*, \dots, D_n^* & \text{are independent;} & D_i^* \sim G_{2x_i g} \end{array}$$

where $L_{x_i g}$ is the estimator of L_{x_i} , as defined in (1.3). $G_{2x_i g}$ has the same definition as in the previous procedure. The ε_i^* come from a Bernoulli distribution with parameter $\gamma_{x_i g}$, which is defined as in (1.4) and is an estimator for γ_{x_i} . The bootstrap sample can now be defined, for $i = 1, \dots, n$, as: $Z_i^* = U_i^* \wedge D_i^*$ and $\delta_i^* = \varepsilon_i^*$ if $U_i^* \leq D_i^*$ and $\delta_i^* = -1$ if $U_i^* > D_i^*$.

Based on the bootstrap sample $(Z_1^*, \delta_1^*), \dots, (Z_n^*, \delta_n^*)$, we define the bootstrap analogue of the partial Koziol-Green estimator given in (1.2) as:

$$1 - \hat{F}_{xhg}^*(t) = (1 - L_{xhg}^*(t))^{\gamma_{xhg}^*}$$

where

$$L_{xhg}^*(t) = 1 - \left\{ \prod_{Z_{(i)}^* \leq t} \left(1 - \frac{w_{n(i)}(x; h_n)}{1 - \sum_{j=1}^{i-1} w_{n(j)}(x; h_n)} \right)^{\varepsilon_{(i)}^*} \right\} I(t < Z_{(n)}^*) \text{ and } \gamma_{xhg}^* = \frac{p_{x1hg}^*}{p_{x0hg}^* + p_{x1hg}^*}$$

with $p_{xkhg}^* = \sum_{i=1}^n w_{ni}(x; h_n) I(\delta_i^* = k)$ ($k = -1, 0, 1$) and $\varepsilon_i^* = I(\delta_i^* \neq -1)$. Here $Z_{(1)}^* \leq \dots \leq Z_{(n)}^*$, and the $\varepsilon_{(i)}^*$ and $w_{n(i)}(x; h_n)$ correspond to $Z_{(i)}^*$.

In the sequel we will use notations P^* , E^* , Var^* , ... for probability, expectation, variance, ... conditionally on the original observations.

The paper is organized as follows. After some notation and regularity conditions in Section 3, we obtain results on an almost sure asymptotic representation in Section 4 and on weak convergence of the bootstrapped process in Section 5. A simulated example will be presented in Section 6 while a real dataset analysis is shown in Section 7. In the appendix we will give the proofs of the theorems and some lemmas which we need in these proofs.

3 Regularity conditions

We will need the following additional notation. For the design points x_1, \dots, x_n we denote $\underline{\Delta}_n = \min_{1 \leq i \leq n} (x_i - x_{i-1})$ and $\overline{\Delta}_n = \max_{1 \leq i \leq n} (x_i - x_{i-1})$. For the kernel K we use $\|K\|_\infty = \sup_{u \in \mathbb{R}} K(u)$, $\|K\|_2^2 = \int_{-\infty}^{+\infty} K^2(u) du$, $\mu_1^K = \int_{-\infty}^{+\infty} u K(u) du$ and $\mu_2^K = \int_{-\infty}^{+\infty} u^2 K(u) du$. For any distribution function \tilde{L} , we denote by $T_{\tilde{L}} = \inf\{t : \tilde{L}(t) = 1\}$ the right endpoint of its support. Note that $T_{F_x} = T_{G_{1x}}$ and $T_{H_x} = T_{F_x} \wedge T_{G_{2x}}$.

We will use the following assumptions on the design points and on the kernel. (Condition (A2') below is slightly stronger than condition (A2) in [3]).

$$(A1) \quad x_n \rightarrow 1, \quad \overline{\Delta}_n = O(n^{-1}), \quad \overline{\Delta}_n - \underline{\Delta}_n = o(n^{-1}).$$

$$(A2') \quad K \text{ is a twice differentiable probability density function with finite support } [-M, M] \text{ for some } M > 0, \mu_1^K = 0, K'' \text{ is continuous and } K(-M) = K'(-M) = K(M) = K'(M) = 0.$$

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

Further we will need typical smoothness conditions on functions like $F_x(t)$, $G_{2x}(t)$ (as functions of x and t) and β_x and p_{xk} (as functions of x).

$$(A3) \quad \text{The second order partial derivatives } \ddot{F}_x(t) = \frac{\partial^2}{\partial x^2} F_x(t), \ddot{G}_{2x}(t) = \frac{\partial^2}{\partial x^2} G_{2x}(t),$$

$F_x''(t) = \frac{\partial^2}{\partial t^2} F_x(t)$, $G_{2x}''(t) = \frac{\partial^2}{\partial t^2} G_{2x}(t)$, $\dot{F}_x'(t) = \frac{\partial^2}{\partial x \partial t} F_x(t)$, $\dot{G}_{2x}'(t) = \frac{\partial^2}{\partial x \partial t} G_{2x}(t)$ exist and are continuous for $0 \leq x \leq 1$ and $0 \leq t \leq T$ with $T < T_{F_x} \wedge T_{G_{2x}}$.

(A4) The second order partial derivatives $\ddot{p}_{xk} = \frac{\partial^2}{\partial x^2} p_{xk}$ ($k = -1, 0, 1$) exist and are continuous for $0 \leq x \leq 1$.

The first order partial derivatives will be denoted by $\dot{F}_x(t)$, $F_x'(t)$, $\dot{G}_{2x}(t)$, $G_{2x}'(t)$, $\dot{\beta}_x$, \dot{p}_{xk} . Further notations are $\|\dot{p}_{xk}\| = \sup_{0 \leq x \leq 1} |\dot{p}_{xk}|$, $\|\ddot{p}_{xk}\| = \sup_{0 \leq x \leq 1} |\ddot{p}_{xk}|$.

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. $g_n/h_n \rightarrow \infty$ in a certain way. This technique of oversmoothing with the initial bandwidth has been successfully used in other resampling schemes in regression (e.g. [1], [6], [10]). Furthermore we will state our results for the fixed bandwidth sequence h_n of optimal rate, i.e. $h_n = Cn^{-1/5}$ for some constant $C > 0$.

4 Almost sure asymptotic representation

Parallel to the result of [3] for the conditional estimator $\hat{F}_{xh}(t)$, we now derive in Theorem 1 an almost sure asymptotic representation for the bootstrap estimator $\hat{F}_{xhg}^*(t)$. It represents the process $\hat{F}_{xhg}^*(t) - \hat{F}_{xg}(t)$ as a weighted sum plus a remainder term with a nice almost sure behaviour. This result will be a key tool in the next section.

Let us introduce the functions $H_x(t) = P(Z_x \leq t)$ and $H_x^{0,1}(t) = P(Z_x \leq t, \varepsilon_x = 1) = P(Z_x \leq t, \delta_x \neq -1)$.

Theorem 1. Assume (A1), (A2'), (A3), (A4) in $[0, T]$ with $T < T_{H_x}$, $h_n = Cn^{-1/5}$ for some $C > 0$, $g_n \rightarrow 0$, $\frac{ng_n^5}{\log n} \rightarrow +\infty$, $\frac{ng_n^5}{\log n} \cdot \frac{h_n}{g_n} = O(1)$. Then, for $t < T_{H_x}$,

$$\hat{F}_{xhg}^*(t) - \hat{F}_{xg}(t) = \sum_{i=1}^n w_{ni}(x; h_n) g_{tx}(Z_i^*, \delta_i^*) - \sum_{i=1}^n w_{ni}(x; g_n) g_{tx}(Z_i, \delta_i) + R_n(x, t)$$

where

$$g_{tx}(Z_i, \delta_i) = \gamma_x(1 - F_x(t)) \left\{ \int_0^t \frac{I(Z_i \leq s) - H_x(s)}{(1 - H_x(s))^2} dH_x^{0,1}(s) \right\}$$

$$\begin{aligned}
& + \frac{I(Z_i \leq t, \delta_i \neq -1) - H_x^{0,1}(t)}{1 - H_x(t)} - \int_0^t \frac{I(Z_i \leq s, \delta_i \neq -1) - H_x^{0,1}(s)}{(1 - H_x(s))^2} dH_x(s) \\
& - \frac{1}{p_{x1}^2} \log(1 - F_x(t)) [p_{x0}I(\delta_i = 1) - p_{x1}I(\delta_i = 0)] \Big\} \quad (4.1)
\end{aligned}$$

and where, for $T < T_{Hx}$, $\sup_{0 \leq t \leq T} |R_n(x, t)| = O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4})$ a.s.

The proof will be given in Section 8.1.

5 Weak convergence

We establish weak convergence to a Gaussian process of the process $(nh_n)^{1/2}(\hat{F}_{xhg}^*(\cdot) - \hat{F}_{xg}(\cdot))$ in the space $l^\infty[0, T]$ of all uniformly bounded real functions on $[0, T]$, endowed with the uniform topology. Since the sample paths of this process are in the space $D[0, T] \subset l^\infty[0, T]$ of right continuous functions with left hand limits, we can easily prove that weak convergence to a Gaussian process in $l^\infty[0, T]$ with uniform topology is equivalent with weak convergence to the same Gaussian process in $D[0, T]$, endowed with the Skorokhod topology. The weak convergence is almost surely, conditionally on the original observations $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$.

We introduce the notation $\tilde{W}_x(\cdot)$ for a Gaussian process with mean function $\beta_x(t)$ and covariance function σ_{st} given by

$$\begin{aligned}
\beta_x(t) &= \frac{1}{2} \mu_2^K \gamma_x(1 - F_x(t)) \left\{ \int_0^t \left[\frac{\ddot{H}_x(s) dH_x^{0,1}(s)}{(1 - H_x(s))^2} + \frac{d\ddot{H}_x^{0,1}(s)}{1 - H_x(s)} \right] \right. \\
&\quad \left. - \frac{1}{p_{x1}^2} \log(1 - F_x(t)) [p_{x0}\ddot{p}_{x1} - p_{x1}\ddot{p}_{x0}] \right\} C^{5/2} \quad (5.1)
\end{aligned}$$

where $C > 0$ is some constant and

$$\begin{aligned}
\sigma_{st} &= \|K\|_2^2 \gamma_x^2(1 - F_x(t))(1 - F_x(s)) \left\{ \int_0^{s \wedge t} \frac{dH_x^{0,1}(y)}{(1 - H_x(y))^2} \right. \\
&\quad \left. + \log(1 - F_x(s)) \log(1 - F_x(t)) \frac{p_{x0}(p_{x0} + p_{x1})}{p_{x1}^3} \right\}. \quad (5.2)
\end{aligned}$$

Theorem 2. Assume (A1), (A2'), (A3), (A4) in $[0, T]$ with $T < T_{Hx}$, $h_n = Cn^{-1/5}$ for some $C > 0$, $g_n \rightarrow 0$, $\frac{ng_n^5}{\log n} \rightarrow +\infty$, $\frac{ng_n^5}{\log n} \cdot \frac{h_n}{g_n} = O(1)$. Then, for $T < T_{Hx}$, as $n \rightarrow \infty$,

$$(nh_n)^{1/2}(\hat{F}_{xhg}^*(\cdot) - \hat{F}_{xg}(\cdot)) \rightarrow \tilde{W}_x(\cdot) \quad \text{in } l^\infty[0, T] \quad \text{almost surely.}$$

The proof of this theorem is given in Section 8.2.

Since $\tilde{W}_x(\cdot)$ above is also the limiting Gaussian process of $(nh_n)^{1/2}(\hat{F}_{xh}(\cdot) - F_x(\cdot))$ (see Theorem 3 in [3]), we conclude that $P^*((nh_n)^{1/2}(\hat{F}_{xhg}^*(t) - \hat{F}_{xg}(t)) \leq y)$ is a strongly consistent estimator for $P((nh_n)^{1/2}(\hat{F}_{xh}(t) - F_x(t)) \leq y)$.

Theorem 3. Assume (A1), (A2'), (A3), (A4) in $[0, T]$ with $T < T_{H_x}$, $h_n = Cn^{-1/5}$ for some $C > 0$, $g_n \rightarrow 0$, $\frac{ng_n^5}{\log n} \rightarrow +\infty$, $\frac{ng_n^5}{\log n} \cdot \frac{h_n}{g_n} = O(1)$. Then, for $t \leq T$, as $n \rightarrow \infty$,

$$\sup_{y \in \mathbb{R}} \left| P^*((nh_n)^{1/2}(\hat{F}_{xhg}^*(t) - \hat{F}_{xg}(t)) \leq y) - P((nh_n)^{1/2}(\hat{F}_{xh}(t) - F_x(t)) \leq y) \right| = o(1) \text{ a.s.}$$

6 Simulated example

The finite sample performance of the normal approximation ([3]) and the bootstrap approximation of the distribution of $(nh_n)^{1/2}(\hat{F}_{xh}(t) - F_x(t))$ will now be compared through some simulations. Let us take fixed and equidistant design points $x_i = i/n$ ($i = 1, \dots, n$). We assume that the survival times Y_i ($i = 1, \dots, n$) are independent random variables and that

$$Y_i \sim \text{Weibull}(a_1 + a_2x_i, b),$$

i.e. $F_{x_i}(t) = 1 - \exp(-(a_1 + a_2x_i)t^b)$, ($t \geq 0$) for some constants a_1 , a_2 and b such that $a_1 > 0 \wedge (-a_2)$ and $b > 0$. Similarly the non informative censoring times D_i ($i = 1, \dots, n$) are independent and for each i , $D_i \sim \text{Weibull}(c_1 + c_2x_i, b)$ for some $c_1 > 0 \wedge (-c_2)$. Furthermore we assume that $\beta_{x_i} = \exp(a_3 + a_4x_i)$ ($i = 1, \dots, n$). By relation (1.1), this gives informative censoring times C_i ($i = 1, \dots, n$) with for each i , $C_i \sim \text{Weibull}((a_1 + a_2x_i) \exp(a_3 + a_4x_i), b)$. We also assume that C_i ($i = 1, \dots, n$) are independent and that for each i , the survival time Y_i , the informative censoring time C_i and the non informative censoring time D_i are independent.

It is now easily verified that the probabilities of uncensored, informatively censored and non informatively censored observations in the data set are given by the following expressions

$$\begin{aligned} p_{x1} &= \frac{a_1 + a_2x}{(a_1 + a_2x) \exp(a_3 + a_4x) + c_1 + c_2x}, \\ p_{x0} &= \frac{(a_1 + a_2x) \exp(a_3 + a_4x)}{(a_1 + a_2x) \exp(a_3 + a_4x) + c_1 + c_2x}, \\ p_{x,-1} &= \frac{c_1 + c_2x}{(a_1 + a_2x) \exp(a_3 + a_4x) + c_1 + c_2x}. \end{aligned}$$

We study two different combinations of the parameters $a_1, a_2, a_3, a_4, c_1, c_2$. Figures 1(a) and 2(a) show how the probabilities uncensored, informatively censored and non informatively censored observations change for the value of the covariate x . On Figure 1, we choose the parameters such that the probability of informative censoring is always smaller then the probability of uncensoring.

[place Figure 1 about here]

In the Koziol-Green model, it is also important to know how the estimates behave when the amount of informatively censored observations in the data set is larger than the amount of uncensored observations. The choices of parameters in Figure 2 is specially designed to obtain such a situation.

[place Figure 2 about here]

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 20 samples were taken and for each sample, 3000 resamples were drawn.

For the bandwidth h_n we take an optimal choice $h_{n,opt} = Cn^{-1/5}$ with $C = 1$. We do not have such an optimal choice for the pilot bandwidth g_n . The only assumption we made, was that the g_n is asymptotically larger than h_n . In this situation we take $g_n = 0.5$ since we have only observations in a window of width 0.5 around $x = 0.5$. By taking the constant $C = 1$, we force the bandwidth h_n to be smaller than the bandwidth g_n .

Figure 1(b)-(d) and Figure 2(c)-(d) show that both the normal and the bootstrap approximation are very close to the true distribution. In both cases the censoring is quite heavy. We see in Figure 2(b) that the quality of the normal and the bootstrap approximation is not good. The true distribution of $(nh_n)^{1/2}(\hat{F}_{xh}(t) - F_x(t))$ makes a vertical jump at $(nh_n)^{1/2}(1 - 0.63) \approx 1.77$. The reason is that all these samples are overestimating the true function $F_{0.5}(0.5) (\approx 0.63)$ by $\hat{F}_{0.5,h}(0.5) = 1$. Since a probability can not be larger then 1, this jump in the picture will always be seen when we are trying to estimate a large probability (like 0.63) of the true distribution function. In this figure it particularly turned up because Figure 2(b) is an exponential distribution. In Figure 2(c)-(d) for the Weibull distribution with scale parameter $b = 2$

or 3, we will also see a vertical jump when we let t increase.

7 Real data example

In this section we illustrate the use of the proposed bootstrap procedure in a real data set: the bone marrow transplantation data of [7]. Bone marrow transplantation is the traditional treatment for patients with acute leukemia. The 137 patients of this dataset were followed after their transplantation. The disease free survival time was recorded for each patient, together with the reason why the patient left the study. In this example we set the indicator of a patient equal to 1 when the patient has relapsed, 0 when the patient died without relapsing and -1 when the patient was still alive at the end of the study. The intuition for this coding, is that we believe that the survival function of time until death has the same form as the survival function of disease free survival time up to some power. The covariate x in this example is the age at transplantation.

[place Figure 3 about here]

In Figure 3, we show 4 situations in which we estimate the conditional distribution $F_x(t)$ by the partial Koziol-Green estimator $\hat{F}_{xh}(t)$. In column 1 of the pictures, we take the covariate age equal to 15 years. This age is selected to show us the behaviour of disease free survival time in adolescents, whose bodies are still growing and developing. In the second column of Figure 3, we take age equal to 40 years, representing people who are fully grown and even start to age. The rows in Figure 3 come from two different choices of the bandwidth h_n , $h_n = 20$ and $h_n = 40$.

In each of the 4 situations, we take 3000 resamples and calculate the bootstrap distribution $\hat{F}_{xhg}^*(t)$. The pilot bandwidth g_n that we use, is twice the bandwidth h_n . To give some simple illustration, we use the area under the curve (AUC) for selecting from these bootstrap estimates, 3 special estimates, the 5%, 50% and 95% curve. The 5% curve is the bootstrap estimate for which 5% of the resamples have an AUC which is less than the AUC of this curve. Similar definitions can be given for the 50% and 95% curves.

For each situation, these curves are drawn in Figure 3. They form each time a band around the partial Koziol-Green estimate $\hat{F}_{xh}(t)$ and give us some idea about how good this estimate is. When the 5% and 95% bootstrap curves are close to the

partial Koziol-Green estimate, we know that our estimate is a rather good estimate for the true distribution $F_x(t)$. The 50% curve forms the center of the bootstrapped band. Comparing this curve with the partial Koziol-Green estimate, gives another way to illustrate the quality of the estimate.

In Figure 3, we see that in each of the four situations the band formed by the 5% and 95% curve, is rather narrow and almost always contains the partial Koziol-Green estimate $\hat{F}_{xh}(t)$. At age 40, it is even more narrow than at age 15.

The construction of proper bootstrap confidence bands for $F_x(t)$ will not be further explored here. It requires a deeper study of the limiting process $\tilde{W}_x(\cdot)$ in Theorem 2 which will be done in the PhD-thesis of Roel Braekers.

8 Appendix

8.1. Proof of almost sure asymptotic representation

First we give some lemmas which we need in the proof of Theorem 1. Let us introduce the empirical estimators $H_{xh}(t)$ and $H_{xh}^{0,1}(t)$ of $H_x(t)$ and $H_x^{0,1}(t)$:

$$H_{xh}(t) = \sum_{i=1}^n w_{ni}(x; h_n) I(Z_i \leq t), \quad H_{xh}^{0,1}(t) = \sum_{i=1}^n w_{ni}(x; h_n) I(Z_i \leq t, \delta_i \neq -1)$$

as well as their bootstrap analogues $H_{xhg}^*(t)$ and $H_{xhg}^{*0,1}(t)$:

$$H_{xhg}^*(t) = \sum_{i=1}^n w_{ni}(x; h_n) I(Z_i^* \leq t), \quad H_{xhg}^{*0,1}(t) = \sum_{i=1}^n w_{ni}(x; h_n) I(Z_i^* \leq t, \delta_i^* \neq -1).$$

Lemma 1. Assume (A1), (A2') and (A4), $h_n = Cn^{-1/5}$ for some $C > 0$, $g_n \rightarrow 0$, $\frac{ng_n^5}{\log n} \rightarrow +\infty$. Then as $n \rightarrow +\infty$,

- (a) $|\text{bias } p_{xkhg}^* - \text{bias } p_{xkh}| = o((nh_n)^{-1/2}) \quad a.s.$
where $\text{bias } p_{xkhg}^* = E^* p_{xkhg}^* - p_{xkg}$ and $\text{bias } p_{xkh} = E p_{xkh} - p_{xk}$.
- (b) $|p_{xkhg}^* - p_{xkg}| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2}) \quad a.s.$
- (c) $|\gamma_{xhg}^* - \gamma_{xg}| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2}) \quad a.s.$
- (d) If (A3) holds instead of (A4), then as $n \rightarrow +\infty$,
 $\sup_{0 \leq t \leq T} |\text{bias } H_{xhg}^*(t) - \text{bias } H_{xh}(t)| = o((nh_n)^{-1/2}) \quad a.s.$
where $\text{bias } H_{xhg}^*(t) = E^* H_{xhg}^*(t) - H_{xg}(t)$ and $\text{bias } H_{xh}(t) = E H_{xh}(t) - H_x(t)$.

Proof. (a) We have,

$$\text{bias } p_{xkhg}^* = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) (p_{xikg} - p_{zkg}) dz + \int_0^{x_n} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) (p_{zkg} - p_{xkg}) dz.$$

To the first term, we apply Lemma 8 of [1]. A second order Taylor series for the second term, gives that,

$$\text{bias } p_{xkhg}^* = \frac{\mu_2^K h_n^2 \ddot{p}_{xkg}}{2} + o(h_n^2) + O\left(\frac{1}{ng_n^2}\right)$$

$$\text{where } \ddot{p}_{xkg} = \sum_{i=1}^n \ddot{w}_{ni}(x; g_n) I(\delta_i = k), \quad \ddot{w}_{nj}(x; g_n) = \frac{1}{g_n^3} \int_{x_{j-1}}^{x_j} K''\left(\frac{x-z}{g_n}\right) dz.$$

The term $\text{bias } p_{xkh}$ can be split up as above and a first order Taylor series on the first and a second order Taylor series on the second, gives that

$$\text{bias } p_{xkh} = \frac{\mu_2^K h_n^2 \ddot{p}_{xk}}{2} + o(h_n^2).$$

This leads to,

$$(nh_n)^{1/2} |\text{bias } p_{xkhg}^* - \text{bias } p_{xkh}| = \frac{\mu_2^K C^{5/2}}{2} |\ddot{p}_{xkg} - \ddot{p}_{xk}| + o(1).$$

Now

$$|\ddot{p}_{xkg} - \ddot{p}_{xk}| \leq |\ddot{p}_{xkg} - E \ddot{p}_{xkg}| + |E \ddot{p}_{xkg} - \ddot{p}_{xk}|.$$

The first term is $o(1)$, using Bernstein's theorem (see e.g. [8]). For the second term we have, for large n ,

$$\begin{aligned} |E \ddot{p}_{xkg} - \ddot{p}_{xk}| &\leq \left| \sum_{i=1}^n \ddot{w}_{ni}(x; g_n) p_{xik} - \frac{1}{g_n^3} \int_0^{x_n} K''\left(\frac{x-z}{g_n}\right) p_{zk} dz \right| \\ &+ \left| \frac{1}{g_n^3} \int_0^{x_n} K''\left(\frac{x-z}{g_n}\right) p_{zk} dz - \ddot{p}_{xk} \right| \\ &\leq \sum_{i=1}^n \frac{1}{g_n^3} \int_{x_{i-1}}^{x_i} \left| K''\left(\frac{x-z}{g_n}\right) \right| |p_{xik} - p_{zk}| dz + \left| \int_{-M}^M K(u) (\ddot{p}_{x-ug_n, k} - \ddot{p}_{xk}) du \right|, \text{ which is } o(1). \end{aligned}$$

(b) Now, $|p_{xkhg}^* - p_{xkg}| \leq |p_{xkhg}^* - E^* p_{xkhg}^*| + |\text{bias } p_{xkhg}^* - \text{bias } p_{xkh}| + |\text{bias } p_{xkh}|.$

The second term is $o((nh_n)^{-1/2})$ a.s. by part (a) of this lemma. The third term is $O(h_n^2 + n^{-1})$ as was shown in the proof of Lemma A.1. in [3]. For the first term we apply Bernstein's inequality where we take $\varepsilon = \varepsilon_n = D(nh_n)^{-1/2}(\log n)^{1/2}$ with D a constant. By the Borel-Cantelli lemma we obtain the result.

(c) It is easily seen that

$$|\gamma_{xhg}^* - \gamma_{xg}| \leq \frac{1}{(p_{x1hg}^* + p_{x0hg}^*)(p_{x1g} + p_{x0g})} \left\{ |p_{x1hg}^* - p_{x1g}| + |p_{x0hg}^* - p_{x0g}| \right\}.$$

By part (b) and Lemma A.1. of [3], we have that $p_{xk hg}^* \xrightarrow{P^*} p_{xk}$ and $p_{xk g} \xrightarrow{P} p_{xk}$. The result now follows from part (b) of this lemma.

(d) The proof of this result is along the same lines as the proof of Lemma 4.1. in [10] (although the resampling scheme is different). An analogous result holds for $H_{xhg}^{*0,1}(t)$.

Lemma 2. Assume (A1), (A2'), (A3), (A4), $h_n = Cn^{-1/5}$ for some $C > 0$, $g_n \rightarrow 0$ and $\frac{ng_n^5}{\log n} \cdot \frac{h_n}{g_n} = O(1)$, $T < T_{H_x}$. Then as $n \rightarrow +\infty$,

- (a) $|p_{xkg} - p_{xk}| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s.
- (b) $|\gamma_{xg} - \gamma_x| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s.
- (c) $\sup_{0 \leq t \leq T} |L_{xg}(t) - L_x(t)| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s.
- (d) $\sup_{0 \leq t \leq T} |\hat{F}_{xg}(t) - F_x(t)| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s.

If moreover $\frac{ng_n^5}{\log n} \rightarrow +\infty$, then as $n \rightarrow +\infty$,

- (e) $\sup_{0 \leq t \leq T} |-\log(1 - L_{xhg}^*(t)) + \log(1 - L_{xg}(t))| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2})$ a.s.
- (f) $\sup_{0 \leq t \leq T} |L_{xhg}^*(t) - L_{xg}(t)| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2})$ a.s.

Proof. (a), (b), (d). In Lemma 1 (resp. Lemma 2 or Theorem 1) of [3], we put $\varepsilon_n = D(nh_n)^{-1/2}(\log n)^{1/2}$ and apply the Borel-Cantelli lemma.

(c), (e). We prove these results along the same lines as in [10].

(f) By a first order Taylor series, we find,

$$L_{xhg}^*(t) - L_{xg}(t) = (1 - L_{xg}(t))[-e^{\theta_n}(\log(1 - L_{xhg}^*(t)) - \log(1 - L_{xg}(t)))]$$

where θ_n lies between 0 and $\log(1 - L_{xhg}^*(t)) - \log(1 - L_{xg}(t))$. Since $e^{\theta_n} \leq \frac{1}{1 - L_{xg}(t)}$, we now obtain the result by parts (c) and (e).

Proof of Theorem 1. Application of a one term Taylor expansion gives

$$\begin{aligned} \hat{F}_{xhg}^*(t) - \hat{F}_{xg}(t) &= \gamma_{xg}(1 - L_{xg}(t))^{\gamma_{xg}-1} [L_{xhg}^*(t) - L_{xg}(t)] \\ &\quad - (1 - \hat{F}_{xg}(t)) \log(1 - L_{xg}(t)) [\gamma_{xhg}^* - \gamma_{xg}] + R_{n1}(x, t) \end{aligned} \quad (8.1)$$

where

$$|R_{n1}(x, t)| \leq (1 - \tilde{L}_{xg}(t))^{-2} (L_{xhg}^*(t) - L_{xg}(t))^2 + \tilde{\gamma}_{xg}^{-2} (\gamma_{xhg}^* - \gamma_{xg})^2 \\ + 2(1 - \tilde{L}_{xg}(t))^{-1} |L_{xhg}^*(t) - L_{xg}(t)| |\gamma_{xhg}^* - \gamma_{xg}|$$

with $\tilde{L}_{xg}(t)$ between $L_{xg}(t)$ and $L_{xhg}^*(t)$ and $\tilde{\gamma}_{xg}$ between γ_{xg} and γ_{xhg}^* .

For the remainder term, we find the following upperbound:

$$|R_{n1}(x, t)| \leq \left[\frac{1}{(1 - L_x(t))^2} + \frac{(\tilde{L}_{xg}(t) - L_x(t))(2 - \tilde{L}_{xg}(t) - L_x(t))}{(1 - \tilde{L}_{xg}(t))^2(1 - L_x(t))^2} \right] (L_{xhg}^*(t) - L_{xg}(t))^2 \\ + \left[\frac{2}{(1 - L_x(t))^2} + \frac{2(\tilde{L}_{xg}(t) - L_x(t))(2 - \tilde{L}_{xg}(t) - L_x(t))}{(1 - \tilde{L}_{xg}(t))^2(1 - L_x(t))^2} \right] (L_{xhg}^*(t) - L_{xg}(t))(\gamma_{xhg}^* - \gamma_{xg}) \\ + \left[\frac{1}{\gamma_x^2} - \frac{\tilde{\gamma}_{xg}^2 - \gamma_x^2}{\tilde{\gamma}_{xg}^2 \gamma_x^2} \right] (\gamma_{xhg}^* - \gamma_{xg})^2.$$

By Lemma 1 and 2, we find that $\tilde{L}_{xg}(T) \xrightarrow{P^*} L_x(T)$ and $\sup_{0 \leq t \leq T} |\tilde{L}_{xg}(t) - L_x(t)| = O_{P^*}((nh_n)^{-1/2}(\log n)^{1/2})$ a.s., so this results in $\sup_{0 \leq t \leq T} |R_{n1}(x, t)| = O_{P^*}((nh_n)^{-1} \log n)$ a.s.

To obtain the asymptotic representation of $L_{xhg}^*(t) - L_{xg}(t)$, we use Theorem 4.1 of [10]:

$$L_{xhg}^*(t) - L_{xg}(t) = (1 - L_{xg}(t)) \left\{ \int_0^t \frac{H_{xhg}^* - H_{xg}}{(1 - H_x)^2} dH_x^{0,1} \right. \\ \left. + \frac{H_{xhg}^{*0,1}(t) - H_{xg}^{0,1}(t)}{1 - H_x(t)} - \int_0^t \frac{H_{xhg}^{*0,1} - H_{xg}^{0,1}}{(1 - H_x)^2} dH_x \right\} + R_{n2}(x, t) \quad (8.2)$$

with $\sup_{0 \leq t \leq T} |R_{n2}(x, t)| = O_{P^*}((nh_n)^{-3/4}(\log n)^{3/4})$ a.s.

To find an asymptotic representation for $\gamma_{xhg}^* - \gamma_{xg}$ we apply a one term Taylor expansion and obtain

$$\gamma_{xhg}^* - \gamma_{xg} = \frac{p_{x1hg}^*}{p_{x1hg}^* + p_{x0hg}^*} - \frac{p_{x1g}}{p_{x1g} + p_{x0g}} \\ = \frac{p_{x0g}}{(p_{x1g} + p_{x0g})^2} [p_{x1hg}^* - p_{x1g}] - \frac{p_{x1g}}{(p_{x1g} + p_{x0g})^2} [p_{x0hg}^* - p_{x0g}] + R_{n3}(x) \quad (8.3)$$

where

$$|R_{n3}(x)| \leq \frac{2}{(\tilde{p}_{x1g} + \tilde{p}_{x0g})^2} [(p_{x1hg}^* - p_{x1g})^2 + (p_{x0hg}^* - p_{x0g})^2]$$

with \tilde{p}_{x0g} between p_{x0hg}^* and p_{x0g} and \tilde{p}_{x1g} between p_{x1hg}^* and p_{x1g} .

By adding and subtracting the same terms, we find that:

$$|R_{n3}(x)| \leq \frac{2}{(p_{x1} + p_{x0})^2} \left[(p_{x1hg}^* - p_{x1g})^2 + (p_{x0hg}^* - p_{x0g})^2 \right] \\ - \frac{2(\tilde{p}_{x1g} - p_{x1g} + p_{x1g} - p_{x1} + \tilde{p}_{x0g} - p_{x0g} + p_{x0g} - p_{x0})(\tilde{p}_{x1g} + \tilde{p}_{x0g} + p_{x1} + p_{x0})}{(\tilde{p}_{x1g} + \tilde{p}_{x0g})^2 (p_{x1} + p_{x0})^2} \times \\ \left[(p_{x1hg}^* - p_{x1g})^2 + (p_{x0hg}^* - p_{x0g})^2 \right]$$

Lemma 1 and 2 imply that $|R_{n3}(x)| = O_{P^*}((nh_n)^{-1} \log n)$.

Combining the results (8.2) and (8.3) into (8.1) gives

$$\hat{F}_{xhg}^*(t) - \hat{F}_{xg}(t) \\ = \gamma_{xg}(1 - \hat{F}_{xg}(t)) \left\{ \left(\int_0^t \frac{H_{xhg}^* - H_{xg}}{(1 - H_x)^2} dH_x^{0,1} + \frac{H_{xhg}^{*0,1}(t) - H_{xg}^{0,1}(t)}{1 - H_x(t)} - \int_0^t \frac{H_{xhg}^{*0,1} - H_{xg}^{0,1}}{(1 - H_x)^2} dH_x \right) \right. \\ \left. - \frac{1}{p_{x1g}^2} \log(1 - \hat{F}_{xg}(t)) \left[p_{x0g}(p_{x1hg}^* - p_{x1g}) - p_{x1g}(p_{x0hg}^* - p_{x0g}) \right] \right\} + R_n(x, t) \\ = \sum_{i=1}^n w_{ni}(x; h_n) g_{tx}(Z_i^*, \delta_i^*) - \sum_{i=1}^n w_{ni}(x; g_n) g_{tx}(Z_i, \delta_i) + R_n(x, t)$$

where g_{tx} is defined in (4.1) and $\sup_{0 \leq t \leq T} |R_n(x, t)| = O_{P^*}((nh_n)^{-3/4} (\log n)^{3/4})$ a.s.

In the last equality we replaced γ_{xg} , \hat{F}_{xg} , p_{x0g} , p_{x1g} by γ_x , F_x , p_{x0} , p_{x1} respectively. This is allowed by Lemma 2.

8.2. Proof of weak convergence

We first establish a lemma on the bootstrap bias which will be needed in the proof of Theorem 2.

Lemma 3. Assume (A1), (A2'), (A3) and (A4), $h_n = Cn^{-1/5}$ for some $C > 0$, $g_n \rightarrow 0$, $\frac{ng_n^5}{\log n} \rightarrow +\infty$ and $T < T_{H_x}$. Then as $n \rightarrow +\infty$,

$$(nh_n)^{1/2} \sup_{0 \leq t \leq T} |b_n^*(t|x) - b_n(t|x)| = o(1) \text{ a.s.}$$

where

$$\begin{aligned} b_n^*(t|x) &= \sum_{i=1}^n w_{ni}(x, h_n) E^* g_{tx}(Z_i^*, \delta_i^*) - \sum_{i=1}^n w_{ni}(x, g_n) g_{tx}(Z_i, \delta_i) \\ b_n(t|x) &= \sum_{i=1}^n w_{ni}(x, h_n) E g_{tx}(Z_i, \delta_i). \end{aligned}$$

Proof. A straightforward calculation gives

$$\begin{aligned} b_n^*(t|x) - b_n(t|x) &= \gamma_x(1 - F_x(t)) \left\{ \int_0^t \frac{\text{bias } H_{xhg}^*(s) - \text{bias } H_{xh}(s)}{(1 - H_x(s))^2} dH_x^{0,1}(s) \right. \\ &\quad + \frac{\text{bias } H_{xhg}^{*0,1}(t) - \text{bias } H_{xh}^{0,1}(t)}{1 - H_x(t)} - \int_0^t \frac{\text{bias } H_{xhg}^{*0,1}(s) - \text{bias } H_{xh}^{0,1}(s)}{(1 - H_x(s))^2} dH_x(s) \\ &\quad \left. - \frac{\log(1 - F_x(t))}{p_{x1}^2} [p_{x0}(\text{bias } p_{x1hg}^* - \text{bias } p_{x1h}) - p_{x1}(\text{bias } p_{x0hg}^* - \text{bias } p_{x0h})] \right\}. \end{aligned}$$

The result now follows from Lemma 1. We also use the inequality $(1 - F_x(t)) \log(1 - F_x(t)) \leq 1$.

Lemma 4. Assume (A1), (A2'), (A3) and (A4) in $[0, T]$ with $T < T_{H_x}$, $h_n = Cn^{-1/5}$ for some $C > 0$, $g_n \rightarrow 0$, $\frac{ng_n^5}{\log n} \rightarrow +\infty$ and $\frac{ng_n^5}{\log n} \cdot \frac{h_n}{g_n} = O(1)$. Then, as $n \rightarrow +\infty$,

$$(a) \quad (nh_n) \max_{1 \leq i \leq n} \sup_{0 \leq t \leq T} |H_{xig}(t) - H_{xi}(t)| \cdot \sum_{j=1}^n w_{nj}^2(x, h_n) = O((nh_n)^{-1/2} (\log n)^{1/2}) \quad a.s.$$

The same holds for the subdistribution $H_{xi}^{0,1}(t)$.

$$(b) \quad (nh_n) \max_{1 \leq i \leq n} |p_{xigk} - p_{xik}| \cdot \sum_{j=1}^n w_{nj}^2(x, h_n) = o(1) \quad a.s. \quad (k = -1, 0, 1)$$

Proof. We have

$$\begin{aligned} &P \left((nh_n) \max_{1 \leq i \leq n} \sup_{0 \leq t \leq T} |H_{xig}(t) - H_{xi}(t)| \cdot \sum_{j=1}^n w_{nj}^2(x, h_n) > \varepsilon \right) \\ &\leq P \left(\|K\|_2^2 \max_{1 \leq i \leq n} \sup_{0 \leq t \leq T} |H_{xig}(t) - H_{xi}(t)| > \varepsilon \right) \\ &\leq \sum_{i=1}^n P \left(\sup_{0 \leq t \leq T} |H_{xig}(t) - H_{xi}(t)| > \frac{\varepsilon}{\|K\|_2^2} \right). \end{aligned}$$

In the same way, we find

$$P \left((nh_n) \max_{1 \leq i \leq n} |p_{xigk} - p_{xik}| \cdot \sum_{j=1}^n w_{nj}^2(x, h_n) > \varepsilon \right) \leq \sum_{i=1}^n P \left(|p_{xigk} - p_{xik}| > \frac{\varepsilon}{\|K\|_2^2} \right).$$

By Lemma A3 and A1b in [10] and Lemma A1 in [3], we get the results.

Proof of Theorem 2. By the asymptotic representation of Theorem 1, we can write

$$\begin{aligned} (nh_n)^{1/2}(\widehat{F}_{xhg}^*(t) - \widehat{F}_{xg}(t)) &= W_{hg}^*(t|x) + (nh_n)^{1/2}(b_n^*(t|x) - b_n(t|x)) \\ &\quad + ((nh_n)^{1/2}b_n(t|x) - \beta_x(t)) + \beta_x(t) + R_n(x, t) \end{aligned}$$

where $W_{hg}^*(t|x) = (nh_n)^{1/2} \sum_{i=1}^n w_{ni}(x, h_n) [g_{tx}(Z_i^*, \delta_i^*) - E^* g_{tx}(Z_i^*, \delta_i^*)]$ and $\beta_x(t)$ is defined in (5.2).

In [3], we already found that $\sup_{0 \leq t \leq T} |(nh_n)^{1/2}b_n(t|x) - \beta_x(t)| = o(1)$ a.s.. So taking Lemma 3 into account, we only need to prove that $W_{hg}^*(\cdot|x)$ converges weakly to $\tilde{W}_x(\cdot)$ in $l^\infty[0, T]$, almost surely. This is done in two steps. First we will show convergence of the finite dimensional distributions and afterwards we will prove tightness. More in detail, this means that first we show that $(W_{hg}^*(t_1|x), \dots, W_{hg}^*(t_q|x)) \xrightarrow{D} N(0; \sigma_{t_j t_k})$ a.s. for any $q = 1, 2, \dots$ and any $0 \leq t_1 \leq \dots \leq t_q \leq T$. This is done by verifying that the following two conditions of [2] hold almost surely for the summands $W_{nik} = (nh_n)^{1/2} w_{ni}(x, h_n) [g_{t_k x}(Z_i^*, \delta_i^*) - E^* g_{t_k x}(Z_i^*, \delta_i^*)]$:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n E^*(W_{nij} W_{nik}) = \sigma_{t_j t_k} \text{ a.s.} \quad (1 \leq j, k \leq q) \quad (8.4)$$

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \int_{\{|W_{ni}| > \varepsilon\}} |W_{ni}|^2 dP^* = 0 \text{ a.s.} \quad (8.5)$$

for every $\varepsilon > 0$, where $|W_{ni}|^2 = \sum_{k=1}^q W_{nik}^2$.

For (8.4) we have:

$$\sum_{i=1}^n E^*(W_{nij} W_{nik}) = nh_n \sum_{i=1}^n w_{ni}^2(x; h_n) \text{Cov}^*(g_{t_j x}(Z_i^*, \delta_i^*), g_{t_k x}(Z_i^*, \delta_i^*)) \quad (8.6)$$

The right hand side of this expression can be calculated as in [3]. We get the same result, but with H_{x_i} , $H_{x_i}^{0,1}$, $p_{x_i k}$ replaced by $H_{x_{ig}}$, $H_{x_{ig}}^{0,1}$, $p_{x_{ig} k}$. With this and Lemma 4, we can write (8.6) as:

$$\sum_{i=1}^n E^*(W_{nij} W_{nik}) = \sigma_{t_j t_k} + o(1) \text{ a.s.} \quad (1 \leq j, k \leq q).$$

For (8.5) we first note that the functions $g_{t_k x}(Z_i^*, \delta_i^*) - E^* g_{t_k x}(Z_i^*, \delta_i^*)$ are bounded

uniformly on $[0, T]$, so that $\max_{1 \leq i \leq n} |W_{ni}| = O_{P^*}((nh_n)^{-1/2})$ a.s., $\sum_{i=1}^n |W_{ni}|^2 = O_{P^*}(1)$ a.s. and hence

$$\begin{aligned} \sum_{i=1}^n \int_{\{|W_{ni}| > \varepsilon\}} |W_{ni}|^2 dP^* &\leq \int_{\{\max_{1 \leq i \leq n} |W_{ni}| > \varepsilon\}} \sum_{i=1}^n |W_{ni}|^2 dP^* \\ &\leq O_{P^*}(1) P^*(\max_{1 \leq i \leq n} |W_{ni}| > \varepsilon) = o_{P^*}(1) \text{ a.s.} \end{aligned}$$

This proves the convergence of the finite dimensional distributions.

To prove tightness, we will check the conditions of Theorem 2.11.9 in [9] (the bracketing central limit theorem). Let us first denote $W_{ni}(t) = (nh_n)^{1/2} w_{ni}(x, h_n) g_{tx}(Z_i^*, \delta_i^*)$ and $\mathcal{F} = [0, T]$. On \mathcal{F} we define the semi-metric $\rho(t, t')$ as

$$\rho(t, t') = \max \left\{ |H_x^{0,1}(t) - H_x^{0,1}(t')|, |F_x(t) - F_x(t')|, \sup_{x^o \in [0,1]} |H_{x^o}^{0,1}(t) - H_{x^o}^{0,1}(t')|^{1/2} \right\}.$$

Before we can verify the conditions of Theorem 2.11.9, we need to divide \mathcal{F} for every n and ε , conditional on the original observations Z_1, \dots, Z_n , in a partition of subintervals $\{\mathcal{F}_{\varepsilon j}^{*n}\}$ such that

$$\sum_{i=1}^n E^* \sup_{t, t' \in \mathcal{F}_{\varepsilon j}^{*n}} |W_{ni}(t) - W_{ni}(t')|^2 \leq \varepsilon^2 \text{ a.s.} \quad (8.7)$$

The smallest number of subintervals in this partition such that (8.7) holds, is called the bracketing number $N_{[\cdot]}^*(\varepsilon, \mathcal{F}, L_2^n)$. Take in $\mathcal{F} = [0, T]$, conditional on the original observations, a sequence of time points $0 = t_0 < t_1 < \dots < t_m = T$ such that $\{Z_1, \dots, Z_n\} \subset \{t_0, \dots, t_m\}$ and $\rho(t, t') \leq C \cdot \varepsilon$ for every $t, t' \in [t_{j-1}, t_j]$, $j = 1, \dots, m$ where C is a constant which will be defined below (see (8.10)).

From Lemma 4, we note that, for some constant $D > 0$,

$$(nh_n) \max_{1 \leq i \leq n} \sup_{0 \leq t \leq T} |H_{x_i}^{0,1}(t) - H_{x_i}^{0,1}(t)| \cdot \sum_{j=1}^n w_{nj}^2(x; h_n) \leq D \left(\frac{\log n}{nh_n} \right)^{1/2} \text{ a.s.} \quad (8.8)$$

Now we define, for all $\varepsilon > 0$, a natural number n_ε which is the largest root of

$$\left(\frac{\log n}{nh_n} \right)^{1/2} = \left(\frac{\log n}{Cn^{4/5}} \right)^{1/2} = \frac{(1 - H_x(T))^2 \varepsilon^2}{12\gamma_x^2 D}. \quad (8.9)$$

For large ε , we note that this equation does not have any roots. In this case we take $n_\varepsilon = 1$.

We define for every ε and n the partition $\{\mathcal{F}_{\varepsilon j}^{*n}\}$.

If $n \leq n_\varepsilon$,

$$\begin{aligned}\mathcal{F}_{\varepsilon j}^{*n} &= \begin{cases} [t_{j-1}, t_j[& \text{if } t_{j-1} \notin \{Z_1, \dots, Z_n\} \\]t_{j-1}, t_j[& \text{if } t_{j-1} \in \{Z_1, \dots, Z_n\} \end{cases} \quad j = 1, \dots, m-1 \\ \mathcal{F}_{\varepsilon m}^{*n} &= \begin{cases} [t_{m-1}, t_m] & \text{if } t_{m-1} \notin \{Z_1, \dots, Z_n\} \\]t_{m-1}, t_m] & \text{if } t_{m-1} \in \{Z_1, \dots, Z_n\} \end{cases} \\ \mathcal{F}_{\varepsilon, m+i}^{*n} &= \{Z_i\}, \quad i = 1, \dots, n.\end{aligned}$$

If $n > n_\varepsilon$,

$$\begin{aligned}\mathcal{F}_{\varepsilon j}^{*n} &= [t_{j-1}, t_j[\quad j = 1, \dots, m-1 \\ \mathcal{F}_{\varepsilon m}^{*n} &= [t_{m-1}, t_m].\end{aligned}$$

We first prove that (8.7) holds for this partition. In this way the bracketing number is given for every ε and n by

$$N_{[]}^*(\varepsilon, \mathcal{F}, L_2^n) = \begin{cases} O(\varepsilon^{-1}) + n & \text{if } n \leq n_\varepsilon \\ O(\varepsilon^{-1}) & \text{if } n > n_\varepsilon \end{cases}.$$

Some easy calculation shows that

$$\begin{aligned}|W_{ni}(t) - W_{ni}(t')| &= (nh_n)^{1/2} w_{ni}(x; h_n) |g_{tx}(Z_i^*, \delta_i^*) - g_{t'x}(Z_i^*, \delta_i^*)| \\ &\leq (nh_n)^{1/2} w_{ni}(x; h_n) \gamma_x \left\{ \frac{|I(Z_i^* \leq t, \delta_i^* \neq -1) - I(Z_i^* \leq t', \delta_i^* \neq -1)|}{1 - H_x(T)} \right. \\ &\quad \left. + \left(\frac{2}{(1 - H_x(T))^2} + \frac{|-1 - \log(1 - F_x(T))|}{p_{x1}^2} \right) |F_x(t) - F_x(t')| + \frac{|H_x^{0,1}(t) - H_x^{0,1}(t')|}{(1 - H_x(T))^2} \right\}\end{aligned}$$

and by the inequality of Cauchy-Schwarz, we get

$$\begin{aligned}|W_{ni}(t) - W_{ni}(t')|^2 &\leq 3(nh_n) w_{ni}^2(x; h_n) \gamma_x^2 \left\{ \frac{|I(Z_i^* \leq t, \delta_i^* \neq -1) - I(Z_i^* \leq t', \delta_i^* \neq -1)|^2}{(1 - H_x(T))^2} \right. \\ &\quad \left. + \left(\frac{2}{(1 - H_x(T))^2} + \frac{|-1 - \log(1 - F_x(T))|}{p_{x1}^2} \right)^2 |F_x(t) - F_x(t')|^2 + \frac{|H_x^{0,1}(t) - H_x^{0,1}(t')|^2}{(1 - H_x(T))^4} \right\}.\end{aligned}$$

Let us note that

$$|I(Z_i^* \leq t, \delta_i^* \neq -1) - I(Z_i^* \leq t', \delta_i^* \neq -1)|^2 = \begin{cases} 1 & t \wedge t' \leq Z_i^* \leq t \vee t', \delta_i^* \neq -1 \\ 0 & \text{otherwise} \end{cases}$$

and hence

$$\sup_{t, t' \in \mathcal{F}_{\varepsilon j}^{*n}} |I(Z_i^* \leq t, \delta_i^* \neq -1) - I(Z_i^* \leq t', \delta_i^* \neq -1)|^2 = \begin{cases} 1 & Z_i^* \in \mathcal{F}_{\varepsilon j}^{*n} \setminus \{\text{left endpoint}\} \\ 0 & \text{otherwise} \end{cases}$$

and

$$E^* \sup_{t, t' \in \mathcal{F}_{\varepsilon j}^{*n}} |I(Z_i^* \leq t, \delta_i^* \neq -1) - I(Z_i^* \leq t', \delta_i^* \neq -1)|^2 = H_{x_{ig}}^{0,1}(\text{right endpoint of } \mathcal{F}_{\varepsilon j}^{*n}) - H_{x_{ig}}^{0,1}(\text{left endpoint of } \mathcal{F}_{\varepsilon j}^{*n}).$$

If $n \leq n_\varepsilon$, we see that this is equal to zero by the construction of the partition. And in this case,

$$\begin{aligned} \sum_{i=1}^n E^* \sup_{t, t' \in \mathcal{F}_{\varepsilon j}^{*n}} |W_{ni}(t) - W_{ni}(t')|^2 &\leq 3(nh_n) \sum_{i=1}^n w_{ni}^2(x; h_n) \gamma_x^2 \times \\ &\quad \left\{ \frac{1}{(1 - H_x(T))^4} + \left(\frac{2}{(1 - H_x(T))^2} + \frac{|-1 - \log(1 - F_x(T))|}{p_{x1}^2} \right)^2 \right\} C^2 \varepsilon^2 \\ &\leq 3\|K\|_2^2 \gamma_x^2 \left\{ \frac{1}{(1 - H_x(T))^4} + \left(\frac{2}{(1 - H_x(T))^2} + \frac{|-1 - \log(1 - F_x(T))|}{p_{x1}^2} \right)^2 \right\} C^2 \varepsilon^2. \end{aligned}$$

If $n > n_\varepsilon$, we get

$$\begin{aligned} \sum_{i=1}^n E^* \sup_{t, t' \in \mathcal{F}_{\varepsilon j}^{*n}} |W_{ni}(t) - W_{ni}(t')|^2 &\leq 3(nh_n) \sum_{i=1}^n w_{ni}^2(x; h_n) \gamma_x^2 \times \\ &\quad \left\{ \frac{1}{(1 - H_x(T))^4} + \left(\frac{2}{(1 - H_x(T))^2} + \frac{|-1 - \log(1 - F_x(T))|}{p_{x1}^2} \right)^2 \right\} C^2 \varepsilon^2 \\ &\quad + \frac{3(nh_n) \gamma_x^2}{(1 - H_x(T))^2} \sum_{i=1}^n w_{ni}^2(x; h_n) \{H_{x_{ig}}^{0,1}(\text{right end of } \mathcal{F}_{\varepsilon j}^{*n}) - H_{x_{ig}}^{0,1}(\text{left end of } \mathcal{F}_{\varepsilon j}^{*n})\}. \end{aligned}$$

Adding and subtracting terms inside the last sum, we obtain

$$\begin{aligned} \sum_{i=1}^n E^* \sup_{t, t' \in \mathcal{F}_{\varepsilon j}^{*n}} |W_{ni}(t) - W_{ni}(t')|^2 &\leq 3\|K\|_2^2 \gamma_x^2 \left\{ \frac{2}{(1 - H_x(T))^4} + \left(\frac{2}{(1 - H_x(T))^2} + \frac{|-1 - \log(1 - F_x(T))|}{p_{x1}^2} \right)^2 \right\} C^2 \varepsilon^2 \\ &\quad + \frac{6\gamma_x^2}{(1 - H_x(T))^2} (nh_n) \max_{1 \leq i \leq n} \sup_{0 \leq t \leq T} |H_{x_{ig}}^{0,1}(t) - H_{x_i}^{0,1}(t)| \cdot \sum_{j=1}^n w_{nj}^2(x; h_n). \end{aligned}$$

Taking the constant C as follows

$$\begin{aligned} C &= \left[6\|K\|_2^2 \gamma_x^2 \left\{ \frac{2}{(1 - H_x(T))^4} \right. \right. \\ &\quad \left. \left. + \left(\frac{2}{(1 - H_x(T))^2} + \frac{|-1 - \log(1 - F_x(T))|}{p_{x1}^2} \right)^2 \right\} \right]^{-1/2} \end{aligned} \quad (8.10)$$

and using (8.8) and (8.9), we have that (8.7) holds for every n , ε and $\{\mathcal{F}_{\varepsilon j}^{*n}\}$.

At this point we can now verify the third condition of the bracketing central limit theorem, which says that for any sequence $\delta_n \downarrow 0$,

$$\int_0^{\delta_n} \sqrt{\log N_{[]}^*(\varepsilon, \mathcal{F}, L_2^n)} d\varepsilon \rightarrow 0 \quad , \text{ as } n \rightarrow \infty. \quad (8.11)$$

We verify this condition in the two possible situations

$$\begin{aligned} (i) \quad & \delta_n \leq F(nh_n)^{-1/4}(\log n)^{1/4} \quad \text{for } n \text{ large} \\ (ii) \quad & \delta_n > F(nh_n)^{-1/4}(\log n)^{1/4} \quad \text{for } n \text{ large} \end{aligned} \quad , \text{ where } F = \left(\frac{12\gamma_x^2 D}{(1 - H_x(T))^2} \right)^{1/2}.$$

In case (i), the integral in (8.11) is bounded above by

$$\int_0^{F\left(\frac{\log n}{nh_n}\right)^{1/4}} \sqrt{\log \left(\frac{G}{\varepsilon} + n \right)} d\varepsilon \quad (8.12)$$

and in case (ii), the integral in (8.11) is bounded above by

$$\int_0^{F\left(\frac{\log n}{nh_n}\right)^{1/4}} \sqrt{\log \left(\frac{G}{\varepsilon} + n \right)} d\varepsilon + \int_{F\left(\frac{\log n}{nh_n}\right)^{1/4}}^{\delta_n} \sqrt{\log \left(\frac{G}{\varepsilon} \right)} d\varepsilon \quad (8.13)$$

where $G > 0$ is some constant. The integral in (8.12) can be rewritten after change of variable and integration by parts as

$$\frac{\sqrt{\log \left[GF^{-1} \left(\frac{nh_n}{\log n} \right)^{1/4} + n \right]}}{\left(\frac{nh_n}{\log n} \right)^{1/4}} + \frac{G}{2} \int_{\log \left[GF^{-1} \left(\frac{nh_n}{\log n} \right)^{1/4} + n \right]}^{+\infty} \frac{1}{(e^u - n)\sqrt{u}} du.$$

The first term obviously tends to zero. The same is true for the second integral by using the bounded convergence theorem (see e.g. [5]). The second integral in (8.13) can be rewritten as

$$G \int_{\log(G/\delta_n)}^{\log \left[GF^{-1} \left(\frac{nh_n}{\log n} \right)^{1/4} \right]} \sqrt{v} e^{-v} dv$$

which clearly tends to zero as $n \rightarrow \infty$. Hence (8.11) is satisfied and this proves the third condition in the bracketing central limit theorem.

The inequalities above also give

$$\sum_{i=1}^n E^*(W_{ni}(t) - W_{ni}(t'))^2 \leq 3\|K\|_2^2 \gamma_x^2 \left\{ \frac{|H_x^{0,1}(t) - H_x^{0,1}(t')|^2}{(1 - H_x(T))^4} + \frac{\max_{1 \leq i \leq n} |H_{xig}^{0,1}(t) - H_{xig}^{0,1}(t')|}{(1 - H_x(T))^2} \right\}$$

$$+ \left(\frac{2}{(1 - H_x(T))^2} + \frac{|-1 - \log(1 - F_x(T))|}{p_{x1}^2} \right)^2 |F_x(t) - F_x(t')|^2 \Bigg\}$$

so that $\sup_{\rho(t,t') < \delta_n} \sum_{i=1}^n E^*(W_{ni}(t) - W_{ni}(t'))^2 \rightarrow 0$ a.s. for every $\delta_n \downarrow 0$. This is the second condition of the bracketing central limit theorem. Also the first condition is satisfied since for all $\eta > 0$,

$$\sup_{0 \leq t \leq T} |W_{ni}(t)| \leq \frac{\|K\|_\infty M}{(nh_n)^{1/2}} < \eta$$

for all n sufficiently large. Here M is the uniform upper bound for $g_{tx}(Z_i, \delta_i)$.

So by the bracketing central limit theorem, we have that the process $W_{hg}^*(\cdot|x)$ is asymptotically tight. Together with the convergence of the finite dimensional distributions, this concludes our proof.

Acknowledgement

This work was supported by the Ministry of the Flemish Community (Project BIL00/28) and by the IAP research network nr P5/24 of the Belgian State.

The authors are also grateful to the associate editor and two anonymous referees for their critical comments which led to an improvement of the paper.

References

- [1] M. Aerts, P. Janssen and N. Veraverbeke, Bootstrapping regression quantiles, *J. Nonparametric Statistics*. **4**, (1994), 1-20.
- [2] A. Araujo and E. Giné, *The central limit theorem for real and Banach valued random variables*, Wiley, New York, 1980.
- [3] R. Braekers and N. Veraverbeke, The partial Koziol-Green model with covariates, *J. Statist. Planning Inf.* **92**, (2001), 55-71.
- [4] B. Efron, Censored data and the bootstrap, *J. Amer. Statist. Assoc.* **76**, (1981), 312-319.
- [5] J. Foran, *Fundamentals of real analysis*, Marcel Dekker, New York, 1991.
- [6] W. Härdle and E. Mammen, Bootstrap methods in nonparametric regression, *Nonparametric Functional Estimation and Related Topics* (ed. G. Roussas), Kluwer, Dordrecht, 1991, pp. 111-123.

- [7] J. P. Klein and M. L. Moeschberger, *Survival analysis: Techniques for censored and truncated data*, Springer-Verlag, New York, 1997.
- [8] R. J. Serfling, *Approximation theorems of mathematical statistics*, Wiley, New York, 1980.
- [9] A. van der Vaart and J. A. Wellner, *Weak convergence and empirical processes*, Springer, New York, 1996.
- [10] I. Van Keilegom and N. Veraverbeke, Estimation and bootstrap with censored data in fixed design nonparametric regression, *Ann. Inst. Statist. Math.* **49**, (1997), 467-491.

Roel Braekers (roel.braekers@luc.ac.be)
 Noël Veraverbeke (noel.veraverbeke@luc.ac.be)
 Limburgs Universitair Centrum
 Universitaire Campus
 B-3590 Diepenbeek (Belgium)

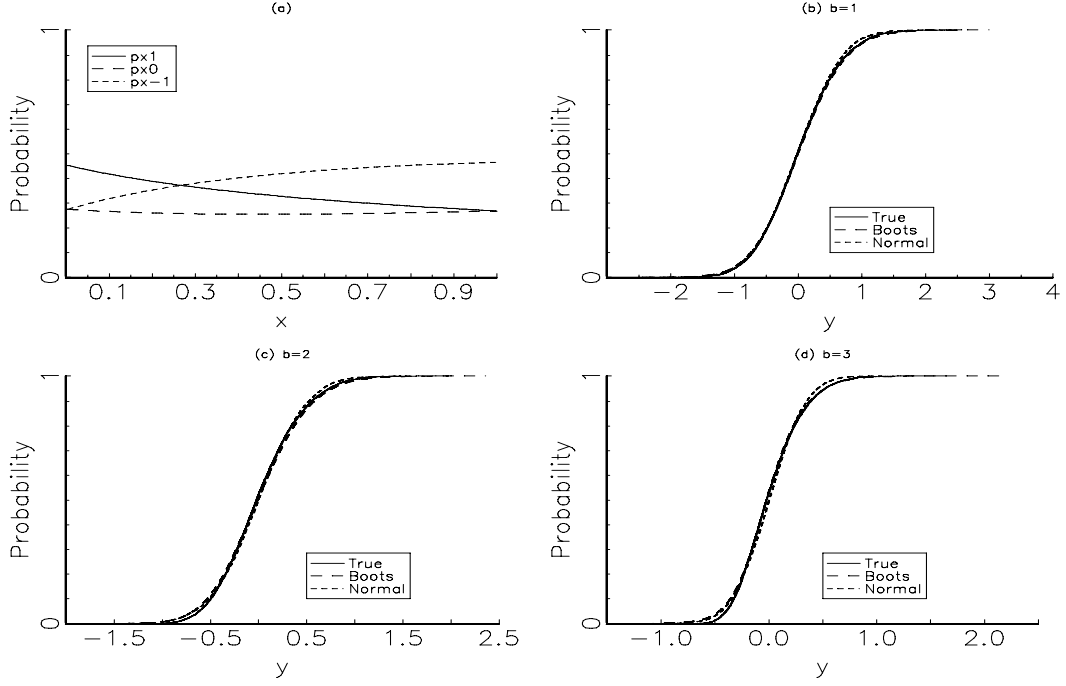


Figure 1: (a) Proportion of uncensored, informatively censored and non informatively censored observations for $a_1 = 1, a_2 = 0.5, a_3 = -0.5, a_4 = 0.5, c_1 = 0.6, c_2 = 2$. (b)-(d) Normal approximation (small dashed curve) and bootstrap approximation (dashed curve) of $P((nh_n)^{1/2}(\hat{F}_{xh}(t) - F_x(t)) \leq y)$ (solid curve) for $a_1 = 1, a_2 = 0.5, a_3 = -0.5, a_4 = 0.5, c_1 = 0.6, c_2 = 2, x = 0.5, t = 0.5$.

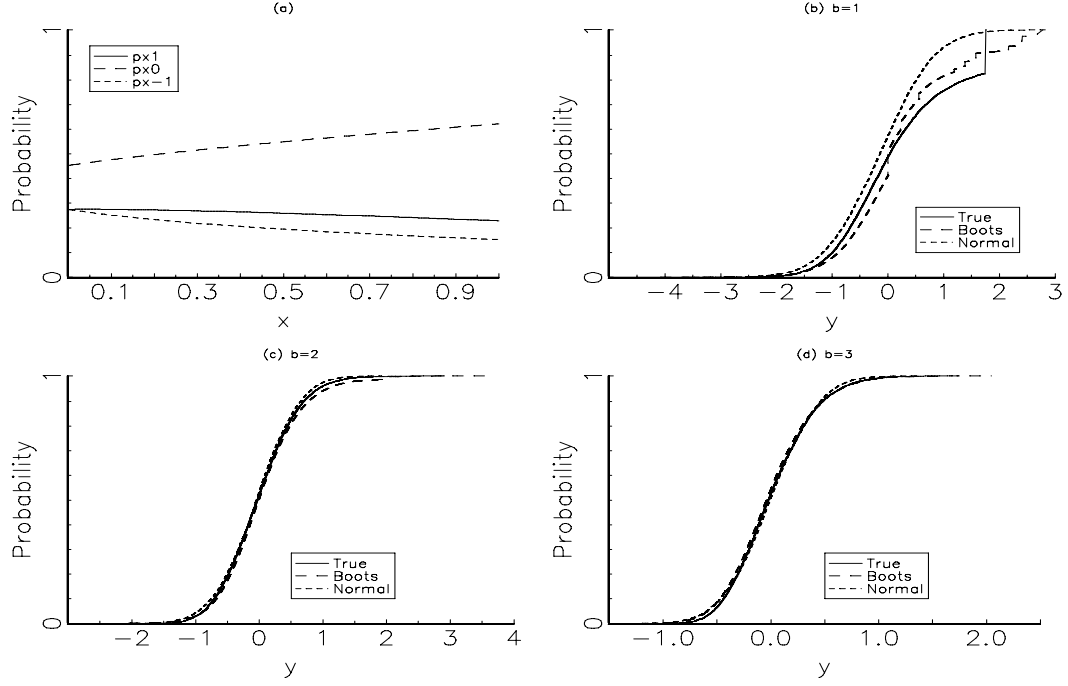


Figure 2: (a) Proportion of uncensored, informatively censored and non informatively censored observations for $a_1 = 1, a_2 = 2, a_3 = 0.5, a_4 = 0.5, c_1 = 1, c_2 = 1$.

(b)-(d) Normal approximation (small dashed curve) and bootstrap approximation (dashed curve) of $P((nh_n)^{1/2}(\hat{F}_{xh}(t) - F_x(t)) \leq y)$ (solid curve) for $a_1 = 1, a_2 = 2, a_3 = 0.5, a_4 = 0.5, c_1 = 1, c_2 = 1, x = 0.5, t = 0.5$.

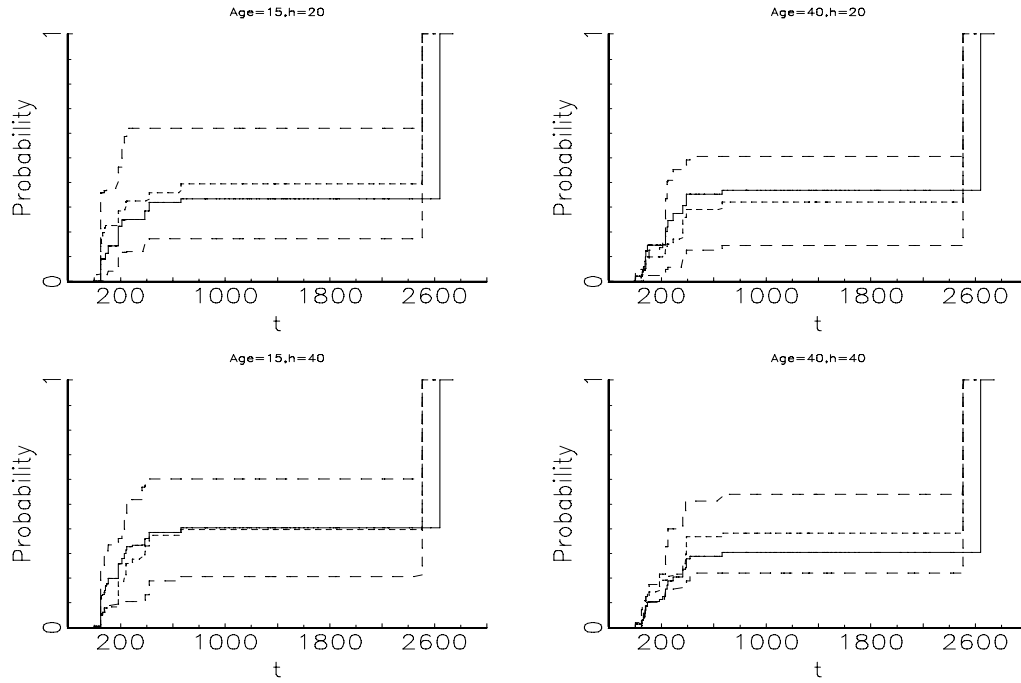


Figure 3: In each of the four pictures, the solid line is the partial Koziol-Green estimator. The long dashed lines are the 5% and 95% curves and the short dashed line is the 50% curve.