On the estimation of the joint distribution in regression models with censored responses

Olivier Lopez

On the estimation of the joint distribution in regression models with censored responses

Olivier Lopez*

June 2007

Abstract

In a regression model with univariate censored responses, a new estimator of the joint distribution function of the covariates and response is proposed, under the assumption that the response and the censoring variable are independent conditionally to the covariates. This estimator is an extension of the multivariate empirical function used in the uncensored case. Furthermore, under some simple additional identifiability assumption, this estimator is not sensible to the "curse of dimensionality", so that it allows to infer on models with multidimensional covariates. Integrals with respect to this empirical measure are considered. Consistency and asymptotic normality of these integrals over a class of functions is obtained, by deriving asymptotic i.i.d. representations. Several applications of the new estimator are proposed.

Key words : distribution function, right censored regression, law of large numbers, asymptotic normality, kernel estimators.

MSC 2000 : 62N01, 62G30, 60F05.

*IRMAR and CREST-ENSAI. Email: lopez@ensai.fr. Address for correspondence: CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal BP 37203, 35172 Bruz cedex, FRANCE

1 Introduction

Under random censoring, estimation of the distribution of a single variable Y is traditionally carried out by using the Kaplan-Meier estimator (see (11)). A vast scope of approaches has been developed to study the theoretical behavior of this estimate, and of Kaplan-Meier integrals ($KM-$integrals in the following). See e.g. Gill (8), Stute and Wang (19), Stute (20), Akritas (1). A crucial identifiability assumption to obtain convergence is the independence of Y and C, the censoring variable. In presence of (uncensored) covariates X, it seems natural to extend Kaplan-Meier's approach, but now to estimate a multivariate distribution function, that is $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$. However, traditional approaches to this kind of problem typically face two major kind of drawbacks, that is either they do not allow to handle multivariate X, or they rely on strong identifiability assumptions which restrain the field of applications. The aim of this paper is to propose a new approach which circumvents these two important limitations, by generalizing the notion of the (multivariate) empirical distribution function.

Indeed, a crucial point in censored regression is to extend the identifiability assumption on the independence of Y and C (used in the univariate case) to the presence of covariates. In the spirit of $KM-$estimator, one may impose that Y and C are independent conditionally to X, which seems to be the slightest identifiability assumption. Under this assumption, Beran (2) provided an estimate of the conditional distribution function $F(y \mid x) = P(Y \leq y \mid X = x)$. In this approach, kernel smoothing is introduced into Kaplan-Meier's approach to account for the information on the interest variable contained in the covariates. Dabrowska (3), (4) studied asymptotics of Beran estimate. Van Keilegom and Akritas (1999) proposed, with some additional assumptions on the regression model, a modification Beran's approach and derived asymptotic properties of their estimate in the case $X \in \mathbb{R}$. A major difficulty in studying this kind of estimate stands in the non-i.i.d. sums that may be involved in. Therefore, several asymptotic i.i.d. representations of the estimated conditional distribution function have been proposed, all in the case where x is univariate, see e.g. Van Keilegom and Akritas (25), Van Keilegom and Veraverberke (26). In particular, Du and Akritas (6) proposed an uniform i.i.d. representation that holds uniformly in y and x.

When it comes to the multivariate distribution function $F(x, y)$, Stute (21) proposed an extension of $KM-$estimator, and furnished asymptotic representation of integrals with respect to this estimator that turned out to have interesting practical applications for regression purpose in some situations, see also Stute (22), Gonzalez-Manteiga, Sanchez-Sellero and Van Keilegom (16), Delecroix, Lopez and Patilea (5), Lopez and Patilea (14). Moreover, in this approach, the covariates do not need to be one-dimensional. Nevertheless, consistency of Stute's estimator relies on assumptions that may be unrealistic in some situations, especially when C and X are not independent. On the other hand, under the more appealing assumption that Y and C are independent conditionally to X, Lo and Singh (13) and Van Keilegom and Akritas (25) used an empirical integral of Beran estimator. Van Keilegom and Akritas (25) also provided some alternative estimate in their so-called "scale-location" model. To our best knowledge, i.i.d. representations of integrals with respect to these estimated distributions have not been provided yet. Moreover, it is particularly disappointing to see that, in the uncensored case, the empirical distribution function of (X', Y) can not be seen as a particular case of these approaches. On the contrary, KM-estimator is a generalization of the (univariate) empirical distribution function. As a large amount of statistical tools are seen to be integrals with respect to the empirical distribution function, it is still of interest to produce some procedure that would generalize this simple and classical way to proceed to the censored framework. In fact, an important preoccupation in the study of censored regression is to extend procedures existing in the uncensored case. For this reason, it is of real interest to use the most natural extension of the uncensored case's concepts.

In this paper, the new estimator is an extension of the notion of the multivariate empirical distribution function, and it can also be seen as a generalization of univariate Kaplan-Meier estimator. Using the results of Dabrowska (4) and Du and Akritas (6), we provide some i.i.d. representation of integrals with respect to this estimator, uniformly over a class of functions. Furthermore, we propose a reasonable modification of the identifiability assumption that may allow us to consider multivariate covariates. The paper is organized as follows. In section 2, we present the model and motivate the introduction of our new estimator of $F(x, y)$. In section 3, we present the asymptotic properties of integrals with respect to this estimate. Section 4 is devoted to some applications of these results, and section 5 to some simulation study, while section 6 gives the proof of some technical results.

2 Model and estimation procedure

2.1 Regression model and description of the methodology

We consider a random vector $(X', Y) \in \mathbb{R}^{d+1}$, and a random variable C which will be referred to as the censoring variable. If variables X and Y are fully observed, and if we dispose on a n-sample of i.i.d. replications $(X_i', Y_i)_{1 \leq i \leq n}$, a traditional way to estimate the joint distribution function $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ is to consider the (multivariate) empirical distribution function,

$$\hat{F}_{emp}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \leq x, Y_i \leq y}, \tag{2.1}$$

where $\mathbf{1}_A$ denotes the indicator function of the set A. If we are interested in estimating $E[\phi(X, Y)] = \int \phi(x, y) dF(x, y)$ for some measurable function ϕ, we can proceed by using

$$\int \phi(x, y) d\hat{F}_{emp}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i, Y_i).$$

Studying the behavior of these integrals is then more general than simply studying the distribution function (2.1). Asymptotic results on these empirical integrals may be derived by applying the classical strong law of large numbers and the central limit theorem. In a censored regression model, the situation is different since the variable Y is not directly available. Indeed, instead of Y, one observes

$$\begin{aligned} T &= Y \wedge C, \\ \delta &= \mathbf{1}_{Y \leq C}. \end{aligned}$$

Observations consist of a $n-$sample $(X_i', T_i, \delta_i)_{1 \leq i \leq n}$. In this framework, the empirical distribution function can not be computed, since it depends on unobserved quantities Y_i. In absence of covariates X, the univariate distribution function $\mathbb{P}(Y \leq y)$ can be estimated computing the Kaplan-Meier estimator,

$$F_{km}(y) = 1 - \prod_{T_i \leq y} \left(1 - \frac{1}{\sum_{j=1}^{n} \mathbf{1}_{T_j \geq T_i}}\right)^{\delta_i}.$$

Asymptotics of F_{km} and of integrals with respect to F_{km} can be found in Stute and Wang (19) and Stute (20). Conditions for convergence are essentially of two kinds : moment conditions (which can be interpreted as assumptions on the "strength" of the censoring in the tail of the distributions, see condition (1.6) in Stute (20), and an identifiability condition that are only needed to ensure that F_{km} converges to the proper function. This identifiability condition, in the univariate case, reduces to

$$Y \text{ and } C \text{ are independent.} \tag{2.2}$$

In a regression framework, an important question is to extend condition (2.2) to the presence of covariates. A first way to proceed would be to assume that

$$(X', Y) \text{ and } C \text{ are independent.} \tag{2.3}$$

However, assumption (2.3) is too restrictive, since, in several frameworks, the censoring variable may depend on X. Stute (21) proposed to replace this assumption by assumption (2.2) and

$$\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y). \tag{2.4}$$

Under these assumption (2.2) and (2.4), Stute (21) studied the asymptotics of the following estimate, that is

$$F_S(x, y) = \sum_{i=1}^{n} W_{in} \mathbf{1}_{X_i \leq x, T_i \leq y},$$

where W_{in} denotes the jump of F_{km} at the i-th observation. Observing that

$$W_{in} = \frac{1}{n} \frac{\delta_i}{1 - G_{km}(T_i-)},$$

where $G_{km}(t)$ denotes the Kaplan-Meier estimate of $G(t) = \mathbb{P}(C \leq t)$ (see Satten and Datta (17)), this estimator may be rewritten as

$$F_S(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - G_{km}(T_i-)}. \tag{2.5}$$

From this writing, one may observe two interesting facts. First, this estimate is a generalization of the empirical distribution function used in the uncensored case. Indeed, in absence of censoring, $1 - G_{km}(t) \equiv 1$ for $t < \infty$, and $\delta = 1$ a.s. Second, F_S can be seen as an approximation of the empirical function

$$\tilde{F}_S(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \mathbf{1}_{X_i \le x, T_i \le y}}{1 - G(T_i -)},$$

function that can not be computed in practice since G is unknown. The identifiability conditions (2.2) and (2.4) (or (2.3)) are needed to ensure that $E[F_S^*(x, y)] = F(x, y)$.

However, conditions (2.3) and (2.4) are still too strong for some applications (see e.g. Beran (2), Dabrowska (3) and (4), Van Keilegom and Akritas (25)). The slightest condition that one may whish to impose, in the spirit of (2.2), is

$$Y \text{ and } C \text{ are independent conditionally to } X. \tag{2.6}$$

Inspired by the empirical distribution function, we are searching for an estimate which puts mass only at the uncensored observations, that is of the form

$$\frac{1}{n} \sum_{i=1}^{n} \delta_i W(X_i, T_i) \mathbf{1}_{X_i \le x, T_i \le y}, \tag{2.7}$$

where $W(X_i, T_i)$ is some weight which has to be chosen in order to compensate the bias due to censoring (F_S is an estimator of the type (2.7), however, under (2.6), it is biased). An "ideal" way to proceed would be to use weights such as, for any function ϕ,

$$E\left[\delta_i W(X_i, T_i)\phi(X_i, T_i)\right] = \int \phi(x, y) dF(x, y),$$

so that integrals with respect to the measure defined by (2.7) converge to the proper limit by the law of large numbers. In this case, (2.7) would appear to be a sum of i.i.d. quantities converging to $F(x, y)$ from the strong law of large numbers. Under (2.6), observe that, for any function W,

$$E[\delta_i W(X_i, T_i)\phi(X_i, T_i)] = E[\{1 - G(Y_i - | X_i)\} W(X_i, Y_i)\phi(X_i, Y_i)], \tag{2.8}$$

where $G(y \mid x)$ denotes $\mathbb{P}(C \le y \mid X = x)$. Hence, a natural choice of W would be

$$W(X_i, T_i) = \frac{1}{1 - G(T_i - \mid X_i)}.$$

This would lead to

$$\tilde{F}(x,y)=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\mathbf{1}_{X_i\le x,T_i\le y}}{1-G(T_i-\mid X_i)}. \tag{2.9}$$

Unfortunately, $G(y\mid x)$ is unknown. However, it can be estimated using Beran's estimator (1981). This estimator is defined, in the case $d=1$, by

$$\hat{G}(y\mid x)=1-\prod_{T_i\le y}\left(1-\frac{w_{in}(x)}{\sum_{j=1}^{n}w_{in}(x)\mathbf{1}_{T_j\ge T_i}}\right)^{1-\delta_i}, \tag{2.10}$$

where, introducing a kernel function K,

$$w_{in}(x)=\frac{K\left(\frac{X_i-x}{h}\right)}{\sum_{j=1}^{n}K\left(\frac{X_j-x}{h}\right)}.$$

The estimator of F that we propose is then

$$\hat{F}(x,y)=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\mathbf{1}_{X_i\le x,T_i\le y}}{1-\hat{G}(T_i-\mid X_i)}. \tag{2.11}$$

This type of approach is quite natural in censored regression, see e.g. van der Laan and Robins (23) or Koul, Susarla, Van Ryzin (12). From this definition, we see that this estimate generalizes the empirical distribution function for the same reasons (2.5) does. Now if we consider a function $\phi(x,y)$, we can estimate $\int\phi(x,y)dF(x,y)$ by

$$\int\phi(x,y)d\hat{F}(x,y)=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\phi(X_i,T_i)}{1-\hat{G}(T_i-\mid X_i)}. \tag{2.12}$$

This estimator is more difficult to study than (2.9), since, as it is the case for Kaplan-Meier integrals, the sums in (2.11) and (2.12) are not i.i.d. In fact, each term depends on the whole sample since $\hat{G}$ is computed from the whole sample. In section 3, we will show that

$$\int\phi(x,y)d\hat{F}(x,y)=\int\phi(x,y)d\tilde{F}(x,y)+S_n(\phi).$$

From (2.8), the classical strong law of large numbers and the central limit theorem, the first integral will converge to $\int\phi(x,y)dF(x,y)$ at rate $n^{-1/2}$, while $S_n(\phi)$, under suitable conditions, can be written as an i.i.d sum which only contributes to the asymptotic variance, but does not modify the limit.

2.2 Comparison with other approaches

Under (2.6), most of the efforts have been concentrated in estimating $F(y \mid x) = \mathbb{P}(Y \leq y \mid X = x)$. Dabrowska (3) and (4) studied uniform consistency and asymptotic normality of Beran's estimator. Van Keilegom and Veraverberke (25), in the case of a fixed design, provided an asymptotic i.i.d. representation of $\hat{F}(y \mid x)$, that is a representation of $\hat{F}(y \mid x)$ as a mean of i.i.d. quantities plus a remainder term which becomes negligible as n grows to infinity. Recently, Du and Akritas (6) provided an analogous representation holding uniformly in y and x for a random X. Van Keilegom and Akritas (25) proposed an alternative to Beran's estimate under some restrictions on the regression model. In particular, they assumed

$$Y = m(X) + \sigma(X)\varepsilon, \tag{2.13}$$

for some location function m, some scale function σ, and ε independent from X.

When it comes to the estimation of the estimation of $F(x, y)$, the only approach that has been used consists of considering

$$\int_{-\infty}^{x} \hat{F}(y \mid u) d\hat{F}(u), \tag{2.14}$$

where $\hat{F}(x)$ denotes the empirical distribution function of X. Instead of (2.10), any other estimate of the conditional distribution function may be used, see for example Van Keilegom and Akritas (25) who provided asymptotic i.i.d. representations for two different estimate based on this principle. To connect another drawback of these procedure with this incapacity to generalize the empirical distribution function, we must mention that none of these approaches has been extended successfully to the case $d > 1$. Of course, the definition of Beran's estimate could be extended to multivariate kernels. But the use of non-parametric regression methods make estimates of the type (2.14) very sensible to the so-called "curse of dimensionality", that is the loss of performance of non-parametric techniques when the number of covariates d increases. This drawback does not affect the estimator (2.1) in the uncensored case. For this reason, parametric estimates which can be written as integrals with respect to (2.1) do not suffer from the curse of dimensionality. It is still the case using the estimate (2.5) under (2.2)-(2.4) (see Stute (22), Delecroix, Lopez and Patilea (5)). Unfortunately, this is not the case if we use (2.14). For this reason, parametric regression has only been considered in the case $d = 1$, see Heuchenne and Van Keilegom (9) and (10).

On the other hand, the estimator proposed in (2.12) is not of the type (2.14). It still relies on Beran's estimator, so that its asymptotical behavior will only be carried out for $d = 1$. However, in section 2.3, we propose a modification of this estimator to handle the case $d > 1$, by slightly strengthening the condition (2.6).

2.3 The case $d > 1$

In (2.11), a non-parametric kernel estimate appears. Therefore, considering a large number of covariates raises theoretical and practical difficulties. For this reason, we propose a slight reasonable modification of the identifiability assumption (2.6) which happens to be a good compromise between (2.6) and (2.3)-(2.4), and under which we will be able to modify the definition of $\hat{F}$ using only univariate kernels. Let $g : \mathbb{R}^d \to \mathbb{R}$ be some known function. The new set of identifiability conditions we propose is

$$Y \text{ and } C \text{ independent conditionally to } g(X), \tag{2.15}$$

$$\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid g(X), Y). \tag{2.16}$$

In particular, condition (2.16) will hold if $\mathcal{L}(C \mid X, Y) = \mathcal{L}(C \mid g(X), Y)$, that is if C depends only on $g(X)$ and Y. As an important example, denote $X = (X^{(1)}, ..., X^{(d)})$. In some practical situations, one may suspect the censoring variable to depend only on $X^{(k)}$ for some k known.

Another interesting advantage of this model is that it may permit us to consider discrete covariates. If we refer to the approach of Van Keilegom and Akritas (25), we can only consider continuous covariates. Here, we will only have to assume that $g(X)$ has a density (but not necessary all component of X). Under this new set of identifiability conditions, we propose to use

$$\tilde{F}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - G(T_i - \mid g(X_i))}, \tag{2.17}$$

$$\hat{F}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - \hat{G}(T_i - \mid g(X_i))}. \tag{2.18}$$

Note that using the set of condition (2.15)-(2.16) does not permit to prevent the estimators of type (2.14) from the curse of dimensionality. In fact, using estimators (2.14), we still need to estimate $F(y \mid x)$, no matter the identifiability conditions.

3 Asymptotics

To simplify the notation, let $z = g(x)$ and $Z_i = g(X_i)$. We provide asymptotic i.i.d. representations of integrals of the type (2.12) which hold uniformly over a class of functions $\mathcal{F}$.

3.1 Assumptions

We list here some assumptions that are needed to ensure consistency and asymptotic normality of our estimate. We will use the following notations to refer to some (sub-)distribution functions,

$$\begin{aligned} H(t) &= \mathbb{P}(T \leq t), \\ H(t \mid z) &= \mathbb{P}(T \leq t \mid Z = z), \\ H_0(t \mid z) &= \mathbb{P}(T \leq t, \delta = 0 \mid Z = z), \\ H_1(t \mid z) &= \mathbb{P}(T \leq t, \delta = 1 \mid Z = z). \end{aligned}$$

Assumptions on the model.

Assumption 1 *The variable $Z = g(X)$ belongs to a compact subset $\mathcal{X} \subset \mathbb{R}$. The distribution function of Z has three bounded derivatives on the interior of $\mathcal{X}$. Furthermore, the density $f_Z(z)$ satisfies*

$$\inf_{z \in \mathcal{X}} f_Z(z) > 0.$$

Assumption 2 *Let $\tau_{H,z} = \inf\{t \mid H(t \mid z) < 1\}$. There exists some real number $\tau < \tau_{H,z}$ for all $z \in \mathcal{X}$.*

Assumption 2 has to be connected with the bad performances of Beran's estimator in the tail of the distribution. This assumption is present in Du and Akritas (6). In Van

Keilegom and Akritas (25), this assumption is avoided only through the specific form of their scale-location regression model.

The important situation that we have in mind in which Assumption 2 holds, is when, for all x, the support of the conditional law $\mathcal{L}(Y \mid Z = z)$ is $[a(z), \tau_H]$, for some $\tau_H \leq \infty$ and $a(z) \geq -\infty$, where the upper bound τ_H does not depend on z and can be finite or not (for example, this condition is fulfilled when Y is gaussian conditionally to $Z = g(X)$). In this case, τ can be chosen arbitrary close to τ_H.

Assumptions on the regularity of the (sub-)distribution functions.

We will assume that the variable $Z = g(X)$ is continuous, but the responses may not be. For any function $J(t \mid z)$ we will denote by $J_c(t \mid z)$ the continuous part, and $J_d(t \mid z) = J(t \mid z) - J_c(t \mid z)$.

Assumption 3 *Functions H and H_c (and consequently H_d) have two derivatives with respect to z. Furthermore, these derivatives are uniformly bounded for $y < \tau$.*

Assumption 4 *For some positive nondecreasing bounded (on $[-\infty;\tau]$) functions L_1, L_2, L_3, we have, for all z,*

$$
\begin{aligned}
|H_c(t_1 \mid z) - H_c(t_2 \mid z)| &\leq |L_1(t_1) - L_1(t_2)|, \\
\left|\frac{\partial H_c}{\partial z}(t_1 \mid z) - \frac{\partial H_c}{\partial z}(t_2 \mid z)\right| &\leq |L_2(t_1) - L_2(t_2)|, \\
\left|\frac{\partial H_{0c}}{\partial z}(t_1 \mid z) - \frac{\partial H_{0c}}{\partial z}(t_2 \mid z)\right| &\leq |L_3(t_1) - L_3(t_2)|,
\end{aligned}
$$

the last two assumptions implying the same kind for $\partial H_{1c}/\partial z$.

Assumption 5 *The jumps of $F(. \mid z)$ and $G(. \mid z)$ are the same for all z. Let $(d_1, d_2, ...)$ be the atoms of G.*

Assumption 6 *$F(. \mid z)$ and $G(. \mid z)$ have two derivatives with respect to z, with the first derivatives uniformly bounded (on $[-\infty;\tau]$). The variation of the functions $\partial_z F(. \mid z)$ and $\partial_z^2 F(. \mid z)$ on $[-\infty;\tau]$ is bounded by a constant not depending on z.*

Assumption 7 *For all d_i, define*

$$\begin{aligned} s_i &= \sup_z \left| F(d_i- \mid z) - F(d_i \mid z) \right|, \\ s_i' &= \sup_z \left| \frac{\partial F}{\partial z}(d_i- \mid z) - \frac{\partial F}{\partial z}(d_i \mid z) \right|, \\ r_i &= \sup_z \left| G(d_i- \mid z) - G(d_i \mid z) \right|, \\ r_i' &= \sup_z \left| \frac{\partial G}{\partial z}(d_i- \mid z) - \frac{\partial G}{\partial z}(d_i \mid z) \right|. \end{aligned}$$

Then $\sum_{d_i \leq \tau} s_i + s_i' + r_i + r_i' < \infty$.

Assumptions on the kernel.

Assumption 8 *The kernel K is a symmetric probability density function with compact support, and K has bounded second derivative.*

Assumption 9 *The bandwidth h satisfies $(\log \log n) n^{-1} h^{-2} = O(1)$, and $nh^5 (\log n)^{-1} = O(1)$.*

Assumptions on the family of functions. To achieve uniform consistency over a class of functions, it is necessary to make assumptions on the class of functions $\mathcal{F}$.

Assumption 10 *The class $\mathcal{F}$ is P-Glivenko-Cantelli (cf. Van der Vaart and Wellner, 1996, page 81) and has an integrable enveloppe Φ satisfying $\Phi(t) = 0$ for $t \geq \tau$, for some τ as defined in Assumption 1.*

For asymptotic normality, we will need more restrictions on the class $\mathcal{F}$. Define $N(\varepsilon, \mathcal{F}, L^2)$ denote the covering number (cf. Van der Vaart and Wellner (24) page 83) of the class $\mathcal{F}$ relatively to the L^2-norm.

Assumption 11 *$N(\varepsilon, \mathcal{F}, L^2) \leq A\varepsilon^{-V}$ for some A and $V > 0$, and $\mathcal{F}$ has a square integrable envelope Φ, satisfying $\Phi(x,t) = 0$ for $t \geq \tau$, for some τ as defined in Assumption 1.*

Particular case of classes satisfying Assumption 11 are $VC-$subgraph classes of functions (see Van der Vaart and Wellner (24)). We also need some differentiability conditions on functions ϕ.

Assumption 12 *Assume that*

- *The envelope* Φ *is square integrable.*
- *Let* $F_Z(x,y) = \mathbb{P}(X \leq x, Y \leq y \mid Z)$, *and for any function* ϕ, *define*

$$\bar{\phi}_z(s) = \int_{\mathcal{X}\times\mathbb{R}} \mathbf{1}_{s\leq y}\phi(x,y)dF_z(x,y).$$

Let $\mathcal{X}_\delta$ *be the set of all points at a distance at least* $\delta > 0$ *from the complementary of* $\mathcal{X}$. *Assume that there is a finite number* $K(\mathcal{F})$ *such as, for all* $\phi \in \mathcal{F}$,

$$\phi(X,Y) = \sum_{i=1}^{K(\mathcal{F})} \phi_i(X,Y)\mathbf{1}_{g(X)\in I_i},$$

where $I_i \subset \mathcal{X}_\delta$, *and* $\bar{\phi}_{i,z}$ *is twice continuously differentiable with respect to* z, *with* $\sup_{s\leq\tau,z} |\partial_z\bar{\phi}_{i,z}(s)| + |\partial_z^2\bar{\phi}_{i,z}(s)| \leq M < \infty$, *for some constant* M *not depending on* ϕ_i.

- $\bar{\Phi}$ *is bounded on* $\mathcal{X}_\delta\times]-\infty;\tau]$, *and has bounded partial derivatives with respect to* z.

The reason for introducing the set $\mathcal{X}_\delta$ is to prevent us from some boundary effects which happen while obtaining uniform convergence rate for kernel estimates, see the proof of our Theorem 3.3 below. Note that it is possible to replace $\mathcal{X}_\delta$ by a set growing with n, that is $\mathcal{X}_{\delta(h)}$. For the sake of simplicity, we do not consider this situation, and prefer to focus on a fixed δ. Consequently, if we consider the case $g(x) = x$, to estimate the distribution function $F(x_0, y_0)$, we should consider the function $\phi(x,y) = \mathbf{1}_{x\leq x_0, y\leq y_0}$. This function does not satisfy Assumption 12, but we can still consider $\mathbf{1}_{x\leq x_0,y\leq y_0}\mathbf{1}_{x\in\mathcal{X}_\delta}$. This will lead to an asymptotically biased estimate, but this bias can be taken arbitrary small, as in the approach of Van Keilegom and Akritas (25).

3.2 Consistency

Theorem 3.1 *Under Assumptions 1, 2, 8, 10, and with* $h \to 0$, *and* $nh \to \infty$,

$$\sup_{\phi\in\mathcal{F}} \left| \int \phi(x,y)\, d\hat{F}(x,y) - \int \phi(x,y)\, dF(x,y) \right| \to_{a.s.} 0.$$

Proof. Write, from the definition (2.12) of $I(\phi)$,

$$\begin{aligned} I(\phi) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\phi\left(X_i,T_i\right)}{1-G\left(T_i-\mid Z_i\right)} \\ &+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\phi(X_i,T_i)[\hat{G}(T_i-|Z_i)-G(T_i-|Z_i)]}{[1-G\left(T_i-|Z_i\right)][1-\hat{G}\left(T_i-|Z_i\right)]} \\ &= I_{0n}+I_{1n}. \end{aligned} \tag{3.1}$$

From the strong law of large numbers, the first term converges almost surely to the expectation $\int \phi\left(x,y\right)dF\left(x,y\right)$ (uniformly over $\mathcal{F}$ since $\mathcal{F}$ is $P-$Glivenko Cantelli), while, for the second,

$$|I_{1n}| \leq O_P(1)\times \sup_{t\leq\tau,x\in\chi}\left|\hat{G}(t-|z)-G\left(t-|z\right)\right|\times\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\left|\Phi(X_i,T_i)\right|}{\left[1-G(T_i-|Z_i)\right]^2}.$$

The empirical sum converges almost surely, while the supremum tends to zero almost surely from Corollary 2.1 of Dabrowska (4). ■

3.3 Asymptotic normality

Theorem 3.1 is not sufficient when it comes to proving asymptotic normality of integrals of type (2.12). As in the case of Kaplan-Meier integrals (see Stute (21)), the i.i.d. expansion introduces an additional term if we need a remainder term decreasing to zero at a sufficiently fast rate. For instance, let us recall the i.i.d. development of Beran's estimator from Du and Akritas (6). Defining

$$\begin{aligned} \xi_z(T_i,\delta_i;t) &= \frac{[1-G(T_i-|z)](1-\delta_i)\mathbf{1}_{T_i\leq t}}{[1-G(T_i|z)][1-H(T_i-|z)]} \\ &\quad -\int_{-\infty}^{t}\frac{\mathbf{1}_{T_i\geq s}[1-G(s-|z)]dH_0(s|z)}{[1-G(s|z)][1-H(s-|z)]^2} \\ &= \psi_{1,z}(T_i,\delta_i)\mathbf{1}_{T_i\leq t}-\int_{-\infty}^{t}\psi_{2,x}(T_i,s)dH_0(s|z), \end{aligned}$$

the authors showed that, under Assumptions 1 to 9 and for $t\leq\tau$,

$$\frac{\hat{G}(t|z)-G(t|z)}{1-G(t|z)}=\frac{1}{n}\sum_{i=1}^{n}w_{in}(z)\xi_z(T_i,\delta_i;t)+R_n^G(z,t),$$

with $\sup_{z\in\mathcal{X}_\delta,t\leq\tau}|R_n^G(z,t)|=O_{a.s.}((\log n)^{3/4}n^{-3/4}h^{-3/4})$. Actually, the authors provide an uniform rate over the whole set $\mathcal{X}$, but this is due to the fact that they overlook the Taylor expansion problem near the boundaries of $\mathcal{X}$. See formulas (A.13) and (A.14) in Du and Akritas (6), where their $O(h_n^3)$ should be $O(h_n^2)$. See also the proof of our Theorem 3.3.

The representation of Du and Akritas can be rewritten, analogously to the expansion of Kaplan-Meier's estimate,

$$\frac{\hat{G}(t|z) - G(t|z)}{1 - G(t|z)} = \int_{-\infty}^{t} \frac{dM_{n,z}(y)}{[1 - G(y|z)][1 - F(y - |z)]} + R_n^G(z, t), \tag{3.2}$$

where

$$M_{n,z}(t) = \sum_{i=1}^{n} w_{ni}(z) \left[(1 - \delta_i)\mathbf{1}_{T_i \le t} - \int_{-\infty}^{t} \frac{\mathbf{1}_{T_i \ge y} dG(y|z)}{1 - G(y - |z)} \right].$$

Observe that, contrary to the i.i.d. representation of Kaplan-Meier estimate, $M_{n,z}$ is not a martingale with respect to the natural filtration $\mathcal{H}_t = \sigma(\{X_i \mathbf{1}_{T_i \le t}, T_i \mathbf{1}_{T_i \le t}, \delta_i \mathbf{1}_{T_i \le t}, i = 1, ..., n\})$, since it is biased. In fact, we have

$$E[\xi_{Z_i}(T_i, \delta_i; t)|X_i] = 0, \tag{3.3}$$

but $E[\xi_z(T_i, \delta_i; t)] \neq 0$. However, from (3.2) and (3.3), it seems natural to define

$$M^i(t) = (1 - \delta_i)\mathbf{1}_{T_i \le t} - \int_{-\infty}^{t} \frac{\mathbf{1}_{T_i \ge y} dG(y|Z_i)}{1 - G(y - |Z_i)},$$

which is a martingale which will naturally appear in the development of the integral of type (2.12).

Theorem 3.2 *Under Assumptions 1 to 12,*

$$\int \phi(x, y) \, d(\hat{F} - \tilde{F})(x, y) = \frac{1}{n} \sum_{i=1}^{n} \int \frac{\bar{\phi}_{Z_i}(s) dM_i(s)}{[1 - F(s- \mid Z_i)][1 - G(s \mid Z_i)]} + R_n(\phi),$$

with $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = O_P((\log n)^{3/4} n^{-3/4} h^{-3/4}) + O(h^2)$, *and* $\bar{\phi}$ *defined in Assumption 12.*

In particular, we see that, choosing h such as $nh^4 \to 0$ and such as $(\log n)^{3/4} nh^3 \to \infty$, the remainder term is $o_P(n^{-1/2})$. Also note that, if we do not wish to restrain ourselves to the set $\mathcal{X}_\delta$ defined in Assumption 12, we should add an $O(h)$ in the remainder term.

Proof. Recalling that $z = g(x)$, write

$$\begin{aligned}
\int \phi(x,y) d(\hat{F} - \tilde{F})(x,y) &= \int \frac{\phi(x,y)[\hat{G}(y-|z) - G(y-|z)]}{[1-G(y-|z)]} d\tilde{F}(x,y) \\
&+ \int \frac{\phi(x,y)[\hat{G}(y-|z) - G(y-|z)]^2}{[1-\hat{G}(y-|z)][1-G(y-|z)]} d\tilde{F}(x,y) \\
&= I_1(\phi) + I_2(\phi).
\end{aligned}$$

For $I_2(\phi)$, observe

$$|I_2(\phi)| \leq C \times \sup_{z \in \mathcal{X}_\delta, y \leq \tau} \left| \frac{\hat{G}(y- \mid z) - G(y- \mid z)}{1 - \hat{G}(y- \mid z)} \right|^2 \int \Phi(x,y) d\tilde{F}(x,y).$$

From Proposition 4.3 in Van Keilegom and Akritas (25), deduce that

$$\sup_{\phi \in \mathcal{F}} |I_2(\phi)| = O_P(n^{-1}h^{-1}[\log(h^{-1})]^{1/2}).$$

Applying the representation (3.2) from Du and Akritas (6),

$$\begin{aligned}
I_1(\phi) &= \frac{1}{n} \sum_{i,j} \frac{\delta_i \phi(X_i, T_i) w_{nj}(Z_i) \xi_{Z_j}(Z_i, T_i-)}{[1 - G(T_i - |Z_i)]} \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \phi(X_i, T_i) R_n^G(Z_i, T_i-)}{[1 - G(T_i - |Z_i)]} \\
&= \int \left[\int_{-\infty}^{y-} \frac{\phi(x,y) dM_{n,z}(y')}{[1 - F(y'|z)][1 - G(y' - |z)]} \right] d\tilde{F}(x,y) \\
&\quad + R_n^{(1)}(\phi).
\end{aligned}$$

Now decompose $I_1(\phi)$ into

$$\begin{aligned}
I_1(\phi) &= \int \left[\int_{-\infty}^{y-} \frac{\phi(x,y) dM_{n,z}(y')}{[1 - F(y'|z)][1 - G(y' - |z)]} \right] dF(x,y) \\
&\quad + \int \left[\int_{-\infty}^{y-} \frac{\phi(x,y) dM_{n,z}(y')}{[1 - G(y' - |x)][1 - F(y'|z)]} \right] d(\tilde{F} - F)(x,y) \\
&\quad + R_n^{(1)}(\phi) \\
&= I_0(\phi) + R_n^{(2)}(\phi) + R_n^{(1)}(\phi).
\end{aligned}$$

From the rate of convergence of R_n^G, and the fact that $|\phi| \leq \Phi$, we obtain the rate $\sup_{\phi \in \mathcal{F}} |R_n^{(1)}(\phi)| = O_P((\log n)^{3/4} n^{-3/4} h^{-3/4})$. In Lemma 6.1, we show that $R_n^{(2)}(\phi) =$

$O_P((\log n)n^{-1}h^{-1}) + O_P(h^2)$, so that only $I_0(\phi)$ remains to be studied. Applying Fubini's theorem, rewrite

$$
\begin{aligned}
I_0(\phi) &= \frac{1}{n}\sum_{j=1}^{n}\int w_{nj}(z)\xi_z(T_j,\delta_j;y-)\phi(x,y)dF(x,y) \\
&= \frac{1}{nh}\sum_{j=1}^{n}\int K\left(\frac{Z_j-z}{h}\right)\frac{\xi_z(T_j,\delta_j;y-)\phi(x,y)dF(x,y)}{f_Z(z)} \\
&\quad \left\{+\frac{1}{nh}\sum_{j=1}^{n}\int \frac{K\left(\frac{Z_j-z}{h}\right)\xi_z(T_j,\delta_j;y-)\phi(x,y)dF(x,y)}{[f_Z(z)-\hat{f}_Z(z)]^{-1}[f_Z(z)]^2}\right. \\
&\quad \left.+\frac{1}{nh}\sum_{j=1}^{n}\int \frac{K\left(\frac{Z_j-z}{h}\right)\xi_z(T_j,\delta_j;y-)\phi(x,y)dF(x,y)}{[f_Z(z)-\hat{f}_Z(z)]^{-2}[f_Z(z)]^2[\hat{f}_Z(z)]}.\right\} \\
&= I_{00}(\phi) + R_n^{(3)}(\phi). \qquad (3.4)
\end{aligned}
$$

We show in Lemma 6.2 that $\sup_{\phi\in\mathcal{F}}|R_n^{(3)}(\phi)| = O_P(n^{-1}h^{-1}\log n + h^2)$. By a change of variable, $I_{00}(\phi)$ can be rewritten as

$$
\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\int K(u)\,\xi_{Z_j+hu}(T_j,\delta_j;y-)\phi(x,y)dF(x,y\mid Z_j+hu)du \\
&\qquad = \frac{1}{n}\sum_{j=1}^{n}\int K(u)\,\psi_{1,Z_j+hu}(T_j,\delta_j)\bar{\phi}_{Z_j+hu}(T_j)du \\
&\qquad - \int K(u)\,\psi_{2,Z_j+hu}(T_j,s)\bar{\phi}_{Z_j+hu}(s)dH_0(s\mid Z_j+hu)du. \qquad (3.5)
\end{aligned}
$$

We now use Assumption 12. By linearity, we only have to consider $\phi(x,y) = \phi_1(x,y)\mathbf{1}_{x\in I_1}$, satisfying Assumption 12. Under Assumption 3, the function ψ_i have two bounded derivatives with respect to z. To use a Taylor expansion in (3.5), we must check that Z_j and $Z_j + hu$ are interior points of $\mathcal{X}$. This is the reason why we introduced the set $\mathcal{X}_\delta$, and why it should appear in Du and Akritas (6) to control the bias of their estimate of the conditional distribution function (for example, observe in Du and Akritas (6), equations (A.13) and (A.14) that the rate is not obtained uniformly in $x_0 \in \mathcal{X}$, because $x_0 + hu$ is not an interior point of $\mathcal{X}$ for all $x_0 \in \mathcal{X}$).

Now consider some j_0 such as $Z_{j_0} \in I_1 \subset \mathcal{X}_\delta$. X_{j_0} is an interior point of $\mathcal{X}$. Furthermore, since u takes values only in a compact interval (K has a compact support), $Z_{j_0} + hu$ is almost surely an interior point of $\mathcal{X}$ for n large enough (only depending on δ). From a

Taylor expansion and Fubini's Theorem, the two integrals appearing in (3.5) corresponding to the index j_0 can be rewritten as

$$\int \frac{\bar{\phi}_{1,Z_{j_0}}(s)dM_{j_0}(s)}{[1-F(s-\mid Z_{j_0})][1-G(s\mid Z_{j_0})]} + O(h^2),$$

where we used $\int uK(u)du = 0$, $\int u^2K(u)du < \infty$, and where the $O(h^2)-$rate depends only on δ.

Now we have to consider the index j such as :

1. $X_j + hu \in I_1$ and $X_j \notin I_1$,

2. $X_j \in I_1$ and $X_j + hu \notin I_1$.

To simplify the discussion, we will assume that $I_1 = [a; b]$. The contribution of these terms to (3.5) is

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{Z_j\in I_1, Z_j+hu\notin I_1} \int \frac{\bar{\phi}_{1,Z_j}(s)dM_j(s)}{[1-F(s-\mid Z_j)][1-G(s\mid Z_j)]}$$
$$+\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{Z_j\notin I_1, Z_j+hu\in I_1}] \int \frac{\bar{\phi}_{1,Z_j}(s)dM_j(s)}{[1-F(s-\mid Z_j)][1-G(s\mid Z_j)]} \quad + \quad R_n^{(4)}(\phi_1),$$

where we can bound

$$|R_n^4(\phi_1)| \leq M \times \frac{1}{n}\sum_{j=1}^{n} \int K(u)[\mathbf{1}_{Z_j\in I_1, Z_j+hu\notin I_1} + \mathbf{1}_{Z_j\notin I_1, Z_j+hu\in I_1}]du,$$

where M is a positive constant, and where we used that $|\phi_1| \leq \Phi$, with $\bar{\Phi}$ bounded on $\mathcal{X}_\delta \times]-\infty; \tau]$. The expectation of the right hand can be bounded by

$$M' \times \int K(u)[F_Z(a+h) - F_Z(a-h) + F_Z(b+h) - F_Z(b-h)]du,$$

where F_Z is the cumulative distribution function of Z. Using Assumption 1, a Taylor expansion shows that this term is $O(h^2)$. ■

4 Applications

4.1 Regression analysis

To simplify, assume that $d = 1$. Consider the following regression model,

$$E[Y \mid X, Y \leq \tau] = f(\theta_0, X),$$

where f is a known function and $\theta_0 \in \Theta \subset \mathbb{R}^k$ an unknown parameter, and τ is as in Assumption 1. Once again, introducing τ is a classical way to proceed for mean-regression under (2.6). See e.g. Heuchenne and Van Keilegom (10). If we assume that θ_0 is the unique minimizer of

$$M(\theta) = E\left[\{Y - f(\theta, X)\}^2 \mathbf{1}_{Y\leq\tau, X\in\mathcal{X}_\delta}\right],$$

we can estimate θ_0 by

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \int_{x\in\mathcal{X}_\delta, y\leq\tau} [y - f(\theta, x)]^2 d\hat{F}(x, y).$$

As a consequence of Theorem 3.1 and Theorem 3.3, the following proposition furnishes the asymptotics for $\hat{\theta}$.

Proposition 4.1 *If $\mathcal{F} = \{x \to f(\theta, x), \theta \in \Theta\}$ is P-Glivenko-Cantelli, we have*

$$\hat{\theta} \to \theta_0 \ a.s. \tag{4.1}$$

Furthermore, let ∇_θ (resp. ∇^2_θ) denotes the vectors of partial derivatives with respect to θ (resp. the Hessian matrix) and assume that $\mathcal{F}' = \{x \to \nabla_\theta f(\theta, x), \theta \in \Theta\}$ and $\mathcal{F}'' = \{x \to \nabla^2_\theta f(\theta, x), \theta \in \Theta\}$ are $P-$Glivenko-Cantelli with an integrable enveloppe. We have, if $nh^4 \to 0$ and under Assumptions 1 to 12 for $\phi(x, y) = \nabla_\theta f(\theta_0, x)[y - f(\theta_0, x)]$,

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow \mathcal{N}(0, \Omega^{-1}V\Omega^{-1}), \tag{4.2}$$

with

$$\begin{aligned}
\Omega &= E\left[\nabla_\theta f(\theta_0, X)\nabla_\theta f(\theta_0, X)'\right], \\
V &= Var\left(\int \phi(x, y)d\tilde{F}(x, y) + \frac{1}{n}\sum_{i=1}^n \int \frac{[1 - G(s|X_i)]^{-1}\bar{\phi}_{X_i}(s)dM_i(s)}{[1 - F(s - |X_i)]}\right).
\end{aligned}$$

Proof. Let

$$M_n(\theta) = \int_{x\in\mathcal{X}_\delta, y\leq\tau} [y - f(\theta, x)]^2 d\hat{F}(x, y).$$

Apply Theorem 3.1 to obtain $\sup_\theta |M_n(\theta) - M(\theta)| = o_P(1)$, and hence (4.1) follows. For (4.2), observe that, from a Taylor expansion,

$$\hat{\theta} - \theta_0 = \nabla^2_\theta M_n(\theta_{1n})^{-1}\nabla_\theta M_n(\theta_0),$$

for some θ_{1n} between θ_0 and $\hat{\theta}$. Apply Theorem 3.1 to see that we have $\nabla^2_\theta M_n(\theta_{1n})^{-1} \to \Omega^{-1} a.s.$, and Theorem 3.3 to obtain that $\nabla_\theta M_n(\theta_0) \Rightarrow \mathcal{N}(0, V)$. ■

4.2 Density estimation

In this section, we consider a random variable Y with Lebesgue density f that we wish to estimate. Estimation of the density of Y received a lot interest in the case where Y and C are independent. See e.g. Mielniczuk (15). This assumption may not hold in several practical situations. In such cases the estimator of Mielniczuk (15) is biased. An alternative is to consider that we are under (2.6) or (2.15)-(2.16), where X represent some auxiliary variables which are observed. In this framework, our estimate $\hat{F}$ will permit us to estimate the density f, for example through the use of kernel smoothing. Let $\tilde{K}$ be a compact support function, h_1 some positive parameter tending to zero, and define

$$\hat{f}_\delta(y) = h_1^{-1} \int_{\mathcal{X}_\delta \times \mathbb{R}} \tilde{K}\left(\frac{y'-y}{h_1}\right) d\hat{F}(x,y'). \tag{4.3}$$

Observe that, since $\tilde{K}$ has compact support, if we choose h_1 small enough, the integral in (4.3) is only on $\mathcal{X}_\delta \times]-\infty;\tau]$ for some $\tau < \tau_H$. Let $\tilde{K}_{h_1,y} = \tilde{K}((y-.)h_1^{-1})$. As an immediate corollary of Theorem 3.3, deduce that

$$\begin{aligned} \hat{f}_\delta(y) &= h_1^{-1} \int_{\mathcal{X}_\delta \times \mathbb{R}} \tilde{K}_{h_1,y}(s) d\tilde{F}(x,y') \\ &+ \frac{1}{nh_1} \sum_{i=1}^n \mathbf{1}_{X_i \in \mathcal{X}_\delta} \int \frac{\bar{\tilde{K}}_{X_i}\left(\frac{\cdot - y}{h_1}\right) dM_i(s)}{[1-F(s- \mid X_i)][1-G(s \mid X_i)]} \\ &+ R_n(y), \end{aligned} \tag{4.4}$$

with

$$\sup_{y \leq \tau} |R_n(y)| = O_P(h_1^{-1}(\log n)^{3/4} n^{-3/4} h^{-3/4}) + O_P(h^2 h_1^{-1}).$$

5 Simulation study

We present in this section a comparison between the behavior of our new estimator $\hat{F}$, and the estimator of Van Keilegom and Akritas ((25)). As already mentioned, their estimator was only proposed for the case $d = 1$, and we would like to emphasize that it is does behaves as well in the multivariate case. For $d = 1$ and $d = 3$, we consider the following regression models,

- $X = (X^{(1)}, ..., X^{(d)}) \sim \mathcal{U}[0,1]^{\otimes d}$.

- $Y = d^{-1}\sum_{i=1}^{d}(-1)^i X^{(i)} + \varepsilon$.
- $\varepsilon \sim \mathcal{N}(m,1)$ independent from X, m was taken 1.7, corresponding to 45% of censoring approximatively.
- $C|X \sim \mathcal{E}(e^{\beta_0' X}/5)$, independent from ε, where $\beta_0 = (1/d, ..., 1/d)$.

In this framework, we are focusing on estimating a truncated mean, that is $E[\phi(Y)]$ where $\phi(y) = y\mathbf{1}_{y\leq\tau}$, we took $\tau = 100$. For each model, we generate 100 samples of size $n = 100$ and we estimate $E[\phi(Y)]$ using either the estimator of the distribution function defined in ((25)), or our new estimator. We then estimate the absolute value of the bias, that is $|E[\hat{\phi}] - E[\phi(Y)]|$, and the variance $E[\hat{\phi}^2] - E[\hat{\phi}]^2$. We consider different values for the smoothing parameter h. In the figure below, model 1 stands for $d = 1$ while model 3 stands for $d = 3$.

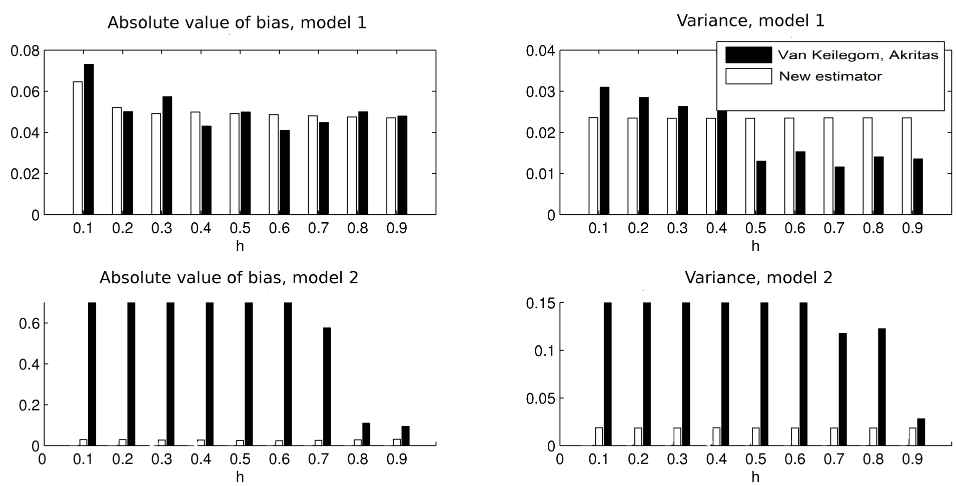

Figure 1: Bias and variance of estimators of $E[Y\mathbf{1}_{Y\leq 100}]$ in function of the smoothing parameter h for $n = 100$ and 45% censoring.

In the one-dimensional case, there is no big difference between both estimators. When it comes to dimension 3, performances of the estimator based on the technique of Van Keilegom and Akritas shrink. It is obvious for the small values of h, and for $h = 0.9$, the bias stays approximatively three times bigger than the one obtained using our new estimator, while the variance is still important.

6 Technical lemmas

Lemma 6.1 *Under Assumptions 1, 2, 6, 9 and 11,*

$$\sup_{\phi\in\mathcal{F}} |R_n^{(2)}(\phi)| = O_P((\log n)n^{-1}h^{-1}) + O_P(h^2).$$

Proof. Let

$$\begin{aligned} U^{i,j}(\phi) &= \frac{\delta_i \phi(X_i,T_i)}{[1-G(T_i-|Z_i)]} K\left(\frac{Z_i-Z_j}{h}\right) \hat{f}_Z(Z_i)^{-1}\xi_{Z_i}(T_j,\delta_j;T_i-) \\ &\quad - \int \phi(x,y) K\left(\frac{z-Z_j}{h}\right)\hat{f}_Z(z)^{-1}\xi_z(T_j,\delta_j;y-)dF(x,y). \end{aligned}$$

Let $W_j = (X_j', Y_j, C_j)$. We can decompose $U^{i,j}$ into $U^{i,j}(\phi) = \sum_{k=1}^4 U_k^{i,j}(\phi)$, where

$$\begin{aligned} U_1^{i,j}(\phi) &= \frac{\delta_i\phi(X_i,T_i)\xi_{Z_i}(T_j,\delta_j;T_i-)}{f_Z(Z_i)[1-G(T_i-|Z_i)]} K\left(\frac{Z_i-Z_j}{h}\right) \\ &\quad -E\left[\frac{\phi(X,Y)\xi_Z(T_j,\delta_j;Y-)}{f_Z(Z)} K\left(\frac{Z-Z_j}{h}\right)|W_j\right], \end{aligned}$$

$$\begin{aligned} U_2^{i,j}(\phi) &= \frac{\delta_i\phi(X_i,T_i)\xi_{Z_i}(T_j,\delta_j;T_i-)K\left(\frac{Z_i-Z_j}{h}\right)}{(\hat{f}_Z(Z_i)-E[\hat{f}_Z(Z_i)|Z_i])^{-1}f_Z(Z_i)^2[1-G(T_i-|Z_i)]} \\ &\quad -E\left[\frac{\phi(X,Y)\xi_Z(T_j,\delta_j;Y-)K\left(\frac{Z-Z_j}{h}\right)}{(\hat{f}_Z(Z)-E[\hat{f}_Z(Z)|Z])^{-1}f_Z(Z)^2}|W_j\right], \end{aligned}$$

$$\begin{aligned} U_3^{i,j}(\phi) &= \frac{\delta_i\phi(X_i,T_i)E[\hat{f}_Z(Z_i)|Z_i]\xi_{Z_i}(T_j,\delta_j;T_i-)}{f_Z(Z_i)^2[1-G(T_i-|Z_i)]} K\left(\frac{Z_i-Z_j}{h}\right) \\ &\quad -E\left[\frac{\phi(X,Y)E[\hat{f}_Z(Z)|Z]\xi_Z(T_j,\delta_j;Y-)}{f_Z(Z)^2} K\left(\frac{Z-Z_j}{h}\right)|W_j\right], \end{aligned}$$

$$\begin{aligned} U_4^{i,j}(\phi) &= \frac{\delta_i\phi(X_i,T_i)[\hat{f}_Z(Z_i)-\hat{f}_Z(Z_i)]^2\xi_{Z_i}(T_j,\delta_j;T_i-)K\left(\frac{Z_i-Z_j}{h}\right)}{f_Z(Z_i)^2\hat{f}_Z(Z_i)[1-G(T_i-|Z_i)]} \\ &\quad -\int \frac{\phi(x,y)[\hat{f}_Z(z)-\hat{f}_Z(z)]^2\xi_x(T_j,\delta_j;y-)K\left(\frac{z-Z_j}{h}\right)}{f_Z(z)^2\hat{f}_Z(z)}dF(x,y). \end{aligned}$$

Observe that, for any $k=1,...,4$, $U_k^{i,i}(\phi)=0$. We have, for some constant M,

$$\begin{aligned} \frac{1}{n^2}\sum_{i,j}|U_4^{i,j}(\phi)| &\leq \frac{M}{n^2}\times\sup_{z\in\mathcal{X}}\frac{|\hat{f}_Z(z)-f_Z(z)|^2}{\hat{f}_Z(z)}\sum_{i,j}K\left(\frac{Z_i-Z_j}{h}\right)\Phi(X_i,T_i) \\ &= O_P(n^{-1}\log n) + O_P(h^2), \end{aligned}$$

from the uniform convergence rate of $\hat{f}_X$, see Einmahl and Mason (7).

Since we have $E[U_1^{i,j}(\phi)|W_j] = 0$, we see that the process defined by

$$\mathbb{U}_1(\phi) = n^{-2} \sum_{i \neq j} \{U_1^{(i,j)}(\phi) - E[U_1^{i,j}(\phi)|W_i]\},$$

is a degenerate $U-$process of order 2. From Lemma 6.3, deduce that this $U-$process is indexed by a class of functions with polynomial covering number. From Corollary 4 in Sherman (18), $\sup_{\phi \in \mathcal{F}} |\mathbb{U}_1(\phi)| = O_P(n^{-1})$. Moreover, from a change of variable and a Taylor expansion,

$$\begin{aligned} E[U_1^{i,j}(\phi)|W_i] &= \frac{h\delta_i \phi(X_i, T_i) \int \xi_{Z_i}(y, \mathbf{1}_{y \leq c}; T_i-) dF(y \mid Z_i) dG(c|Z_i)}{f_Z(Z_i)[1 - G(T_i - |Z_i)]} \\ &\quad + R_1^{i,j}(\phi), \end{aligned} \tag{6.1}$$

where, for some constant M and using Assumption 6 and $\int uK(u)du = 0$,

$$|R_1^{i,j}| \leq M h^3 \frac{\delta_i \Phi(X_i, T_i)}{1 - G(T_i - |Z_i)}.$$

The first term in (6.1) is zero from (3.3). Finally, we have obtained

$$\frac{1}{n^2 h} \sum_{i,j} U_1^{i,j}(\phi) = O_P(n^{-1}h^{-1}) + O_P(h^2).$$

Using the same arguments, the terms $n^{-1}h^{-1} \sum_{i,j} U_k^{i,j}$ for $k = 2, 3$ can be decomposed into a degenerate $U-$process of order greater than 2 indexed by a polynomial class, plus a "bias" term of order $O_P(h^2)$ uniformly over $\mathcal{F}$. Hence, for $k = 2, 3$,

$$\frac{1}{n^2 h} \sum_{i,j} U_k^{i,j}(\phi) = O_P(n^{-1}h^{-1}) + O_P(h^2).$$

Finally, $R_n^{(2)}(\phi) = n^{-2}h^{-1} \sum_{i,j} \sum_{k=1}^{4} U_k^{i,j}(\phi)$. ■

Lemma 6.2 *Under Assumptions 1, 2, 6, 9, and 11,*

$$\sup_{\phi \in \mathcal{F}} R_n^{(3)}(\phi) = O_P(n^{-1}h^{-1} \log n) + O_P(h^2).$$

Proof. From (3.4), we see that the second term of $R_n^{(3)}(\phi)$ has order $O_P(n^{-1}h^{-1} \log n)$ from Lemma 4.3 of Van Keilegom and Akritas (25), and the fact that $|\phi| \leq \Phi$. The first term is

$$
\begin{aligned}
&\frac{1}{nh}\int\sum_{j=1}^{n}\frac{K\left(\frac{Z_j-z}{h}\right)\xi_z(T_j,\delta_j;y-)\phi(x,y)\{\hat{f}_Z(z)-E[\hat{f}_Z(z)]\}dF(x,y)}{[f_Z(z)]^2}\\
+&\frac{1}{nh}\int\sum_{j=1}^{n}\frac{K\left(\frac{Z_j-z}{h}\right)\xi_z(T_j,\delta_j;y-)\phi(x,y)(E[\hat{f}_Z(z)]-f_Z(z))dF(x,y)}{[f_Z(z)]^2}.
\end{aligned}
$$

The first part can be written as

$$
\begin{aligned}
\frac{1}{n^2h^2}\int\left[\sum_{j,i}K(\frac{Z_j-z}{h})\xi_z(T_j,\delta_j;y-)\{K(\frac{Z_i-z}{h})-E[K(\frac{Z-z}{h})]\}\right]&\\
\times\frac{\phi(x,y)dF(x,y)}{[f_Z(z)]^2}&.
\end{aligned}
$$

Observe that the terms for $i=j$ are negligible, since

$$
\begin{aligned}
&\left|\frac{1}{n^2h^2}\sum_{i=1}^{n}\int K\left(\frac{Z_i-z}{h}\right)^2\frac{\xi_z(T_i,\delta_i;y-)\phi(x,y)dF(x,y)}{[f_Z(z)]^2}\right|\\
&\leq\frac{M}{n^2h^2}\sum_{i=1}^{n}\int K\left(\frac{Z_i-z}{h}\right)^2\Phi(x,y)dF(x,y)=O_P(n^{-1}h^{-1}),
\end{aligned}
$$

$$
\begin{aligned}
&\left|\frac{1}{n^2h^2}\sum_{i=1}^{n}\int K\left(\frac{Z_i-z}{h}\right)\frac{E\left[K\left(\frac{Z-z}{h}\right)\right]\xi_z(T_i,\delta_i;y-)\phi(x,y)dF(x,y)}{[f_Z(z)]^2}\right|\\
&\leq\frac{M}{n^2h^2}\sum_{i=1}^{n}\int K\left(\frac{Z_i-z}{h}\right)E\left[K\left(\frac{Z-z}{h}\right)\right]\Phi(x,y)dF(x,y)\\
&=O_P(n^{-1}h^{-1}).
\end{aligned}
$$

Let $N_j(z,y)=K\left(\frac{Z_j-z}{h}\right)\xi_z(T_j,\delta_j;y-)-E\left[K\left(\frac{Z_j-z}{h}\right)\xi_z(T_j,\delta_j;y-)\right]$. We have to consider

$$
\left|\frac{1}{n^2h^2}\int\sum_{j\neq i}^{n}N_j(z,y)\frac{\left(K\left(\frac{Z_i-z}{h}\right)-E\left[K\left(\frac{Z-z}{h}\right)\right]\right)\phi(x,y)dF(x,y)}{f_Z(z)^2}\right|, \tag{6.2}
$$

$$
\left|\frac{1}{nh^2}\int\sum_{i=1}^{n}\frac{E[N_1(x,y)]\left(K\left(\frac{Z_i-z}{h}\right)-E\left[K\left(\frac{Z-z}{h}\right)\right]\right)\phi(x,y)dF(x,y)}{f_Z(z)^2}\right|. \tag{6.3}
$$

From a Taylor expansion, and from Assumptions 1, 6, and 9, we have $h^{-1}E\left[N_j(z,y)\right]=h^2C(z,y)$, with $C(z,y)$ bounded for $z\in\mathcal{X}$ and $y\leq\tau$. Consequently one readily sees that, uniformly over $\mathcal{F}$, (6.3) is $O_P(h^2)$. For (6.2), by Cauchy-Schwarz inequality, the absolute value is bounded by

$$\left(\int\left[\frac{1}{n^2h^2}\sum_{i\neq j}N_j(z,y)\{K(\frac{Z_i-z}{h})-E[K(\frac{Z-z}{h})]\}\right]^2 dF(x,y)\right)^{1/2} \times\left(\int\frac{\Phi(x,y)^2dF(x,y)}{f_Z(z)^4}\right)^{1/2}.$$

Take the expectation of the first parenthesis to see that this expectation is $O(n^{-2}h^{-2})$, while the second parenthesis is finite from the square integrability of Φ and Assumptions 1 and 2. ■

Lemma 6.3 *Let $w=(x,D,t)$, and define the set of functions indexed by $\phi\in\mathcal{F}$, and $h>0$,*

$$\mathcal{G}_1=\left\{\psi_{\phi,h}:(w_1,w_2)\rightarrow\frac{D_1\phi(x_1,t_1)\xi_{x_1}(t_2,D_2;t_1-)K\left(\frac{x_1-x_2}{h}\right)}{f_X(x_1)[1-G(t_1-|x_1)]^2}\right\}.$$

Under Assumptions 1, 2, 9, and 12, for all $\varepsilon>0$ and for some $V_1<\infty$,

$$N(\varepsilon,\mathcal{G}_1,L^2)\leq A_1\varepsilon^{-V_1}.$$

Proof. Let

$$\tilde{g}=\frac{D_1\xi_{x_1}(t_2,D_2;t_1-)}{f_X(x_1)[1-G(t_1-|x_1)]^2}.$$

The class of functions $\tilde{g}\times\mathcal{F}$ as a polynomial covering number, from Lemma A.1 in Einmahl and Mason (7) (the function g is bounded). Let $\mathcal{G}'_1=\{K\left(\frac{x_1-x_2}{h}\right),h>0\}$. This class is uniformly bounded. Apply Lemma A.1 of Einmahl and Mason (7) to conclude on the covering number of $\mathcal{G}_1=\mathcal{G}'_1\times(g\times\mathcal{F})$. ■

References

[1] Akritas, M. (2000). The Central Limit Theorem with Censored Data. *Bernoulli* **6**, 1109–1120.

[2] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

[3] Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Statist.* **14**, 181–197.

[4] Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.* **17**, 1157–1167.

[5] Delecroix, M., Lopez, O. & Patilea, V. (2006). Nonlinear Censored Regression Using Synthetic Data. Document de travail CREST, 2006-10. www.crest.fr.

[6] Du, Y. & Akritas, M. G. (2002). I.i.d representations of the conditional Kaplan-Meier process for arbitrary distributions. *Math. Methods Statist.* **11**, 152–182.

[7] Einmahl, U. & Mason, D. M. (2000). An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators. *J. Theoret. Probab.* **13**, 1–37.

[8] Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 44–58.

[9] Heuchenne, C. & Van Keilegom, I. (2007a). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Inst. Statist. Math.* (to appear).

[10] Heuchenne, C. & Van Keilegom, I. (2007b). Nonlinear censored regression. *Technometrics* (to appear).

[11] Kaplan, E. M., & Meier, P. (1958). Nonparametric estimator from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457–481.

[12] Koul, H., Susarla, V. & Van Ryzin, J. (1981). Regression analysis with randomly right censored data, *Ann. Statist.* **9**, 1276–1288.

[13] Lo, S.-H. & Singh, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Probab. Theory Related Fields* **71**, 455–465.

[14] Lopez, O. & Patilea, V. (2007). Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. Document Crest 2007-01. http://www.crest.fr/doctravail/document/2007-01.pdf.

[15] Mielniczuk, J. (1986). Some asymptotic properties of kernel estimation of a density function in case of censored data. *Ann. Statist.* **14**, 766-Ű773.

[16] Sánchez Sellero, C., González Manteiga, W. & Van Keilegom, I. (2005). Uniform representation of product-limit integrals with applications. *Scand. J. Statist.*, **32**, 563–581.

[17] Satten, G. A. & Datta, S. (2001). The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.* **55**, 207–210.

[18] Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22**, 439–459.

[19] Stute, W. & Wang. J. L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, 1591–1607.

[20] Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422–439.

[21] Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23**, 461–471.

[22] Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica* **9**, 1089–1102.

[23] van der Laan, M. J. & Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality.* Springer-Verlag, New-York.

[24] van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New-York.

[25] Van Keilegom, I. & Akritas, M.G. (1999) Transfer of tail information in censored regression models. *Ann. Statist.* **27**, 1745–1784.

[26] Van Keilegom, I. & Veraverbeke, N. (1997). Estimation and Bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Stat. Math.* **49**, 467–491.