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

0106

ESTIMATION OF THE BIVARIATE AND MARGINAL DISTRIBUTIONS WITH CENSORED DATA

M.G. AKRITAS and I. VAN KEILEGOM

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

Estimation of the Bivariate and Marginal Distributions with Censored Data

Michael G. Akritas
Department of Statistics
Pennsylvania State University
U.S.A.

Ingrid Van Keilegom[1]
Institut de Statistique
Université Catholique de Louvain
Belgium

February 12, 2001

Abstract

New estimators for the bivariate and marginal distributions when both variables are subject to censoring are proposed under two censoring schemes. The estimators of the marginal distributions are not the marginals of the corresponding estimator of the bivariate distribution. Under the usual scheme which assumes that the censoring variables are independent from the variables of interest, the proposed estimator of the marginal distribution outperforms the Kaplan-Meier estimator. Under the assumption that one of the censoring variables is conditionally independent of the survival time it censors given the other survival time, the proposed estimator of the marginal distribution generalizes the estimator of Cheng (1989). All estimators require estimation of the conditional distribution when the conditioning variable is subject to censoring. Such a method of estimation is proposed. The weak convergence of the proposed estimators is obtained. Simulation results suggest that they perform well relative to existing estimators.

KEY WORDS: Kernel estimation; Conditional distribution; Bivariate distribution; Marginal distribution; Right censoring; Asymptotic representation; Weak convergence.

[1] This research was supported by 'Projet d'Actions de Recherche Concertées', No. 98/03–217, from the Belgian government.

1 Introduction

The problem of estimating the bivariate distribution of $\boldsymbol{T} = (T_1, T_2)$ when both variables are subject to random censoring has received considerable attention in the statistical literature. Nonparametric estimators have been proposed by (among others) Campbell and Földes (1982), Tsai, Leurgans and Crowley (1986), Dabrowska (1988), Pruitt (1991), Prentice and Cai (1992) and van der Laan (1996), while e.g. Oakes (1989) and Wang and Wells (1999) proposed semiparametric estimators. All estimators in these papers are restricted to the case where the bivariate censoring variable $\boldsymbol{C} = (C_1, C_2)$ is independent from $\boldsymbol{T}$. Many of the estimators are not proper bivariate distributions (see Pruitt 1993b) and, though certain efficiency results have been obtained under additional restrictions (see Gill, van der Laan and Wellner 1993, van der Laan 1996), none is shown to be efficient under the full generality allowed by the aforementioned independent censoring scheme. Finally, estimation of the marginal distribution is not discussed explicitly in this literature giving the impression that the Kaplan-Meier estimator is still the estimator of choice.

In this paper we consider estimation of the bivariate and marginal distributions under two scenaria. First when $\boldsymbol{C}$ and $\boldsymbol{T}$ are independent and second when T_2 is conditionally independent from C_2 given T_1. One situation where the second scenario is relevant is when T_2 is the survival time of the 'dependent' of two related individuals, 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. The proposed bivariate estimator under the first scenario leads to an estimator of the marginal distribution that outperforms the Kaplan-Meier estimator. Under the second scenario, the resulting estimator of the marginal distribution generalizes the estimator of Cheng (1989).

When there is no censoring, the usual empirical distribution function is the estimator of choice for estimating the bivariate and marginal distributions. Under independent censoring, however, the Kaplan-Meier estimator of T_2 can be improved by taking advantage of the available information regarding T_1, and vice-versa. To see this, suppose that the second coordinate of the i-th observation is censored. In this case the Kaplan-Meier estimator for the distribution of T_2 redistributes the mass that it would have received to all observations (of the second coordinate) known to be larger than it. This scheme fails to take advantage of the information contained in the first coordinate of the i-th observation. Thus, if the correlation of the two coordinates is positive, if the first coordinate of the i-th observation is large (in relation to values of T_1), and if the second coordinate is censored

Table 1: *Mean squared error of the Kaplan-Meier estimator $\tilde{F}(y)$ and the new estimator $\hat{F}(y)$.*

$P(\Delta_2 = 0)$	ρ_T	n	$y = F^{-1}(0.25)$		$y = F^{-1}(0.50)$		$y = F^{-1}(0.75)$	
			$\tilde{F}(y)$	$\hat{F}(y)$	$\tilde{F}(y)$	$\hat{F}(y)$	$\tilde{F}(y)$	$\hat{F}(y)$
0.3	0.3	50	.00384	.00383	.00589	.00579	.00582	.00532
		100	.00200	.00199	.00308	.00304	.00268	.00258
	0.6	50	.00388	.00384	.00624	.00537	.00584	.00525
		100	.00191	.00189	.00284	.00229	.00279	.00265
	0.9	50	.00414	.00405	.00620	.00224	.00597	.00544
		100	.00202	.00200	.00276	.00094	.00279	.00254
0.6	0.3	50	.00499	.00499	.00969	.00936	.01623	.01300
		100	.00250	.00249	.00524	.00513	.00811	.00727
	0.6	50	.00485	.00483	.00999	.00909	.01585	.01243
		100	.00238	.00235	.00496	.00444	.00802	.00718
	0.9	50	.00508	.00482	.01061	.00708	.01761	.01355
		100	.00259	.00237	.00462	.00279	.00820	.00719

near its origin, then it makes sense that more of its mass be redistributed to the large (in relation to values of T_2) observations of the second coordinate. The estimator proposed in this paper achieves precisely this. Table 1 presents the results of a small simulation study comparing the proposed and the Kaplan-Meier estimators. In this simulation $\boldsymbol{T}$ was generated from a bivariate normal distribution with zero mean, variance one for both coordinates and correlation (denoted by ρ_T in the table) of 0.3, 0.6 or 0.9. Only T_2 is subject to censoring and we considered estimation of the marginal distribution of T_2. The distribution of the censoring variable C_2 is normal, but with mean m chosen to achieve the two considered censoring probabilities (0.3 and 0.6). The results are based on 1000 simulation runs. It is readily seen that the advantage of the proposed estimator is more pronounced in the right tail and increases as the proportion of censored data and the correlation ρ_T increase.

The proposed estimators of the bivariate distribution use the method of Akritas (1994), who considered the case where one of the survival times is uncensored (this is a typical situation in regression problems arising in biostatistics). The method consists of averaging conditional distributions over an appropriate range, produces estimators which are proper distribution functions, allows the censoring variable to depend on the uncensored survival time and was shown to be efficient.

As a first step in implementing this approach we need to consider estimation of the conditional distribution of, say, T_2 given $T_1 = t$ when T_1 is also subject to censoring. This poses the problem of defining suitable windows, since it is not clear when a censored value of T_1 should or should not be included in the window. To our knowledge, this has not been addressed in the statistical literature. We adopted the approach of allowing in the window only pairs of observations for which the conditioning variable is uncensored. That this is a valid approach is justified in Section 2. The second step for estimating the bivariate distribution consists of averaging the conditional distributions. Unlike the estimator in Akritas (1994) where the conditioning variable is uncensored, we now use the Kaplan-Meier estimator of the marginal distribution of the conditioning variable in order to average the conditional distributions. This is one way that information contained in the censored values of the conditioning variable is utilized. The proposed estimator for the scenario that T_2 is conditionally independent from C_2 given T_1 is precisely such an average of conditional distributions. Under the stronger assumption that $\boldsymbol{C}$ and $\boldsymbol{T}$ are independent, the estimator of the bivariate distribution should not depend on which variable we consider as the conditioning variable. Therefore, we treat the two variables symmetrically by interchanging roles. Thus, the proposed estimator for the bivariate distribution in this case is a weighted average of the two bivariate estimators obtained by considering each of the two variables as the conditioning variable. Treating the two variables symmetrically is another way of utilizing information contained in the censored values. The estimators under both scenaria have intuitive, closed form expressions and are easy to compute.

The proposed estimators for the marginal distributions are not the marginals of the corresponding estimators for the bivariate distribution. This is because the unassigned mass of the estimated bivariate distribution renders its marginal distributions inconsistent estimators. It turns out that, under the assumption of conditional independence, our estimator generalizes the estimator of Cheng (1989) to the case where the conditioning variable is subject to censoring. Under the assumption of independence, the proposed estimator is a weighted average of the Kaplan-Meier estimator and the generalized Cheng estimator. The simulation reported above shows that it outperforms the Kaplan-Meier estimator.

The results in this paper have several applications. First, the new estimators of the marginal distribution can be used for constructing improved matched pairs tests. A second application is to regression problems arising in astronomy, where it is common for both variables to be subject to censoring. One of the attractive features of the adopted approach of Akritas (1994) is that it allows for transfer of tail information under an assumed nonparametric regression model (see Van Keilegom and Akritas 1999). The transfer of tail information results in better estimation of the tails of the conditional distributions, reducing thus the region of unassigned mass of the estimator of the bivariate distribution. Such reduction is useful for a practical extension of the least squares estimator to censored data. Finally, estimation of the bivariate distribution has recently been applied to ROC curves for censored data in Heagerty, Lumley and Pepe (2000). The proposed estimator for the case that T_2 is conditionally independent from C_2 given T_1 permits a useful extension of their methodology to the case where the diagnostic marker is also subject to censoring. Some of these applications will be explored elsewhere.

The next section gives a precise description of the estimators. Section 3 presents the main results regarding the asymptotic behavior of the estimators. Section 4 presents results from a simulation study comparing the present estimator of the bivariate distribution with the estimator of Pruitt (1991) and the nonparametric maximum likelihood estimator (NPMLE) of van der Laan (1996). Finally, Section 5 contains the proofs of the main results.

2 Description of the estimators

Let $\boldsymbol{T} = (T_1, T_2)$ be the pair of survival times and $\boldsymbol{C} = (C_1, C_2)$ the pair of censoring times. We observe $\tilde{\boldsymbol{T}} = (\tilde{T}_1, \tilde{T}_2) = (T_1 \wedge C_1, T_2 \wedge C_2)$ and $\boldsymbol{\Delta} = (\Delta_1, \Delta_2) = (I(T_1 \leq C_1), I(T_2 \leq C_2))$. In this section we will describe the proposed estimators for F, the bivariate distribution function of $\boldsymbol{T}$, and F_1, F_2, the marginal distributions of T_1, T_2, respectively.

The proposed estimators are all based on estimators of $F_{2|1}(y_2|y_1)$ and $F_{1|2}(y_1|y_2)$, the conditional distribution functions of T_2 given $T_1 = y_1$, and of T_1 given $T_2 = y_2$, respectively. Thus we first describe such estimators, which are obtained by generalizing Beran's (1981) estimator to the case where the conditioning variable is subject to censoring. The main problem for doing so is that when for a pair of observations the value of the conditioning variable is censored, it is not clear whether or not it belongs to a particular window. We adopted the approach of considering only those pairs in the window for which the value of the conditioning variable is uncensored, and computed the usual kernel estimator of the survival function in each such window (Beran 1981). Note that by doing so we are really estimating $P(T_2 \leq y_2 | T_1 = y_1, \Delta_1 = 1)$. However, if T_2 is independent of C_1, these two probabilities are equal, since

$$P(T_2 \leq y_2 | T_1 = y_1, \Delta_1 = 1) \;=\; P(T_2 \leq y_2 | T_1 = y_1, C_1 \geq y_1) \;=\; P(T_2 \leq y_2 | T_1 = y_1).$$

Letting K be a known probability density function (kernel) and $\{h_n\}$ a sequence of positive constants tending to zero as n tends to infinity (bandwidth sequence), the proposed extension of the Beran (1981) estimator is

$$\hat{F}_{2|1}(y_2|y_1) = 1 - \prod_{\tilde{T}_{2i} \leq y_2, \Delta_{2i}=1} \left(1 - \frac{W_{n1i}(y_1; h_n)}{\sum_{j=1}^{n} W_{n1j}(y_1; h_n) I(\tilde{T}_{2j} \geq \tilde{T}_{2i})}\right), \tag{2.1}$$

where

$$W_{n1i}(y; h_n) = \begin{cases} \dfrac{K\left(\frac{y - T_{1i}}{h_n}\right)}{\sum_{\Delta_{1j}=1} K\left(\frac{y - T_{1j}}{h_n}\right)} & \text{if } \Delta_{1i} = 1 \\ 0 & \text{if } \Delta_{1i} = 0, \end{cases}$$

and similarly for $\hat{F}_{1|2}(y_1|y_2)$. We now proceed with the description of the estimators of the bivariate and marginal distributions.

2.1 Estimation of the bivariate distribution

The proposed estimators will be based on the relations

$$F(\boldsymbol{y}) = \int_0^{y_1} F_{2|1}(y_2|z_1)\, dF_1(z_1), \;\; F(\boldsymbol{y}) = \int_0^{y_2} F_{1|2}(y_1|z_2)\, dF_2(z_2). \tag{2.2}$$

As relation (2.2) suggests, the estimator of the bivariate distribution is obtained by averaging out the conditioning variable. For this purpose, F_1 and F_2 are estimated by the usual Kaplan-Meier estimator. We next give detailed descriptions of the estimators for each of the two cases under consideration.

2.1.1 T_2 independent of C_2 given T_1; T_1, T_2 independent of C_1

The proposed estimator of the bivariate distribution is

$$\hat{F}(\boldsymbol{y}) = \int_0^{y_1} \hat{F}_{2|1}(y_2|z_1)\, d\tilde{F}_1(z_1), \tag{2.3}$$

where

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

is the Kaplan-Meier (1958) estimator of $F_1(y)$. It should be pointed out that, depending on the censoring, it may not be possible to obtain a consistent estimator of $F(y_1, y_2)$ for all $\boldsymbol{y} = (y_1, y_2)$. As is evident from relation (2.3) consistent estimation of $F(y_1, y_2)$ requires consistent estimation of $\hat{F}_{2|1}(y_2|z_1)$ for all $z_1 \le y_1$, as well as consistent estimation of $F_1(y_1)$.

2.1.2 $\boldsymbol{T}$ independent of $\boldsymbol{C}$

Recall that in this case we want to treat the two variables symmetrically. This is accomplished by expressing $F(\boldsymbol{y})$ as a linear combination of the two expressions in (2.2),

$$F(\boldsymbol{y}) = w(\boldsymbol{y}) \int_0^{y_2} F_{1|2}(y_1|z_2)\, dF_2(z_2) + (1 - w(\boldsymbol{y})) \int_0^{y_1} F_{2|1}(y_2|z_1)\, dF_1(z_1). \tag{2.4}$$

The equality in (2.4) holds with $w(\boldsymbol{y})$ any real number. The choice of $w(\boldsymbol{y})$ is discussed in Remark 3.1. Let $\hat{F}_{2|1}(y_2|y_1)$ be as defined in (2.1), $\hat{F}_{1|2}(y_1|y_2)$ be defined in a similar way (simply by interchanging the role of the first and second variable), and let $\tilde{F}_1(y_1)$, $\tilde{F}_2(y_2)$ be the marginal Kaplan-Meier estimators. Then the proposed estimator of the bivariate distribution is

$$\hat{F}(\boldsymbol{y}) = w(\boldsymbol{y}) \int_0^{y_2} \hat{F}_{1|2}(y_1|z_2)\, d\tilde{F}_2(z_2) + (1 - w(\boldsymbol{y})) \int_0^{y_1} \hat{F}_{2|1}(y_2|z_1)\, d\tilde{F}_1(z_1). \tag{2.5}$$

The remark made following relation (2.3) regarding consistent estimation of F applies here too.

2.2 Estimation of the marginal distributions

In principle, estimation of the marginal distribution follows from that of the bivariate distribution by setting one of the arguments to infinity. In the presence of censoring, however, one needs to be more careful. Indeed, this principle suggests that an estimator of the marginal distribution F_2 is obtained (under the first censoring scenario) by setting $y_1 = \infty$ in relation (2.3). According to the remark made following (2.3), however, this would be consistent only if the Kaplan-Meier estimator $\tilde{F}_1$ is consistent on the entire real line. Thus, even if the ordinary Kaplan-Meier estimator $\tilde{F}_2(y_2)$ is consistent for $F_2(y_2)$, setting $y_1 = \infty$ in relation (2.3) yields an inconsistent estimator in most realistic situations.

To be able to utilize the information contained in T_1 in order to achieve a more efficient estimation of F_2 we use the relation

$$F_2(y_2) = \int_0^{\tau_1-} F_{2|1}(y_2|z_1)\, dF_1(z_1) + (1 - F_1(\tau_1-))P(T_2 \le y_2|T_1 \ge \tau_1), \tag{2.6}$$

which is true for any τ_1. It is clear that if τ_1 is chosen so that a) $F_{2|1}(y_2|z_1)$ is consistently estimated by the generalized Beran estimator for all $z_1 < \tau_1$, b) $F_1(\tau_1)$ is consistently estimated by the Kaplan-Meier estimator and c) $P(T_2 \le y_2|T_1 \ge \tau_1)$ is consistently estimated, then (2.6) does yield an estimator of the marginal distribution which utilizes information contained in T_1. Note that the obvious estimator of $P(T_2 \le y_2|T_1 \ge \tau_1)$ is the Kaplan-Meier estimator based on all observations for which $\tilde{T}_1 \ge \tau_1$. To utilize the maximum amount of information contained in T_1, τ_1 should be chosen as the maximum number satisfying a), b) and c) above. With $\tau_1 = 0$ (2.6) becomes $F_2(y_2) = F_2(y_2)$ and the estimator described below reduces to the ordinary Kaplan-Meier estimator $\tilde{F}_2(y_2)$.

The description of the estimator under the two censoring scenaria is given next.

2.2.1 T_2 independent of C_2 given T_1; T_1, T_2 independent of C_1

When T_2 is only conditionally independent of C_2 given T_1, the Kaplan-Meier estimator $\tilde{F}_2$ is not consistent for F_2. The proposed estimator for $F_2(y_2)$, which generalizes the estimator of Cheng (1989), is

$$\hat{F}_2(y_2) = \int_0^{\tau_1-} \hat{F}_{2|1}(y_2|z_1)\, d\tilde{F}_1(z_1) + (1 - \tilde{F}_1(\tau_1-))\hat{P}(T_2 \le y_2|T_1 \ge \tau_1), \tag{2.7}$$

where

$$\hat{P}(T_2 \le y_2|T_1 \ge \tau_1) = 1 - \prod_{\tilde{T}_{2i} \le y, \Delta_{2i}=1, \tilde{T}_{1i} \ge \tau_1} \left(1 - \frac{1}{n_{\tau_1} - i + 1}\right), \tag{2.8}$$

where n_{τ_1} is the number of i's for which $\tilde{T}_{1i} \ge \tau_1$.

Since this censoring mechanism is not symmetric in the two variables, estimation of the marginal distribution $F_1(y_1)$ is necessarily different. Thus, the Kaplan-Meier estimator $\tilde{F}_1$ is consistent for F_1, while $\hat{F}_{1|2}(y_1|y_2)$ is inconsistent for $F_{1|2}(y_1|y_2)$ 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 $\tilde{F}_1$ is the estimator of choice for F_1.

2.2.2 $\boldsymbol{T}$ independent of $\boldsymbol{C}$

This censoring scheme is symmetric in the two variables and thus we will present only the estimator of the marginal distribution of T_2. The derivation for T_1 is completely similar. As with the estimation of the bivariate distribution, we again like to treat the two variables symmetrically. Therefore, we start from equation (2.4) and adjust this equation such that it can be used for the marginal distribution. This means that we take y_1 equal to ∞ and we replace the integral in the second term of (2.4) by the right hand side of (2.6). This leads to the following estimator :

$$\begin{aligned}\hat{F}_2(y_2) &= w_2(y_2)\tilde{F}_2(y_2) \\ &\quad +(1-w_2(y_2))\left[\int_0^{\tau_1-} \hat{F}_{2|1}(y_2|z_1)\,d\tilde{F}_1(z_1) + (1-\tilde{F}_1(\tau_1-))\hat{P}(T_2 \le y_2|T_1 \ge \tau_1)\right]\end{aligned} \tag{2.9}$$

(for any real number $w_2(y_2)$), so it is a weighted average of the traditional Kaplan-Meier estimator and the estimator proposed for case 1.

3 Main results

In this section we present the main asymptotic results for the proposed estimators. Some additional notation and the assumptions used to derive the results are given first.

3.1 Notation and assumptions

Let $H(\boldsymbol{y})$ denote the distribution function of $\tilde{\boldsymbol{T}} = (\tilde{T}_1, \tilde{T}_2) = (T_1 \wedge C_1, T_2 \wedge C_2)$ and set $H(\boldsymbol{y}, \boldsymbol{\delta}) = P(\tilde{T}_1 \le y_1, \tilde{T}_2 \le y_2, \Delta_1 = \delta_1, \Delta_2 = \delta_2)$, $G_j(y) = P(C_j \le y)$, $H_j(y) = P(\tilde{T}_j \le y)$ and $H_j^u(y) = P(\tilde{T}_j \le y, \Delta_j = 1)$, $j = 1, 2$. In addition define the following conditional (sub-)distributions $H_{2|1}(y_2|y_1) = P(\tilde{T}_2 \le y_2|\tilde{T}_1 = y_1, \Delta_1 = 1)$, $H_{2|1}^u(y_2|y_1) = P(\tilde{T}_2 \le y_2, \Delta_2 = 1|\tilde{T}_1 = y_1, \Delta_1 = 1)$, and similarly for $H_{1|2}$, $H_{1|2}^u$. Moreover, we define $F_{2|1}(y_2| \ge y_1) = P(T_2 \le y_2|T_1 \ge y_1)$, and similarly for $H_{2|1}^u(s| \ge t_1)$ and $H_{2|1}(s| \ge t_1)$. The probability density functions of the distributions defined above will be denoted with lower case letters. Finally, we set

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

$$\xi_{2|1}(t_2,\delta_2,y|t_1) = (1-F_{2|1}(y|t_1))\left\{-\int_0^{t_2\wedge y} \frac{dH_{2|1}^u(s|t_1)}{(1-H_{2|1}(s|t_1))^2} + \frac{I(t_2 \le y, \delta_2 = 1)}{1-H_{2|1}(t_2|t_1)}\right\},$$

$$\xi_{2|1}(t_2,\delta_2,y| \ge t_1) = (1-F_{2|1}(y| \ge t_1))\left\{-\int_0^{t_2\wedge y} \frac{dH_{2|1}^u(s| \ge t_1)}{(1-H_{2|1}(s| \ge t_1))^2} + \frac{I(t_2 \le y, \delta_2 = 1)}{1-H_{2|1}(t_2| \ge t_1)}\right\}.$$

Thus, $\xi_1(t,\delta,y)$ is the function that appears in Lo and Singh's (1986) iid representation of the product limit estimator, $\xi_{2|1}(t_2,\delta_2,y|t_1)$ is a conditional version of it given in Akritas (1994), and $\xi_{2|1}(t_2,\delta_2,y|\geq t_1)$ is the corresponding version when we condition on $T_1 \geq t_1$.

The assumptions we need for the proofs of the main results are listed below for convenient reference.

(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 $\int uK(u)\,du = 0$.

Let $\tau_2(y_1) < \inf\{y_2 : H_{2|1}(y_2|y_1) = 1\}$ and similarly for $\tau_1(y_2)$. Define $0 \leq \tau_j < \infty$ such that $\sup_y \tau_j(y) \leq \tau_j < \inf\{y : H_j(y) = 1\}$, $j = 1, 2$. Also, let $\tau_2(\geq y_1) < \inf\{y_2 : H_{2|1}(y_2| \geq y_1) = 1\}$ and define in a similar way $\tau_1(\geq y_2)$.

(A2)(i) $H_1(y)$ and $H_1^u(y)$ are three times continuously differentiable with respect to y.
(ii) $H_{2|1}(y_2|y_1)$ and $H_{2|1}^u(y_2|y_1)$ are twice continuously differentiable with respect to y_1 and y_2, and for $(y_1, y_2) \in [0, \tau_1] \times [0, \tau_2]$ all derivatives are uniformly bounded.

(A3)(i) $H_2(y)$ and $H_2^u(y)$ are three times continuously differentiable with respect to y.
(ii) $H_{1|2}(y_1|y_2)$ and $H_{1|2}^u(y_1|y_2)$ are twice continuously differentiable with respect to y_1 and y_2, and for $(y_1, y_2) \in [0, \tau_1] \times [0, \tau_2]$ all derivatives are uniformly bounded.

3.2 Results for the estimators of the bivariate distribution

3.2.1 T_2 independent of C_2 given T_1; T_1, T_2 independent of C_1

The following two results pertain to the estimator $\hat{F}$ of the bivariate distribution given by (2.3).

Theorem 3.1 *Let*

$$\tau_{2|1}(t_2, \delta_2, y|t_1) = \frac{\xi_{2|1}(t_2, \delta_2, y|t_1)}{1 - G_1(t_1)},$$

and denote by $\tau'_{2|1}(t_2, \delta_2, y|t_1)$ and $\tau''_{2|1}(t_2, \delta_2, y|t_1)$ the first, respectively second derivative of $\tau_{2|1}(t_2, \delta_2, y|t_1)$ with respect to t_1. Also set

$$\begin{aligned}\eta_{2|1}(\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{y}) = &\int_0^{y_1} F_{2|1}(y_2|z_1)\, d\xi_1(t_1, \delta_1, z_1) \\ &+ I(\delta_1 = 1) I(t_1 \leq y_1) \tau_{2|1}(t_2, \delta_2, y_2|t_1),\end{aligned}$$

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

Then, under assumptions $(A1),(A2)$, we have

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

where

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

and $\Omega = \{\boldsymbol{y} : y_1 \le \tau_1$ *and* $y_2 \le \inf_{y\le y_1}\tau_2(y)\}$.

Theorem 3.2 *Let $\eta_{2|1}(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y})$, $\beta_{2|1}(\boldsymbol{y})$ be as defined in Theorem 3.1. Then, under assumptions $(A1),(A2)$,*

- *If $nh_n^4 \to 0$, then $n^{1/2}(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}))$ $(\boldsymbol{y}\in\Omega)$ converges weakly to a zero mean Gaussian process $Z(\boldsymbol{y})$ with covariance function*

$$Cov(Z(\boldsymbol{y}),Z(\boldsymbol{y}')) = E\left[\int_0^{y_1}\int_0^{y'_1} F_{2|1}(y_2|z_1)F_{2|1}(y'_2|z'_1)\,d\xi_1(\tilde{T}_1,\Delta_1,z_1)\,d\xi_1(\tilde{T}_1,\Delta_1,z'_1)\right]$$
$$+E[I(\Delta_1=1)I(\tilde{T}_1\le y_1\wedge y'_1)\tau_{2|1}(\tilde{T}_2,\Delta_2,y_2|\tilde{T}_1)\tau_{2|1}(\tilde{T}_2,\Delta_2,y'_2|\tilde{T}_1)].$$

- *If $nh_n^4 = K$ (for some $K>0$), then $n^{1/2}(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}))$ $(\boldsymbol{y}\in\Omega)$ converges weakly to a Gaussian process $Z'(\boldsymbol{y})$ with mean function*

$$E(Z'(\boldsymbol{y})) = K^{1/2}\beta_{2|1}(\boldsymbol{y})$$

  and the same covariance function as $Z(\boldsymbol{y})$.

3.2.2 $\boldsymbol{T}$ independent of $\boldsymbol{C}$

The following two results pertain to the estimator $\hat{F}$ of the bivariate distribution given by (2.5).

Theorem 3.3 *Let $\eta_{2|1}(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y})$, $\beta_{2|1}(\boldsymbol{y})$ be as defined in Theorem 3.1. Then, under assumptions $(A1)-(A3)$,*

$$\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y})$$
$$= n^{-1}\sum_{i=1}^n[w(\boldsymbol{y})\eta_{1|2}(\tilde{\boldsymbol{T}}_i,\boldsymbol{\Delta}_i,\boldsymbol{y}) + (1-w(\boldsymbol{y}))\eta_{2|1}(\tilde{\boldsymbol{T}}_i,\boldsymbol{\Delta}_i,\boldsymbol{y})]$$
$$+h_n^2[w(\boldsymbol{y})\beta_{1|2}(\boldsymbol{y}) + (1-w(\boldsymbol{y}))\beta_{2|1}(\boldsymbol{y})] + R_n(\boldsymbol{y}),$$

where

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

and $\Omega' = \{\boldsymbol{y} : y_1 \leq \inf_{y\leq y_2} \tau_1(y)$ *and* $y_2 \leq \inf_{y\leq y_1} \tau_2(y)\}$.

Theorem 3.4 *Let* $\eta_{2|1}(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y})$*,* $\beta_{2|1}(\boldsymbol{y})$ *be as defined in Theorem 3.1. Then, under assumptions* $(A1)-(A3)$*,*

- *If* $nh_n^4 \to 0$*, then* $n^{1/2}(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}))$ $(\boldsymbol{y} \in \Omega')$ *converges weakly to a zero mean Gaussian process* $W(\boldsymbol{y})$ *with covariance function*

$$\begin{aligned} &Cov(W(\boldsymbol{y}), W(\boldsymbol{y}')) \\ &= E[\{w(\boldsymbol{y})\eta_{1|2}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\boldsymbol{y}) + (1-w(\boldsymbol{y}))\eta_{2|1}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\boldsymbol{y})\} \\ &\quad \times\{w(\boldsymbol{y}')\eta_{1|2}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\boldsymbol{y}') + (1-w(\boldsymbol{y}'))\eta_{2|1}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\boldsymbol{y}')\}]. \end{aligned}$$

- *If* $nh_n^4 = K$ *(for some* $K > 0$*), then* $n^{1/2}(\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}))$ $(\boldsymbol{y} \in \Omega')$ *converges weakly to a Gaussian process* $W'(\boldsymbol{y})$ *with mean function*

$$E(W'(\boldsymbol{y})) = K^{1/2}[w(\boldsymbol{y})\beta_{1|2}(\boldsymbol{y}) + (1-w(\boldsymbol{y}))\beta_{2|1}(\boldsymbol{y})]$$

  and the same covariance function as $W(\boldsymbol{y})$*.*

Remark 3.1 The above result enables us to choose the weights $w(\boldsymbol{y})$ in such a way that the mean squared error of the estimator $\hat{F}(\boldsymbol{y})$ is minimal. Denote the j-th integral in the definition of $\hat{F}(\boldsymbol{y})$ given in (2.5) by $T_j(\boldsymbol{y})$, $j = 1, 2$. Simple calculations show that the weights

$$w(\boldsymbol{y}) = \frac{\sigma_{22} - \sigma_{12} + \mu_2^2 - \mu_1\mu_2}{\sigma_{11} + \sigma_{22} - 2\sigma_{12} + \mu_1^2 + \mu_2^2 - 2\mu_1\mu_2}$$

minimize the mean squared error of $\hat{F}(\boldsymbol{y})$, where σ_{ij} $(i, j = 1, 2)$ equals the asymptotic covariance of $T_i(\boldsymbol{y})$ and $T_j(\boldsymbol{y})$ and μ_i, $i = 1, 2$, denotes the asymptotic bias of $T_i(\boldsymbol{y})$.

3.3 Results for the estimators of the marginal distributions

3.3.1 T_2 independent of C_2 given T_1; T_1, T_2 independent of C_1

The following two results pertain to the estimator $\hat{F}_2$ of the marginal distribution given by (2.7).

Theorem 3.5 *Let $\eta_{2|1}(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y})$, $\beta_{2|1}(\boldsymbol{y})$ be as defined in Theorem 3.1. In addition, define*

$$\zeta_{2|1}(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y}) = -F_{2|1}(y_2| \geq y_1)\xi_1(t_1,\delta_1,y_1-) + (1-F_1(y_1-))\xi_{2|1}(t_2,\delta_2,y_2| \geq y_1).$$

Then, under assumptions $(A1),(A2)$,

$$\begin{aligned}\hat{F}_2(y_2) - F_2(y_2) = n^{-1}\sum_{i=1}^{n}&[\eta_{2|1}(\tilde{\boldsymbol{T}}_i,\boldsymbol{\Delta}_i,\tau_1-,y_2) + \zeta_{2|1}(\tilde{\boldsymbol{T}}_i,\boldsymbol{\Delta}_i,\tau_1,y_2)]\\ &+h_n^2\beta_{2|1}(\tau_1-,y_2) + R_n(y_2),\end{aligned}$$

where

$$\sup_{y_2\in\Omega_2}|R_n(y_2)| = o_P(n^{-1/2}) + o_P(h_n^2)$$

and $\Omega_2 = \{y_2 : y_2 \leq \inf_{y<\tau_1}\tau_2(y)$ and $y_2 \leq \tau_2(\geq \tau_1)\}$.

Theorem 3.6 *Let $\eta_{2|1}(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y})$, $\beta_{2|1}(\boldsymbol{y})$ and $\zeta_{2|1}(\boldsymbol{t},\boldsymbol{\delta},\boldsymbol{y})$ be as in Theorem 3.5. Then, under assumptions $(A1),(A2)$,*

- *If $nh_n^4 \to 0$, then $n^{1/2}(\hat{F}_2(y_2) - F_2(y_2))$ $(y_2 \in \Omega_2)$ converges weakly to a zero mean Gaussian process $Z_2(y_2)$ with covariance function*

$$\begin{aligned}Cov(Z_2(y_2),Z_2(y_2')) = E[&\{\eta_{2|1}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\tau_1-,y_2) + \zeta_{2|1}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\tau_1,y_2)\}\\ &\times\{\eta_{2|1}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\tau_1-,y_2') + \zeta_{2|1}(\tilde{\boldsymbol{T}},\boldsymbol{\Delta},\tau_1,y_2')\}].\end{aligned}$$

- *If $nh_n^4 = K$ (for some $K > 0$), then $n^{1/2}(\hat{F}_2(y_2)-F_2(y_2))$ $(y_2 \in \Omega_2)$ converges weakly to a Gaussian process $Z_2'(y_2)$ with mean function*

$$E(Z_2'(y_2)) = K^{1/2}\beta_{2|1}(\tau_1-,y_2)$$

  and the same covariance function as $Z_2(y_2)$.

3.4 $\boldsymbol{T}$ independent of $\boldsymbol{C}$

The following two results pertain to the estimator $\hat{F}_2$ of the marginal distribution given by (2.9)

Theorem 3.7 *Assume $(A1)-(A3)$. Then,*

$$\begin{aligned}\hat{F}_2(y_2) - F_2(y_2) = n^{-1}\sum_{i=1}^{n}&[w_2(y_2)\xi_2(\tilde{T}_{2i},\Delta_{2i},y_2)\\ &+(1-w_2(y_2))\{\eta_{2|1}(\tilde{\boldsymbol{T}}_i,\boldsymbol{\Delta}_i,\tau_1-,y_2) + \zeta_{2|1}(\tilde{\boldsymbol{T}}_i,\boldsymbol{\Delta}_i,\tau_1,y_2)\}]\\ &+h_n^2(1-w_2(y_2))\beta_{2|1}(\tau_1-,y_2) + R_n(y_2),\end{aligned}$$

where

$$\sup_{y_2 \in \Omega_2} |R_n(y_2)| = o_P(n^{-1/2}) + o_P(h_n^2).$$

Theorem 3.8 *Assume* $(A1) - (A3)$.

- *If* $nh_n^4 \to 0$, *then* $n^{1/2}(\hat{F}_2(y_2) - F_2(y_2))$ $(y_2 \in \Omega_2)$ *converges weakly to a zero mean Gaussian process* $W_2(y_2)$ *with covariance function*

$$\begin{aligned} & Cov(W_2(y_2), W_2(y_2')) \\ & = E[\{w_2(y_2)\xi_2(\tilde{T}_2, \Delta_2, y_2) + (1 - w_2(y_2))[\eta_{2|1}(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \tau_1-, y_2) + \zeta_{2|1}(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \tau_1, y_2)]\} \\ & \quad \times \{w_2(y_2')\xi_2(\tilde{T}_2, \Delta_2, y_2') + (1 - w_2(y_2'))[\eta_{2|1}(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \tau_1-, y_2') + \zeta_{2|1}(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \tau_1, y_2')]\}]. \end{aligned}$$

- *If* $nh_n^4 = K$ *(for some* $K > 0$*), then* $n^{1/2}(\hat{F}_2(y_2) - F_2(y_2))$ $(y_2 \in \Omega_2)$ *converges weakly to a Gaussian process* $W_2'(y_2)$ *with mean function*

$$E(W_2'(y_2)) = K^{1/2}(1 - w_2(y_2))\beta_{2|1}(\tau_1-, y_2)$$

  and the same covariance function as $W_2(y_2)$.

Remark 3.2 The above result enables us to choose the weights $w_2(y_2)$ in such a way that the mean squared error of the estimator $\hat{F}_2(y_2)$ is minimal. Denote $\tilde{F}_2(y_2)$ by $T_1(y_2)$ and the expression between brackets in (2.9) by $T_2(y_2)$. Simple calculations show that the weights

$$w_2(y_2) = \frac{\sigma_{22} - \sigma_{12} + \mu_2^2}{\sigma_{11} + \sigma_{22} - 2\sigma_{12} + \mu_2^2}$$

minimize the mean squared error of $\hat{F}_2(y_2)$, where σ_{ij} $(i, j = 1, 2)$ equals the asymptotic covariance of $T_i(y_2)$ and $T_j(y_2)$ and μ_2, denotes the asymptotic bias of $T_2(y_2)$ ($T_1(y_2)$ is asymptotically unbiased).

4 Simulations

In this section we carry out a small simulation study to compare the finite sample behavior of three estimators of the bivariate distribution : the estimator $\hat{F}(\boldsymbol{y})$ proposed in (2.5), the estimator of Pruitt (1991) and the nonparametric maximum likelihood estimator (NPMLE) of van der Laan (1996). Although other estimators of the bivariate distribution have been proposed in the literature, we restrict attention to the above estimators, since

Table 2: *Bias and variance of Pruitt's estimator* $\tilde{F}_P(\boldsymbol{y})$*, the NPMLE* $\tilde{F}_L(\boldsymbol{y})$ *and the new estimator* $\hat{F}(\boldsymbol{y})$*.*

$P(\Delta_{1,2}=0)$	ρ_T	n	Bias			Variance		
			$\tilde{F}_P(\boldsymbol{y})$	$\tilde{F}_L(\boldsymbol{y})$	$\hat{F}(\boldsymbol{y})$	$\tilde{F}_P(\boldsymbol{y})$	$\tilde{F}_L(\boldsymbol{y})$	$\hat{F}(\boldsymbol{y})$
0.2	0	30	.00128	.00909	.00007	.01030	.01097	.00969
		80	-.00083	.00115	-.00101	.00381	.00431	.00359
		150	.00038	-.00202	.00010	.00209	.00230	.00201
	0.3	30	-.00253	.00593	-.00207	.00967	.01004	.00977
		80	-.00258	-.00242	-.00050	.00364	.00415	.00374
		150	-.00331	-.00173	-.00117	.00199	.00217	.00203
0.3	0	30	.00593	.01679	-.00055	.01402	.01466	.01275
		80	-.00086	-.00630	-.00091	.00476	.00576	.00455
		150	-.00688	.01253	.00026	.00288	.00292	.00249
	0.3	30	.00176	.01453	.02038	.01291	.01252	.01158
		80	-.00387	-.01016	-.00063	.00448	.00527	.00460
		150	-.00284	-.00956	.00006	.00235	.00271	.00241

simulations (see Pruitt (1993a) and van der Laan (1996)) have shown that other estimators do not behave as well in practice.

Consider a bivariate log-normal distribution for the vector $\boldsymbol{T}$ of survival times. Each $\log(T_j)$ $(j = 1, 2)$ is defined to have zero mean and variance equal to one and the correlation between $\log(T_1)$ and $\log(T_2)$, denoted by ρ_T, will be either 0 or 0.3. The distribution of the vector $(\log(C_1), \log(C_2))$ is also bivariate normal, but with mean $(m, m)'$ for a certain real number m, and with variance-covariance matrix equal to the identity matrix. The value of m is determined in such a way that the probability of censoring in each of the two marginals equals 0.2 or 0.3. For the situation where $\rho_T = 0$, this means that 36 % respectively 51 % of the data have at least one censored component. We carry out simulations for samples of size 30, 80 and 150. The results are based on 1000 simulation runs and the probability that is calculated is $P(T_1 \leq 2, T_2 \leq 2)$. This probability equals 0.5714 for $\rho_T = 0$ and 0.6033 for $\rho_T = 0.3$. Since all three estimators depend on a bandwidth, we have to start with selecting an appropriate bandwidth before we make any comparisons. We choose to work with the bandwidth for which the mean squared error of the estimator (obtained from the 1000 runs) is minimal. Table 2 shows the bias and variance of the three estimators, for the different settings of the parameters and for the bandwidth that minimizes this mean squared error. The table shows that in most cases, the proposed estimator behaves the best, followed by Pruitt's estimator, while the performance of the NPMLE is usually the worst. In a few situations, Pruitt's estimator behaves the best, but the proposed one behaves only slightly worse in these cases. It is also worth noting that all three estimators have a very small bias, which contributes only little to the mean squared error.

5 Proofs

We will restrict us to proving Theorems 3.1, 3.2 and 3.5. Theorems 3.3 and 3.7 follow readily from Theorems 3.1 and 3.5 respectively, while the proofs of Theorems 3.4, 3.6 and 3.8 are similar to that of Theorem 3.2.

We start with a lemma which is needed in the proof of Theorem 3.1.

Lemma 5.1 *Under the same assumptions as in Theorem 3.1,*

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

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

$$\mathcal{F} = \Big\{ I(\Delta_1 = 1)I(y_1 - L\delta \le T_1 \le y_1)\tau_{2|1}(\tilde{T}_2, \Delta_2, y_2|T_1)\int_{\frac{y_1-T_1}{\delta}}^{L} K(u)\,du : \\ \boldsymbol{y} \in \Omega, 0 < \delta < 1\Big\}.$$

In a first step we will show that the class $\mathcal{F}$ is Donsker. From Theorem 2.5.6 in van der Vaart and Wellner (1996), it follows that it suffices to show that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\,d\varepsilon < \infty, \tag{5.1}$$

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

$$I(y_1 - L\delta \le T_1 \le y_1) = I(T_1 \le y_1) - I(T_1 \le y_1 - L\delta),$$

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

$$
\begin{aligned}
&\tau_{2|1}(\tilde{T}_2, \Delta_2, y_2|T_1) \\
&= (1 - F_{2|1}(y_2|T_1)) \left\{ - \int_0^{\tilde{T}_2 \wedge y_2} \frac{dH_{2|1}^u(s|T_1)}{(1 - H_{2|1}(s|T_1))^2} + \frac{I(\tilde{T}_2 \le y_2, \Delta_2 = 1)}{1 - H_{2|1}(\tilde{T}_2|T_1)} \right\} \frac{f_1(T_1)}{h_1^u(T_1)},
\end{aligned}
$$

note that $1 - F_{2|1}(y_2|T_1)$ 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 $O(\exp(K\varepsilon^{-1}))$. On the integral in $\tau_{2|1}(\tilde{T}_2, \Delta_2, y_2|T_1)$ we apply the same corollary, but since 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 $\tau_{2|1}(\tilde{T}_2, \Delta_2, y_2|T_1)$, 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 (5.1) is satisfied, since for $\varepsilon > 2M$ one bracket suffices (where M is an upper bound for ! $\tau_{2|1}(\tilde{T}_2, \Delta_2, y_2|T_1)$). This shows that the class $\mathcal{F}$ is Donsker. Next, some straightforward calculations show that

$$
E[\xi_{2|1}(\tilde{T}_2, \Delta_2, y_2|T_1)|T_1]^2 = (1 - F_{2|1}(y_2|T_1))^2 \int_0^{y_2} \frac{dH_{2|1}^u(s|T_1)}{(1 - H_{2|1}(s|T_1))^2}
$$

and hence,

$$
\begin{aligned}
&\mathrm{Var}\left[I(\Delta_1 = 1) I(y_1 - Lh_n \le T_1 \le y_1) \tau_{2|1}(\tilde{T}_2, \Delta_2, y_2|T_1) \int_{\frac{y_1-T_1}{h_n}}^{L} K(u)\,du \right] \\
&\le E\Bigg[I(y_1 - Lh_n \le T_1 \le y_1)(1 - F_{2|1}(y_2|T_1))^2 \int_0^{y_2} \frac{dH_{2|1}^u(s|T_1)}{(1 - H_{2|1}(s|T_1))^2} \frac{f_1^2(T_1)}{h_1^{u2}(T_1)} \\
&\qquad\qquad \times \left\{ \int_{\frac{y_1-T_1}{h_n}}^{L} K(u)\,du \right\}^2 \Bigg] \\
&\le K h_n
\end{aligned}
$$

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.

Proof of Theorem 3.1. Write

$$\begin{aligned}
&\hat{F}(\boldsymbol{y}) - F(\boldsymbol{y}) \\
&= \int_0^{y_1} \{\hat{F}_{2|1}(y_2|z_1) - F_{2|1}(y_2|z_1)\}\, dF_1(z_1) \\
&\quad + \int_0^{y_1} \{\hat{F}_{2|1}(y_2|z_1) - F_{2|1}(y_2|z_1)\}\, d(\tilde{F}_1(z_1) - F_1(z_1)) \\
&\quad + \int_0^{y_1} F_{2|1}(y_2|z_1)\, d(\tilde{F}_1(z_1) - F_1(z_1)) \\
&= A_1(\boldsymbol{y}) + A_2(\boldsymbol{y}) + A_3(\boldsymbol{y}).
\end{aligned}$$

On $A_3(\boldsymbol{y})$ we apply the i.i.d. representation for the Kaplan-Meier estimator given in Lo and Singh (1986) :

$$A_3(\boldsymbol{y}) = n^{-1} \sum_{i=1}^{n} \int_0^{y_1} F_{2|1}(y_2|z_1)\, d\xi_1(\tilde{T}_{1i}, \Delta_{1i}, z_1) + o_P(n^{-1/2}).$$

Next, using the notation $\dot{F}_{2|1}(y_2|y_1)$ for the partial derivative of $F_{2|1}(y_2|y_1)$ with respect to y_1 and similarly for $\dot{\hat{F}}_{2|1}(y_2|y_1)$,

$$\begin{aligned}
A_2(\boldsymbol{y}) &= \{\hat{F}_{2|1}(y_2|y_1) - F_{2|1}(y_2|y_1)\}\{\tilde{F}_1(y_1) - F_1(y_1)\} \\
&\quad - \int_0^{y_1} \{\tilde{F}_1(z_1) - F_1(z_1)\}\, \{\dot{\hat{F}}_{2|1}(y_2|z_1) - \dot{F}_{2|1}(y_2|z_1)\}\, dz_1
\end{aligned}$$

and this is $o_P(n^{-1/2})$, because $\sup_{y \le \tau_1} |\tilde{F}_1(y) - F_1(y)| = O(n^{-1/2}(\log n)^{1/2})$ a.s. (see e.g. Lo and Singh (1986)), $\sup_{y_1} \sup_{y_2 \le \tau_2(y_1)} |\hat{F}_{2|1}(y_2|y_1) - F_{2|1}(y_2|y_1)| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. and $\sup_{y_1} \sup_{y_2 \le \tau_2(y_1)} |\dot{\hat{F}}_{2|1}(y_2|y_1) - \dot{F}_{2|1}(y_2|y_1)| = O((nh_n^3)^{-1/2}(\log n)^{1/2})$ a.s. (see Van Keilegom and Akritas (1999), applied to the restricted data set $\{(\tilde{\boldsymbol{T}}_i, \boldsymbol{\Delta}_i); \Delta_{1i} = 1\}$).

In the rest of the proof we consider $A_1(\boldsymbol{y})$, on which we apply 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),

$$\begin{aligned}
&A_1(\boldsymbol{y}) \\
&= (nh_n)^{-1} \sum_{\Delta_{1i}=1} \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1) I(z_1 \le y_1)\, dz_1 + o_P(n^{-1/2})
\end{aligned}$$

$$= (nh_n)^{-1} \sum_{\Delta_{1i}=1} \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1)[I(z_1 \le y_1) - I(T_{1i} \le y_1)]\, dz_1$$

$$+(nh_n)^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \le y_1) \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1)\, dz_1 + o_P(n^{-1/2})$$

$$= A_{11}(\boldsymbol{y}) + A_{12}(\boldsymbol{y}) + o_P(n^{-1/2}).$$

We start with $A_{11}(\boldsymbol{y})$.

$$A_{11}(\boldsymbol{y})$$

$$= -(nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \le T_{1i} \le y_1) \int_{y_1}^{T_{1i}+Lh_n} K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1)\, dz_1$$

$$+(nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 \le T_{1i} \le y_1 + Lh_n) \int_{T_{1i}-Lh_n}^{y_1} K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1)\, dz_1$$

$$= -A_{111}(\boldsymbol{y}) + A_{112}(\boldsymbol{y}).$$

The derivation for the two terms above is similar. We consider $A_{111}(\boldsymbol{y})$.

$$A_{111}(\boldsymbol{y})$$

$$= (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \le T_{1i} \le y_1)\tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) \int_{y_1}^{T_{1i}+Lh_n} K\left(\frac{z_1 - T_{1i}}{h_n}\right) dz_1$$

$$+(nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \le T_{1i} \le y_1)\tau'_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i})$$

$$\times \int_{y_1}^{T_{1i}+Lh_n} K\left(\frac{z_1 - T_{1i}}{h_n}\right) (z_1 - T_{1i})\, dz_1$$

$$+\frac{1}{2}(nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \le T_{1i} \le y_1) \int_{y_1}^{T_{1i}+Lh_n} \tau''_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|\xi_{1i})$$

$$\times K\left(\frac{z_1 - T_{1i}}{h_n}\right) (z_1 - T_{1i})^2\, dz_1$$

$$= A_{1111}(\boldsymbol{y}) + A_{1112}(\boldsymbol{y}) + A_{1113}(\boldsymbol{y}),$$

where ξ_{1i} is between z_1 and T_{1i}. The term $A_{1111}(\boldsymbol{y})$ equals

$$n^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \le T_{1i} \le y_1)\tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) \int_{\frac{y_1 - T_{1i}}{h_n}}^{L} K(u)\, du$$

and this is $o_P(n^{-1/2})$, uniformly in $\boldsymbol{y} \in \Omega$, by Lemma 5.1. Also $A_{1113}(\boldsymbol{y})$ is $o_P(n^{-1/2})$, since it is easily seen that this term is bounded by

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

which is $O_P(h_n^3) = o_P(n^{-1/2})$. The term $A_{1112}(\boldsymbol{y})$ equals

$$h_n n^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \le T_{1i} \le y_1) \tau'_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) \int_{\frac{y_1 - T_{1i}}{h_n}}^{L} K(u) u \, du.$$

In a very similar way as was shown in Lemma 5.1 for the term $A_{1111}(\boldsymbol{y})$ it can be shown that $\sup_{\boldsymbol{y} \in \Omega} |A_{1112}(\boldsymbol{y}) - EA_{1112}(\boldsymbol{y})| = o_P(n^{-1/2})$. It follows that

$$\begin{aligned}
& A_{111}(\boldsymbol{y}) \\
&= EA_{1112}(\boldsymbol{y}) + o_P(n^{-1/2}) \\
&= h_n \int_{y_1 - Lh_n}^{y_1} E[\tau'_{2|1}(\tilde{T}_2, \Delta_2, y_2|t)|T_1 = t] \int_{\frac{y_1 - t}{h_n}}^{L} K(u) u \, du \, dH_1^u(t) + o_P(n^{-1/2}) \\
&= h_n^2 \int_0^L E[\tau'_{2|1}(\tilde{T}_2, \Delta_2, y_2|y_1 - vh_n)|T_1 = y_1 - vh_n] \int_v^L K(u) u \, du \, h_1^u(y_1 - vh_n) \, dv \\
&\quad + o_P(n^{-1/2}) \\
&= h_n^2 \int_0^L \{Q(y_1 - vh_n) - Q(y_1)\} \int_v^L K(u) u \, du \, dv \\
&\quad + h_n^2 Q(y_1) \int_0^L \int_v^L K(u) u \, du \, dv + o_P(n^{-1/2}) \\
&= h_n^2 \int_0^L Q'(\xi_v)(-vh_n) \int_v^L K(u) u \, du \, dv + h_n^2 Q(y_1) \int_0^L \int_v^L K(u) u \, du \, dv + o_P(n^{-1/2}) \\
&= h_n^2 h_1^u(y_1) E[\tau'_{2|1}(\tilde{T}_2, \Delta_2, y_2|y_1)|T_1 = y_1] \int_0^L \int_v^L K(u) u \, du \, dv + o_P(n^{-1/2}),
\end{aligned}$$

where $Q(y) = h_1^u(y) E[\tau'_{2|1}(\tilde{T}_2, \Delta_2, y_2|y)|T_1 = y]$ and $y_1 - vh_n \le \xi_v \le y_1$. It remains to consider

$$\begin{aligned}
& A_{12}(\boldsymbol{y}) \\
&= (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \le y_1) \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1) \, dz_1 \\
&= (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \le y_1) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) dz_1 \\
&\quad + (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \le y_1) \tau'_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) (z_1 - T_{1i}) \, dz_1 \\
&\quad + \frac{1}{2}(nh_n)^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \le y_1) \tau''_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) (z_1 - T_{1i})^2 \, dz_1 \\
&\quad + o_P(h_n^2) \\
&= n^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \le y_1) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i})
\end{aligned}$$

$$+\frac{1}{2}\mu_2^K h_n^2 n^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \le y_1)\tau''_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) + o_P(h_n^2)$$

$$= n^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \le y_1)\tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i})$$

$$+\frac{1}{2}\mu_2^K h_n^2 \int I(z_1 \le y_1)\tau''_{2|1}(z_2, \delta_2, y_2|z_1)I(\delta_1 = 1)\, dH(\boldsymbol{z}, \boldsymbol{\delta})$$

$$+\frac{1}{2}\mu_2^K h_n^2 \int I(z_1 \le y_1)\tau''_{2|1}(z_2, \delta_2, y_2|z_1)I(\delta_1 = 1)\, d(\hat{H}(\boldsymbol{z}, \boldsymbol{\delta}) - H(\boldsymbol{z}, \boldsymbol{\delta})) + o_P(h_n^2),$$

where $\hat{H}(\boldsymbol{z}, \boldsymbol{\delta}) = n^{-1}\sum_{i=1}^n I(\tilde{T}_{1i} \le z_1, \tilde{T}_{2i} \le z_2, \Delta_{1i} = \delta_1, \Delta_{2i} = \delta_2)$. 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 $\mathcal{F} = \{I(\tilde{T}_1 \le y_1)\tau''_{2|1}(\tilde{T}_2, \Delta_2, y_2|\tilde{T}_1)I(\Delta_1 = 1) : \boldsymbol{y} \in \Omega\}$ 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 5.1), which can be done in a very similar way as in the proof of Lemma 5.1. This finishes the proof.

Proof of Theorem 3.2. To prove the weak convergence of the given process, we will make use of Theorem 2.5.6 in van der Vaart and Wellner (1996), i.e. we will show that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty, \tag{5.2}$$

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 $\mathcal{F} = \{\eta_{2|1}(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \boldsymbol{y}) : \boldsymbol{y} \in \Omega\}$. 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 $\eta_{2|1}(\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{y})$. Write

$$\int_0^{y_1} F_{2|1}(y_2|z_1)\, d\xi_1(t_1, \delta_1, z_1)$$

$$= F_{2|1}(y_2|y_1)\xi_1(t_1, \delta_1, y_1) - \int_0^{y_1} \xi_1(t_1, \delta_1, z_1)\, dF_{2|1}(y_2|z_1)$$

$$= T_1(\boldsymbol{y}) + T_2(\boldsymbol{y}).$$

Since $F_{2|1}(y_2|y_1)$ 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

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

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 $\eta_{2|1}(\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{y})$. In the proof of Lemma 5.1, it was shown that also this term requires $O(\exp(K\varepsilon^{-1}))$ brackets. Hence, (5.2) is satisfied, since for $\varepsilon > 2M$ one bracket suffices (where M is an upper bound for $\eta_{2|1}(\tilde{\boldsymbol{T}}, \boldsymbol{\Delta}, \boldsymbol{y})$). This finishes the proof.

Proof of Theorem 3.5. The derivation for the first term on the right hand side of (2.7) follows from Theorem 3.1 by letting $y_1 = \tau_1-$. It suffices therefore to consider

$$
\begin{aligned}
& (1 - \tilde{F}_1(\tau_1-))\hat{F}_{2|1}(y_2| \ge \tau_1) - (1 - F_1(\tau_1-))F_{2|1}(y_2| \ge \tau_1) \qquad (5.3) \\
& = (1 - F_1(\tau_1-))[\hat{F}_{2|1}(y_2| \ge \tau_1) - F_{2|1}(y_2| \ge \tau_1)] \\
& \quad -[\tilde{F}_1(\tau_1-) - F_1(\tau_1-)]F_{2|1}(y_2| \ge \tau_1) \\
& \quad -[\tilde{F}_1(\tau_1-) - F_1(\tau_1-)][\hat{F}_{2|1}(y_2| \ge \tau_1) - F_{2|1}(y_2| \ge \tau_1)].
\end{aligned}
$$

Since $\sup_{y \le \tau_1} |\tilde{F}_1(y) - F_1(y)| = O(n^{-1/2}(\log n)^{1/2})$ a.s. (see e.g. Lo and Singh (1986)) and similarly for $\hat{F}_{2|1}(y| \ge \tau_1)$, the last term of (5.3) is $o_P(n^{-1/2})$. Using the representation given in Theorem 1 in Lo and Singh (1986), the second term above equals

$$
-n^{-1} \sum_{i=1}^{n} F_{2|1}(y_2| \ge \tau_1)\xi_1(\tilde{T}_{1i}, \Delta_{1i}, \tau_1-) + o_P(n^{-1/2})
$$

and the first term can be decomposed in a similar way. This finishes the proof.

Acknowledgment. The authors are grateful to Ronald Pruitt, Mark van der Laan and the Statlib library (lib.stat.cmu.edu) for providing the program of their estimators.

References

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

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

Campbell, G. and Földes, A. (1982). Large sample properties of nonparametric bivariate estimators with censored data. In: *Nonparametric Statistical Inference, Colloquia Mathematica-Societatis János Bolyai* (B.V. Gnedenko, M.L. Puri and I. Vincze, eds.). North Holland, Amsterdam.

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

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.*, **16**, 1475-1489.

Gill, R. D., van der Laan, M. J. and Wellner, J. A. (1993). Inefficient estimators of the bivariate survival function for three models. *Ann. Inst. H. Poincaré Probab. Statist.*, **31**, 547-597.

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.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.*, **84**, 487-493.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, **79**, 495-512.

Pruitt, R. C. (1991). Strong consistency of self-consistent estimators : general theory and an application to bivariate survival analysis. Technical report 543, Univ. Minnesota.

Pruitt, R. C. (1993a). Small sample comparison of six bivariate survival curve estimators. *J. Statist. Comput. Simul.*, **45**, 147-167.

Pruitt, R. C. (1993b). Identifiability of bivariate survival curves from censored data. *J. Amer. Statist. Assoc.*, **88**, 573-579.

Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review*, **58**, 263-277.

Tsai, W.-Y., Leurgans, S. and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.*, **14**, 1351-1365.

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.

Wang, W. and Wells, M.T. (1999). Semi-parametric estimation of the bivariate survival function. Technical report, Institute of statistics, Academia Sinica.