Nonparametric estimation of the multivariate distribution function in a censored regression model with applications

Olivier LOPEZ[1]

October 21, 2009

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 based on the conditional Kaplan-Meier estimator of Beran (1981), and happens to be an extension of the multivariate empirical distribution function used in the uncensored case. We derive asymptotic i.i.d. representations for the integrals with respect to the measure defined by this estimated distribution function. These representations hold even in the case where the covariates are multidimensional under some additional assumption on the censoring. Applications to censored regression and to density estimation are considered.

Key words: Multivariate distribution; Right censoring; Parametric regression; Survival analysis; Dimension reduction.

Short title: Multivariate distribution function for censored regression

[1] Laboratoire de Statistique Théorique et Appliquée, Université Paris VI, 175 rue du Chevaleret, 75013 Paris, France, E-mail: olivier.lopez0@upmc.fr.

1 Introduction

Under random censoring, estimation of the distribution of a single variable Y is traditionally carried out using the Kaplan-Meier estimator (Kaplan and Meier, 1958). A vast scope of approaches has been developed to study the theoretical behavior of this estimator, and of Kaplan-Meier integrals ($KM-$integrals in the following). See e.g. Gill (1983), Stute and Wang (1993), Stute (1995), Akritas (2000). 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, and to provide a version of the uniform strong law of large numbers and of the uniform central limit theorem which apply to this framework of censored regression.

Indeed, a crucial point in censored regression is to extend the identifiability assumption on the independence of Y and C (needed to ensure the convergence of $KM-$integrals in absence of covariates) to the case where explanatory variables are present. 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 (1981) provided an estimator 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. Asymptotic behavior of Beran type estimators has been studied by Dabrowska (1987,1989, 1992), Mielniczuk (1987, 1991), Mc Keague and Utikal (1990), Gonzalez Manteiga and Cardoso Suarez (1994), Li and Doss (1995), Li (1997), Van Keilegom and Veraverbeke (1997). See also Linton and Nielsen (1995) who use similar tools for estimating the cumulative hazard function. Van Keilegom and Akritas (1999) proposed, with some additional assumptions on the regression model, a modification of Beran's approach and derived asymptotic properties of their estimator in the case $X \in \mathbb{R}$. A major difficulty in studying this kind of estimator stands in the non-i.i.d. sums that may be involved in. To circumvent this problem in the case of the (unconditional) product-limit estimator, Lo and Singh (1986) provided a representation as a sum of i.i.d. terms with some remainder term which is asymptotically negligible. Their result was then extended by Major and Retjo (1988) to a more general setting. For the conditional Kaplan-Meier estimator, representations similar to Lo and Singh (1986) were derived, all in the case where x is univariate, see e.g. Van Keilegom and Akritas (1999), Van Keilegom and Veraverberke (1997). In particular, Du and Akritas (2002) 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 (1993, 1996) 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 (1999), Sanchez-Sellero, Gonzalez-Manteiga and Van Keilegom (2005), Delecroix, Lopez and Patilea (2008), Lopez and Patilea (2009). 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, Van Keilegom and Akritas (1999) used an integrated version of Beran's estimator to estimate $F(x, y)$. Van Keilegom and Akritas (1999) also provided some alternative estimator 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 related to 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.

The main contribution of this present paper is to propose a new estimator of $F(x, y)$ which is an extension of the notion of the multivariate empirical distribution function, and can also be seen as a generalization of the univariate Kaplan-Meier estimator. To perform the asymptotic analysis of this estimator, we rely on the asymptotic representation derived by Du and Akritas (2002), but the new proof of this result that we provide improves the convergence rate of the remainder term. Our main theoretical result (Theorem 3.5) provides a general asymptotic representation of our estimator of $F(x, y)$. Unlike the other existing results on this issue, our results apply not only to the estimation of $F(x, y)$, but also to the more general case of integrals with respect to the underlying probability measure. Therefore Theorem 3.5 can be seen as a generalization of the uniform CLT in this censored regression framework. Furthermore, we propose a reasonable modification of the identifiability assumption of the model 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 estimator. Section 4 is devoted to some applications of these results, 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

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

and we will assume throughout this paper that $\mathbb{P}(Y = C) = 0$. 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,

$$F_{km}\left(y\right) = 1 - \prod_{T_i \leq y} \left(1 - \frac{d\hat{H}_1(T_i)}{n^{-1}\sum_{j=1}^n \mathbf{1}_{T_j \geq T_i}}\right)^{\delta_i}, \tag{2.2}$$

where $\hat{H}_1(t) = n^{-1}\sum_{i=1}^n \delta_i \mathbf{1}_{T_i\leq t}$. The definition (2.2) is valid only for y less than the largest observation, conventions must be adopted if one wishes to define it on the whole real line. Asymptotics of F_{km} and of integrals with respect to F_{km} can be found in Stute and Wang (1993) and Stute (1995). Since F_{km} is a piecewise constant function, this estimator can be rewritten as $F_{km}(y) = \sum_{i=1}^n W_{in}\mathbf{1}_{T_i\leq y}$, where W_{in} denotes the jump of F_{km} at the $i-$th observation. 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, 1995), and an identifiability condition that is only needed to ensure that F_{km} converges to the proper function. This identifiability condition, in the univariate case, reduces to

$$Y \text{ and } C \text{ are independent.} \tag{2.3}$$

In a regression framework, an important question is to extend condition (2.3) to the presence of covariates. A first way to proceed would be to assume that

$$(X', Y) \text{ and } C \text{ are independent.} \tag{2.4}$$

However, assumption (2.4) is too restrictive, since, in several frameworks, the censoring variable may depend on X. Stute (1996) proposed to replace this assumption by assumption (2.3) and

$$\mathbb{P}\left(Y \leq C \mid X, Y\right) = \mathbb{P}\left(Y \leq C \mid Y\right). \tag{2.5}$$

Under (2.3) and (2.5), Stute (1996) studied the asymptotics of an estimator based on the jumps W_{in} of F_{km}, that is

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

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

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

where $G_{km}(t)$ denotes the Kaplan-Meier estimator of $G(t) = \mathbb{P}(C \leq t)$ (see e.g. Satten and Datta, 2001), this estimator may be rewritten as

$$F_S(x,y) = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - G_{km}(T_i-)}. \tag{2.6}$$

From this writing, one may observe two interesting facts. First, this estimator 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

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

which can not be computed in practice since G is unknown. The identifiability conditions (2.3) and (2.5) (or (2.4)) are needed to ensure that $E[F_S^*(x,y)] = F(x,y)$.

However, conditions (2.4) and (2.5), which impose restrictions on the conditional law of the censoring, are still too strong for some applications. The slightest condition that one may wish to impose, in the spirit of (2.3), is

$$Y \text{ and } C \text{ are independent conditionally to } X, \tag{2.7}$$

see e.g. Beran (1981), Dabrowska (1987, 1989, 1992), Van Keilegom and Akritas (1999). Inspired by the empirical distribution function, we are searching for an estimator which puts mass only at the uncensored observations, that is of the form

$$\frac{1}{n}\sum_{i=1}^{n} \delta_i W(X_i, T_i) \mathbf{1}_{X_i \leq x, T_i \leq y}, \tag{2.8}$$

where $W(X_i, T_i)$ is some weight which has to be chosen in order to compensate the bias due to censoring ($\tilde{F}_S$ is an estimator of the type (2.8), however, under (2.7), it is biased). An "ideal" way to proceed would be to use weights such as, for any function ϕ,

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

so that integrals with respect to the measure defined by (2.8) converge to the proper limit by the law of large numbers. In this case, (2.8) would appear to be a sum of i.i.d. quantities converging to $F(x,y)$ from the strong law of large numbers. Under (2.7), observe that, for any function ϕ,

$$E[\delta_i W(X_i, T_i)\phi(X_i, T_i)] = E[\{1 - G(Y_i - |X_i)\} W(X_i, Y_i)\phi(X_i, Y_i)], \tag{2.9}$$

where $G(y \mid x)$ denotes $\mathbb{P}(C \leq y \mid X = x)$. 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 the analog of F_S defined in (2.6),

$$\tilde{F}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - G(T_i - \mid X_i)}. \tag{2.10}$$

Unfortunately, $G(y \mid x)$ is unknown. In the uncensored case, conditional distribution function estimation has been considered by Stute (1986) and Horvath and Yandell (1988) among others. In this censored framework, $G(y \mid x)$ can be estimated using Beran's kernel estimator (1981). This estimator is defined, in the case $d = 1$, by

$$\hat{G}(y \mid x) = 1 - \prod_{T_i \leq y} \left(1 - \frac{w_{in}(x) d\hat{H}_0(T_i)}{\sum_{j=1}^{n} w_{jn}(x) \mathbf{1}_{T_j \geq T_i}} \right)^{1-\delta_i}, \tag{2.11}$$

for y less than the largest observation of the sample $(T_i)_{1 \leq i \leq n}$, where, introducing a positive kernel function K,

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

and with $\hat{H}_0(t) = n^{-1} \sum_{i=1}^{n} (1 - \delta_i) \mathbf{1}_{T_i \leq t}$. The definition in equation (2.11) is valid adopting the convention $0/0 = 0$, since the denominator can not be zero unless it is also the case for the numerator. As for the Kaplan-Meier estimator, a convention has to be adopted to define it above the largest observation. The results we provide are valid for any kind of convention adopted for dealing with this issue, since they do not focus on the behavior of Beran's estimator near the tail of the distribution of Y (Assumption 2 below). With at hand the estimator (2.11), the estimator of F that we propose is then

$$\hat{F}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - \hat{G}(T_i - \mid X_i)}, \tag{2.12}$$

for $y \leq \max_{1 \leq i \leq n}(T_i)$. This type of approach is quite natural in censored regression, see e.g. van der Laan and Robins (2003) or Koul, Susarla, Van Ryzin (1981). From this definition, we see that this estimator generalizes the empirical distribution function for the same reasons (2.6) does. Now if we consider a function $\phi(x, y)$, we can estimate $\int \phi(x, y) dF(x, y)$ by

$$\int \phi(x, y) d\hat{F}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \phi(X_i, T_i)}{1 - \hat{G}(T_i - \mid X_i)}. \tag{2.13}$$

This estimator is more difficult to study than (2.10), since, as it is the case for Kaplan-Meier integrals, the sums in (2.12) and (2.13) are not i.i.d. In fact, each term depends on the whole sample since $\hat{G}$ itself is computed from the whole sample. In section 3, we will show that

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

where, from (2.9), the first integral converges to $\int \phi(x,y)dF(x,y)$ at rate $n^{-1/2}$ (consequence of the strong law of large numbers and the central limit theorem), while $S_n(\phi)$, under suitable conditions, is equivalent to a centered i.i.d sum which only contributes to the asymptotic variance.

2.2 Comparison with other approaches

Under (2.7), most of the efforts have been concentrated in estimating $F(y \mid x) = \mathbb{P}(Y \leq y \mid X = x)$. Dabrowska (1987, 1989) studied uniform consistency and asymptotic normality of Beran's estimator. Van Keilegom and Veraverberke (1997), 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. More recently, Du and Akritas (2002) provided an analogous representation holding uniformly in y and x for a random variable X. Van Keilegom and Akritas (1999) proposed an alternative to Beran's estimator under some restrictions on the regression model. In particular, they assumed

$$Y = m(X) + \sigma(X)\varepsilon, \tag{2.14}$$

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

$$\int_{-\infty}^{x} \hat{F}(y \mid u)d\hat{F}(u), \tag{2.15}$$

where $\hat{F}(x)$ denotes the empirical distribution function of X. Instead of (2.11), any other estimator of the conditional distribution function may be used, see for example Van Keilegom and Akritas (1999) who provided asymptotic i.i.d. representations for two different estimators 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 estimator could be extended to multivariate kernels. But the use of non-parametric regression methods make estimators of the type (2.15) very sensitive 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 estimators 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 estimator (2.6) under (2.3)-(2.5) (see Stute, 1999, Delecroix, Lopez and Patilea, 2008). Unfortunately, this is not the case if we use (2.15). For this reason, parametric regression has only been considered in the case $d = 1$, see e.g. Heuchenne and Van Keilegom (2007a, 2007b).

On the other hand, the estimator proposed in (2.13) is not of the type (2.15). 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 below, we propose a modification of this estimator to handle the case $d > 1$, by slightly strengthening the condition (2.7).

2.3 The case $d > 1$

In (2.12), a non-parametric kernel estimator 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 (2.7) which happens to be a good compromise between (2.7) and (2.4)-(2.5), 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

$$Y \text{ and } C \text{ independent conditionally to } g(X), \tag{2.16}$$

$$\mathbb{P}(Y \le C \mid X, Y) = \mathbb{P}(Y \le C \mid g(X), Y). \tag{2.17}$$

In particular, condition (2.17) 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 $g(X) = X^{(k)}$ 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 (1999), we can only consider continuous covariates. Here, we will only have to assume that $g(X)$ has a density (but not necessary all components of X). Under this new set of identifiability conditions, we propose to use

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

$$\hat{F}(x,y) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i \mathbf{1}_{X_i\le x, T_i\le y}}{1-\hat{G}(T_i- \mid g(X_i))}. \tag{2.19}$$

Note that using the set of condition (2.16)-(2.17) does not permit to prevent the estimators of type (2.15) from the curse of dimensionality. In fact, using estimators (2.15), we still need to estimate $F(y \mid x)$, no matter the identifiability conditions.

3 Asymptotic properties

In this section, we present the asymptotic properties of integrals with respect to $\hat{F}$. In section 3.1, we discuss the different assumptions under which our asymptotic results hold. Section 3.2 presents an uniform strong law of large numbers for $\hat{F}$, while section 3.3 provides some new results on the asymptotic i.i.d. representation of Beran's estimator by improving the convergence rate of the remainder term. In section 3.4, we provide uniform asymptotic representations of integrals with respect to $\hat{F}$, with different rate of convergence for the remainder term, depending on the assumptions on the class of functions considered. These representations can be used to obtain uniform CLT results for general classes of functions.

3.1 Assumptions

To simplify the notations, we denote $Z_i = g(X_i)$ in the following.

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

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

Assumptions on the model.

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

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

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 $z \in \mathcal{Z}$.*

Assumption 2 has to be connected with the bad performances of Beran's estimator in the tail of the distribution of Y. This assumption is present in Du and Akritas (2002). In Van Keilegom and Akritas (1999), this assumption is avoided only through the specific form of their scale-location regression model (2.14).

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 $[a(z), \tau_H] \subset]-\infty, +\infty[$, 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.

Assumption 3

- *Functions H_0 and H_1 have four derivatives with respect to z. Furthermore, these derivatives are uniformly bounded for $y < \tau$.*
- *Let $H_f(t|z) = H(t|z)f_Z(z)$, and let*

$$\tilde{\Phi}(s|z) = \frac{\psi_\tau(s,z)}{f_Z(z)^2}\frac{\partial^2 H_f(s|z)}{\partial z^2},$$

where $\psi_t(s,z) = \mathbf{1}_{s\leq t}[1 - H(s - |x)]^{-2}$. Assume that $\sup_z V(\tilde{\Phi}(\cdot|z)) < \infty$, where $V(g(\cdot))$ denotes the variation norm of a function g defined on $]-\infty;\tau]$.

Assumptions on the kernel.

Assumption 4 *The kernel K is a symmetric probability density function with compact support, and K is twice continuously differentiable.*

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 5 *The class* $\delta[1 - G(T - |X|)]^{-1} \times \mathcal{F}$ *is P-Glivenko-Cantelli (cf. Van der Vaart and Wellner, 1996, page 81) and has an integrable envelope* Φ *satisfying* $\Phi(t, z) = 0$ *for* $t \geq \tau$, *for some* τ *as defined in Assumption 1.*

For asymptotic normality, we will need more restrictions on the class $\mathcal{F}$. First, we need some additional moment condition on the envelope, and some differentiability with respect to z.

Assumption 6 *Assume that*

- *The envelope* Φ *is square integrable and* $\sup_{z\in\mathcal{Z}_\delta} \int \Phi(x, y)d|\partial_z^2 F_z(x, y)| < \infty$, *where* $F_z(x, y) = \mathbb{P}(X \leq x, Y \leq y|Z = z)$.
- *We have* $\Phi(x, t) = 0$ *for* $t \geq \tau$ *or* $g(x) \in \mathcal{Z} - \mathcal{Z}_\delta$, *for some* τ *as defined in Assumption 1, and* $\mathcal{Z}_\delta$ *the set of all points of* $\mathcal{Z}$ *at distance at least* δ, *for some* $\delta \geq 0$.
- *Let* $K_\phi(z, t) = \int_{x,y\leq t} \phi(x, y)dF_z(x, y)$. *The functions* K_ϕ *are twice differentiable with respect to* z, *and*

$$\sup_{\phi\in\mathcal{F},y\leq\tau,x\in\mathcal{Z}} |\partial_z K_\phi(z, y)| + |\partial_z^2 K_\phi(z, y)| < \infty.$$

The reason for introducing the set $\mathcal{Z}_\delta$ is to prevent us from some boundary effects which happen while obtaining uniform convergence rate for kernel estimators, see the proof of our Theorem 3.5 and Theorem 3.6 below. The order of the bias terms corresponding to our kernel estimators will be $O(h^2)$ if we restrain ourselves to points in $\mathcal{Z}_\delta$, while considering the boundaries leads to increase the order of the bias. However, δ can be taken as small as required.

For the sake of simplicity, we restrain ourselves to a class of smooth functions with respect to z (but not necessarily smooth with respect to y). But our result can easily be generalized to the case of functions with a finite number of discontinuities. Indeed, for some fixed $K \geq 0$, let $(I_i)_{1\leq i\leq K}$ be a sequence of subsets of $\mathcal{Z}_\delta$, and define $\phi(X, Y) = \sum_{i=1}^K \phi_i(X, Y)\mathbf{1}_{g(X)\in I_i}$, where $I_i \subset \mathcal{Z}_\delta$, and $\phi_i \in \mathcal{F}_i$, where $\mathcal{F}_i$ satisfies the same assumptions as the class $\mathcal{F}$.

We also need an assumption on the complexity of the class of functions. We consider two cases, the case where $\mathcal{F}$ is related to some Donsker class of functions, and the more restrictive assumption on its uniform covering number under which we obtain a faster rate for the remainder term in the asymptotic expansion. Let $N(\varepsilon, \mathcal{F}, L^2)$ denote the covering number (cf. Van der Vaart and Wellner, 1996 page 83) of the class $\mathcal{F}$ relatively to the L^2-norm.

Assumption 7 *Let C_1 be the set of all functions from $]-\infty;\tau]$ to with variation bounded by M. Let C_2 be the set of the functions f from $\mathcal{Z}_\delta$ to $[-M,M]$ with twice continuous derivatives bounded by M, for some $M \geq 0$, with $\sup_{z,z'\in\mathcal{Z}_\delta} |f''(z)-f''(z')||z-z'|^{-\eta} \leq M$ for some $0<\eta<1$. Let $\mathcal{G}=\{(z,t)\rightarrow g(z,t) : \forall z \in \mathcal{Z}_\delta, g(z,\cdot)\in C_1, \partial_z g(z,\cdot) \in C_1 \forall t \leq \tau, g(\cdot,t)\in C_2\}$. Assume that $(z,t)\rightarrow G(t|z)\in\mathcal{G}$ and that the class of functions $\mathcal{F}\mathcal{G}$ is Donsker.*

It is natural to assume that, to obtain our uniform CLT, the class $\mathcal{F}$ has to be Donsker. Indeed, by definition of a Donsker class, the uniform CLT will not hold in absence of censoring if $\mathcal{F}$ is not Donsker. Our Assumption 7 is a little bit more restrictive, but is close to optimality. Indeed, the class of functions $\mathcal{G}$ contains the functions G and $\hat{G}$, seen as functions of two variables. As pointed out in Lemma 6.1, the bracketing number of this class can be obtained, which shows that $\mathcal{G}$ is a uniformly bounded Donsker class. Hence, Assumption 7 will be fulfilled in a large number of situations, for example if $\mathcal{F}$ is an uniformly bounded Donsker class (see Van der Vaart and Wellner, 1996 page 192 for additional permanence properties of Donsker classes), or by elementary arguments if the bracketing number of $\mathcal{F}$ can be computed.

We now give a more restrictive assumption under which the convergence of our estimators can be improved.

Assumption 8 *For all $\varepsilon>0$, and for a class of functions $\mathcal{C}$, define $N(\varepsilon,\mathcal{C},L^r(\mathbb{P}))$ be the smallest number of balls with respect to the norm $\|\cdot\|_{2,\mathbb{P}}$ in $L^2(\mathbb{P})$. We say that a class of function $\mathcal{C}$ with a square integrable envelope C is a VC-class of functions if $\sup_{\mathbb{P}} N(\varepsilon\|C\|_{2,\mathbb{P}},\mathcal{C},L^2(\mathbb{P})) \leq A\varepsilon^{-V}$ for some A and $V>0$, where the supremum is taken on all probability measures such as $\|C\|_{2,\mathbb{P}}<\infty$. Assume that $\mathcal{F}$ is VC.*

3.2 Uniform Strong Law of Large Numbers

Theorem 3.1 *Under Assumptions 1 to 5, and with $h\rightarrow 0$, and $nh/[\log n]^{1/2}\rightarrow\infty$,*

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

Remark 3.2 *In all the proofs below, we will assume that we are on the set $A=\{\hat{f}_Z(z)\neq 0, \hat{H}(t|z)\neq 1, \hat{G}(t|z)\neq 1, \forall t\leq\tau, z\in\mathcal{Z}\}$. This allows us not to discuss any problem caused* *by the presence of these functions at the denominator. From the uniform convergence of nonparametric estimators (see Einmahl and Mason, 2005, and Van Keilegom and Akritas, 1999) and our Assumptions 1, 2, and 5, there exists some n_0 (almost surely) such as we are on the set A for all $n \geq n_0$. This is sufficient for deriving our asymptotic representations.*

Proof. Write, from the definition (2.13) 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 $\int \phi\left(x, y\right) dF\left(x, y\right)$ (uniformly over $\mathcal{F}$ from Assumption 5), while, for the second,

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

The empirical sum converges almost surely, while the supremum tends to zero almost surely from Proposition 4.3 in Van Keilegom and Akritas (1999). ■

3.3 An i.i.d. representation of the conditional Kaplan-Meier estimator

Theorem 3.3 *Let*

$$\begin{aligned} \eta_{\Lambda,i}(t, z) &= w_{in}(z)\left(\frac{[1 - \delta_i]\mathbf{1}_{T_i \leq t}}{1 - H(T_i - |z)} - \int_{-\infty}^{t} \frac{\mathbf{1}_{T_i \geq s} dH_0(s|z)}{[1 - H(T_i - |z)]^2}\right), \\ \eta_{G,i}(t, z) &= w_{in}(z)\left(\frac{[1 - \delta_i][1 - G(Y_i - |z)]\mathbf{1}_{Y_i \leq t}}{[1 - G(Y_i)][1 - H(T_i|z)]} - \int_{-\infty}^{T_i} \frac{[1 - G(s - |z)]\mathbf{1}_{T_i \geq s} dH_0(s|z)}{[1 - G(s|z)][1 - H(T_i - |z)]^2}\right). \end{aligned}$$

Under Assumptions 1 to 4, for h such as $nh^4 \to 0$,

$$\hat{\Lambda}_G(t|z) - \Lambda_G(t|z) = \frac{1}{n}\sum_{i=1}^{n} \eta_{\Lambda,i}(t, z) + R_\Lambda(t, z), \tag{3.2}$$

$$\hat{G}(t|z) - G(t|z) = \frac{1}{n}\sum_{i=1}^{n} \eta_{G,i}(t, z) + R_G(t, z), \tag{3.3}$$

with $\sup_{t \leq \tau, z \in \mathcal{Z}_\delta} |R_G(t, z)| = O_P([\log n] n^{-1} h^{-1})$.

The proof of this result follows the lines of Du and Akritas (2002), but using our Lemma 3.4 below which replaces their Proposition 4.1. The difference between the results of Du and Akritas (2002) and the present one only comes from the faster convergence rate for the remainder term, but the main terms of the expansion are rigorously the same. The proof of Lemma 3.4 is technical and is postponed to the Appendix section.

Lemma 3.4 *Let* $\psi_t(s,z) = \mathbf{1}_{s\leq t}[1-H(s-|z)]^{-2}$, *and* $\Psi_t(s,z) = \psi_t(s,z)f_Z(z)^{-2}$, *for* $t\leq\tau$, *and assume that* $\sup_z V(\Psi_\tau(\cdot,z)) < \infty$. *Let*

$$I_n(\psi_t, z) = \int \psi_t(s,z)[\hat{H}(s-|z) - H(s|z)]d[\hat{H}_0 - H_0](s|z)$$

Under Assumptions 1 to 4, for h such as $nh^4 \to 0$,

$$\sup_{t\leq\tau, z\in\mathcal{Z}_\delta} |I_n(\psi_t, z)| = O_P([\log n]n^{-1}h^{-1}).$$

Proof of Theorem 3.3. As in the proof of Theorem 3.1 in Du and Akritas (2002), we first decompose $\hat{\Lambda}_G(t|x) - \Lambda_G(t|x)$ into

$$\begin{aligned}\hat{\Lambda}_G(t|z) - \Lambda_G(t|z) &= \frac{1}{n}\sum_{i=1}^n \eta_{\Lambda,i}(t,z) + \int_{-\infty}^t \frac{[\hat{H}(t-|z) - H(t-|z)]^2 d\hat{H}_0(t|z)}{[1-\hat{H}(t-|z)][1-H(t-|z)]^2} \\ &\quad + \int_{-\infty}^t \frac{\hat{H}(t-|z) - H(t-|z)}{[1-H(t-|z)]^2} d\{\hat{H}_0(t|z) - H_0(t|z)\}.\end{aligned}$$

The second term is of rate $O_P([\log n]n^{-1}h^{-1})$ uniformly over $(t,z)\in[-\infty;\tau]\times\mathcal{Z}_\delta$, from Lemma 4.2 in Du and Akritas (2002). On the other hand, the last term is covered by Lemma 3.4, so that $\sup_{(t,z)\in[-\infty;\tau]\times\mathcal{Z}_\delta} |R_\Lambda(t,z)| = O_P([\log n]n^{-1}h^{-1})$, which proves (3.2). To prove (3.3), we follow the proof Theorem 3.2 in Du and Akritas (2002), but using our Lemma 3.4 instead of their Proposition 4.1. Using a representation of conditional Kaplan-Meier estimator which is the conditional version of Lemma 2.4 in Gill (1983) and the decomposition proposed in the proof of Theorem 3.2 in Du and Akritas (2002), we get

$$\hat{G}(t|z) - G(t|z) = [1-G(t|z)]\int_{-\infty}^t \frac{[1-G(t-|z)]}{[1-G(t|z)]} d(\hat{\Lambda}_G(t|z) - \Lambda_G(t|z)) + R_G(t|z),$$

with $\sup_{(t,z)\in[-\infty;\tau]\times\mathcal{Z}_\delta} R_G(t|z) = O_P([\log n]n^{-1}h^{-1})$. Then (3.3) follows from integration by parts and (3.2). ∎

3.4 Uniform Central Limit Theorem

Theorem 3.1 is not sufficient when it comes to proving asymptotic normality of integrals of type (2.13). As in the case of Kaplan-Meier integrals (see Stute, 1996), the i.i.d. expansion introduces an additional term if we need a remainder term decreasing to zero at a sufficiently fast rate. The representation of Du and Akritas (2002) which has been recalled in Theorem 3.3 can be rewritten, analogously to the expansion of Kaplan-Meier's estimator in absence of covariates,

$$\frac{\hat{G}(t|z)-G(t|z)}{1-G(t|z)} = \int_{-\infty}^{t} \frac{dM_{n,z}(y)}{[1-G(y|z)][1-F(y-|z)]} + R_n^G(z,t), \tag{3.4}$$

where

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

As a by-product of representation (3.4), the weak convergence of the process $[\hat{G}(t|z) - G(t|z)][1-G(t|z)]^{-1}$ can be obtained, retrieving the results of Theorem 1 in Dabrowska (1992). The rate of convergence is slower than $n^{-1/2}$ as for usual kernel estimators. The representation (3.4) can be compared to the representation derived by Stute (1995) and Akritas (2000) for the Kaplan-Meier estimator. Indeed, (3.4) is similar to equation (13) in Theorem 6 of Akritas (2000), except for the presence of the kernel weights, and of conditional distribution functions. On the other hand, a crucial difference with the asymptotic representation of Kaplan-Meier estimator comes from the fact that the integral in (3.4) does not have zero expectation. Indeed, $M_{n,z}$ is not a martingale with respect to the natural filtration $\mathcal{H}_t = \sigma(\{X_i\mathbf{1}_{T_i\leq t}, T_i\mathbf{1}_{T_i\leq t}, \delta_i\mathbf{1}_{T_i\leq t}, i=1,...,n\})$, since it is biased. In fact, we have, for all t,

$$E[\eta_{G,i}(Z_i,t)|Z_i] = 0, \tag{3.5}$$

but $E[\eta_{i,G}(z,t)] \neq 0$. However, there exists a martingale which is be naturally related to the asymptotics of $\hat{F}$, as shown in our Theorems 3.5 and Theorem 3.6 below. Define

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

which can be seen as a corrected version of $M_{n,z}$.

Theorem 3.5 *Under Assumptions 1 to 7, assume that* $h \in \mathcal{H}_\eta = [h_{\min}, h_{\max}]$, *where* $[\log n]^{-1}nh_{\min}^{3+\eta} \to \infty$, *and* $nh_{\max}^4 \to 0$,

$$\int \phi(x,y)\, d(\hat{F}-\tilde{F})(x,y) = \frac{1}{n}\sum_{i=1}^{n} \Psi_i(\phi) + R_n(\phi),$$

with

$$\Psi_i(\phi) = \int \frac{\bar{\phi}_{Z_i}(s)dM_i(s)}{[1-F(s- \mid Z_i)][1-G(s \mid Z_i)]},$$
$$\bar{\phi}_z(s) = \int_{\mathcal{Z}\times\mathbb{R}} \mathbf{1}_{s\leq y}\phi(x,y)dF_z(x,y),$$

and where $\sup_{\phi\in\mathcal{F}}|R_n(\phi)| = o_P(n^{-1/2})$.

Proof. First, using the uniform convergence rate of $\hat{G}$, observe that

$$\int \phi(x,y)d[\hat{F}(x,y)-\tilde{F}(x,y)] = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i\phi(X_i,T_i)[\hat{G}(T_i-|Z_i)-G(T_i-|Z_i)]}{[1-G(T_i-|Z_i)]^2} + O_P([\log n]n^{-1}h^{-1}).$$

The function $\delta\phi(X,T)G(T-|Z)[1-G(T-|Z)]^{-2}$ belongs to a Donsker class $\mathcal{C}$, from Assumption 7 and a permanence property of Donsker class (see Example 2.10.10 in Van der Vaart and Wellner, 1996). It is also the case for $\delta\phi(X,T)\hat{G}(T-|Z)[1-G(T-|Z)]^{-2}$ with probability tending to one from Lemma 6.2. Using the asymptotic equicontinuity of Donsker classes (see e.g. Van der Vaart and Wellner, 1996), we get, from Assumption 7 that

$$\int \phi(x,y)d[\hat{F}(x,y)-\tilde{F}(x,y)] = \int \frac{\phi(x,y)[\hat{G}(y-|g(x))-G(y-|g(x))]dF(x,y)}{1-G(t-|g(x))} + R_{F,\phi}(t,x),$$

where $\sup_{(t,x)\in[-\infty;\tau]\times\mathcal{Z},\phi\in\mathcal{F}}|R_{F,\phi}(t,x)| = o_P(n^{-1/2})$. Now using the i.i.d. representation of $\hat{G}$ from Theorem 3.3, it follows that

$$\int \frac{\phi(x,y)[\hat{G}(y-|g(x))-G(y-|g(x))]dF(x,y)}{1-G(t-|g(x))} =$$
$$\sum_{i=1}^n \int \frac{K\left(\frac{Z_i-g(x)}{h}\right)}{\sum_{j=1}^n K\left(\frac{Z_j-g(x)}{h}\right)}\phi(x,y)\eta_{G,i}(g(x),y-)dF(x,y) + R_\phi,$$

where $\sup_{\phi\in\mathcal{F}}|R_\phi| \leq E[|\Phi|]\times O_P(n^{-1}h^{-1})$. Let $\hat{f}_Z(z) = n^{-1}h^{-1}\sum_{j=1}^n K([Z_j-z]h^{-1})$. The main term can be decomposed into three parts, that is

$$\frac{1}{nh}\sum_{i=1}^n \int K\left(\frac{Z_i-g(x)}{h}\right)\frac{\phi(x,y)\eta_{G,i}(x,y-)}{f_Z(g(x))}dF(x,y) \quad (3.6)$$
$$+\frac{1}{nh}\sum_{i=1}^n \int K\left(\frac{Z_i-g(x)}{h}\right)\frac{\phi(x,y)\eta_{G,i}(g(x),y-)[\hat{f}_Z(g(x))-f_Z(g(x))]}{f_Z(g(x))^2}dF(x,y) \quad (3.7)$$
$$+\frac{1}{nh}\sum_{i=1}^n \int K\left(\frac{Z_i-g(x)}{h}\right)\frac{\phi(x,y)\eta_{G,i}(g(x),y-)[\hat{f}_Z(g(x))-f_Z(g(x))]^2}{\hat{f}_Z(g(x))f_Z(g(x))^2}dF(x,y). \quad (3.8)$$

Using the uniform convergence rate of $\hat{f}_Z$ (see Einmahl and Mason 2005), (3.8) is easily seen to be $O_P([\log n]n^{-1}h^{-1})$ uniformly over $\phi \in \mathcal{F}$. On the other hand, (3.7) can be rewritten as $\int \phi(x,y)U_n(g(x),y)dF(x,y)$, where

$$\begin{aligned} U_n(z,y) &= \frac{1}{n^2h^2}\sum_{i,j} K\left(\frac{Z_i - z}{h}\right)\frac{\eta_{G,i}(z,y-)}{f_Z(z)^2}\left\{K\left(\frac{Z_j - z}{h}\right) - f_Z(z)\right\} \\ &= \left(\frac{1}{nh}\sum_{i=1}^n K\left(\frac{Z_i - z}{h}\right)\frac{\eta_{G,i}(z,y-)}{f_Z(z)^2}\right)\left(\frac{1}{nh}\sum_{j=1}^n K\left(\frac{Z_j - z}{h}\right) - f_Z(z)\right). \end{aligned}$$

The right parenthesis is $O_P(h^2 + [\log n]^{1/2}n^{-1/2}h^{-1/2})$ uniformly over $z \in \mathcal{Z}_\delta$ using the uniform rate of consistency of $\hat{f}_Z$. On the other hand, the first parenthesis is $E[\eta_{G,1}(z,t-)|Z_1 = z]f_Z(z)^{-1} + R(z,t)$, with $\sup_{z,t}|R(z,t)| = O_P(h^2 + [\log n]^{1/2}n^{-1/2}h^{-1/2})$ (see Einmahl and Mason, 2005). Moreover, as pointed out in equation (3.5), $E[\eta_{G,1}(z,t-)|Z_1 = z] = 0$. Hence (3.7) can be bounded by $E[\Phi(X,Y)] \times O_P(h^4 + [\log n]n^{-1}h^{-1})$. Now let $w_i(x,y) = \phi(x,y)\eta_{G,i}(g(x),y-)$. From Assumption 6, w_i is twice differentiable with respect to $z = g(x)$, and $\sup_{y\leq\tau, g(x)\in\mathcal{Z}_\delta}|\partial_z^2 w_i(x,y)| \leq M|\Phi(x,y) + \sup_{\phi\in\mathcal{F}, y\leq\tau, g(x)\in\mathcal{Z}_\delta}|\partial_z\phi(x,y) + \partial_z^2\phi(x,y)|$. A change of variables and a second order Taylor expansion, Assumption 4 and Assumption 6 lead to rewrite (3.6) as

$$\frac{1}{n}\sum_{i=1}^n \int \phi(X_i,y)\eta_{G,i}(g(X_i),y-)dF_{Z_i}(x,y) + O_P(h^2).$$

■

In Theorem 3.5, the remainder term is $o_P(n^{-1/2})$. However, one could be interested in a more precise rate for the remainder term. This can be obtained if we strengthen the assumptions on the class of functions $\mathcal{F}$, that is if we replace Assumption 7 by the more restrictive assumption 4.

Theorem 3.6 *Under Assumptions 1 to 6, and Assumption 8, assume that* $h \in \mathcal{H}_\eta = [h_{\min}, h_{\max}]$, *where* $[\log n]^{-1}nh_{\min}^{3+\eta} \to \infty$, *and* $nh_{\max}^4 \to 0$,

$$\begin{aligned} \int \phi(x,y)\, d(\hat{F} - \tilde{F})(x,y) &= \frac{1}{n}\sum_{i=1}^n \Psi_i(\phi) \\ &\quad + R_n(\phi), \end{aligned}$$

with $\Psi_i(\phi)$ *defined in Theorem 3.5 and* $\sup_{\phi\in\mathcal{F}}|R_n(\phi)| = O_P([\log n]n^{-1}h^{-1} + h^2)$.

Proof. Write

$$\int \phi(x,y)d[\hat{F}(x,y) - \tilde{F}(x,y)] = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i\phi(X_i,T_i)[\hat{G}(T_i - |Z_i) - G(T_i - |Z_i)]}{[1 - G(T_i - |Z_i)]^2} + R(\phi),$$

where $\sup_{\phi\in\mathcal{F}}|R(\phi)| \leq O_P(1)\times\sup_{t\leq\tau,z\in\mathcal{Z}_\delta}|\hat{G}(t-|z)-G(t-|z)|^2 = O_P([\log n]n^{-1}h^{-1})$. Using the i.i.d. representation of $\hat{G}$, this can be rewritten as

$$\frac{1}{n^2h}\sum_{i,j}K\left(\frac{Z_i-Z_j}{h}\right)\frac{\delta_i\phi(X_i,T_i)\eta_{G,j}(Z_i,T_i-)}{[1-G(T_i-|Z_i)]\hat{f}_Z(Z_i)}+\tilde{R}(\phi),$$

with $\sup_{\phi\in\mathcal{F}}|\tilde{R}(\phi)| = O_P([\log n]n^{-1}h^{-1})$. The main term can be decomposed into

$$\frac{1}{n^2h}\sum_{i,j}K\left(\frac{Z_i-Z_j}{h}\right)\frac{\delta_i\phi(X_i,T_i)\eta_{G,j}(Z_i,T_i-)}{[1-G(T_i-|Z_i)]f_Z(Z_i)} \tag{3.9}$$

$$+\frac{1}{n^2h}\sum_{i,j}K\left(\frac{Z_i-Z_j}{h}\right)\frac{\delta_i\phi(X_i,T_i)\eta_{G,j}(Z_i,T_i-)[\hat{f}_Z(Z_i)-f_Z(Z_i)]}{[1-G(T_i-|Z_i)]f_Z^2(Z_i)} \tag{3.10}$$

$$+\frac{1}{n^2h}\sum_{i,j}K\left(\frac{Z_i-Z_j}{h}\right)\frac{\delta_i\phi(X_i,T_i)\eta_{G,j}(Z_i,T_i-)[\hat{f}_Z(Z_i)-f_Z(Z_i)]^2}{[1-G(T_i-|Z_i)]\hat{f}_Z(Z_i)f_Z^2(Z_i)}. \tag{3.11}$$

Using the uniform rate of convergence of the kernel estimator of the density, it is easy to see that (3.11) is of rate $O_P([\log n]n^{-1}h^{-1})$. We now consider (3.9). First observe that the sum of the terms for $i=j$ is negligible. Consider the $U-$process

$$\begin{aligned} U(\phi,h) &= \frac{1}{n(n-1)}\sum_{i\neq j}K\left(\frac{Z_i-Z_j}{h}\right)\frac{\delta_i\phi(X_i,T_i)\eta_{G,j}(Z_i,T_i-)}{f_Z(Z_i)[1-G(T_i-|Z_i)]} \\ &= \frac{1}{n(n-1)}\sum_{i\neq j}\psi_\phi(X_i,T_i,\delta_i;X_j,T_j,\delta_j). \end{aligned}$$

The class of functions $\{z\to K([z-u]/h) : h\geq 0, u\in\mathcal{Z}_\delta\}$ is VC from Assumption 4 (see Pakes and Pollard, 1989). Hence, using Lemma A.1 in Einmahl and Mason (2000), the class of functions ψ_ϕ for $\phi\in\mathcal{F}$ is VC. the $U-$process $U(\phi,h)$ is indexed by a VC-class of functions. Denote by $P(\phi,h)$ the Hajek projection of $U(\phi,h)$. Defining $R(\phi,h) = U(\phi,h)-P(\phi,h)$, it follows from Sherman (1994) that $\sup_{\phi,h}|R(\phi,h)| = O_P(n^{-1})$. To compute the Hajek projection, observe that

$$\begin{aligned} E[U(\phi,h)] &= O_P(h^3), \\ \sup_{x,t,\delta}\left|E\left[K\left(\frac{Z_i-Z_j}{h}\right)\frac{\delta_i\phi(X_i,T_i)\eta_{G,j}(Z_i,T_i-)}{f_Z(Z_i)[1-G(T_i-|Z_i)]}|X_i,T_i,\delta_i=(x,t,\delta)\right]\right| &= O_P(h^3), \end{aligned}$$

which can be proved by similar arguments as those developed in the proof of Theorem 3.5 (change of variables and Taylor expansion, using the regularity Assumptions 6). Hence, we can write

$$\begin{aligned} (3.9) \quad = \quad & \frac{1}{nh}\sum_{j=1}^n E\left[K\left(\frac{Z_i-Z_j}{h}\right)\frac{\delta_i\phi(X_i,T_i)\eta_{G,j}(Z_i,T_i-)}{f_Z(Z_i)[1-G(T_i-|Z_i)]}|X_j,T_j,\delta_j\right] \\ & +R'(\phi), \end{aligned}$$

with $\sup_{\phi\in\mathcal{F}}|R'(\phi)| = O_P(h^2+[\log n]n^{-1}h^{-1})$. Using again the same arguments as in Theorem 3.5, the conditional expectations are equal to $\Psi_j(\phi)+\tilde{R}_j(\phi)$, where $\sup_{j,\phi}|\tilde{R}_j| = O_P(h^2)$. The negligibility of (3.10) can be obtained using similar arguments, but with a third order $U-$process. ■

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 (2.7). See e.g. Heuchenne and Van Keilegom (2007b). If we assume that θ_0 is the unique minimizer of

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

we can estimate θ_0 by

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

As a consequence of Theorem 3.1 and Theorem 3.5, the following proposition furnishes the asymptotics for $\hat{\theta}$.

Proposition 4.1 *Assume that the Assumption of Theorem 3.1 hold with $\mathcal{F} = \{x \to f(\theta, x), \theta \in \Theta\}$. 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 $\mathcal{F}' = \{x \to \nabla^2_\theta f(\theta, x), \theta \in \Theta\}$ satisfies the Assumptions for Theorem 3.1. We have, under Assumptions 1 to 6 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

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

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 development,

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

for some θ_{1n} between θ_0 and $\hat{\theta}$. Apply Theorem 3.1 to see that we have $\nabla_\theta^2 M_n(\theta_{1n})^{-1} \to \Omega^{-1} a.s.$, and Theorem 3.5 to obtain that $\nabla_\theta M_n(\theta_0) \Rightarrow \mathcal{N}(0, V)$. ■

4.2 Density estimation

In this section, we assume that Y has a Lebesgue density f that we wish to estimate. Estimation of the density of Y has received a lot interest in the case $Y \perp C$. See e.g. Mielniczuk (1986). This assumption may not hold in several practical situations. In such cases the estimator of Mielniczuk (1986) is biased. An alternative is to consider that we are under (2.7) or (2.16)-(2.17), where X represent some auxiliary variables which are observed. In this framework, our estimator $\hat{F}$ 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

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

Observe that, since $\tilde{K}$ has compact support, if we choose h_1 small enough, the integral in (4.3) is only on $\mathcal{Z}_\delta\times] - \infty; \tau]$ for some $\tau < \tau_H$. Let $\tilde{K}_{h_1,y} = \tilde{K}((y - .)h_1^{-1})$. As an immediate corollary of Theorem 3.6, deduce that

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

with

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

5 Conclusion

We developed a new estimator for the distribution function $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ in the case where the variable Y is randomly right-censored, with the censoring variable allowed to depend from the covariates X. Our estimator can be seen as a generalized version of the empirical distribution function, and we provided asymptotic expansions uniform over quite general classes of functions. These representations allows us to obtain uniform law of large numbers and uniform central limit theorem for integrals with respect to the estimator we propose. Moreover, we indicated how this estimator can be slightly modified in order to be used in the case where X is multivariate. This becomes achievable by reinforcing the identifiability assumption of the regression model, by assuming that the censoring depends from X only through a univariate term $g(X)$. This property seems quite interesting, in the sense that most result on regression in this framework were provided only in the case where $X \in \mathbb{R}$. Estimation of the function g from the data is the next step for improving this technique, and will be investigated elsewhere.

6 Technicalities

6.1 Proof of Lemma 3.4

Let us introduce some notations.

$$\hat{H}_{0,f}(u|z) = \hat{H}_0(u|z)\hat{f}_Z(z), \qquad H_{0,f}(u|z) = H_0(u|z)f_Z(z),$$
$$\hat{H}_f(u|z) = \hat{H}(u-|z)\hat{f}_Z(z), \qquad H_f(u|z) = H(u-|z)f_Z(z),$$

where

$$\hat{f}_Z(z) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{Z_i - z}{h}\right).$$

We shall decompose

$$\begin{aligned} \hat{H}(u|z) - H(u|z) &= \frac{\hat{H}_f(u|z) - H_f(u|z)}{f_Z(z)} + \frac{[f_Z(z) - \hat{f}_Z(z)][\hat{H}_f(u|z) - H_f(u|z)]}{f_Z(z)\hat{f}_Z(z)} \\ &\quad + \frac{[f_Z(z) - \hat{f}_Z(z)]H_f(u|z)}{f_Z(z)^2} + \frac{[f_Z(z) - \hat{f}_Z(z)]^2 H_f(u|z)}{\hat{f}_Z(z)f_Z(z)^2}, \end{aligned}$$

and $\hat{H}_0(u|z) - H_0(u|z)$ similarly. $I_n(\psi_t, z)$ can then be decomposed into 16 terms, of which the following ones are the hardest to bound,

$$\int \frac{\psi_t(u,z)}{f_Z(z)^2}[\hat{H}_f(u|z) - H_f(u|z)]d[\hat{H}_{0,f}(u|z) - H_{0,f}(u|z)], \tag{6.1}$$

$$\int \frac{\psi_t(u,z)H_f(u|z)}{f_Z(z)^3}[f_Z(z) - \hat{f}_Z(z)]d[\hat{H}_{0,f}(u|z) - H_{0,f}(u|z)]. \tag{6.2}$$

For the others, it suffices to observe that $\sup_{z\in\mathcal{Z}_\delta} |[\hat{f}_Z(z) - f_Z(z)]\hat{f}_Z(z)^{-1}| = O_P([\log n]^{1/2}n^{-1/2}h^{-1/2})$, $\sup_{u,z\in\mathcal{Z}_\delta} |\hat{H}_f(u|z) - H_f(u|z)| = O_P([\log n]^{1/2}n^{-1/2}h^{-1/2})$ (see Einmahl and Mason, 2005). Consider the term in (6.1). Let $\tilde{H}_f(u|z) = E[\hat{H}_f(u|z)]$, and $\tilde{H}_{0,f}(u|z) = E[\hat{H}_{0,f}(u|z)]$. Decompose (6.1) into

$$\begin{aligned}
&\int \frac{\psi_t(u,z)}{f_Z(z)^2}[\tilde{H}_f(u|z) - H_f(u|z)]d[\hat{H}_{0,f}(u|z) - H_{0,f}(u|z)] \\
&+ \int \frac{\psi_t(u,z)}{f_Z(z)^2}[\hat{H}_f(u|z) - \tilde{H}_f(u|z)]d[\hat{H}_{0,f}(u|z) - H_{0,f}(u|z)] \\
&\qquad = A_1(t,z) + A_2(t,z).
\end{aligned}$$

To study $A_1(t,z)$, by Taylor expansion,

$$\tilde{H}_f(u|z) - H_f(u|z) = \frac{h^2 \int v^2 K(v)dv}{2}\frac{\partial^2 H_f(u|z)}{\partial z^2} + h^4 \int \frac{v^4 K(v)}{24}\frac{\partial^4 H_f(u|\tilde{z}(v,u))}{\partial z^4}dv. \tag{6.3}$$

A similar expansion holds for $\tilde{H}_{0,f}(u|z) - H_{0,f}(u|z)$. Since $\int d[\hat{H}_{0,f}(u|z) + H_{0,f}(u|z)] \leq 2$,

$$\sup_{z,t}\left|\int \frac{\psi_t(u,z)}{f_Z(z)^2}\frac{\partial^4 H_f(u|\tilde{z}(v,u))}{\partial z^4}d[\hat{H}_{0,f}(u|z) - H_{0,f}(u|z)]\right| \leq M < \infty.$$

Next we have to study

$$\begin{aligned}
\int_{-\infty}^t \tilde{\Phi}(u|z)d[\hat{H}_{0,f}(u|z) - H_{0,f}(u|z)] &= \int_{-\infty}^t \tilde{\Phi}(u|z)d[\hat{H}_{0,f}(u|z) - \tilde{H}_{0,f}(u|z)] \\
&+ \int_{-\infty}^t \tilde{\Phi}(u|z)d[\tilde{H}_{0,f}(u|z) - H_{0,f}(u|z)] \\
&= A_{11}(s,z) + A_{12}(s,z),
\end{aligned}$$

where

$$\tilde{\Phi}(u|z) = \frac{\psi_t(u,z)}{f_Z(z)^2}\frac{\partial^2 H_f(u|z)}{\partial z^2}.$$

By integration by parts, and expansion like in (6.3), $A_{12}(t,z)$ can be uniformly bounded by

$$\sup_{u,z}\left|[\tilde{H}_{0,f}(t|z) - H_{0,f}(t|z)]\tilde{\Phi}(t|z) - \int_{-\infty}^t [\tilde{H}_{0,f}(u|z) - H_{0,f}(u|z)]d\tilde{\Phi}(u|z)\right| = O(h^2). \tag{6.4}$$

On the other hand, $A_{11}(t,z)$ can be rewritten as

$$A_{11}(t,z) = \frac{1}{nh}\sum_{i=1}^{n}\left\{(1-\delta_i)\mathbf{1}_{T_i\le t}\Phi(T_i|z)K\left(\frac{Z_i-z}{h}\right) - E\left[(1-\delta_i)\mathbf{1}_{T_i\le t}\Phi(T_i|z)K\left(\frac{Z_i-z}{h}\right)\right]\right\}.$$

By Theorem 4 in Einmahl and Mason (2005), $\sup_{t,z}|A_{11}(t,z)| = O_P([\log n]^{1/2}n^{-1/2}h^{-1/2})$. Gathering the results, deduce that $\sup_{t,z}|A_1(t,z)| = O_P([\log n]n^{-1}h^{-1})$.

To study $A_2(t,z)$, decompose into

$$\begin{aligned}
&\int \frac{\psi_t(u,z)}{f_Z(z)^2}[\hat{H}_f(u|z)-\tilde{H}_f(u|z)]d[\tilde{H}_{0,f}(u|z)-H_{0,f}(u|z)] \\
&+\int \frac{\psi_t(u,z)}{f_Z(z)^2}[\hat{H}_f(u|z)-\tilde{H}_f(u|z)]d[\hat{H}_{0,f}(u|z)-\tilde{H}_{0,f}(u|z)] \\
&\qquad = A_{21}(t,z)+A_{22}(t,z).
\end{aligned}$$

By integration by parts,

$$\begin{aligned}
A_{21}(t,z) &= \Psi_t(t,z)[\hat{H}_f(t|z)-\tilde{H}_f(t|z)][\tilde{H}_{0,f}(t|z)-H_{0,f}(t|z)] \\
&\quad -\int_{-\infty}^{t}[\hat{H}_f(u|z)-\tilde{H}_f(u|z)][\tilde{H}_{0,f}(u|z)-H_{0,f}(u|z)]d\Psi_t(u,z) \\
&\quad -\int_{-\infty}^{t}\Psi_t(u,z)[\tilde{H}_{0,f}(u|z)-H_{0,f}(u|z)]d[\hat{H}_f(u|z)-\tilde{H}_f(u|z)],
\end{aligned}$$

where $\Psi_t(u,z) = \psi_t(u,z)f_Z(z)^{-2}$. The first two terms are uniformly bounded by $O_P([\log n]n^{-1}h^{-1})$. Using the analog of (6.3) for $\tilde{H}_{0,f}-H_{0,f}$, the third term in the decomposition of $A_{21}(t,z)$ can be rewritten as

$$\begin{aligned}
&\frac{h^2\int v^2K(v)dv}{nh_1}\sum_{i=1}^{n}\left(\Psi_t(T_i,z)\frac{\partial^2 H_{0,f}(T_i|z)}{\partial z^2}K\left(\frac{Z_i-z}{h}\right)\mathbf{1}_{T_i\le t}\right. \\
&\left.-E\left[\Psi_t(T,z)\frac{\partial^2 H_{0,f}(T|z)}{\partial z^2}K\left(\frac{Z-z}{h}\right)\mathbf{1}_{T\le t}\right]\right)+h^4R_n(t,z),
\end{aligned}$$

where $\sup_{t,z}|R_n(t,z)| = O_P(1)$. Using again Theorem 4 in Einmahl and Mason (2005), the main term is seen to be $O_P([\log n]^{1/2}h^{5/2}n^{-1/2})$ uniformly in t and z, which leads to $\sup_{t,z}|A_{21}(t,z)| = O_P([\log n]n^{-1}h^{-1})$.

The last term to study is $A_{22}(t,z)$ which can be rewritten as

$$\frac{1}{n^2h_1^2}\sum_{i\neq j}\xi_{z,t,h}(W_i,W_j)+\frac{1}{nh_1}\left(\frac{1}{nh_1}\sum_{i=1}^{n}\xi_{z,t,h_1}(W_i,W_i)\right) = U_{n1}(z,t,h)+U_{n2}(z,t,h),$$

where $W_i = (Z_i, T_i, \delta_i)$, and

$$\begin{aligned}\xi_{x,t,h}(w_1, w_2) &= (1-\delta_1)\Psi_t(t_1, z)\mathbf{1}_{t_1\leq s}K\left(\frac{z_1 - z}{h}\right)\Gamma_{z,h}(z_2, t_2, t_1)\\ &\quad - \int \mathbf{1}_{u\leq t}(1-\delta)\Psi_t(u, z)K\left(\frac{v-z}{h}\right)\Gamma_{z,h_1}(z_2, t_2, u)d\mathbb{P}_W(v, u, \delta),\end{aligned}$$

where $\mathbb{P}_W$ denotes the law of W_i, and where

$$\Gamma_{z,h_1}(z_2, t_2, u) = \mathbf{1}_{t_2\leq u}K\left(\frac{z_2 - z}{h}\right) - \int \mathbf{1}_{t'\leq u}K\left(\frac{v-z}{h}\right)d\mathbb{P}_W(v, t', \delta).$$

By elementary calculus, we have

$$\sup_{t,x}|U_{n2}(z, t, h)| \leq C\frac{1}{nh}\left(\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{Z_i - z}{h}\right)\right) = O_P(n^{-1}h^{-1}).$$

On the other hand, $h^2U_{1n}(z, t, h_1)$ is a degenerate second order $U-$process indexed by a $VC-$class of bounded functions. Moreover,

$$\begin{aligned}E\left[\xi_{z,t,h}(W_1, W_2)^2\right] &\leq CE\left[K^2\left(\frac{Z_1 - z}{h}\right)\Gamma^2_{z,t,h}(Z_2, T_2, T_1)\right],\\ &\leq 2C\left(E\left[K^2\left(\frac{Z_1 - z}{h}\right)K^2\left(\frac{Z_2 - z}{h}\right)\right]\right.\\ &\quad \left.+E\left[K^2\left(\frac{Z_1 - z}{h}\right)\right]\left\{\int K\left(\frac{s-z}{h_1}\right)d\mathbb{P}_W(s, u, \delta)\right\}^2\right)\\ &\leq \tilde{C}h^2.\end{aligned}$$

Hence, by Theorem 2 of Major (2006),

$$\sup_{t,z}|U_{n1}(z, t, h)| = O_P([\log n]n^{-1}h^{-1}).$$

6.2 The class $\mathcal{G}$

In this section, we give some results on the bracketing number of the class of functions $\mathcal{G}$ defined in Assumption 7. Define the bracket $[f, g]$ as the set of functions h such as $f \leq h \leq g$, and, for all $\varepsilon > 0$, define the $\varepsilon-$bracketing number of a class of functions $\mathcal{C}$ in L^r, as the smallest number of brackets $[f_i, g_i]$ such as $\|f_i - g_i\|_r \leq \varepsilon$ needed to cover $\mathcal{C}$. This number is denoted by $N_{[]}(\varepsilon, \mathcal{C}, L^r)$. We recall that the class $\mathcal{C}$ will be Donsker as soon as $\int_0^1[\log N_{[]}(\varepsilon, \mathcal{C}, L^2)]^{1/2}d\varepsilon < \infty$ (see e.g. Van der Vaart and Wellner, 1996). We recall some bounds on the bracketing numbers of the classes C_1 and C_2 defined in Assumption 7, see Corollary 2.7.2 and Theorem 2.7.5 in Van der Vaart and Wellner (1996),

$$\begin{aligned}\log N_{[]}(\varepsilon, C_1, L^2) &\leq O(\varepsilon^{-1}),\\ \log N_{[]}(\varepsilon, C_2, L^\infty) &\leq O(\varepsilon^{-1}).\end{aligned}$$

Lemma 6.1 *Let $\mathcal{G}$ be the class of functions defined in Assumption 7. We have, for any $\varepsilon > 0$,*

$$\log N_{[]}(\varepsilon, \mathcal{G}) \leq O(\varepsilon^{-2/(1+\eta)}).$$

Proof. To shorten the notation, let $N_1(\varepsilon) = \log N_{[]}(\varepsilon, C_1, L^2)$, $N_2(\varepsilon) = \log N_{[]}(\varepsilon, C_2, L^\infty)$. Assume, without loss of generality, that $M = 1$, where M is the constant involved in Assumption 7. Let $\{[\psi_{1,j}, \psi_{2,j}]\}_{1 \leq j \leq N_1(\varepsilon)}$, and $\{[\phi_{1,k}, \phi_{2,k}]\}_{1 \leq k \leq N_1(\varepsilon^{1/(1+\eta)})}$ be respectively a set of $\varepsilon-$ and $\varepsilon^{1/(1+\eta)}-$brackets covering C_1. Let $z_i = i\varepsilon^{1/(1+\eta)}$, for $0 \leq i \leq \varepsilon^{-1/(1+\eta)}$. Define

$$\tilde{\psi}_{1,i,j,k}(z, y) \quad = \quad \psi_{1,j}(y) + (z - z_i)\phi_{1,k}(y).$$

Let f be a function defined on $\{0, ..., \lfloor \varepsilon^{-1/(1+\eta)} \rfloor + 1\}$ taking values in $\{1, ..., N_1(\varepsilon)\} \times \{1, ..., N_1(\varepsilon^{1/(1+\eta)})\}$, and define the lower bracket indexed by f,

$$\tilde{\psi}^f(z, y) \quad = \quad \tilde{\psi}_{1,i,f(i)}(z, y) \text{ for} z \in [z_i; z_{i+1}[.$$

As in the proof of Theorem 2.7.1 and Corollary 2.7.2 in Van der Vaart and Wellner, it is clear that for any $\chi \in \mathcal{G}$, there exists some f such as $f \leq \tilde{\psi}_1^f$ and $\chi - \tilde{\psi}_1^f \preceq \varepsilon$. Hence the number of functions f will provide the order of the bracketing number, that is

$$\exp(\varepsilon^{-1} + \varepsilon^{-1/(1+\eta)})^{\varepsilon^{-1/(1+\eta)}} = O(\exp(\varepsilon^{-2/(1+\eta)})).$$

■

As a consequence, $\mathcal{G}$ is a Donsker class of functions, since $0 < \eta < 1$. Lemma 6.2 below shows that, if G belongs to $\mathcal{G}$, $\hat{G}$ also belongs to $\mathcal{G}$ with probability tending to one (which allows to use the asymptotic equicontinuity argument of the proof of Theorem 3.5.

Lemma 6.2 *Let $Z = g(X)$ and $\hat{G}(t|z)$ denote Beran's estimator of $G(t|z) = \mathbb{P}(C \leq t|Z = z)$. Under the Assumptions of Theorem 3.5, $(t, z) \to \hat{G}(t|z) \in \mathcal{G}$ with probability tending to one if $h \in \mathcal{H}_\eta$.*

Proof. This Lemma follows directly from the uniform convergence of the $\hat{G}$ and its derivative towards G, see Proposition 4.3 in Akritas and Van Keilegom (1999), and from Proposition 4.4 in Van Keilegom and Akritas (1999) which concerns $|\partial_z \hat{G}(t|z) - \partial_z G(t|z')||z - z'|^{-\eta}$. ■

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] Dabrowska, D. M. (1992). Variable bandwidth conditional Kaplan-Meier estimate. *Scand. J. Stat.* **19**, 351–361.

[6] Delecroix, M., Lopez, O. & Patilea, V. (2008). Nonlinear Censored Regression Using Synthetic Data. *Scand. J. Stat.* **35**, 248–265.

[7] 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.

[8] 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.

[9] Einmahl, U. & Mason, D. M., (2005). Uniform in bandwidth consistency of kerneltype function estimators. *Ann. Statist.*, **33**, 1380–1403.

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

[11] Gonzalez-Manteiga, W. & Cadarso-Suarez, C. (1994). Asymptotic properties of generalized Kaplan-Meier estimator with some applications. *J. Nonparametric Statistics* **4**, 65–78.

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

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

[14] Horvath, L. & Yandell, B.S. (1988). Asymptotics of conditional empirical processes, *J. Multivariate Anal.* **26**, 184–206.

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

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

[17] Li, G. & Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting, *Ann. Statist.* **23**, 787–823.

[18] Li, G. (1997). Optimal rate local smoothing in a multiplicative intensity counting process model, *Math. Methods Statist.* **6**, 224–244.

[19] Linton, O.B. & Nielsen, J.P. (1995). Kernel estimation in a nonparametric marker dependent hazard model, *Ann. Statist.* **23**, 1735–1748.

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

[21] Lopez, O. & Patilea, V. (2009). Nonparametric lack-of-fit tests for parametric mean-regression models with censored data, *J. Multivariate Anal.* **100**, 210–230.

[22] Major, P. & Retjo, L. (1988). Strong embedding of the estimator of the distribution function under random censoring, *Ann. Statist.* **16**, 1113–1132.

[23] McKeague, I.W. & Utikal, K.J. (1990). Inference for a nonlinear counting process regression model, *Ann. Statist.* **18**, 1172–1187.

[24] Mielniczuk, J. (1986). Some asymptotic properties of kernel estimation of a density function in case of censored data. *Ann. Statist.* **14**, 766-Ũ773.

[25] Mielniczuk, J. (1987). A remark concerning strong uniform consistency of the conditional Kaplan-Meier estimator. *Statist. Probability Letters* **5**, 333–337.

[26] Mielniczuk, J. (1991). Some asymptotic properties of nonparametric regression estimators in case of randomly censored data. *Statistics* **22**, 85–93.

[27] Pakes, A. & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027Ű-1057.

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

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

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

[31] Stute, W. (1986). On almost sure convergence of conditional empirical distribution functions, *Ann. Probab.*, **14**, 891–901.

[32] Stute, W. (1993). Consistent estimation under random censorship when covariates are present. *J. Multiv. Analysis* **45**, 89–103.

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

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

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

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

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

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

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

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