Estimation of the bivariate and marginal distributions with censored data

Citation for published version (APA):

Akritis, M. G., & Van Keilegom, I. (2000). *Estimation of the bivariate and marginal distributions with censored data*. (SPOR-Report : reports in statistics, probability and operations research; Vol. 200009). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2000

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Estimation of the Bivariate and Marginal Distributions with Censored Data

Michael Akritas and Ingrid Van Keilegom

*Penn State University and
Eindhoven University of Technology*

May 22, 2000

Abstract

Two new estimators for the bivariate distribution when both variables are subject to censoring are proposed. One of the estimators is valid under the assumption that the censoring variables are independent from the variables of interest, and leads to an estimator of the marginal distribution that outperforms the Kaplan-Meier estimator. The other assumes only that one of the censoring variables is conditionally independent of the survival time it censors given the other survival time. Both 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 Introduction

The problem of estimating the bivariate distribution of $\mathbf{T} = (T_1, T_2)$ when both variables are subject to random censoring has received considerable attention in the statistical literature. See Pruitt (1993a) and van der Laan (1996) for a review of most of the available approaches. All estimators in these papers are restricted to the case where the bivariate censoring variable $\mathbf{C} = (C_1, C_2)$ is independent from $\mathbf{T}$. 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 scenarios. First when $\mathbf{C}$ and $\mathbf{T}$ are independent and second when T_2 is conditionally independent from C_2 given T_1 . 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 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 $\mathbf{T}$

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

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. From the practical point of view, one of the attractive

features of this approach 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.

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 the next section. 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. When the scenario that $\mathbf{C}$ and $\mathbf{T}$ are independent can be assumed, 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.

As estimators of the marginal distributions we propose the marginals of the estimates of the bivariate distributions. Thus, 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, and this is in agreement with the asymptotic results of Section 3.

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. As mentioned above, the transfer of tail information allows a practical extension of the least squares estimator to censored data. Regression problems where both variables are subject to censoring arise commonly in astronomy. 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. Finally, the present method can be applied to situations where the pairs of censoring and survival times are dependent but this dependence can be explained by another variable. Some of these applications will be explored elsewhere.

The next section sets the notation and provides the explicit expressions for the proposed estimators. The main results regarding the asymptotic behavior of the estimators are presented in Section 3. 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 The Proposed Estimators

In this section we describe the proposed estimators. We first introduce some needed notation and discuss the main ideas before presenting the estimators for each of the two cases under consideration.

Let $\mathbf{T} = (T_1, T_2)$ be the pair of survival times and $\mathbf{C} = (C_1, C_2)$ the pair of censoring times. The bivariate distribution functions of $\mathbf{T}$ and $\mathbf{C}$ will be denoted by $F(\mathbf{y})$ and $G(\mathbf{y})$, respectively. We are interested in estimating $F(\mathbf{y})$. We observe $\tilde{\mathbf{T}} = (\tilde{T}_1, \tilde{T}_2) = (T_1 \wedge C_1, T_2 \wedge C_2)$ and $\Delta = (\Delta_1, \Delta_2) = (I(T_1 \leq C_1), I(T_2 \leq C_2))$. Let $H(\mathbf{y})$ denote the distribution function of $\tilde{\mathbf{T}}$ and set $H(\mathbf{y}, \delta) = P(\tilde{T}_1 \leq y_1, \tilde{T}_2 \leq y_2, \Delta_1 = \delta_1, \Delta_2 =$

δ_2) for the sub-distribution functions. Moreover we will set $H_j(y) = P(\tilde{T}_j \leq y)$ and $H_j^u(y) = P(\tilde{T}_j \leq y, \Delta_j = 1)$, $j = 1, 2$, $H_{2|1}(y_2|y_1) = P(\tilde{T}_2 \leq y_2 | \tilde{T}_1 = y_1, \Delta_1 = 1)$, $H_{2|1}^u(y_2|y_1) = P(\tilde{T}_2 \leq y_2, \Delta_2 = 1 | \tilde{T}_1 = y_1, \Delta_1 = 1)$, $F_j(y) = P(T_j \leq y)$, $j = 1, 2$, $F_{2|1}(y_2|y_1) = P(T_2 \leq y_2 | T_1 = y_1)$, and similarly for $H_{1|2}$, $H_{1|2}^u$, $F_{1|2}$. The probability density functions of the distributions defined above will be denoted with lower case letters.

The proposed estimators will be based on the relations

$$F(\mathbf{y}) = \int_0^{y_1} F_{2|1}(y_2|z_1) dF_1(z_1), \quad F(\mathbf{y}) = \int_0^{y_2} F_{1|2}(y_1|z_2) dF_2(z_2). \quad (2.1)$$

Thus, as a first step, we must estimate the conditional distributions $F_{2|1}(y_2|z_1)$, $F_{1|2}(y_1|z_2)$ by adapting the Beran (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

$$\begin{aligned} P(T_2 \leq y_2 | T_1 = y_1, \Delta_1 = 1) &= P(T_2 \leq y_2 | T_1 = y_1, C_1 \geq y_1) \\ &= P(T_2 \leq y_2 | T_1 = y_1). \end{aligned}$$

Finally, as relation (2.1) suggests, the estimator of the bivariate distribution is obtained by averaging out the conditioning variable. For this purpose, F_1 , 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 Case 1 : T_2 independent of C_2 given T_1 ; T_1, T_2 independent of C_1

Let 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) and let

$$\hat{F}_{2|1}(y_2|y_1) = 1 - \prod_{\tilde{T}_{2i} \leq y_2, \Delta_{2i}=1} \left(1 - \frac{W_{n1i}(y_1; h_n)}{\sum_{j=1}^n W_{n1j}(y_1; h_n) I(\tilde{T}_{2j} \geq \tilde{T}_{2i})} \right), \quad (2.2)$$

where

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

be the proposed extension of the Beran (1981) estimator to the case where the conditioning variable is subject to censoring. Then the proposed estimator of the bivariate distribution is

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

where

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

is the Kaplan-Meier estimator of $F_1(y)$. The corresponding estimator for the marginal distribution of T_2 is $\hat{F}_2(y_2) = \hat{F}(\infty, y_2)$. Thus, it is given by (2.3) with $y_1 = \infty$.

2.2 Case 2 : T independent of C

Recall that in this case we want to treat the two variables symmetrically. This is accomplished by expressing $F(\mathbf{y})$ as a linear combination of the two expressions in (2.1),

$$F(\mathbf{y}) = w(\mathbf{y}) \int_0^{y_2} \hat{F}_{1|2}(y_1|z_2) d\tilde{F}_2(z_2) + (1 - w(\mathbf{y})) \int_0^{y_1} \hat{F}_{2|1}(y_2|z_1) d\tilde{F}_1(z_1). \quad (2.4)$$

The equality in (2.4) holds with $w(\mathbf{y})$ any real number. The choice of $w(\mathbf{y})$ is discussed in Remark 3.1. Let $\hat{F}_{2|1}(y_2|y_1)$ be as defined in (2.2), $\hat{F}_{1|2}(y_1|y_2)$ 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

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

The corresponding estimators for the marginal distributions of T_1 and T_2 are $\hat{F}_1(y_1) = \hat{F}(y_1, \infty)$ and $\hat{F}_2(y_2) = \hat{F}(\infty, y_2)$, respectively. Thus, each is a linear combination of the corresponding Kaplan-Meier estimator and another estimator obtained by averaging out the conditioning variable. For example,

$$\hat{F}_1(y_1) = w_1(y_1) \int_0^\infty \hat{F}_{1|2}(y_1|z_2) d\tilde{F}_2(z_2) + (1 - w_1(y_1))\tilde{F}_1(y_1), \quad (2.6)$$

where $w_1(y_1) = w(y_1, \infty)$.

3 Main results

The statements and proofs of the main results require the following additional notations. Let

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

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

$$\begin{aligned}\xi_1(t, \delta, y) &= (1 - F_1(y)) \left\{ - \int_0^{t \wedge y} \frac{dH_1^u(s)}{(1 - H_1(s))^2} + \frac{I(t \leq y, \delta = 1)}{1 - H_1(t)} \right\}, \\ \xi_{2|1}(t_2, \delta_2, y|t_1) &= (1 - F_{2|1}(y|t_1)) \left\{ - \int_0^{t_2 \wedge y} \frac{dH_{2|1}^u(s|t_1)}{(1 - H_{2|1}(s|t_1))^2} + \frac{I(t_2 \leq y, \delta_2 = 1)}{1 - H_{2|1}(t_2|t_1)} \right\}, \\ \tau_{2|1}(t_2, \delta_2, y|t_1) &= \xi_{2|1}(t_2, \delta_2, y|t_1) \frac{f_1(t_1)}{h_1^u(t_1)}\end{aligned}$$

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

We use the notation $\tau_j < \inf\{y : H_j(y) = 1\}$ ($j = 1, 2$), $\tau_1(y_2) < \inf\{y_1 : H_{1|2}(y_1|y_2) = 1\}$ and similarly for $\tau_2(y_1)$. 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 $\log n(nh_n)^{-1} \rightarrow 0$.
- (ii) The probability density function K has compact support $[-L, L]$, K is twice continuously differentiable and $\int uK(u) du = 0$.
- (iii) For $j = 1, 2$, $\sup_y \tau_j(y) \leq \tau_j < \infty$.
- (A2)(i) $H_1(y)$ and $H_1^u(y)$ are three times continuously differentiable with respect to y and $\inf_{y \leq \tau_1} h_1^u(y) > 0$.

(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 all derivatives are bounded uniformly on $[0, \tau_1] \times [0, \tau_2]$.

(A3)(i) $H_2(y)$ and $H_2^u(y)$ are three times continuously differentiable with respect to y and $\inf_{y \leq \tau_2} h_2^u(y) > 0$.

(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 all derivatives are bounded uniformly on $[0, \tau_1] \times [0, \tau_2]$.

3.1 Case 1 : T_2 independent of C_2 given T_1 ; T_1, T_2 independent of C_1

Theorem 3.1 Assume (A1) – (A2). Then,

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

where

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

and $\Omega = \{\mathbf{y} : y_1 \leq \tau_1 \text{ and } y_2 \leq \tau_2(y) \text{ for all } y \leq y_1\}$.

Theorem 3.2 Assume (A1) – (A2).

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

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

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

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

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

Because the asymptotic theory for the estimator $\hat{F}_2(y_2)$ of the marginal distribution of T_2 is based on an i.i.d. representation for

$\tilde{F}_1(y_1)$ which is valid up to τ_1 , we need to work with a slightly modified version of $\hat{F}_2(y_2)$. Namely, the asymptotic result shown below pertains to

$$\hat{F}_{2,\tau}(y_2) = \int_0^{\tau_1} \hat{F}_{2|1}(y_2|z_1) d\tilde{F}_1(z_1).$$

This is actually an estimator of $F_{2,\tau}(y_2) = \int_0^{\tau_1} F_{2|1}(y_2|z_1) dF_1(z_1)$, which can become arbitrarily close to $F_2(y_2)$ if $\inf\{y : F_1(y) = 1\} \leq \inf\{y : G_1(y) = 1\}$.

Corollary 3.3 *Assume (A1) – (A2).*

- *If $nh_n^4 \rightarrow 0$, then $n^{1/2}(\hat{F}_{2,\tau}(y_2) - F_{2,\tau}(y_2))$ ($y_2 \leq \inf_{y \leq \tau_1} \tau_2(y)$) converges weakly to a zero mean Gaussian process $Z(y_2)$ with covariance function*

$$\begin{aligned} \text{Cov}(Z(y_2), Z(y'_2)) &= E \left[\int_0^{\tau_1} \int_0^{\tau_1} F_{2|1}(y_2|z_1) F_{2|1}(y'_2|z'_1) d\xi_1(\tilde{T}_1, \Delta_1, z_1) d\xi_1(\tilde{T}_1, \Delta_1, z'_1) \right] \\ &\quad + E[I(\Delta_1 = 1)I(\tilde{T}_1 \leq \tau_1)\tau_{2|1}(\tilde{T}_2, \Delta_2, y_2|\tilde{T}_1)\tau_{2|1}(\tilde{T}_2, \Delta_2, y'_2|\tilde{T}_1)]. \end{aligned}$$

- *If $nh_n^4 = K$ (for some $K > 0$), then $n^{1/2}(\hat{F}_{2,\tau}(y_2) - F_{2,\tau}(y_2))$ ($y_2 \leq \inf_{y \leq \tau_1} \tau_2(y)$) converges weakly to a Gaussian process $Z'(y_2)$ with mean function*

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

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

This result follows readily from Theorem 3.2.

3.2 Case 2 : T independent of C

Theorem 3.4 *Assume (A1) – (A3). Then,*

$$\begin{aligned} &\hat{F}(\mathbf{y}) - F(\mathbf{y}) \\ &= n^{-1} \sum_{i=1}^n [w(\mathbf{y})\eta_{1|2}(\tilde{T}_i, \Delta_i, \mathbf{y}) + (1 - w(\mathbf{y}))\eta_{2|1}(\tilde{T}_i, \Delta_i, \mathbf{y})] \\ &\quad + h_n^2 [w(\mathbf{y})\beta_{1|2}(\mathbf{y}) + (1 - w(\mathbf{y}))\beta_{2|1}(\mathbf{y})] + R_n(\mathbf{y}), \end{aligned}$$

where

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

and $\Omega' = \{\mathbf{y} : y_1 \leq \tau_1(y) \text{ for all } y \leq y_2 \text{ and } y_2 \leq \tau_2(y) \text{ for all } y \leq y_1\}$.

Theorem 3.5 Assume (A1) – (A3).

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

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

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

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

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

For similar reasons as in Section 3.1, we have to consider the following slightly different versions of the estimators $\hat{F}_1(y_1)$ and $\hat{F}_2(y_2)$ defined in Section 2.2 :

$$\hat{F}_{1,\tau}(y_1) = w_1(y_1) \int_0^{\tau_2} \hat{F}_{1|2}(y_1|z_2) d\tilde{F}_2(z_2) + (1 - w_1(y_1))\tilde{F}_1(y_1)$$

and analogously, $\hat{F}_{2,\tau}(y_2)$ is defined.

Corollary 3.6 Assume (A1) – (A3).

- If $nh_n^4 \rightarrow 0$, then $n^{1/2}(\hat{F}_{1,\tau}(y_1) - F_{1,\tau}(y_1))$ ($y_1 \leq \inf_{y \leq \tau_2} \tau_1(y)$) converges weakly to a zero mean Gaussian process $W(y_1)$ with covariance function

$$\begin{aligned} & \text{Cov}(W(y_1), W(y'_1)) \\ &= E[\{w_1(y_1)\eta_{1|2}(\tilde{\mathbf{T}}, \Delta, y_1, \tau_2) + (1 - w_1(y_1))\xi_1(\tilde{T}_1, \Delta_1, y_1)\} \\ & \quad \times \{w_1(y'_1)\eta_{1|2}(\tilde{\mathbf{T}}, \Delta, y'_1, \tau_2) + (1 - w_1(y'_1))\xi_1(\tilde{T}_1, \Delta_1, y'_1)\}]. \end{aligned}$$

- If $nh_n^4 = K$ (for some $K > 0$), then $n^{1/2}(\hat{F}_{1,\tau}(y_1) - F_{1,\tau}(y_1))$ ($y_1 \leq \inf_{y \leq \tau_2} \tau_1(y)$) converges weakly to a Gaussian process $W'(y_1)$ with mean function

$$E(W'(y_1)) = K^{1/2}w_1(y_1)\beta_{1|2}(y_1, \tau_2)$$

and the same covariance function as $W(y_1)$.

The corresponding weak convergence result for $\hat{F}_{2,\tau}(y_2)$ can be obtained in exactly the same manner (simply interchange the role of the first and second variable in the above result). The proofs of Theorems 3.4, 3.5 and Corollary 3.6 are very similar to those of Theorems 3.1, 3.2 and Corollary 3.3 respectively, and will therefore be omitted.

Remark 2.1. The above result enables us to choose the weights $w(\mathbf{y})$ in such a way that the mean squared error of the estimator $\hat{F}(\mathbf{y})$ is minimal. Denote the j -th integral in the definition of $\hat{F}(\mathbf{y})$ given in (2.5) by $T_j(\mathbf{y})$ ($j = 1, 2$). Simple calculations show that the weights

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

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

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 $\hat{F}(\mathbf{y})$ 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 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 $\mathbf{T}$ 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

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

$P(\Delta_{1,2} = 0)$	ρ_T	n	Bias			Variance		
			$\tilde{F}_P(\mathbf{y})$	$\tilde{F}_L(\mathbf{y})$	$\hat{F}(\mathbf{y})$	$\tilde{F}_P(\mathbf{y})$	$\tilde{F}_L(\mathbf{y})$	$\hat{F}(\mathbf{y})$
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

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 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,*

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

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

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

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

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

where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of $(\tilde{\mathbf{T}}, \mathbf{\Delta})$, 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

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

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,

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

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

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

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

$$\hat{F}(\mathbf{y}) - F(\mathbf{y})$$

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

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

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

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 $\dot{F}_{2|1}(y_2|y_1)$,

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

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

In the rest of the proof we consider $A_1(\mathbf{y})$, 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),

$$\begin{aligned}
&A_1(\mathbf{y}) \\
&= (nh_n)^{-1} \sum_{\Delta_{1i}=1} \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1) I(z_1 \leq y_1) dz_1 + o_P(n^{-1/2}) \\
&= (nh_n)^{-1} \sum_{\Delta_{1i}=1} \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1) [I(z_1 \leq y_1) - I(T_{1i} \leq y_1)] dz_1 \\
&\quad + (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \leq y_1) \int K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1) dz_1 + o_P(n^{-1/2}) \\
&= A_{11}(\mathbf{y}) + A_{12}(\mathbf{y}) + o_P(n^{-1/2}).
\end{aligned}$$

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

$$\begin{aligned}
& A_{11}(\mathbf{y}) \\
&= -(nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \leq T_{1i} \leq y_1) \int_{y_1}^{T_{1i}+Lh_n} K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1) dz_1 \\
&\quad + (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 \leq T_{1i} \leq y_1 + Lh_n) \int_{T_{1i}-Lh_n}^{y_1} K\left(\frac{z_1 - T_{1i}}{h_n}\right) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|z_1) dz_1 \\
&= -A_{111}(\mathbf{y}) + A_{112}(\mathbf{y}).
\end{aligned}$$

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

$$\begin{aligned}
& A_{111}(\mathbf{y}) \\
&= (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \leq T_{1i} \leq y_1) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) \int_{y_1}^{T_{1i}+Lh_n} K\left(\frac{z_1 - T_{1i}}{h_n}\right) dz_1 \\
&\quad + (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \leq T_{1i} \leq y_1) \tau'_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|T_{1i}) \\
&\quad \quad \times \int_{y_1}^{T_{1i}+Lh_n} K\left(\frac{z_1 - T_{1i}}{h_n}\right) (z_1 - T_{1i}) dz_1 \\
&\quad + \frac{1}{2} (nh_n)^{-1} \sum_{\Delta_{1i}=1} I(y_1 - Lh_n \leq T_{1i} \leq y_1) \int_{y_1}^{T_{1i}+Lh_n} \tau''_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2|\xi_{1i}) \\
&\quad \quad \times K\left(\frac{z_1 - T_{1i}}{h_n}\right) (z_1 - T_{1i})^2 dz_1 \\
&= A_{1111}(\mathbf{y}) + A_{1112}(\mathbf{y}) + A_{1113}(\mathbf{y}),
\end{aligned}$$

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
& + \frac{1}{2} \mu_2^K h_n^2 n^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \leq y_1) \tau_{2|1}''(\tilde{T}_{2i}, \Delta_{2i}, y_2 | T_{1i}) + o_P(h_n^2) \\
& = n^{-1} \sum_{\Delta_{1i}=1} I(T_{1i} \leq y_1) \tau_{2|1}(\tilde{T}_{2i}, \Delta_{2i}, y_2 | T_{1i}) \\
& + \frac{1}{2} \mu_2^K h_n^2 \int I(z_1 \leq y_1) \tau_{2|1}''(z_2, \delta_2, y_2 | z_1) I(\delta_1 = 1) dH(\mathbf{z}, \boldsymbol{\delta}) \\
& + \frac{1}{2} \mu_2^K h_n^2 \int I(z_1 \leq y_1) \tau_{2|1}''(z_2, \delta_2, y_2 | z_1) I(\delta_1 = 1) d(\hat{H}(\mathbf{z}, \boldsymbol{\delta}) - H(\mathbf{z}, \boldsymbol{\delta})) + o_P(h_n^2),
\end{aligned}$$

where $\hat{H}(\mathbf{z}, \boldsymbol{\delta}) = n^{-1} \sum_{i=1}^n I(\tilde{T}_{1i} \leq z_1, \tilde{T}_{2i} \leq z_2, \Delta_{1i} = \delta_1, \Delta_{2i} = \delta_2)$. 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 $\mathbf{y} \in \Omega$. Proving this is equivalent to showing that the class $\mathcal{F} = \{I(\tilde{T}_1 \leq y_1) \tau_{2|1}''(\tilde{T}_2, \Delta_2, y_2 | \tilde{T}_1) I(\Delta_1 = 1) : \mathbf{y} \in \Omega\}$ 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 $(\tilde{\mathbf{T}}, \boldsymbol{\Delta})$, $L_2(P)$ is the L_2 -norm, and $\mathcal{F} = \{\eta_{2|1}(\tilde{\mathbf{T}}, \boldsymbol{\Delta}, \mathbf{y}) : \mathbf{y} \in \Omega\}$.

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 $\eta_{2|1}(\mathbf{t}, \boldsymbol{\delta}, \mathbf{y})$. Write

$$\begin{aligned}
& \int_0^{y_1} F_{2|1}(y_2 | z_1) d\xi_1(t_1, \delta_1, z_1) \\
& = F_{2|1}(y_2 | y_1) \xi_1(t_1, \delta_1, y_1) - \int_0^{y_1} \xi_1(t_1, \delta_1, z_1) dF_{2|1}(y_2 | z_1) \\
& = T_1(\mathbf{y}) + T_2(\mathbf{y}).
\end{aligned}$$

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 $I(t_1 \leq y_1)$ 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 $T_1(\mathbf{y})$ is $O(\exp(K\varepsilon^{-1}))$. Next, write

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

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 $T_2(\mathbf{y})$ is also $O(\exp(K\varepsilon^{-1}))$. It remains to calculate the bracketing number of the second term of $\eta_{2|1}(\mathbf{t}, \boldsymbol{\delta}, \mathbf{y})$. 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 $\eta_{2|1}(\tilde{\mathbf{T}}, \boldsymbol{\Delta}, \mathbf{y})$). 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.
- Cheng, P. E. (1989). Nonparametric estimation of survival curve under dependent censorship. *J. Statist. Plan. Infer.*, **23**, 181-191.
- 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*, in press.
- Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap : some asymptotic representations. *Probab. Th. Rel. Fields*, **71**, 455-465.
- 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.
- 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.