I N S T I T U T   D E

S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N

P A P E R

0314

A NOTE ON THE NONPARAMETRIC ESTIMATION OF THE BIVARIATE DISTRIBUTION UNDER DEPENDENT CENSORING

I. VAN KEILEGOM

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

A note on the nonparametric estimation of the bivariate distribution under dependent censoring

Ingrid Van Keilegom[1]

Institut de Statistique

Université catholique de Louvain

August 11, 2003

Abstract

Consider the random vector (T_1, T_2), and assume that both T_1 and T_2 are subject to random right censoring. We propose new estimators of the bivariate and marginal distributions of T_1 and T_2. The estimators do not require the common assumption of independence between the vector of survival and censoring times, but allow for a certain type of dependent censoring. The proposed estimator of the marginal distribution generalizes the estimator of Cheng (1989). The estimators have intuitive, closed form expressions and are easy to compute. The weak convergence of the estimators is obtained. As an application we discuss the estimation of the regression coefficients in a polynomial regression model, when both the response and the covariate are subject to censoring.

KEY WORDS: Bivariate censoring; Bivariate distribution; Censored covariates; Kernel estimation; Least squares estimation; Marginal distribution; Nonparametric regression; Right censoring.

[1] This research was supported by 'Projet d'Actions de Recherche Concertées', No. 98/03–217, and by the IAP research network nr. P5/24 of the Belgian government.

1 Introduction

Let $\boldsymbol{T} = (T_1, T_2)$ be a bivariate lifetime vector with bivariate distribution function $F(\boldsymbol{y}) = P(T_1 \leq y_1, T_2 \leq y_2)$ and let $\boldsymbol{C}$ be a bivariate censoring vector with distribution $G(\boldsymbol{y})$. The problem of estimating the function $F(\boldsymbol{y})$ has been well studied in the literature (see e.g. van der Laan (1996) or Akritas and Van Keilegom (2003) for an overview). Most of the proposed estimators are based on the assumption that the vector of survival times $\boldsymbol{T}$ and the vector of censoring times $\boldsymbol{C}$ are independent.

In a number of practical situations, this is however not an acceptable assumption. Think e.g. of the situation where T_1 and T_2 are successive times and observed over a time period of length C, where C is random and known. Further assume that C is independent of T_1 and T_2. In this situation the sum $T_1 + T_2$ is subject to censoring, but not each survival time separately. It follows that $C_1 = C$ and that $C_2 = (C - T_1)I(T_1 \leq C)$. Hence, $\boldsymbol{T}$ and $\boldsymbol{C}$ are clearly dependent. Another example comes from the context of censored regression, where one often prefers to make use of the additional information given by the covariate, to weaken the assumption of independence between the response and its censoring time. The covariate is sometimes subject to censoring itself. Think e.g. of a regression problem in astronomy, where it is common for both variables to be subject to censoring, or of a medical study, where the covariate is the concentration of a certain product in the blood, and that this concentration cannot be detected below or above a certain detection limit. For a last example where $\boldsymbol{T}$ and $\boldsymbol{C}$ are dependent, consider the situation of two related individuals, one of which is dependent of the other, such as mother and child. When the mother dies, it may be more likely that the child will be lost to follow-up, whereas it may be appropriate to assume that death of the child will not affect the chances that the mother will be lost to follow-up. Hence, the censoring time of the child will depend here on the survival time of the mother.

In order to be able to estimate the bivariate distribution in the above type of situations, we weaken the common assumption of independence between $\boldsymbol{T}$ and $\boldsymbol{C}$ to

(A0) T_2 is independent of C_2 given T_1; and T_1 and T_2 are independent of C_1.

It is readily seen that it is not unrealistic to assume that $(A0)$ is satisfied in the above situations.

Let $F(y_2|y_1) = P(T_2 \leq y_2|T_1 = y_1)$ and $F_j(y) = P(T_j \leq y)$ $(j = 1, 2)$. The idea behind the proposed estimator of the bivariate distribution is as follows. The method consists of writing $F(y_1, y_2)$ as an average of $F(y_2|y)$ over all $y \leq y_1$ and estimate the so obtained expression. To estimate the conditional distributions, use is made of a generalization of a conditional Kaplan-Meier (1958) type estimator, introduced by Beran (1981), based on kernel smoothing around $T_1 = y$. Note however that T_1 is subject to censoring here, and that it is not clear to which window a censored observation belongs. We adopt the approach of allowing in the window only observations for which T_1 is uncensored. The second step for estimating the bivariate distribution consists of averaging the conditional distributions. Unlike the estimator in Akritas (1994), where T_1 is uncensored, we now use the Kaplan-Meier estimator of the marginal distribution of T_1 for this purpose.

For the estimation of the marginal distribution $F_2(y)$ of T_2, first note that the Kaplan-Meier estimator of $F_2(y)$ cannot be used here, since the independence of T_2 and C_2 is not assumed under $(A0)$. Instead, we use the marginal of the estimator of the bivariate distribution of T_1 and T_2. This has the additional advantage that the so-obtained estimator makes use of information given by the variable T_1. The proposed estimator is a generalization of the estimator of Cheng (1989) to the case where T_1 is subject to censoring.

The results in this paper have several applications. First, the estimation of the bivariate distribution has recently been applied to ROC curves for censored data in Heagerty, Lumley and Pepe (2000). The proposed estimator permits a useful extension of their methodology to the case where the diagnostic marker is also subject to censoring. Second, in a regression context the estimation of the regression function when both the response and the covariate are subject to censoring can be accomplished using the results of this paper. We will discuss this issue when the regression function is polynomial by estimating the regression coefficients by an extension of the classical least squares method. The case of semi- or nonparametric regression could also be dealt with.

In the next section we define the estimator of the bivariate distribution and state its asymptotic properties. Section 3 presents the marginal distribution estimator in detail, while an application to least squares regression is discussed in Section 4. The Appendix collects the proofs of the main results.

2 Estimation of the bivariate distribution

Let $(\boldsymbol{T}_i, \boldsymbol{C}_i)$ $(i = 1, \ldots, n)$ be n independent copies of $(\boldsymbol{T}, \boldsymbol{C})$. We observe $(\tilde{\boldsymbol{T}}_i, \boldsymbol{\Delta}_i)$ $(i = 1, \ldots, n)$ where $\tilde{T}_{ij} = T_{ij} \wedge C_{ij}$, $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ $(j = 1, 2)$. First write $F(\boldsymbol{y})$ in the following way :

$$F(\boldsymbol{y}) = \int_0^{y_1} F(y_2|z_1)\, dF_1(z_1). \tag{2.1}$$

We will estimate $F(\boldsymbol{y})$ by plugging in appropriate estimators of $F_1(y_1)$ and $F(y_2|y_1)$. For the estimation of F_1 use is made of the classical Kaplan-Meier (1958) estimator (we assume T_1 is continuous) :

$$\hat{F}_1(y) = 1 - \prod_{\tilde{T}_{i1} \le y, \Delta_{i1}=1} \left(1 - \frac{1}{n-i+1}\right).$$

For the estimation of $F(y_2|y_1)$, Beran (1981) proposed a conditional Kaplan-Meier type estimator restricted to the case where T_1 is uncensored, which is based on smoothing over the variable T_1. In the present situation of censoring in T_1, the situation is more complicated, since it is not clear to which window a censored observation belongs, as its true lifetime is unknown. However, the following equation, which is valid under assumption $(A0)$, shows that $F(y_2|y_1)$ can be estimated by using only the pairs of observations for which T_1 is uncensored :

$$P(T_2 \le y_2|T_1 = y_1, \Delta_1 = 1) = P(T_2 \le y_2|T_1 = y_1, C_1 \ge y_1) = P(T_2 \le y_2|T_1 = y_1).$$

This leads to the following estimator :

$$\hat{F}(y_2|y_1) = 1 - \prod_{\tilde{T}_{i2} \le y_2, \Delta_{i2}=1} \left(1 - \frac{W_{ni}(y_1; h_n)}{\sum_{j=1}^n W_{nj}(y_1; h_n) I(\tilde{T}_{j2} \ge \tilde{T}_{i2})}\right), \tag{2.2}$$

where

$$W_{ni}(y; h_n) = \begin{cases} \dfrac{K\left(\frac{y - T_{i1}}{h_n}\right)}{\sum_{\Delta_{j1}=1} K\left(\frac{y-T_{j1}}{h_n}\right)} & \text{if } \Delta_{i1} = 1 \\ 0 & \text{if } \Delta_{i1} = 0, \end{cases}$$

K is a kernel function (with support $[-L, L]$) and $\{h_n\}$ a bandwidth sequence. See also Akritas and Van Keilegom (2003), where $\hat{F}(y_2|y_1)$ has been proposed in the construction of an estimator of the bivariate distribution under the more restrictive assumption of independence between $\boldsymbol{T}$ and $\boldsymbol{C}$.

Then, the proposed estimator of the bivariate distribution is

$$\hat{F}(\boldsymbol{y}) = \int_0^{y_1} \hat{F}(y_2|z_1)\, d\hat{F}_1(z_1). \tag{2.3}$$

We will next develop the asymptotic theory for this estimator. Some notations need therefore to be introduced. Let $H(\boldsymbol{y}) = P(\tilde{T}_1 \le y_1, \tilde{T}_2 \le y_2)$, $H(\boldsymbol{y}, \boldsymbol{\delta}) = P(\tilde{T}_1 \le y_1, \tilde{T}_2 \le$

$y_2, \Delta_1 = \delta_1, \Delta_2 = \delta_2)$, $G_j(y) = P(C_j \leq y)$, $H_j(y) = P(\tilde{T}_j \leq y)$, $H_j^u(y) = P(\tilde{T}_j \leq y, \Delta_j = 1)$ $(j = 1, 2)$, $G(y_2|y_1) = P(C_2 \leq y_2|\tilde{T}_1 = y_1, \Delta_1 = 1)$, and similarly for $H(y_2|y_1)$ and $H^u(y_2|y_1)$. The probability density functions of the (sub)distributions defined above will be denoted with lower case letters. Further, we set

$$\xi_1(t,\delta,y) = (1-F_1(y))\left\{-\int_0^{t\wedge y} \frac{dH_1^u(s)}{(1-H_1(s))^2} + \frac{I(t \leq y, \delta = 1)}{1-H_1(t)}\right\},$$

$$\xi(t_2,\delta_2,y|t_1) = (1-F(y|t_1))\left\{-\int_0^{t_2\wedge y} \frac{dH^u(s|t_1)}{(1-H(s|t_1))^2} + \frac{I(t_2 \leq y, \delta_2 = 1)}{1-H(t_2|t_1)}\right\},$$

$$\tau(t_2,\delta_2,y|t_1) = \frac{\xi(t_2,\delta_2,y|t_1)}{1-G_1(t_1)},$$

and denote by $\tau'(t_2,\delta_2,y|t_1)$ and $\tau''(t_2,\delta_2,y|t_1)$ the first, respectively second derivative of $\tau(t_2,\delta_2,y|t_1)$ with respect to t_1. Also set

$$\eta(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y}) = \int_0^{y_1} F(y_2|z_1)\,d\xi_1(t_1,\delta_1,z_1) + I(\delta_1 = 1)I(t_1 \leq y_1)\tau(t_2,\delta_2,y_2|t_1),$$

$$\beta(\boldsymbol{y}) = \frac{1}{2}\int K(u)u^2\,du \int I(z_1 \leq y_1)\tau''(z_2,\delta_2,y_2|z_1)I(\delta_1 = 1)\,dH(\boldsymbol{z},\boldsymbol{\delta})$$
$$-2h_1^u(y_1)E[\tau'(\tilde{T}_2,\Delta_2,y_2|y_1)|T_1 = y_1]\int_0^L\int_v^L K(u)u\,du\,dv.$$

Let $\tau_2(y_1) < \inf\{y_2 : H(y_2|y_1) = 1\}$. Define $0 \leq \tau_1 < \infty$ such that $\tau_1 < \inf\{y : H_1(y) = 1\}$ and let $\Omega = \{\boldsymbol{y} : y_1 \leq \tau_1$ and $y_2 \leq \inf_{y \leq y_1}\tau_2(y)\}$. The following assumptions will be required in the results below :

(A1)(i) The sequence h_n satisfies $nh_n^5(\log n)^{-1} = O(1)$ and $\log n(nh_n)^{-1} \to 0$.
(ii) The probability density function K has compact support $[-L, L]$, K is twice continuously differentiable and $\int uK(u)\,du = 0$.
(A2)(i) $H_1(y)$ and $H_1^u(y)$ are three times continuously differentiable with respect to y.
(ii) $H(y_2|y_1)$ and $H^u(y_2|y_1)$ are twice continuously differentiable with respect to y_1 and y_2, and for $(y_1, y_2) \in \Omega$ all derivatives are uniformly bounded.

Theorem 2.1 *Under assumptions* $(A0)-(A2)$, *we have*

$$\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}) = n^{-1}\sum_{i=1}^n \eta(\tilde{\boldsymbol{T}}_i,\boldsymbol{\Delta}_i,\boldsymbol{y}) + h_n^2\beta(\boldsymbol{y}) + R_n(\boldsymbol{y}),$$

where

$$\sup_{\boldsymbol{y}\in\Omega}|R_n(\boldsymbol{y})| = o_P(n^{-1/2}) + o_P(h_n^2).$$

Theorem 2.2 *Assume (A0) − (A2). Then, if* $nh_n^4 \to K$ $(K \geq 0)$*,* $n^{1/2}(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}))$ $(\boldsymbol{y} \in \Omega)$ *converges weakly to a Gaussian process* $Z(\boldsymbol{y})$ *with covariance function*

$$Cov(Z(\boldsymbol{y}), Z(\boldsymbol{y}')) = E\left[\int_0^{y_1}\int_0^{y_1'} F(y_2|z_1)F(y_2'|z_1')\, d\xi_1(\tilde{T}_1, \Delta_1, z_1)\, d\xi_1(\tilde{T}_1, \Delta_1, z_1')\right]$$
$$+E[I(\Delta_1 = 1)I(\tilde{T}_1 \leq y_1 \wedge y_1')\tau(\tilde{T}_2, \Delta_2, y_2|\tilde{T}_1)\tau(\tilde{T}_2, \Delta_2, y_2'|\tilde{T}_1)]$$

and mean function

$$E(Z(\boldsymbol{y})) = K^{1/2}\beta(\boldsymbol{y}).$$

Note that the constant K is allowed to be 0, in which case the limiting process has zero mean.

Remark 2.1. We refer to Akritas and Van Keilegom (2003) for a bootstrap procedure that can be used to select an appropriate bandwidth. Note that contrary to the estimation of the conditional distribution, the choice of the bandwidth is not all that crucial for the estimator $\hat{F}(\boldsymbol{y})$ of the bivariate distribution. This is because the effect of the bandwidth, which enters the estimator $\hat{F}(y_2|y_1)$ of the conditional distribution, is averaged out over a range of y_1-values.

3 Estimation of the marginal distribution

Unlike the marginal distribution F_1, F_2 cannot be estimated by the Kaplan-Meier estimator. This is because the independence of T_2 and C_2 is not assumed under $(A0)$. An alternative could be to estimate $F_2(y)$ by $\hat{F}(\infty, y)$. However, the estimator $\hat{F}(y_1, y_2)$ is not consistent for large values of y_1, due to the inconsistency of the estimator $\hat{F}_1$ in the right tail. Hence, we will work with

$$\tilde{F}_2(y) = \int_0^{\tau_1} F(y|z)\, dF_1(z),$$

which can be made arbitrarily close to $F_2(y)$, provided $\inf\{y; H_1(y) = 1\} = \inf\{y; F_1(y) = 1\}$. The corresponding estimator is now defined by

$$\hat{F}_2(y) = \int_0^{\tau_1} \hat{F}(y|z)\, d\hat{F}_1(z). \tag{3.1}$$

This estimator generalizes the estimator of Cheng (1989) to the case where T_1 is subject to censoring.

Since the censoring mechanism is not symmetric in the two variables, estimation of the marginal distribution $F_1(y)$ is necessarily different. The Kaplan-Meier estimator $\hat{F}_1$ is consistent for F_1, while the estimator obtained by interchanging the role of the first and second variable in formula (3.1) is inconsistent for F_1, due to the lack of independence between T_1 and C_2; see the discussion regarding the generalized Beran estimator early in Section 2. Therefore, the Kaplan-Meier estimator $\hat{F}_1$ is the estimator of choice for F_1.

We continue with the asymptotic properties of $\hat{F}_2(y)$. Let $\Omega_2 = \{y_2 : y_2 \leq \inf_{y_1 \leq \tau_1} \tau_2(y_1)\}$.

Theorem 3.1 *Under assumptions* $(A0) - (A2)$, *we have*

$$\hat{F}_2(y) - \tilde{F}_2(y) = n^{-1} \sum_{i=1}^{n} \eta(\tilde{\boldsymbol{T}}_i, \boldsymbol{\Delta}_i, \tau_1, y) + h_n^2 \beta(\tau_1, y) + R_n(y),$$

where

$$\sup_{y \in \Omega_2} |R_n(y)| = o_P(n^{-1/2}) + o_P(h_n^2).$$

Theorem 3.2 *Assume* $(A0) - (A2)$. *Then, if* $nh_n^4 \to K$ $(K \geq 0)$, $n^{1/2}(\hat{F}_2(y) - \tilde{F}_2(y))$ $(y \in \Omega_2)$ *converges weakly to a Gaussian process* $W(y)$ *with covariance function*

$$Cov(W(y), W(y')) = E[\eta(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \tau_1, y)\eta(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \tau_1, y')]$$

and mean function

$$E(W(y)) = K^{1/2}\beta(\tau_1, y).$$

These results follow readily from Theorems 2.1 and 2.2.

4 Application to polynomial regression

Consider the following polynomial regression model :

$$T_2 = \beta_0 + \beta_1 T_1 + \ldots + \beta_p T_1^p + \sigma(T_1)\varepsilon,$$

where $\sigma^2(T_1) = \text{Var}(T_2|T_1)$, $E(\varepsilon|T_1) = 0$. As before, both T_1 and T_2 are subject to censoring. The variable T_2 will often represent a transformation of the (positive) survival time to the whole real line (like the logarithmic transformation). The methods developed in Section 2 will now be used in order to construct estimators of the coefficients $\beta_0, \ldots, \beta_p$. To our knowledge the estimation of the regression coefficients in polynomial regression when both the covariate and the response are subject to censoring, has not yet been considered in the literature. Let $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$. Then,

$$\boldsymbol{\beta} = \boldsymbol{C}^{-1} \begin{pmatrix} \int x^0 y \, dF(x,y) \\ \vdots \\ \int x^p y \, dF(x,y) \end{pmatrix},$$

where $\boldsymbol{C} = (\int x^{i+j-2}\, dF_1(x))$ $(i,j = 1,\ldots,p+1)$. The idea is now to replace the distributions $F_1(x)$ and $F(x,y)$ by respectively $\hat{F}_1(x)$ and $\hat{F}(x,y)$. However, these estimators are inconsistent for large values of x and y. Hence, what is estimated is actually

$$\boldsymbol{\beta}_\Omega = \boldsymbol{C}_\Omega^{-1} \begin{pmatrix} \int_\Omega x^0 y\, dF(x,y) \\ \vdots \\ \int_\Omega x^p y\, dF(x,y) \end{pmatrix},$$

where $\boldsymbol{C}_\Omega = (c_{ij}) = (\int_{-\infty}^{\tau_1} x^{i+j-2}\, dF_1(x))$ $(i,j = 1,\ldots,p+1)$. The vector $\boldsymbol{\beta}_\Omega$ can be made arbitrarily close to $\boldsymbol{\beta}$ provided $b_{F_1} \leq b_{G_1}$ and for any x, $b_{F(\cdot|x)} \leq b_{G(\cdot|x)}$, where for any distribution F, b_F denotes the upper bound of the support of F. The proposed estimator is now

$$\hat{\boldsymbol{\beta}}_\Omega = \hat{\boldsymbol{C}}_\Omega^{-1} \begin{pmatrix} \int_\Omega x^0 y\, d\hat{F}(x,y) \\ \vdots \\ \int_\Omega x^p y\, d\hat{F}(x,y) \end{pmatrix},$$

where $\hat{\boldsymbol{C}}_\Omega = (\int_{-\infty}^{\tau_1} x^{i+j-2}\, d\hat{F}_1(x))$ $(i,j = 1,\ldots,p+1)$. Note that if T_2 is some monotone transformation of the survival time, say $T_2 = g(Y_2)$, then $F(x,y) = P(T_1 \leq x, T_2 \leq y) = P(T_1 \leq x, Y_2 \leq g^{-1}(y))$. We estimate $F(x,y)$ then from the latter expression, since $Y_2 \geq 0$ (which is required for the asymptotic theory of Section 2). The estimator $\hat{\boldsymbol{\beta}}_\Omega$ is an extension of the estimator proposed by Akritas (1994) to the case where the covariate is subject to censoring. The asymptotic expansion given by Theorem 5.1 in the latter paper can easily be generalized to the present setup, by using arguments similar to those in the proof of Lemma A.2 and Theorem 2.1. It suffices to replace the empirical distribution function of T_1 by the Kaplan-Meier estimator $\hat{F}_1$, and to use the representation for $\hat{F}_1 - F_1$ given by Theorem 1 in Lo and Singh (1986). This leads to the following representation, provided the distribution of T_1 has bounded support, the distribution of the vector $(T_1, T_1^2, \ldots, T_1^p)$ is non-singular and $nh_n^4 \to 0$:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_\Omega - \boldsymbol{\beta}_\Omega = {} & n^{-1}\sum_{i=1}^n \boldsymbol{C}_\Omega^{-1}\boldsymbol{T}_{i1} I(T_{i1} \leq \tau_1, \Delta_{i1} = 1) \int_{-\infty}^{\tau_2(T_{i1})} y\, d\tau(\tilde{T}_{i2}, \Delta_{i2}, y|T_{i1}) \\ & + n^{-1}\sum_{i=1}^n \boldsymbol{C}_\Omega^{-1} \int_\Omega \boldsymbol{x} y\, dF(y|x)\, d\xi_1(\tilde{T}_{i1}, \Delta_{i1}, x) \\ & + n^{-1}\sum_{i=1}^n \sum_{k=1}^{p+1}\sum_{\ell=1}^{p+1} G_{k\ell}(\boldsymbol{C}_\Omega) \int_\Omega \boldsymbol{x} y\, dF(x,y) \int_{-\infty}^{\tau_1} x^{k+\ell-2}\, d\xi_1(\tilde{T}_{i1}, \Delta_{i1}, x) \\ & + o_P(n^{-1/2}),\end{aligned}$$

where $\boldsymbol{T}_{i1} = (T_{i1}^0, T_{i1}^1, \ldots, T_{i1}^p)'$, $\boldsymbol{x} = (x^0, x^1, \ldots, x^p)'$, $G_{k\ell}(\boldsymbol{C}_\Omega) = (g_{k\ell rs}(\boldsymbol{C}_\Omega))$ $(r, s = 1, \ldots, p+1)$ and $g_{k\ell rs}(\boldsymbol{C}_\Omega)$ is the partial derivative with respect to $c_{k\ell}$ of the element (r, s) of the matrix $\boldsymbol{C}_\Omega^{-1}$.

The asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}_\Omega$ follows now immediately from the above representation together with the central limit theorem.

Appendix

We start with two lemmas needed in the proofs.

Lemma A.1 *Under assumptions* $(A0) - (A2)$,

$$\sup_{\boldsymbol{y}\in\Omega} \left| n^{-1} \sum_{\Delta_{i1}=1} I(y_1 - Lh_n \leq T_{i1} \leq y_1)\tau(\tilde{T}_{i2}, \Delta_{i2}, y_2|T_{i1}) \int_{\frac{y_1 - T_{i1}}{h_n}}^{L} K(u)\, du \right| = o_P(n^{-1/2}).$$

Proof. The proof is based on results in van der Vaart and Wellner (1996). Let

$$\mathcal{F} = \left\{ I(\Delta_1 = 1) I(y_1 - L\delta \leq T_1 \leq y_1)\tau(\tilde{T}_2, \Delta_2, y_2|T_1) \int_{\frac{y_1 - T_1}{\delta}}^{L} K(u)\, du : \boldsymbol{y} \in \Omega, 0 < \delta < 1 \right\}.$$

In a first step we will show that the class $\mathcal{F}$ is Donsker. By Theorem 2.5.6 in van der Vaart and Wellner (1996), it suffices to show that $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty$, where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of $(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta})$, and $L_2(P)$ is the L_2-norm. We start with $I(y_1 - L\delta \leq T_1 \leq y_1) = I(T_1 \leq y_1) - I(T_1 \leq y_1 - L\delta)$, which is the difference of two monotone and bounded functions and hence, by Theorem 2.7.5 in van der Vaart and Wellner (1996), these functions require $O(\exp(K\varepsilon^{-1}))$ brackets. Also $\int_{\frac{y_1 - T_1}{\delta}}^{L} K(u)\, du$ is monotone and bounded and hence it requires the same amount of brackets. For $\tau(\tilde{T}_2, \Delta_2, y_2|T_1)$, note that $1 - F(y_2|T_1)$ as well as its first derivative with respect to T_1 is bounded and hence by Corollary 2.7.2 in van der Vaart and Wellner (1996), its bracketing number is $O(\exp(K\varepsilon^{-1}))$. We apply the same corollary to the integral in $\tau(\tilde{T}_2, \Delta_2, y_2|T_1)$, but as the integral depends on both T_1 and $\tilde{T}_2$, it has to be twice continuously differentiable with respect to both T_1 and $\tilde{T}_2$ in order to make sure that $O(\exp(K\varepsilon^{-1}))$ brackets are sufficient. Finally, the second term between brackets in the expression of $\tau(\tilde{T}_2, \Delta_2, y_2|T_1)$ is the ratio of a monotone function and a function which is twice continuously differentiable. Hence its bracketing number is also $O(\exp(K\varepsilon^{-1}))$. This shows that the bracketing number of the class $\mathcal{F}$ is $O(\exp(K\varepsilon^{-1}))$ and hence $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty$, since for $\varepsilon > 2M$ one bracket suffices (where M is an upper bound for $|\tau(\tilde{T}_2, \Delta_2, y_2|T_1)|$). This shows that the class $\mathcal{F}$ is Donsker. Next, some straightforward calculations show that

$$E[\xi(\tilde{T}_2, \Delta_2, y_2|T_1)|T_1]^2 = (1 - F(y_2|T_1))^2 \int_0^{y_2} \frac{dH^u(s|T_1)}{(1 - H(s|T_1))^2}$$

and hence,

$$\begin{aligned}
&\mathrm{Var}\left[I(\Delta_1 = 1)I(y_1 - Lh_n \le T_1 \le y_1)\tau(\tilde{T}_2, \Delta_2, y_2|T_1) \int_{\frac{y_1 - T_1}{h_n}}^{L} K(u)\,du\right] \\
&\le E\left[I(y_1 - Lh_n \le T_1 \le y_1)\frac{(1 - F(y_2|T_1))^2}{(1 - G_1(T_1))^2} \int_0^{y_2} \frac{dH^u(s|T_1)}{(1 - H(s|T_1))^2} \{ \int_{\frac{y_1 - T_1}{h_n}}^{L} K(u)\,du\}^2\right]
\end{aligned}$$

which is bounded by Kh_n for some $K > 0$, since the integrals in the above expression are bounded uniformly over all $\boldsymbol{y} \in \Omega$ and $T_1 \le y_1$. Since the class $\mathcal{F}$ is Donsker, it follows from Corollary 2.3.12 in van der Vaart and Wellner (1996) that

$$\lim_{\alpha \downarrow 0} \limsup_{n \to \infty} P\left(\sup_{f \in \mathcal{F}, \mathrm{Var}(f) < \alpha} n^{-1/2} \left| \sum_{i=1}^n f(X_i) \right| > \varepsilon \right) = 0,$$

for each $\varepsilon > 0$. By restricting the supremum inside this probability to the elements in $\mathcal{F}$ corresponding to $\delta = h_n$, the result follows.

Lemma A.2 *Under assumptions* $(A0) - (A2)$,

$$\begin{aligned}
&\int_0^{y_1} \{\hat{F}(y_2|z_1) - F(y_2|z_1)\}\,dF_1(z_1) \\
&= n^{-1} \sum_{i=1}^n I(T_{i1} \le y_1, \Delta_{i1} = 1)\tau(\tilde{T}_{i2}, \Delta_{i2}, y_2|T_{i1}) + h_n^2 \beta(\boldsymbol{y}) + R_n(\boldsymbol{y}), \qquad \text{(A.1)}
\end{aligned}$$

where

$$\sup_{\boldsymbol{y} \in \Omega} |R_n(\boldsymbol{y})| = o_P(n^{-1/2}) + o_P(h_n^2).$$

Proof. We apply to $\hat{F}(y_2|z_1) - F(y_2|z_1)$ the representation for the Beran estimator given in Van Keilegom and Veraverbeke (1997). (Note that the representation in that paper is valid for fixed design; however, a very similar proof can be given for the present situation of random design.) Hence, using the fact that $(1 - G_1(z_1))^{-1} = f_1(z_1)/h_1^u(z_1)$, the left hand side of (A.1) equals

$$\begin{aligned}
&n^{-1}{\textstyle\sum_i^u} \int K_h(z_1 - T_{i1})\tau_i(y_2|z_1)I(z_1 \le y_1)\,dz_1 + o_P(n^{-1/2}) \\
&= n^{-1}{\textstyle\sum_i^u} \int K_h(z_1 - T_{i1})\tau_i(y_2|z_1)[I(z_1 \le y_1) - I(T_{i1} \le y_1)]\,dz_1 \\
&\quad + n^{-1}{\textstyle\sum_i^u} I(T_{i1} \le y_1) \int K_h(z_1 - T_{i1})\tau_i(y_2|z_1)\,dz_1 + o_P(n^{-1/2}) \\
&= A_1(\boldsymbol{y}) + A_2(\boldsymbol{y}) + o_P(n^{-1/2}),
\end{aligned}$$

where $\sum_i^u = \sum_{\Delta_{i1}=1}$, $K_h(u) = h_n^{-1}K(u/h_n)$ and $\tau_i(y_2|y_1) = \tau(\tilde{T}_{i2}, \Delta_{i2}, y_2|y_1)$. Now,

$$\begin{aligned}
A_1(\boldsymbol{y}) &= -n^{-1}{\sum}_i^u I(y_1 - Lh_n \le T_{i1} \le y_1) \int_{y_1}^{T_{i1}+Lh_n} K_h(z_1 - T_{i1})\tau_i(y_2|z_1)\, dz_1 \\
&\quad + n^{-1}{\sum}_i^u I(y_1 \le T_{i1} \le y_1 + Lh_n) \int_{T_{i1}-Lh_n}^{y_1} K_h(z_1 - T_{i1})\tau_i(y_2|z_1)\, dz_1 \\
&= -A_{11}(\boldsymbol{y}) + A_{12}(\boldsymbol{y}).
\end{aligned}$$

The derivation for the terms $A_{11}(\boldsymbol{y})$ and $A_{12}(\boldsymbol{y})$ is similar. Consider e.g.

$$\begin{aligned}
&A_{11}(\boldsymbol{y}) \\
&= n^{-1}{\sum}_i^u I(y_1 - Lh_n \le T_{i1} \le y_1)\tau_i(y_2|T_{i1}) \int_{y_1}^{T_{i1}+Lh_n} K_h(z_1 - T_{i1})\, dz_1 \\
&\quad + n^{-1}{\sum}_i^u I(y_1 - Lh_n \le T_{i1} \le y_1)\tau_i'(y_2|T_{i1}) \int_{y_1}^{T_{i1}+Lh_n} K_h(z_1 - T_{i1})(z_1 - T_{i1})\, dz_1 \\
&\quad + \frac{1}{2}n^{-1}{\sum}_i^u I(y_1 - Lh_n \le T_{i1} \le y_1) \int_{y_1}^{T_{i1}+Lh_n} \tau_i''(y_2|\xi_{i1})K_h(z_1 - T_{i1})(z_1 - T_{i1})^2\, dz_1 \\
&= A_{111}(\boldsymbol{y}) + A_{112}(\boldsymbol{y}) + A_{113}(\boldsymbol{y}),
\end{aligned}$$

where ξ_{i1} is between z_1 and T_{i1}. The term $A_{111}(\boldsymbol{y})$ equals

$$n^{-1}{\sum}_i^u I(y_1 - Lh_n \le T_{i1} \le y_1)\tau_i(y_2|T_{i1}) \int_{\frac{y_1 - T_{i1}}{h_n}}^{L} K(u)\, du$$

and this is $o_P(n^{-1/2})$, uniformly in $\boldsymbol{y} \in \Omega$, by Lemma A.1. Also $A_{113}(\boldsymbol{y})$ is $o_P(n^{-1/2})$, since it is easily seen that this term is bounded by

$$\frac{1}{2}h_n^3(nh_n)^{-1} \sup_{\boldsymbol{y}\in\Omega, z\le y_1, t, \delta} |\tau''(t, \delta, y_2|z)| {\sum}_i^u I(y_1 - Lh_n \le T_{i1} \le y_1) \int K(u)u^2\, du$$

which is $O_P(h_n^3) = o_P(n^{-1/2})$. The term $A_{112}(\boldsymbol{y})$ equals

$$h_n n^{-1}{\sum}_i^u I(y_1 - Lh_n \le T_{i1} \le y_1)\tau_i'(y_2|T_{i1}) \int_{\frac{y_1 - T_{i1}}{h_n}}^{L} K(u)u\, du.$$

In a very similar way as was shown in Lemma A.1 for the term $A_{111}(\boldsymbol{y})$, it can be shown that $\sup_{\boldsymbol{y}\in\Omega} |A_{112}(\boldsymbol{y}) - EA_{112}(\boldsymbol{y})| = o_P(n^{-1/2})$. It follows that $A_{11}(\boldsymbol{y})$ equals

$$\begin{aligned}
&EA_{112}(\boldsymbol{y}) + o_P(n^{-1/2}) \\
&= h_n \int_{y_1 - Lh_n}^{y_1} E[\tau_1'(y_2|t)|T_1 = t] \int_{\frac{y_1 - t}{h_n}}^{L} K(u)u\, du\, dH_1^u(t) + o_P(n^{-1/2}) \\
&= h_n^2 \int_0^L E[\tau_1'(y_2|y_1 - vh_n)|T_1 = y_1 - vh_n] \int_v^L K(u)u\, du\, h_1^u(y_1 - vh_n)\, dv + o_P(n^{-1/2})
\end{aligned}$$

$$
\begin{aligned}
&= h_n^2 \int_0^L \{Q(y_1 - vh_n) - Q(y_1)\} \int_v^L K(u)u\,du\,dv + h_n^2 Q(y_1) \int_0^L \int_v^L K(u)u\,du\,dv + o_P(n^{-1/2}) \\
&= h_n^2 \int_0^L Q'(\xi_v)(-vh_n) \int_v^L K(u)u\,du\,dv + h_n^2 Q(y_1) \int_0^L \int_v^L K(u)u\,du\,dv + o_P(n^{-1/2}) \\
&= h_n^2 h_1^u(y_1) E[\tau_1'(y_2|y_1)|T_1 = y_1] \int_0^L \int_v^L K(u)u\,du\,dv + o_P(n^{-1/2}),
\end{aligned}
$$

where $Q(y) = h_1^u(y)E[\tau_1'(y_2|y)|T_1 = y]$ and $y_1 - vh_n \leq \xi_v \leq y_1$. It remains to consider $A_2(\boldsymbol{y})$, which equals

$$
\begin{aligned}
& n^{-1}\sum\nolimits_i^u I(T_{i1} \leq y_1)\tau_i(y_2|T_{i1}) \int K_h(z_1 - T_{i1})\,dz_1 \\
& + n^{-1}\sum\nolimits_i^u I(T_{i1} \leq y_1)\tau_i'(y_2|T_{i1}) \int K_h(z_1 - T_{i1})(z_1 - T_{i1})\,dz_1 \\
& + \frac{1}{2} n^{-1}\sum\nolimits_i^u I(T_{i1} \leq y_1)\tau_i''(y_2|T_{i1}) \int K_h(z_1 - T_{i1})(z_1 - T_{i1})^2\,dz_1 + o_P(h_n^2) \\
& = n^{-1}\sum\nolimits_i^u I(T_{i1} \leq y_1)\tau_i(y_2|T_{i1}) + \frac{1}{2}\mu_2^K h_n^2 n^{-1}\sum\nolimits_i^u I(T_{i1} \leq y_1)\tau_i''(y_2|T_{i1}) + o_P(h_n^2) \\
& = n^{-1}\sum\nolimits_i^u I(T_{i1} \leq y_1)\tau_i(y_2|T_{i1}) + \frac{1}{2}\mu_2^K h_n^2 \int I(z_1 \leq y_1)\tau''(z_2, \delta_2, y_2|z_1) I(\delta_1 = 1)\,dH(\boldsymbol{z}, \boldsymbol{\delta}) \\
& \quad + \frac{1}{2}\mu_2^K h_n^2 \int I(z_1 \leq y_1)\tau''(z_2, \delta_2, y_2|z_1) I(\delta_1 = 1)\,d(\hat{H}(\boldsymbol{z}, \boldsymbol{\delta}) - H(\boldsymbol{z}, \boldsymbol{\delta})) + o_P(h_n^2),
\end{aligned}
$$

where $\mu_2^K = \int K(u)u^2\,du$ and $\hat{H}(\boldsymbol{z}, \boldsymbol{\delta}) = n^{-1}\sum_{i=1}^n I(\tilde{T}_{i1} \leq z_1, \tilde{T}_{i2} \leq z_2, \Delta_{i1} = \delta_1, \Delta_{i2} = \delta_2)$. To show that the last term of the above sum is $o_P(h_n^2)$, we will show that the integral in this term tends to zero, uniformly in $\boldsymbol{y} \in \Omega$. Proving this is equivalent to showing that the class $\mathcal{F} = \{I(\tilde{T}_1 \leq y_1)\tau''(\tilde{T}_2, \Delta_2, y_2|\tilde{T}_1)I(\Delta_1 = 1) : \boldsymbol{y} \in \Omega\}$ is Glivenko-Cantelli (see p. 81 in van der Vaart and Wellner (1996)). Using Theorem 2.4.1 in the same book, we therefore need to show that the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$ for all $\varepsilon > 0$ (using the same notations as in Lemma A.1), which can be done in a very similar way as in the proof of Lemma A.1.

Proof of Theorem 2.1. Write

$$
\begin{aligned}
\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}) &= \int_0^{y_1} \{\hat{F}(y_2|z_1) - F(y_2|z_1)\}\,dF_1(z_1) + \int_0^{y_1} F(y_2|z_1)\,d(\hat{F}_1(z_1) - F_1(z_1)) \\
&\quad + \int_0^{y_1} \{\hat{F}(y_2|z_1) - F(y_2|z_1)\}\,d(\hat{F}_1(z_1) - F_1(z_1)) \\
&= A_1(\boldsymbol{y}) + A_2(\boldsymbol{y}) + A_3(\boldsymbol{y}).
\end{aligned}
$$

For the proof of $A_1(\boldsymbol{y})$ we refer to Lemma A.2. To $A_2(\boldsymbol{y})$ we apply the representation for the Kaplan-Meier estimator given by Theorem 1 of Lo and Singh (1986):

$$A_2(\boldsymbol{y}) = n^{-1}\sum_{i=1}^{n}\int_0^{y_1} F(y_2|z_1)\,d\xi_1(\tilde{T}_{i1},\Delta_{i1},z_1) + o_P(n^{-1/2}).$$

Next, using the notation $\dot{F}(y_2|y_1)$ for the partial derivative of $F(y_2|y_1)$ with respect to y_1 and similarly for $\dot{\hat{F}}(y_2|y_1)$,

$$\begin{aligned}A_3(\boldsymbol{y}) &= \{\hat{F}(y_2|y_1) - F(y_2|y_1)\}\{\hat{F}_1(y_1) - F_1(y_1)\}\\ &\quad - \int_0^{y_1}\{\hat{F}_1(z_1) - F_1(z_1)\}\,\{\dot{\hat{F}}(y_2|z_1) - \dot{F}(y_2|z_1)\}\,dz_1\end{aligned}$$

and this is $o_P(n^{-1/2})$, because $\sup_{y\le\tau_1}|\hat{F}_1(y) - F_1(y)| = O(n^{-1/2}(\log n)^{1/2})$ a.s. (see e.g. Lo and Singh (1986)),

$$\sup_{y_1}\sup_{y_2\le\tau_2(y_1)}|\hat{F}(y_2|y_1) - F(y_2|y_1)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \quad a.s.$$

$$\sup_{y_1}\sup_{y_2\le\tau_2(y_1)}|\dot{\hat{F}}(y_2|y_1) - \dot{F}(y_2|y_1)| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \quad a.s.$$

(see Van Keilegom and Akritas (1999), applied to the restricted data set $\{(\tilde{\boldsymbol{T}}_i,\boldsymbol{\Delta}_i);\ \Delta_{i1}=1\}$).

Proof of Theorem 2.2. Using Theorem 2.5.6 in van der Vaart and Wellner (1996), we need to show that $\int_0^\infty \sqrt{\log N_{[\,]}(\varepsilon,\mathcal{F},L_2(P))}\,d\varepsilon < \infty$, where $N_{[\,]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of $(\tilde{\boldsymbol{T}},\boldsymbol{\Delta})$, $L_2(P)$ is the L_2-norm, and $\mathcal{F} = \{\eta(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\boldsymbol{y}) : \boldsymbol{y}\in\Omega\}$. Proving this entails that the class $\mathcal{F}$ is Donsker and hence the weak convergence of the given process follows from p. 81-82 in van der Vaart and Wellner's book. We start with the first term of $\eta(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y})$. Write

$$\begin{aligned}\int_0^{y_1} F(y_2|z_1)\,d\xi_1(t_1,\delta_1,z_1) &= F(y_2|y_1)\xi_1(t_1,\delta_1,y_1) - \int_0^{y_1}\xi_1(t_1,\delta_1,z_1)\,dF(y_2|z_1)\\ &= T_1(\boldsymbol{y}) + T_2(\boldsymbol{y}).\end{aligned}$$

Since $F(y_2|y_1)$ is deterministic and bounded, it requires only $O(\varepsilon^{-1})$ brackets. The first term of $\xi_1(t_1,\delta_1,y_1)$ is uniformly bounded over y_1, as well as its first derivative with respect to t_1. Hence, this term can be covered by $O(\exp(K\varepsilon^{-1}))$ brackets by Corollary 2.7.2 in van der Vaart and Wellner (1996). The indicator $I(t_1\le y_1)$ is obviously monotone and bounded, and its bracketing number is therefore $O(\exp(K\varepsilon^{-1}))$ by Theorem 2.7.5 in the same book. Since the remaining factors of the second term of $\xi_1(t_1,\delta_1,y_1)$ do not depend on y_1 and y_2 and are bounded, this shows that the bracketing number of $T_1(\boldsymbol{y})$ is $O(\exp(K\varepsilon^{-1}))$. Next, write

$$T_2(\boldsymbol{y}) = \int_0^{y_1} (1 - F_1(z_1)) \int_0^{t_1 \wedge z_1} \frac{dH_1^u(s)}{(1 - H_1(s))^2} \, dF(y_2|z_1) \\ - \frac{I(t_1 \leq y_1, \delta_1 = 1)}{1 - H_1(t_1)} \int_{t_1}^{y_1} (1 - F_1(z_1)) dF(y_2|z_1).$$

The integrals in the first and second term above are bounded and have bounded derivatives with respect to t_1 and hence they need $O(\exp(K\varepsilon^{-1}))$ brackets. The expression in front of the integral in the second term also requires $O(\exp(K\varepsilon^{-1}))$ brackets since it consists of monotone functions. This shows that the bracketing number of $T_2(\boldsymbol{y})$ is also $O(\exp(K\varepsilon^{-1}))$. It remains to calculate the bracketing number of the second term of $\eta(\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{y})$. In the proof of Lemma A.1, it was shown that also this term requires $O(\exp(K\varepsilon^{-1}))$ brackets. Hence, $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))} \, d\varepsilon < \infty$, since for $\varepsilon > 2M$ one bracket suffices (where M is an upper bound for $|\eta(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \boldsymbol{y})|$).

Acknowledgments. The author would like to thank Michael Akritas for helpful discussions and useful suggestions.

References

Akritas, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann. Statist.*, **22**, 1299-1327.

Akritas, M. G. and Van Keilegom, I. (2003). Estimation of the bivariate and marginal distributions with censored data. *J. Royal Statist. Soc. - Series B*, **65**, 457-471.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.

Cheng, P. E. (1989). Nonparametric estimation of survival curve under dependent censorship. *J. Statist. Plan. Infer.*, **23**, 181-191.

Heagerty, P. J., Lumley, T. and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, **56**, 337-344.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457-481.

Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap : some asymptotic representations. *Probab. Th. Rel. Fields*, **71**, 455-465.

van der Laan, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.*, **24**, 596-627.

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

Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.*, **27**, 1745-1784.

Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, **49**, 467-491.