T E C H N I C A L

R E P O R T

0408

UNIFORM REPRESENTATION OF PRODUCT-LIMIT INTEGRALS WITH APPLICATIONS

C. SANCHEZ SELLERO, W. GONZALEZ MANTEIGA and I. VAN KEILEGOM

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Uniform representation of product-limit integrals with applications

César Sánchez Sellero[1]

Universidad de Santiago de Compostela

Wenceslao González Manteiga[1]

Universidad de Santiago de Compostela

Ingrid Van Keilegom[2]

Université catholique de Louvain

March 3, 2004

Abstract

Let X be a d-variate random vector that is completely observed, and let Y be a random variable that is subject to right censoring and left truncation. For arbitrary functions φ we consider expectations of the form $E[\varphi(X,Y)]$, which appear in many statistical problems, and we estimate these expectations by using a product-limit estimator for censored and truncated data, extended to the context where covariates are present. An almost sure representation for these estimators is obtained, with a remainder term that is of a certain negligible order, uniformly over a class of φ-functions. This uniformity is important for the application to goodness-of-fit testing in regression and to inference for the regression depth, which we consider in more detail. A bootstrap approximation is proposed to approximate the distribution of the product-limit integrals.

Key Words: Bootstrap; Censoring; Goodness-of-fit; Nonparametric regression; Regression depth; Resampling; Truncation.

[1] Financial support from the Spanish Ministry of Science and Technology (with additional European FEDER support) through project BFM2002-03213 and from Xunta de Galicia through project PGIDIT03PXIC20702PN are gratefully acknowledged.
[2] Financial support from the IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy) is gratefully acknowledged.

1 Introduction

Given a variable of interest Y, with distribution function F, integrals of the type

$$\int \varphi(y)\, F(dy)$$

where φ is an F-measurable function, appear quite often in statistical problems related to Y. That is the case for the common mean, which can be written as $E(Y) = \int y\, F(dy)$, or in general for moments of the variable of interest.

From a sample of independent observations of Y, $Y_1, \ldots, Y_n$, estimation of the integral $\int \varphi(y)\, F(dy)$ can be obtained by the sample mean $n^{-1} \sum_{i=1}^{n} \varphi(Y_i)$, which can be seen as an integral with respect to the empirical distribution function $F_n(y) = n^{-1} \sum_{i=1}^{n} I_{\{Y_i \leq y\}}$. Statistical properties of the sample mean, such as consistency or distributional convergence, can be obtained from classical results in probability theory, such as the law of large numbers or the central limit theory.

However, in many survival or reliability studies, observations of the variable of interest are not directly available, because the data are typically truncated and/or censored. In particular, left truncation and right censoring together naturally occur in cohort follow-up studies. In this case, instead of the variable of interest Y, one can observe the random vector (T, Z, δ) only if $T \leq Z$, where $Z = \min(Y, C)$ and $\delta = I(Y \leq C)$. Here T is the truncating variable and C is the censoring variable, which are assumed to be independent of Y.

From n independent observations $(T_i, Z_i, \delta_i),\ i \in \{1, \ldots, n\}$, the distribution function of Y can be estimated by the product-limit estimator

$$\hat{F}(y) = 1 - \prod_{\{i: Z_i \leq y, \delta_i = 1\}} \left(1 - \frac{1}{nC_n\left(Z_i\right)}\right)$$

where $C_n(y) = n^{-1} \sum_{i=1}^{n} I_{\{T_i \leq y \leq Z_i\}}$. This estimator was introduced by Tsai et al. (1987). Then, the integral $\int \varphi(y)\, F(dy)$ can be estimated by

$$\int \varphi(y)\, \hat{F}(dy)$$

which will be called the product-limit integral. This integral can be written as a weighted average of the Z-values:

$$\int \varphi(y)\, \hat{F}(dy) = \sum_{i=1}^{n} W_{in} \varphi\left(Z_i\right)$$

where $W_{in} = (\delta_i / (nC_n(Z_i))) \prod_{\{j: Z_j < Z_i, \delta_j = 1\}} (1 - 1/(nC_n(Z_j))),\ i \in \{1, \ldots, n\}$ are the product-limit weights.

When there is only censoring, it is common to consider the weights attached by the product-limit estimator to the Z-order statistics, because they have a simple structure, but when truncation is present this is no longer the case. For this reason, here we consider the weights attached to the usual Z-values.

In applications, a d-variate vector X will be typically recorded for each subject. In the context of survival in medical trials, this X informs about characteristics of the individuals, such as sex, age, blood measure, state of the disease under study at a given date, etc. In general, X is considered as a vector of observable covariables correlated with the variable of interest.

Then, the available information is given by $(X_i, T_i, Z_i, \delta_i),\ i \in \{1, \ldots, n\}$ and the main goal becomes the estimation of the joint distribution function of (X, Y), say $F_{X,Y}$. Under the assumption that T, C and (X, Y) are mutually independent, one has the consistent estimator

$$\hat{F}_{X,Y}(x, y) = \sum_{i=1}^{n} W_{in} I_{\{X_i \leq x, Z_i \leq y\}}$$

where W_{in} are the already-mentioned product-limit weights, and where $X_i \leq x$ stands for $X_{ij} \leq x_j$ for $j = 1, \ldots, d$. This estimator was introduced by Gürler and Gijbels (1996). Note that the weights are the same as in the case without covariables. In fact, this estimator reduces to the estimator studied by Stute (1996) in the case of only censoring.

Finally, the integral $\int \varphi(x, y)\, F_{X,Y}(dx, dy)$, which appears in statistical problems related to (X, Y), can be estimated by

$$\int \varphi(x, y)\, \hat{F}_{X,Y}(dx, dy) = \sum_{i=1}^{n} W_{in} \varphi\left(X_i, Z_i\right)$$

which will be called the product-limit integral (with covariables).

Numerous examples where integrals of this kind naturally appear, can be found in the context of linear regression (see Gross and Lai (1996)), or more recently in the context of regression depth (see Park and Hwang (2003)), associated to robust methodology. Applications to real data, as the Stanford heart transplant data and AIDS incubation data, can be found in the mentioned papers.

The main purpose of this paper is to provide an almost sure representation of integrals of this type, uniformly over a certain class of functions φ, and with a negligibility rate for the remainders that is fast enough for the common applications, like the goodness-of-fit test and the regression depth developed in Section 4. Such representation is given in Theorem 3.1. The uniformity of the representation over a certain class of functions is essential for the applications to hold and is in fact an important contribution of this paper.

Some representations related to the context of this paper can be found in the literature. Zhou and Yip (1999) provide a representation for the product-limit estimator itself under censoring and truncation and without covariables. Gürler and Gijbels (1996) obtain a similar representation with covariables. In all cases the order of neglibility of the remainder is a logarithmic function of n divided by n, almost surely, and the uniformity is established in an interval of the real line, which can be interpreted in terms of product-limit integrals as uniformity with respect to a family of indicators. In the case of censoring without truncation and with covariables, Stute (1996) obtained a representation for product-limit integrals for a fixed φ-function.

The paper is organised as follows. In the next section we provide some notation and some basic assumptions necessary for the rest of the paper. Section 3 contains the main result, that is, the asymptotic uniform representation of a family of product-limit integrals with covariables. In Section 4 some applications to goodness-of-fit tests and regression depth are discussed. Finally, the Appendix contains the proofs.

2 Basic assumptions and notations

The distribution functions of Y, T and C will be denoted by F, G and D, respectively. We assume throughout that F, G and D are continuous. For any one-dimensional distribution function L, let $a_L = \inf\{x : L(x) > 0\}$ denote the largest lower bound for the support of L, and $b_L = \sup\{x : L(x) < 1\}$ denote the smallest upper bound for the support of L.

Due to truncation, one can only observe data on the Z-values larger than $\max\{a_F, a_G\}$. Then, if $a_F < a_G$, there will never be observations at the intersection of the support of F with the interval (a_F, a_G). For this reason, an identifiability assumption will be

(a) $a_G \leq a_F$.

Similarly, due to censoring, one can only observe data on the Z-values smaller than $\min\{b_F, b_D\}$. Then, if $b_D < b_F$, the region (b_D, b_F) is not identifiable. For this reason, an additional identifiability assumption will be

(b) $b_F \leq b_D$.

We like to note that condition (a) (and similarly for (b)) is requiring slightly more than what we actually need for the results in this paper. In fact, we could instead assume that either $a_G \leq a_F$ or, in case $a_F < a_G$, $\varphi(x, y) = 0$ for $y \in (a_F, a_G)$. To keep the presentation clear, we prefer however to work under assumption (a).

Further, we assume throughout that

(c) (T, C) is independent of (X, Y), and C is independent of T.

When there are no covariables, the first assumption reduces to the common assumption that (T, C) and Y are independent. In the situation with covariables, there exist two common ways to extend this assumption : either one assumes condition (c) mentioned above (see e.g. Gürler and Gijbels (1996) and Gijbels and Gürler (1998)), or one assumes that (T, C) is independent of Y given X. Although the latter assumption is less restrictive, it requires the use of smoothing techniques to estimate the bivariate distribution of (X, Y), which can be avoided under assumption (c).

Finally, we assume that the probability of observing (T, Z, δ) is non-zero, i.e.

(d) $\alpha := P(T \leq Z) > 0$.

Denote

$$W^1(y) = P(Z \leq y, \delta = 1/T \leq Z),$$

$$W^1_{X,Z}(x, y) = P(X \leq x, Z \leq y, \delta = 1/T \leq Z),$$

$$C(y) = P(T \leq y \leq Z/T \leq Z),$$

and

$$\gamma_0(y) = \frac{1}{C(y)} \exp\left\{ - \int_{-\infty}^{y} \frac{1}{C(z)}\, W^1(dz) \right\},$$

and for each φ defined over $\mathbb{R}^{d+1}$ consider

$$\gamma_1^\varphi(y) = \frac{1}{C(y)} \int I_{\{y<\omega\}} \varphi(x, \omega) \gamma_0(\omega)\, W^1_{X,Z}(dx, d\omega),$$

$$\gamma_2^\varphi(t, z) = \int I_{\{v<\omega, t\leq v\leq z\}} \varphi(x, \omega) \gamma_0(\omega) \frac{1}{C^2(v)}\, W^1(dv)\, W^1_{X,Z}(dx, d\omega),$$

$$\gamma_3^\varphi(t, z) = \int I_{\{t\leq\omega\leq z\}} \varphi(x, \omega) \gamma_0(\omega) \frac{1}{C(\omega)}\, W^1_{X,Z}(dx, d\omega).$$

Finally consider

$$\eta_i^\varphi = \varphi\left(X_i, Z_i\right) \gamma_0\left(Z_i\right) \delta_i - \gamma_1^\varphi\left(Z_i\right) \delta_i + \gamma_2^\varphi\left(T_i, Z_i\right) - \gamma_3^\varphi\left(T_i, Z_i\right).$$

3 Main results

In this section an almost sure representation of product-limit integrals with covariables is given for a general class of functions $\{\varphi\}$. First, the assumptions needed for this result are stated. We introduce the notation $\Lambda(y) = P(T \le y \le C)$, and M will denote the distribution function of $Z = \min(Y, C)$. Note that $1 - M = (1 - F)(1 - D)$ by the independence of Y and C.

(φ 1) The family of functions $\{\varphi\}$ is a measurable VC-subgraph class of functions.

(φ 2) The family of functions $\{\varphi\}$ admits an envelope Φ, such that

$$E\left[\Phi(X,Y)\Lambda^{-2}(Y)(1-F(Y))^{-5}\right] < \infty,$$
$$E\left[\Phi^2(X,Y)\Lambda^{-2}(Y)(1-F(Y))^{-3}\right] < \infty.$$

(Λ) $E\left[\Lambda^{-2}(Y)\right] < \infty$.

Note that condition (Λ) is satisfied when $a_G < a_F$ and $b_F < b_D$.

(We refer to van der Vaart and Wellner (1996) for the definition of VC-subgraphs (page 41) and envelope functions (page 84).)

Theorem 3.1 *Under (a), (b), (c), (d), (φ 1), (φ 2) and (Λ), we have*

$$\int \varphi \, d\hat{F}_{X,Y} - \int \varphi \, dF_{X,Y} = n^{-1} \sum_{i=1}^{n} \eta_i^{\varphi} + R_n(\varphi),$$

where

$$\sup_{\varphi} |R_n(\varphi)| = O\left(\frac{\ln^3 n}{n}\right) \qquad \textit{almost surely.}$$

Some straightfoward calculations show that the main term of this representation reduces to the one in Stute (1996) when no truncation is present and when one single φ-function is considered. Using this almost sure representation, one can easily obtain the uniform consistency and the weak convergence of the process of product limit integrals. See e.g. Section 4, where, for the application to goodness-of-fit testing, use is made of the classical theory of Billingsley (1968), and where, for the application to regression depth, modern Donsker-theory based on bracketing numbers is used (see e.g. van der Vaart and Wellner (1996)).

The variance-covariance function of the limiting Gaussian process might however be hard to estimate. In these occasions it will be useful to approximate the distribution of the product-limit integrals by means of the bootstrap. We accomplish this by applying a bootstrap procedure to the dominant terms of the above representation. To do this, the functions γ_0, γ_1^{φ}, γ_2^{φ} and γ_3^{φ} have to be estimated by their empirical analogues:

$$\gamma_{0n}(y) = \frac{1}{C_n(y)} \exp\left\{ -\int_{-\infty}^{y^-} \frac{1}{C_n(z)} \, W_n^1(dz) \right\},$$

$$\gamma_{1n}^{\varphi}(y) = \frac{1}{C_n(y)} \int I_{\{y<\omega\}} \varphi(x,\omega)\gamma_{0n}(\omega)\, W_{X,Z,n}^1(dx,d\omega),$$

$$\gamma_{2n}^{\varphi}(t,z) = \int I_{\{v<\omega, t\le v\le z\}} \varphi(x,\omega)\gamma_{0n}(\omega) \frac{1}{C_n^2(v)}\, W_n^1(dv)\, W_{X,Z,n}^1(dx,d\omega),$$

$$\gamma_{3n}^{\varphi}(t,z) = \int I_{\{t\le\omega\le z\}} \varphi(x,\omega)\gamma_{0n}(\omega) \frac{1}{C_n(\omega)}\, W_{X,Z,n}^1(dx,d\omega),$$

where $C_n(y) = n^{-1}\sum_{i=1}^n I_{\{T_i\le y\le Z_i\}}$, $W_n^1(y) = n^{-1}\sum_{i=1}^n I_{\{Z_i\le y,\delta_i=1\}}$ and $W_{X,Z,n}^1(y) = n^{-1}\sum_{i=1}^n I_{\{X_i\le x, Z_i\le y,\delta_i=1\}}$ are the empirical versions of the corresponding functions, based on the available data.

Let

$$\eta_{in}^{\varphi} = \varphi(X_i,Z_i)\,\gamma_{0n}(Z_i)\,\delta_i - \gamma_{1n}^{\varphi}(Z_i)\,\delta_i + \gamma_{2n}^{\varphi}(T_i,Z_i) - \gamma_{3n}^{\varphi}(T_i,Z_i)$$

be the empirical version of the expression η_i^{φ} in the representation. An approximation of the distribution of the product-limit integrals is then given by the bootstrap distribution of

$$n^{-1}\sum_{i=1}^n \eta_{in}^{\varphi} V_i^*$$

where $\{V_1^*,\ldots,V_n^*\}$ is a sample of i.i.d. random variables with zero mean and unit variance, being also independent of the original sample $\{(X_i,T_i,Z_i,\delta_i)\}_{i=1}^n$.

4 Applications

In this section we show how the uniform representation of the product-limit integrals obtained in Section 3, can be used to develop statistical methods with truncated and censored data and to derive their properties. In particular, the goodness-of-fit testing of regression models and the regression depth methodology are treated here.

4.1 Goodness-of-fit testing of regression models

Let $m(x) = E(Y/X = x)$ be the regression function of Y on X. In what follows we suppose for simplicity that X is one-dimensional. It is often reasonable to propose that the regression function belongs to a certain parametric family, that is, $m \in \mathcal{M} = \{m_\theta : \theta \in \Theta\}$ where $\Theta \subset R^p$ is the space of parameters. When $m \in \mathcal{M}$, one has $Y = m_{\theta_0}(X) + \varepsilon$, where θ_0 is the true parameter and ε is such that $E(\varepsilon/X) = 0$. Since there may be different competing models and to prevent wrong specification, it is necessary to test

$$H_0 : m \in \mathcal{M} \qquad \text{versus} \qquad H_1 : m \notin \mathcal{M}$$

This problem has received much attention during the last decade. One of the main approaches is based on the integrated regression function $I(x) = \int_{-\infty}^{x} m\,dF_X = E(YI_{\{X\le x\}})$, where F_X is the (marginal) distribution function of X. By a measure theoretic argument, for a given F_X, the function I uniquely determines m.

With an i.i.d. sample $\{(X_i, Y_i)\}_{i=1}^n$, I can be estimated by its empirical analogue $I_n(x) = n^{-1}\sum_{i=1}^n Y_i I_{\{X_i\le x\}}$ and also by $\hat{I}(x) = n^{-1}\sum_{i=1}^n m_{\theta_n}(X_i)I_{\{X_i\le x\}}$ under the null hypothesis, where θ_n is a convenient estimator of the parameter θ_0. Then, many tests can be constructed based on the process $D_n(x) = n^{1/2}(I_n(x) - \hat{I}(x)) = n^{-1/2}\sum_{i=1}^n (Y_i - m_{\theta_n}(X_i))I_{\{X_i\le x\}}$. For a complete presentation of this approach and its main properties we refer to Stute (1997). A Cramér-von Mises test (based on the L_2 norm of the process) or Kolmogorov-Smirnov test (based on the supremum norm) can be used to carry out the test in practice. Note that Stute et al. (2000) studied the practical behavior of an extension of these tests to right censored data (without truncation).

We will now show that the theory of product-limit integrals developed in Section 3 can be used to extend this testing procedure to the case of right censored and left truncated data. Because of lack of information in the right tail (which is typical for censored/truncated data problems), we are unable to test hypotheses concerning the function $m(x)$ itself, and we need to consider instead the function $m_b(x) = E(Y \le y|X = x, Y \le b)$ for $b < b_F$. Note that $m_b(x)$ can be made arbitrarily close to $m(x)$ by taking b close enough to b_F.

Consider for simplicity the simple null hypothesis $H_0 : m_b = m_0$, and define the following process :

$$D_n^0(x_0) = n^{1/2}\sum_{i=1}^n W_{in}(Z_i - m_0(X_i))\, I_{\{X_i\le x_0\}}I_{\{Z_i\le b\}}$$

($x_0 \in I\!R$), which can be expressed as a family of product-limit integrals

$$D_n^0(x_0) = n^{1/2}\int \varphi_{x_0}(x,y)\,\hat{F}_{X,Y}(dx,dy)$$

where $\varphi_{x_0}(x,y) = I_{\{x\le x_0\}}I_{\{y\le b\}}(y - m_0(x))$. In Theorem 4.1 below a uniform representation is obtained for the process D_n^0, as an application of Theorem 3.1 to the class of functions $\{\varphi_{x_0}\}_{x_0\in R}$. We will denote $\eta_i(x_0) = \eta_i^{\varphi_{x_0}}$. Let

(M) $E[(Y - m_0(X))^2\Lambda^{-2}(Y)I(Y \le b)] < \infty$.

Note that (M) is satisfied when $a_G < a_F$ and $E[(Y - m_0(X))^2] < \infty$.

Theorem 4.1 *Under (a), (b), (c), (d), (M) and (Λ), we have*

$$D_n^0(x_0) = n^{-1}\sum_{i=1}^n \eta_i(x_0) + R_n(x_0)$$

where

$$\sup_{x_0 \in R} |R_n(x_0)| = O\left(\frac{\ln^3 n}{n}\right) \qquad \textit{almost surely.}$$

For the distributional convergence of the process the assumption (Λ) needs to be replaced by the slightly stronger condition
$(\Lambda 3)$ $E\left[\Lambda^{-3}(Y)\right] < \infty$.

Theorem 4.2 *Under (a), (b), (c), (d), (M) and $(\Lambda 3)$, the process D_n^0 converges in distribution to a Gaussian process with mean zero and covariance function $K(x_0, x_1) = E[\eta(x_0)\eta(x_1)]$.*

4.2 Regression depth

A second application stems from the context of regression depth estimation. Let us start by explaining how the regression depth is defined for completely observed data. Suppose a random vector (X, Y) satisfies the model

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_d X_d + \varepsilon,$$

where $X = (X_1, \ldots, X_d)'$, $\beta = (\beta_0, \ldots, \beta_d)'$, and $\varepsilon = \varepsilon(\beta)$ is the error, and let (X_i, Y_i) be n i.i.d. observations from the same distribution as (X, Y). (Note that no assumptions are made on the distribution of the error). Rousseeuw and Hubert (1999) call the vector β a 'non-fit' if there exists an affine hyperplane V in the X-space such that no X_i belongs to V and such that $\varepsilon_i(\beta) \geq 0$ for all X_i in one of the open halfspaces and $\varepsilon_i(\beta) \leq 0$ for all X_i in the other halfspace. Figure 1 in Rousseeuw and Hubert (1999) gives a nice illustration of this concept when $d = 1$. The authors then define the (sample) regression depth of β by (modulo a factor n^{-1}) :

$$\text{rdepth}_n(\beta) = n^{-1} \min_V [\min\{L^+(V, \beta) + R^-(V, \beta), R^+(V, \beta) + L^-(V, \beta)\}],$$

where the minimum is taken over all hyperplanes V in the X-space,

$$L^+(V, \beta) = \sum_{j=1}^n I(X_j \in L(V), \varepsilon_j(\beta) \geq 0),$$

$$R^-(V, \beta) = \sum_{j=1}^n I(X_j \in R(V), \varepsilon_j(\beta) \leq 0),$$

$R^+(V, \beta)$ and $L^-(V, \beta)$ are defined analogously, and $L(V)$ and $R(V)$ denote the two halfspaces corresponding to the hyperplane V. In other words, the regression depth represents the smallest proportion of data points that need to be removed to make β a non-fit. The concept of regression depth is useful for defining the 'deepest regression line' (line with largest regression depth), which generalizes the univariate median.

Note that $\text{rdepth}_n(\beta)$ can be equivalently written as

$$\text{rdepth}_n(\beta) = \min_V \Big[\min\Big\{ \iint \varphi_{V,\beta,1}(x,y)\, F_n(dx,dy), \iint \varphi_{V,\beta,2}(x,y)\, F_n(dx,dy) \Big\} \Big],$$

where

$$F_n(x,y) = n^{-1} \sum_{i=1}^n I(X_i \le x, Y_i \le y),$$

$$\varphi_{V,\beta,1}(x,y) = L^+_{V,\beta}(x,y) + R^-_{V,\beta}(x,y),$$

$$\varphi_{V,\beta,2}(x,y) = R^+_{V,\beta}(x,y) + L^-_{V,\beta}(x,y),$$

$$L^+_{V,\beta}(x,y) = I(x \in L(V), y - (1,x')\beta \ge 0),$$

$$R^-_{V,\beta}(x,y) = I(x \in R(V), y - (1,x')\beta \le 0),$$

and similarly for $R^+_{V,\beta}(x,y)$ and $L^-_{V,\beta}(x,y)$.

Next, suppose the data are subject to right censoring and left truncation. Instead of observing (X_i, Y_i), we thus observe $(X_i, T_i, Z_i, \delta_i)$ $(i = 1, \ldots, n)$. We then replace the multivariate empirical distribution $F_n(x,y)$ by the estimator $\hat{F}_{X,Y}(x,y)$, leading to

$$\widehat{\text{rdepth}}(\beta) = \min_V \Big[\min\Big\{ \iint_A \varphi_{V,\beta,1}(x,y)\, \hat{F}_{X,Y}(dx,dy), \iint_A \varphi_{V,\beta,2}(x,y)\, \hat{F}_{X,Y}(dx,dy) \Big\} \Big], \quad (4.1)$$

where $A = (a_F, b]$ and $b < b_F$. The population regression depth is now defined by

$$\text{rdepth}(\beta) = \min_V \Big[\min\Big\{ \iint_A \varphi_{V,\beta,1}(x,y)\, F_{X,Y}(dx,dy), \iint_A \varphi_{V,\beta,2}(x,y)\, F_{X,Y}(dx,dy) \Big\} \Big].$$

As for the application to goodness-of-fit testing, integration over y is limited to the set A, because of the usual tail problems that occur with censored and truncated data (see also Park and Hwang (2003)). In the next theorem, we prove the consistency of the regression depth estimator, and we also establish its rate of convergence.

Theorem 4.3 *Under (a),(b),(c),(d) and (Λ), we have*

$$\sup_{\beta \in R^{d+1}} |\widehat{rdepth}(\beta) - rdepth(\beta)| = O_P(n^{-1/2}).$$

Appendix: Proofs

In the proofs below, any probability or expected value with respect to the observed random variables (X, T, Z, δ) is understood to be conditionally on the event that T is less than or equal to Z.

Proof of Theorem 3.1. This proof consists of a sequence of decompositions of the product-limit integral. At each step we obtain either components of the dominant part or negligible terms. Here we give the main steps, while the necessary proofs for each step are deferred to lemmas below.

First, the denominator $nC_n(Z_j)$ is replaced by $nC_n(Z_j)+1$,

$$\int \varphi d\hat{F}_{X,Y} = \sum_{i=1}^{n} \varphi\left(X_i, Z_i\right) \frac{\delta_i}{nC_n\left(Z_i\right)} \prod_{\{j: Z_j < Z_i, \delta_j = 1\}} \left(1 - \frac{1}{nC_n\left(Z_j\right)+1}\right) + O\left(\frac{\ln n}{n}\right)$$

almost surely and uniformly in φ, where the negligibility of the remainder can be obtained from Lemma A.2. Then, the product can be expressed in terms of natural logarithms. Let us denote it by

$$E_i = \prod_{\{j: Z_j < Z_i, \delta_j = 1\}} \left(1 - \frac{1}{nC_n\left(Z_j\right)+1}\right) = \exp\left\{\sum_{\{j: Z_j < Z_i, \delta_j = 1\}} \ln\left(1 - \frac{1}{nC_n\left(Z_j\right)+1}\right)\right\}.$$

Now, expanding

$$\frac{1}{C_n\left(Z_i\right)} = \frac{1}{C\left(Z_i\right)} - \frac{C_n\left(Z_i\right) - C\left(Z_i\right)}{C^2\left(Z_i\right)} + \frac{\left(C_n\left(Z_i\right) - C\left(Z_i\right)\right)^2}{C^2\left(Z_i\right) C_n\left(Z_i\right)},$$

the product-limit integral can be decomposed into

$$\int \varphi d\hat{F}_{X,Y} = P_1(\varphi) - P_2(\varphi) + P_3(\varphi) + O\left(\frac{\ln n}{n}\right), \tag{A.1}$$

where

$$P_1(\varphi) = n^{-1} \sum_{i=1}^{n} \varphi\left(X_i, Z_i\right) \frac{1}{C\left(Z_i\right)} \delta_i E_i,$$

$$P_2(\varphi) = n^{-1} \sum_{i=1}^{n} \varphi\left(X_i, Z_i\right) \frac{C_n\left(Z_i\right) - C\left(Z_i\right)}{C^2\left(Z_i\right)} \delta_i E_i,$$

$$P_3(\varphi) = n^{-1} \sum_{i=1}^{n} \varphi\left(X_i, Z_i\right) \frac{\left(C_n\left(Z_i\right) - C\left(Z_i\right)\right)^2}{C^2\left(Z_i\right) C_n\left(Z_i\right)} \delta_i E_i.$$

To deal with P_1, we begin with a Taylor expansion for E_i,

$$E_i = \exp\left\{-\int_{-\infty}^{Z_i} \frac{1}{C(\omega)} W^1(d\omega)\right\}\left[1 + n\int_{-\infty}^{Z_i^-} \ln\left(1 - \frac{1}{nC_n(\omega)+1}\right) W_n^1(d\omega) + \int_{-\infty}^{Z_i} \frac{1}{C(\omega)} W^1(d\omega)\right]$$
$$+\frac{1}{2}e^{\Delta_i}\left[n\int_{-\infty}^{Z_i^-} \ln\left(1 - \frac{1}{nC_n(\omega)+1}\right) W_n^1(d\omega) + \int_{-\infty}^{Z_i} \frac{1}{C(\omega)} W^1(d\omega)\right]^2,$$

where Δ_i is between $n\int_{-\infty}^{Z_i^-} \ln\left(1 - \frac{1}{nC_n(\omega)+1}\right) W_n^1(d\omega)$ and $(-1)\cdot\int_{-\infty}^{Z_i} \frac{1}{C(\omega)} W^1(d\omega)$. Substituting into the expression for $P_1(\varphi)$, we have

$$P_1(\varphi) = n^{-1}\sum_{i=1}^{n} \varphi(X_i, Z_i)\,\gamma_0(Z_i)\,\delta_i\,(1 + B_{ni} - C_{ni}) + R_{1n}(\varphi),$$

where

$$B_{ni} = n\int_{-\infty}^{Z_i^-} \ln\left(1 - \frac{1}{nC_n(\omega)+1}\right) W_n^1(d\omega) + \int_{-\infty}^{Z_i^-} \frac{1}{C_n(\omega)} W_n^1(d\omega),$$
$$C_{ni} = \int_{-\infty}^{Z_i^-} \frac{1}{C_n(\omega)} W_n^1(d\omega) - \int_{-\infty}^{Z_i} \frac{1}{C(\omega)} W^1(d\omega),$$

and

$$R_{1n}(\varphi) = \frac{1}{2}n^{-1}\sum_{i=1}^{n} \varphi(X_i, Z_i)\,\frac{1}{C(Z_i)}\delta_i e^{\Delta_i}(B_{ni} - C_{ni})^2.$$

By Lemmas A.3, A.4 and A.5 we obtain

$$P_1(\varphi) = n^{-1}\sum_{i=1}^{n} \left[\varphi(X_i, Z_i)\,\gamma_0(Z_i)\,\delta_i - \gamma_1^{\varphi}(Z_i)\,\delta_i + \gamma_2^{\varphi}(T_i, Z_i)\right] + O\left(\frac{\ln^3 n}{n}\right). \tag{A.2}$$

For P_2 we make use of similar arguments, but now a Taylor expansion of order one for E_i is sufficient,

$$E_i = \exp\left\{-\int_{-\infty}^{Z_i} \frac{1}{C(\omega)} W^1(d\omega)\right\}$$
$$+e^{\rho_i}\left[n\int_{-\infty}^{Z_i^-} \ln\left(1 - \frac{1}{nC_n(\omega)+1}\right) W_n^1(d\omega) + \int_{-\infty}^{Z_i} \frac{1}{C(\omega)} W^1(d\omega)\right],$$

where ρ_i is between $n\int_{-\infty}^{Z_i^-} \ln\left(1 - \frac{1}{nC_n(\omega)+1}\right) W_n^1(d\omega)$ and $(-1)\cdot\int_{-\infty}^{Z_i} \frac{1}{C(\omega)} W^1(d\omega)$. Then, we have

$$P_2(\varphi) = n^{-1}\sum_{i=1}^{n} \varphi(X_i, Z_i)\,\gamma_0(Z_i)\,\delta_i \frac{C_n(Z_i) - C(Z_i)}{C(Z_i)} + R_{2n}(\varphi),$$

where

$$R_{2n}(\varphi) = n^{-1} \sum_{i=1}^{n} \varphi\left(X_i, Z_i\right) \delta_i \frac{C_n\left(Z_i\right) - C\left(Z_i\right)}{C^2\left(Z_i\right)} e^{\rho_i} \left[B_{ni} - C_{ni}\right],$$

B_{ni} and C_{ni} denoting the same quantities as before.

By Lemmas A.6 and A.7 we obtain

$$P_2(\varphi) = n^{-1} \sum_{i=1}^{n} \gamma_3^{\varphi}\left(T_i, Z_i\right) - E\left[\varphi(X, Z)\gamma_0(Z)\delta\right] + O\left(\frac{\ln^3 n}{n}\right). \tag{A.3}$$

Straightforward calculations show that $E[\varphi(X, Z)\gamma_0(Z)\delta] = \int \varphi\, dF_{X,Y}$.

Finally, from expressions (A.1), (A.2) and (A.3), and Lemma A.8 the proof of this theorem is completed.

Lemma A.1

$$\sup_{i\in\{1,\ldots,n\}} \frac{C\left(Z_i\right)}{C_n\left(Z_i\right)} = O(\ln n) \qquad \textit{almost surely.}$$

Proof. Let $U_i = 1 - M(Z_i)$ and $V_i = G(T_i)$ for $i \in \{1, \ldots, N\}$, where N is the (unknown) size of the complete sample, which satisfies $n/N \to \alpha$ a.s. Then U_i and V_i are uniformly distributed on the unit interval and are independent. Then,

$$\sup_{i\in\{1,\ldots,n\}} \frac{C\left(Z_i\right)}{C_n\left(Z_i\right)} = \frac{n}{N}\frac{1}{\alpha} \sup_{i\in\{1,\ldots,n\}} \frac{U_i G\left(Z_i\right)}{H_N\left(U_i, G\left(Z_i\right)\right)} \tag{A.4}$$

where $H_N(s,t) = N^{-1}\sum_{i=1}^{N} I_{\{U_i\le s, V_i\le t\}}$ is the empirical distribution function of (U_1, V_1), $\ldots, (U_N, V_N)$.

There is $z_0 \in [a_M, b_M]$ and $c \in (0,1)$ such that $1 - M(z) \ge c$ for all $z \in [a_M, z_0]$ and $G(z) \ge c$ for all $z \in (z_0, b_M]$. Then,

$$\sup_{i\in\{1,\ldots,n\}} \frac{U_i G\left(Z_i\right)}{H_N\left(U_i, G\left(Z_i\right)\right)} = \max\left\{ \sup_{i:Z_i\le z_0} \frac{U_i G\left(Z_i\right)}{H_N\left(U_i, G\left(Z_i\right)\right)}, \sup_{i:Z_i> z_0} \frac{U_i G\left(Z_i\right)}{H_N\left(U_i, G\left(Z_i\right)\right)} \right\}. \tag{A.5}$$

Now, since $T_i \le Z_i$ for all $i \in \{1, \ldots, n\}$, and then $V_{1:n} \le V_i = G(T_i) \le G(Z_i)$ for all $i \in \{1, \ldots, n\}$, we have

$$\sup_{i:Z_i\le z_0} \frac{U_i G\left(Z_i\right)}{H_N\left(U_i, G\left(Z_i\right)\right)} \le \sup_{\substack{i\,:\,Z_i\le z_0 \\ V_{1:n}\le t\le 1}} \frac{U_i t}{H_N\left(U_i, t\right)}. \tag{A.6}$$

Now we obtain a probability bound for the right-hand side of (A.6):

$$P\left(\sup_{\substack{i\,:\,Z_i\le z_0 \\ V_{1:n}\le t\le 1}} \frac{U_i t}{H_N\left(U_i, t\right)} \ge \lambda \right) \le P\left(\sup_{\substack{c\le s\le 1 \\ \lambda/N\le t\le 1}} \frac{st}{H_N\left(s, t\right)} \ge \lambda \right)$$

$$= P\left(\sup_{\substack{c\le s\le 1\\ \lambda/N\le t\le 1}} \frac{-H_N(s,t)}{st} \ge -\frac{1}{\lambda}\right) \le e^{1+c}\lambda^c e^{-c\lambda}.$$

The first inequality comes from the fact that if $st/H_N(s,t) \ge \lambda$ then $t \ge \lambda s^{-1} H_N(s,t) \ge \lambda/N$ (because $s^{-1} \ge 1$ and $H_N(s,t) \ge 1/N$), and from the restriction $Z_i \le z_0$, which implies $U_i = 1 - M(Z_i) \ge c$. The last inequality can be derived with the same arguments as in Lemma 1.1 in Stute (1993), applied to the reverse martingale $-H_N(s,t)/(st)$.

Similar arguments lead to the same probability bounding for the last supremum at the right-hand side in (A.5).

Finally, taking $\lambda = 2c^{-1}\ln N$ and applying Borel-Cantelli Lemma, we obtain

$$\sup_{i\in\{1,\ldots,n\}} \frac{U_i G(Z_i)}{H_N(U_i, G(Z_i))} = O(\ln N)$$

with probability one, as N goes to infinity. Since $n/N \to \alpha$ with probability one, then using the expression in (A.4), we also have

$$\sup_{i\in\{1,\ldots,n\}} \frac{C(Z_i)}{C_n(Z_i)} = O(\ln n)$$

with probability one.

Lemma A.2 *Assuming* $\int \Phi(x,z) C^{-1}(z)(1-F(z))^{-2}\, W^1_{X,Z}(dx,dz) < \infty$ *and* $\int (1-F(z))^2 C^{-2}(z)\, W^1(dz) < \infty$, *we have*

$$n^{-1}\sum_{i=1}^n \Phi(X_i,Z_i)\frac{\delta_i}{C_n(Z_i)}\left|\prod_{\{j:Z_j<Z_i,\delta_j=1\}}\left(\frac{nC_n(Z_j)-1}{nC_n(Z_j)}\right)\right.$$
$$\left. - \prod_{\{j:Z_j<Z_i,\delta_j=1\}}\left(\frac{nC_n(Z_j)}{nC_n(Z_j)+1}\right)\right| = O\left(\frac{\ln^3 n}{n}\right) \quad \textit{almost surely.}$$

Proof. As in Stute (1993, p. 154), since $|\prod_{j=1}^n a_j - \prod_{j=1}^n b_j| \le \sum_{j=1}^n |a_j - b_j|$, for $|a_j|, |b_j| \le 1$, then the difference of products in this lemma can be bounded by $n^{-1}\int_{-\infty}^{Z_i^-} C_n^{-2}(z)\, W_n^1(dz)$, and the whole expression can be bounded by

$$\left[\sup_i \frac{C(Z_i)}{C_n(Z_i)}\right]^3 n^{-1}\int_{-\infty}^{+\infty}(1-F(z))^2 C^{-2}(z)\, W_n^1(dz)$$
$$\times\, n^{-1}\textstyle\sum_{i=1}^n \Phi(X_i,Z_i)(1-F(Z_i))^{-2} C^{-1}(Z_i)\delta_i.$$

Lemma A.1 and the moment assumptions in this lemma lead to an almost sure bounding of order $O(n^{-1}\ln^3 n)$.

Lemma A.3 *Assuming* $\int \Phi(x,z)C^{-1}(z)(1-F(z))^{-1}\, W^1_{X,Z}(dx,dz) < \infty$ *and* $\int(1-F(z))^2$ $C^{-2}(z)\, W^1(dz) < \infty$*, we have*

$$\sup_{\varphi}\left|n^{-1}\sum_{i=1}^{n}\varphi\left(X_i,Z_i\right)\gamma_0\left(Z_i\right)\delta_i B_{ni}\right| = O\left(\frac{\ln^2 n}{n}\right) \qquad \textit{almost surely.}$$

Proof. First, we decompose B_{ni} into

$$\begin{aligned} B_{ni} = n\int_{-\infty}^{Z_i^-}&\left(\ln\left(1-\frac{1}{nC_n(\omega)+1}\right)+\frac{1}{nC_n(\omega)+1}\right)W_n^1(d\omega)\\ -n\int_{-\infty}^{Z_i^-}&\left(\frac{1}{nC_n(\omega)+1}-\frac{1}{nC_n(\omega)}\right)W_n^1(d\omega). \end{aligned}$$

Using that for $x \in [0,1/2]$, $-2x^2 \leq \ln(1-x)+x \leq -x^2/2$, the first integral can be bounded in absolute value by $2n\int_{-\infty}^{Z_i^-}(nC_n(\omega)+1)^{-2}W_n^1(d\omega)$, and then

$$|B_{ni}| \leq 3n^{-1}\int_{-\infty}^{Z_i^-}\frac{1}{C_n^2(\omega)}W_n^1(d\omega).$$

Substituting this into the expression in this lemma, we can obtain the bound

$$\begin{aligned} 3n^{-1}\left[\sup_i\frac{C\left(Z_i\right)}{C_n\left(Z_i\right)}\right]^2&\int_{-\infty}^{+\infty}(1-F(z))^2C^{-2}(z)\, W_n^1(dz)\\ &\times n^{-1}\textstyle\sum_{i=1}^{n}\Phi\left(X_i,Z_i\right)\left(1-F\left(Z_i\right)\right)^{-1}C^{-1}(Z_i)\delta_i. \end{aligned}$$

Lemma A.1 and the moment assumptions in this lemma are sufficient to conclude the proof.

Lemma A.4 *Assuming* $\int \Phi(x,z)C^{-1}(z)(1-F(z))^{-2}\, W^1_{X,Z}(dx,dz) < \infty$*,* $\int \Phi^2(x,z)C^{-2}(z)$ $(1-F(z))^{-1}\, W^1_{X,Z}(dx,dz) < \infty$ *and* $\int(1-F(z))^3C^{-3}(z)\, W^1(dz) < \infty$*, we have*

$$n^{-1}\sum_{i=1}^{n}\varphi\left(X_i,Z_i\right)\gamma_0\left(Z_i\right)\delta_i C_{ni} = n^{-1}\sum_{i=1}^{n}\left[\gamma_1^{\varphi}\left(Z_i\right)-\gamma_2^{\varphi}\left(T_i,Z_i\right)\right]+O\left(\frac{\ln^2 n}{n}\right)$$

almost surely and uniformly in φ*.*

Proof. C_{ni} can be decomposed into

$$\begin{aligned} C_{ni} = &-\int_{-\infty}^{Z_i^-}\frac{C_n(\omega)}{C^2(\omega)}W_n^1(d\omega)+\int_{-\infty}^{Z_i^-}\frac{2}{C(\omega)}W_n^1(d\omega)-\int_{-\infty}^{Z_i}\frac{1}{C(\omega)}W^1(d\omega)\\ &+\int_{-\infty}^{Z_i^-}\frac{\left[C_n(\omega)-C(\omega)\right]^2}{C^2(\omega)C_n(\omega)}W_n^1(d\omega), \end{aligned}$$

which, by substitution, gives place to

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\varphi\left(X_i,Z_i\right)\gamma_0\left(Z_i\right)\delta_i C_{ni} \\
& =-\int\frac{\varphi(x,\omega)\gamma_0(\omega)}{C^2(v)}I_{\{v<\omega\}}I_{\{t\le v\le u\}}\ H_n(dt,du)W_n^1(dv)W_{n,X,Z}^1(dx,d\omega) \\
& \quad +2\int\frac{\varphi(x,\omega)\gamma_0(\omega)}{C(v)}I_{\{v<\omega\}}\ W_n^1(dv)W_{n,X,Z}^1(dx,d\omega) \\
& \quad -\int\frac{\varphi(x,\omega)\gamma_0(\omega)}{C(v)}I_{\{v<\omega\}}\ W^1(dv)W_{n,X,Z}^1(dx,d\omega)+Q_{n1}(\varphi),
\end{aligned}
\tag{A.7}
$$

where $H_n(t,u)=n^{-1}\sum_{i=1}^{n}I(T_i\le t,Z_i\le u)$ and where $\sup_\varphi|Q_{n1}(\varphi)|$ is of the order $O(\ln^2 n/n)$ almost surely, by arguments similar to those given in the preceding lemmas, and by using that $\sup_t|C_n(t)-C(t)|=O(n^{-1/2}\ln^{1/2}n)$ almost surely.

The third integral in (A.7) is already a sum of i.i.d terms, that will contribute to the dominant part of the representation. The second integral in (A.7) is a U-process of order two, where

$$
\varphi u\left(\left(X_j,Z_j,\delta_j\right),\left(X_k,Z_k,\delta_k\right)\right)=\frac{\varphi\left(X_j,Z_j\right)\gamma_0\left(Z_j\right)}{C\left(Z_k\right)}I_{\{Z_k<Z_j\}}\delta_j\delta_k
$$

is the family of kernel functions. This is a VC-subgraph class of functions with envelope

$$
\Phi u\left(\left(X_j,Z_j,\delta_j\right),\left(X_k,Z_k,\delta_k\right)\right)=\frac{\Phi\left(X_j,Z_j\right)\gamma_0\left(Z_j\right)}{C\left(Z_k\right)}I_{\{Z_k<Z_j\}}\delta_j\delta_k,
$$

which is square integrable under the moment conditions assumed in this lemma. Then, by the law of the iterated logarithm for U-processes, as given in Theorem 5.4.1 in De la Peña and Giné (1999), applied to the remainder of the Hájek decomposition, we have

$$
\begin{aligned}
& \int\frac{\varphi(x,\omega)\gamma_0(\omega)}{C(v)}I_{\{v<\omega\}}\ W_n^1(dv)W_{n,X,Z}^1(dx,d\omega) \\
& \quad =\int\frac{\varphi(x,\omega)\gamma_0(\omega)}{C(v)}I_{\{v<\omega\}}\ \Big[W_n^1(dv)W_{X,Z}^1(dx,d\omega)+W^1(dv)W_{n,X,Z}^1(dx,d\omega) \\
& \qquad -W^1(dv)W_{X,Z}^1(dx,d\omega)\Big]+Q_{n2}(\varphi),
\end{aligned}
$$

where $\sup_\varphi|Q_{n2}(\varphi)|=O(\ln\ln n/n)$ almost surely.

The first integral in (A.7) can be written as a triple sum,

$$
n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}I_{\{Z_k<Z_j,T_i\le Z_k\le Z_i\}}\frac{\varphi\left(X_j,Z_j\right)\gamma_0\left(Z_j\right)}{C^2\left(Z_k\right)}\delta_j\delta_k,
$$

that can be decomposed in three terms: the summands with $i = k$ give place to a U-process of order two multiplied by n^{-1}, the summands with $i = j$ give place to another U-process of order two multiplied by n^{-1}, and the summands with $i \neq k$ and $i \neq j$ make a U-process of order three. With respect to the U-processes of order two multiplied by n^{-1}, consistency is sufficient for our purposes, and this consistency can be obtained by applying Corollary 5.2.3 of De la Peña and Giné (1999) under the integrability of the corresponding envelope functions. For the U-process of order three, we employ the same argument based on the Hájek decomposition and the law of iterated logarithm for the remainder as used before, where the square integrability of the envelope is a consequence of the moment assumptions.

Finally, recapitulation of the dominant terms leads to the representation in this lemma.

Lemma A.5 *Assuming* $\int \Phi(x,z)C^{-1}(z)(1-F(z))^{-4}\, W^1_{X,Z}(dx,dz) < \infty$ *and* $\int (1-F(z))^2 C^{-2}(z)\, W^1(dz) < \infty$*, we have*

$$\sup_{\varphi} |R_{1n}(\varphi)| = O\left(\frac{\ln^3 n}{n}\right) \qquad \textit{almost surely.}$$

Proof. Since $\Delta_i \leq 0$, it follows that $e^{\Delta_i} \leq 1$, and hence

$$|R_{1n}(\varphi)| \leq \frac{1}{2} n^{-1} \sum_{i=1}^{n} \Phi(X_i, Z_i) \frac{1}{C(Z_i)} \delta_i (B_{ni} - C_{ni})^2 . \tag{A.8}$$

From the proof of Lemma A.3, it is derived that $\sup_i (1 - F(Z_i))^2 |B_{ni}| = O(\ln^2 n/n)$ almost surely.

C_{ni} can be decomposed into $C_{ni} = C^1_{ni} + C^2_{ni}$, where

$$C^1_{ni} = \int_{-\infty}^{Z_i^-} \left(\frac{1}{C_n(\omega)} - \frac{1}{C(\omega)}\right) W^1_n(d\omega),$$

$$C^2_{ni} = \int_{-\infty}^{Z_i^-} \frac{1}{C(\omega)} \left(W^1_n(d\omega) - W^1(d\omega)\right).$$

Since

$$\left|C^1_{ni}\right| \leq \sup_z |C_n(z) - C(z)| \sup_i \frac{C(Z_i)}{C_n(Z_i)} \int_{-\infty}^{Z_i^-} \frac{1}{C^2(\omega)} W^1_n(d\omega)$$

and with similar arguments as above, we have $\sup_i (1 - F(Z_i))^2 |C^1_{ni}| = O(\ln^{3/2} n/n^{1/2})$ almost surely.

For C^2_{ni}, we apply Theorem 5.4.1 of De la Peña and Giné (1999) to obtain

$$\sup_t \left| \int_{-\infty}^{t} \frac{1 - F(\omega)}{C(\omega)} \left(W^1_n(d\omega) - W^1(d\omega)\right) \right| = O\left(\frac{\ln \ln^{1/2} n}{n^{1/2}}\right)$$

almost surely. Hence, $\sup_i (1 - F(Z_i))|C_{ni}^2| = O(\ln\ln^{1/2} n/n^{1/2})$ almost surely.

The proof is finished applying the bounds obtained for B_{ni} and C_{ni} to the expression (A.8), and making use of the moment assumptions.

Lemma A.6 *Assuming $\int \Phi^2(x,z) C^{-3}(z)(1-F(z))^2 \, W^1_{X,Z}(dx,dz) < \infty$, we have*

$$n^{-1}\sum_{i=1}^{n} \varphi\left(X_i, Z_i\right)\gamma_0\left(Z_i\right)\delta_i \frac{C_n\left(Z_i\right) - C\left(Z_i\right)}{C\left(Z_i\right)}$$
$$= n^{-1}\sum_{i=1}^{n} \gamma_3^{\varphi}\left(T_i, Z_i\right) - E\left[\varphi(X,Z)\gamma_0(Z)\delta\right] + O\left(\frac{\ln\ln n}{n}\right)$$

almost surely and uniformly in φ.

Proof. This proof is similar to the proof of Lemma A.4. Here,

$$n^{-1}\sum_{i=1}^{n} \varphi\left(X_i, Z_i\right)\gamma_0\left(Z_i\right)\delta_i \frac{C_n\left(Z_i\right)}{C\left(Z_i\right)} = \int \frac{\varphi(x,\omega)\gamma_0(\omega)}{C(\omega)} I_{\{t\le\omega\le u\}} \, H_n(dt,du) W^1_{n,X,Z}(dx,d\omega),$$

where $H_n(t,u) = n^{-1}\sum_{i=1}^n I(T_i \le t, Z_i \le u)$, which is a V-process of order two. The diagonal terms constitute a sum of i.i.d. random variables, which is bounded of order $O(n^{-1})$ uniformly in φ. The non-diagonal terms constitute a U-process of order two. Finally, the assertion in this lemma is a consequence of the Hájek decomposition of that U-process, where the remainder term is a canonical (degenerate) U-process of order two.

Lemma A.7 *Assuming $\int \Phi(x,z) C^{-2}(z)(1-F(z))^{-2} \, W^1_{X,Z}(dx,dz) < \infty$ and $\int (1-F(z))^2 \, C^{-2}(z) \, W^1(dz) < \infty$, we have*

$$\sup_{\varphi} |R_{2n}(\varphi)| = O\left(\frac{\ln^2 n}{n}\right) \qquad \textit{almost surely.}$$

Proof. This proof is similar to that of Lemma A.5. Here, since $\rho_i \le 0$, it follows that $e^{\rho_i} \le 1$, and hence

$$|R_{2n}(\varphi)| \le n^{-1}\sum_{i=1}^{n} \Phi\left(X_i, Z_i\right)\delta_i \frac{\left|C_n\left(Z_i\right) - C\left(Z_i\right)\right|}{C^2\left(Z_i\right)} \left|B_{ni} - C_{ni}\right|.$$

Applying the bounds obtained in the proof of Lemma A.5 for B_{ni} and C_{ni}, the fact that $\sup_z |C_n(z) - C(z)| = O(\ln^{1/2} n/n^{1/2})$, and making use of the moment assumptions, the proof is concluded.

Lemma A.8 *Assuming $\int \Phi(x,z) C^{-3}(z) \, W^1_{X,Z}(dx,dz) < \infty$, we have*

$$\sup_{\varphi} |P_3(\varphi)| = O\left(\frac{\ln^2 n}{n}\right) \qquad \textit{almost surely.}$$

Proof. Since $0 \leq E_i \leq 1$,

$$\sup_{\varphi} |P_3(\varphi)| \leq \frac{1}{n} \sum_{i=1}^{n} \Phi(X_i, Z_i) \frac{(C_n(Z_i) - C(Z_i))^2}{C^2(Z_i) C_n(Z_i)} \delta_i.$$

The expression on the right hand side can be bounded as in the preceding lemmas.

Proof of Theorem 4.1. This proof is just an application of Theorem 3.1 to the class of functions $\{\varphi_{x_0}\}_{x_0 \in R}$. To this end, it is easily shown that $\{\varphi_{x_0}\}_{x_0 \in R}$ is a VC-subgraph class of functions, that admits the envelope $\Phi(x,y) = |y - m_0(x)| I_{\{y \leq b\}}$, which satisfies the moment conditions required in Theorem 3.1.

Proof of Theorem 4.2. By Theorem 4.1, it is enough to prove the convergence of the dominant term. This will be done by showing the convergence of the finite dimensional distributions and the tightness of the process. The convergence of the finite dimensional distributions can be derived from a multivariate central limit theorem, so the rest of the proof is devoted to the proof of tightness.

Let us decompose

$$n^{-1} \sum_{i=1}^{n} \eta_i(x_0) = D_n^{0a}(x_0) + D_n^{0b}(x_0) - D_n^{0c}(x_0)$$

where

$$D_n^{0a}(x_0) = n^{-1/2} \sum_{i=1}^{n} I_{\{X_i \leq x_0\}} (Z_i - m_0(X_i)) \gamma_0(Z_i) \delta_i I_{\{Z_i \leq b\}}$$

$$D_n^{0b}(x_0) = n^{-1/2} \sum_{i=1}^{n} \left(\gamma_2^{\varphi_{x_0}}(T_i, Z_i) - \delta_i \gamma_1^{\varphi_{x_0}}(Z_i) \right)$$

$$D_n^{0c}(x_0) = n^{-1/2} \sum_{i=1}^{n} \gamma_3^{\varphi_{x_0}}(T_i, Z_i)$$

We will prove the tightness separately for D_n^{0a}, D_n^{0b} and D_n^{0c}. In the three cases, we will use Theorem 15.6 of Billingsley (1968).

Let us start with D_n^{0a}. Consider $x_0 \leq x_1 \leq x_2$ and let us denote

$$\alpha_i^a = I_{\{x_0 < X_i \leq x_1\}} (Z_i - m_0(X_i)) \gamma_0(Z_i) \delta_i I_{\{Z_i \leq b\}}$$

$$\beta_i^a = I_{\{x_1 < X_i \leq x_2\}} (Z_i - m_0(X_i)) \gamma_0(Z_i) \delta_i I_{\{Z_i \leq b\}}$$

Since $(\alpha_1^a, \beta_1^a), \ldots, (\alpha_n^a, \beta_n^a)$ are independent, $E(\alpha_i^a) = 0$, $E(\beta_i^a) = 0$ and $\alpha_i^a \beta_i^a = 0$ for all $i \in \{1, \ldots, n\}$, we have

$$E\left[\left(D_n^{0a}(x_1) - D_n^{0a}(x_0) \right)^2 \left(D_n^{0a}(x_2) - D_n^{0a}(x_1) \right)^2 \right] = n^{-2} E\left[\left(\sum_{i=1}^{n} \alpha_i^a \right)^2 \left(\sum_{i=1}^{n} \beta_i^a \right)^2 \right]$$

$$= n^{-2}n(n-1)E\left((\alpha_1^a)^2\right)E\left((\beta_2^a)^2\right) \leq \left\{E\left[I_{\{x_0<X_1\leq x_2\}}\left(Z_i - m_0\left(X_i\right)\right)^2\gamma_0^2\left(Z_i\right)\delta_i I_{\{Z_i\leq b\}}\right]\right\}^2$$
$$= \left[S_a\left(x_2\right) - S_a\left(x_0\right)\right]^2$$

where $S_a(x) = E[I_{\{X\leq x\}}(Z - m_0(X))^2\gamma_0^2(Z)\delta I_{\{Z\leq b\}}]$ is a non-decreasing function and continuous function of x. Then, Theorem 15.6 of Billingsley (1968) can be applied to prove the tightness of D_n^{0a}.

For D_n^{0b}, consider again $x_0 \leq x_1 \leq x_2$ and use the notation

$$\alpha_i^b = \int \nabla\left(\omega, \delta_i, T_i, Z_i\right) I_{\{x_0<x\leq x_1\}}\left(\omega - m_0(x)\right) I_{\{\omega\leq b\}}\gamma_0(\omega)\, W_{X,Z}^1(dx, d\omega)$$
$$\beta_i^b = \int \nabla\left(\omega, \delta_i, T_i, Z_i\right) I_{\{x_1<x\leq x_2\}}\left(\omega - m_0(x)\right) I_{\{\omega\leq b\}}\gamma_0(\omega)\, W_{X,Z}^1(dx, d\omega),$$

where $\nabla(\omega, \delta_i, T_i, Z_i) = \int I(v < \omega, T_i \leq v \leq Z_i)C^{-2}(v)W^1(dv) - \delta_i I(Z_i < \omega)C^{-1}(Z_i)$. Then, the situation is similar to the preceding one, but now $\alpha_i\beta_i$ do not have to be zero. Then with the same arguments as before, and also a Cauchy-Schwartz inequality, we have

$$E\left[\left(D_n^{0b}\left(x_1\right) - D_n^{0b}\left(x_0\right)\right)^2\left(D_n^{0b}\left(x_2\right) - D_n^{0b}\left(x_1\right)\right)^2\right] = n^{-2}E\left[\left(\sum_{i=1}^n \alpha_i^b\right)^2\left(\sum_{i=1}^n \beta_i^b\right)^2\right]$$
$$= n^{-2}\left[nE\left(\left(\alpha_1^b\right)^2\left(\beta_1^b\right)^2\right) + n(n-1)E\left(\left(\alpha_1^b\right)^2\right)E\left(\left(\beta_2^b\right)^2\right) + 2n(n-1)\left[E\left(\alpha_1^b\beta_1^b\right)\right]^2\right]$$
$$\leq n^{-2}\left[nE\left(\left(\alpha_1^b\right)^2\left(\beta_1^b\right)^2\right) + 3n(n-1)E\left(\left(\alpha_1^b\right)^2\right)E\left(\left(\beta_2^b\right)^2\right)\right]$$
$$\leq 64\left(E\left[\Lambda^{-1}(Y)\right]\right)^4 E\left[\Lambda^{-3}(Y)\right]\left[S_b\left(x_2\right) - S_b\left(x_0\right)\right]^4$$

where $S_b(x) = E[I_{\{X\leq x\}}|Z - m_0(X)|\gamma_0(Z)\delta I_{\{Z\leq b\}}]$ is a non-decreasing and continuous function of x. Then, Theorem 15.6 of Billingsley (1968) can be applied to prove the tightness of D_n^{0b}. The same arguments can be applied to prove the tightness of D_n^{0c}, so the details are omitted.

Proof of Theorem 4.3. Let, for $k = 1, 2$,

$$I_k(V, \beta) = \int \varphi_{V,\beta,k}(x, y)I_A(y)\, F_{X,Y}(dx, dy),$$
$$\hat{I}_k(V, \beta) = \int \varphi_{V,\beta,k}(x, y)I_A(y)\, \hat{F}_{X,Y}(dx, dy),$$

where $I_A(y) = I(y \in A)$, and let $J(V, \beta) = \min(I_1(V, \beta), I_2(V, \beta))$ and $\hat{J}(V, \beta) = \min(\hat{I}_1(V, \beta), \hat{I}_2(V, \beta))$. We first show that conditions (φ 1) and (φ 2) are satisfied for the functions $\varphi_{V,\beta,1}I_A$ and $\varphi_{V,\beta,2}I_A$. The envelope for the classes $\varphi_{V,\beta,1}I_A$ and $\varphi_{V,\beta,2}I_A$

is given by I_A, and hence condition (φ 2) follows from condition (Λ). The class of half-spaces $L(V)$ for all possible hyperplanes V in R^d is a VC-class (see e.g. van der Vaart and Wellner (1996), p. 152). It follows that the class of half-spaces $L(V)\times R$ in R^{d+1} is also VC, as well as the half-spaces formed by the (x, y) values in R^{d+1} satisfying $y-(1,x')\beta \geq 0$. Finally, the intersection of two VC-classes is VC (see van der Vaart and Wellner (1996), Lemma 2.6.17) and the family of indicator functions of the sets in a VC-class is also VC (see De la Peña and Giné, p. 224), which implies that $\{L^+_{V,\beta}\}$ is a VC-class. In a similar way it can be shown that $R^-_{V,\beta}$, $R^+_{V,\beta}$ and $L^-_{V,\beta}$ are VC-classes, and hence condition (φ 1) holds. It now follows from Theorem 3.1 that $(k = 1, 2)$

$$\hat{I}_k(V,\beta) - I_k(V,\beta) = n^{-1}\sum_{i=1}^{n} \eta_{i,k}(V,\beta) + R_{nk}(V,\beta),$$

where $\eta_{i,k}(V,\beta) = \eta_i^{\varphi_{V,\beta,k} I_A}$ and $\sup_{V,\beta} |R_{nk}(V,\beta)| = O(\ln^3 n/n)$ a.s. Hence,

$$\begin{aligned}
\sup_{\beta} |\widehat{\text{rdepth}}(\beta) - \text{rdepth}(\beta)| &= \sup_{\beta} |\min_{V} \hat{J}(V,\beta) - \min_{V} J(V,\beta)| \\
&\leq \sup_{V,\beta} |\hat{J}(V,\beta) - J(V,\beta)| \\
&\leq \sup_{V,\beta} \max_{k\in\{1,2\}} |\hat{I}_k(V,\beta) - I_k(V,\beta)| \\
&\leq \sum_{k=1}^{2} \sup_{V,\beta} |\hat{I}_k(V,\beta) - I_k(V,\beta)| \\
&\leq \sum_{k=1}^{2} \sup_{V,\beta} |n^{-1}\sum_{i=1}^{n} \eta_{i,k}(V,\beta)| + \sum_{k=1}^{2} \sup_{V,\beta} |R_{nk}(V,\beta)|.
\end{aligned}$$

It remains to prove that the class $\mathcal{F} = \{\eta_k(V,\beta)\}$ is Donsker $(k = 1, 2)$, from which the weak convergence of the process $n^{-1/2}\sum_{i=1}^{n} \eta_{i,k}(V,\beta)$ follows (and hence the above expression is $O_P(n^{-1/2})$). First, by Theorem 2.10.6 in van der Vaart and Wellner (1996) it suffices to show that the classes of functions $\{\varphi_{V,\beta,k}\}$, $\{\gamma_0 I_A\}$ and $\{\gamma_j^{\varphi_{V,\beta,k} I_A}\}$ $(j = 1, 2, 3)$ are Donsker. Since $\{\varphi_{V,\beta,k}\}$ is a VC-class with a square integrable envelope function, it is Donsker (see e.g. Theorem 2.6.8 in van der Vaart and Wellner (1996)). The function $\gamma_0(Z)I_A(Z)$ does not depend on β nor on V, and so it is clearly Donsker. We next consider $\gamma_2^{\varphi_{V,\beta,k} I_A}$ (the derivation for $\gamma_1^{\varphi_{V,\beta,k} I_A}$ and $\gamma_3^{\varphi_{V,\beta,k} I_A}$ is similar). By Theorem 2.5.6 in van der Vaart and Wellner (1996), it suffices to show that

$$\int_0^\infty \sqrt{\ln N_{[]}(\varepsilon, \{\gamma_2^{\varphi_{V,\beta,k} I_A}\}, L_2(P))}\, d\varepsilon < \infty, \tag{A.9}$$

where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of (X, T, Z, δ), and $L_2(P)$ is the L_2-norm. Fix $\varepsilon > 0$. Write

$$\begin{aligned}\gamma_2^{\varphi_{V,\beta,k} I_A}(T, Z) = &\int I_{\{v<\omega, v\leq Z\}} \varphi_{V,\beta,k}(x, \omega) I_A(\omega) \gamma_0(\omega) \frac{1}{C^2(v)}\, W^1(dv)\, W^1_{X,Z}(dx, d\omega) \\ &- \int I_{\{v<\omega, v\leq T\}} \varphi_{V,\beta,k}(x, \omega) I_A(\omega) \gamma_0(\omega) \frac{1}{C^2(v)}\, W^1(dv)\, W^1_{X,Z}(dx, d\omega),\end{aligned}$$

which is the difference of two monotone, bounded functions, and hence it requires $\exp(K\varepsilon^{-1})$ brackets (see Theorem 2.7.5 in van der Vaart and Wellner (1996)). However, when $\varepsilon > M^2$ (where M is an upper bound for $\int_{-\infty}^{b} C^{-2}(z) W^1(dz) < \infty$), 1 bracket suffices. Hence the integral in (A.9) is finite.

References

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

De la Peña, V.H. and Giné, E. (1999). *Decoupling: from dependence to independence.* Springer, New York.

Gijbels, I. and Gürler, Ü. (1998). Covariance function of a bivariate distribution function estimator for left truncated and right censored data. *Statistica Sinica*, **8**, 1219–1232.

Gross, S.T. and Lai, T.L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. *Journal of the American Statistical Association*, **91**, 1166–1180.

Gürler, Ü. and Gijbels, I. (1996). A bivariate distribution function estimator and its variance under left truncation and right censoring. *Discussion Paper* #9702, Institute of Statistics, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Park, J. and Hwang, J. (2003). Regression depth with censored and truncated data. *Communications in Statistics - Theory and Methods*, **32**, 997–1008.

Rousseeuw, P.J. and Hubert, M. (1999). Regression depth (with discussion). *Journal of the American Statistical Association*, **94**, 388–433.

Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *The Annals of Statistics*, **21**, 146–156.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics*, **23**, 461–471.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, **25**, 613–641.

Stute, W., González-Manteiga, W. and Sánchez-Sellero, C. (2000). Nonparametric model checks in censored regression. *Communications in Statistics – Theory and Methods*, **29**, 1611–1629.

Tsai, W.Y., Jewell, N.P. and Wang, M.C. (1987). A note on the product limit estimator under right censoring and left truncation. *Biometrika*, **74**, 883–886.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer, New York.

Zhou, Y. and Yip, P.S.F. (1999). A strong representation of the product-limit estimator fot left truncated and right censored data. *Journal of Multivariate Analysis*, **60**, 261–280.