



Bivariate Estimation With Right-Truncated Data

Author(s): Ülkü Gürler

Source: *Journal of the American Statistical Association*, Vol. 91, No. 435 (Sep., 1996), pp. 1152-1165

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: <https://www.jstor.org/stable/2291734>

Accessed: 04-02-2019 06:01 UTC

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/2291734?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

American Statistical Association, Taylor & Francis, Ltd. are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

Bivariate Estimation With Right-Truncated Data

Ülkü GÜRLER

Bivariate estimation with survival data has received considerable attention recently; however, most of the work has focused on random censoring models. Another common feature of survival data, random truncation, is considered in this study. Truncated data may arise if the time origin of the events under study precedes the observation period. In a random right-truncation model, one observes the iid samples of (Y, T) only if $(Y \leq T)$, where Y is the variable of interest and T is an independent variable that prevents the complete observation of Y . Suppose that (Y, X) is a bivariate vector of random variables, where Y is subject to right truncation. In this study the bivariate reverse-hazard vector is introduced, and a nonparametric estimator is suggested. An estimator for the bivariate survival function is also proposed. Weak convergence and strong consistency of this estimator are established via a representation by iid variables. An expression for the limiting covariance function is provided, and an estimator for the limiting variance is presented. Alternative methods for estimating the bivariate distribution function are discussed. Obtaining large-sample results for the bivariate distribution functions present more technical difficulties, and thus their performances are compared via simulation results. Finally, an application of the suggested estimators is presented for transfusion-related AIDS (TR-AIDS) data on the incubation time.

KEY WORDS: Bivariate distribution; Nonparametric estimation; Reverse hazard; Weak convergence.

1. INTRODUCTION

Randomly truncated data frequently arise in medical studies; other application areas include economics, insurance and astronomy. In a broad sense, random truncation corresponds to biased sampling, where only partial or incomplete data are available about the variable of interest. A typical realization can occur as follows: Suppose that individuals/items experience two consecutive events in time, an initiating event at t and a terminating event at s . Usually, statistical interest is in the duration between the two. Random truncation may occur, if the observation period starts after the initiating event. Consider the following example: Suppose that an individual is infected with human immuno-deficiency virus (HIV) at time t and diagnosed with acquired immune deficiency syndrome (AIDS) at time s . If the observation period is terminated at T_e , then only those individuals for whom the incubation time $Y = t - s \leq T = T_e - s$ can be observed, and right truncation occurs. In AIDS cohort studies, a group of patients who are infected with HIV but have not yet developed AIDS are selected. If the recruitment starts at T_0 and the follow-up is terminated at T_e , then only those individuals for whom $Y = t - s \geq T = T_0 - s$ are observed, and random left truncation occurs, which could also give rise to right censoring. In this situation, the bivariate lifetime data could occur if one is also interested in pediatric AIDS, as suggested by a referee. In particular, in a sample of pregnant women with HIV-infected babies, the incubation times of the mothers and the time from birth to development of AIDS for the babies constitute a bivariate data where one component is subject to left truncation. More complicated situations would occur if the lifetimes could then be censored. Both components would be under a left truncation

effect if only those women are selected who already have an HIV-positive child. Earlier onset of AIDS would then be a truncating force for both. The bivariate data may also correspond to a lifetime and a covariate; for example, age or gender. An application with AIDS data considering age as a covariate is illustrated in Section 4. This incomplete structure of data induced by truncation obviously creates bias in estimation; in epidemics, this is particularly important at the early stages of the disease, when sufficient historical data have not yet accumulated.

Consider first the univariate truncation model, where one observes the iid pairs $(Y_i, T_i), i = 1, \dots, n$, only if $(Y_i \leq T_i)$, where the main interest is in Y . Let F and G be the distribution functions of Y and T . Woodroffe (1985) supplied the following identifiability condition: F and G can be estimated completely only if $(F, G) \in R_o$, where $R_o = \{(F, G): a_G \leq a_F; b_G \leq b_F\}$, with a_W and b_W denoting the lower and upper endpoints of the support of any distribution function W . Then $(F, G) \in R_o$ implies $\alpha \equiv P(T \leq Y) > 0$. Here α is the proportion of the bivariate population (Y, T) , which can be observed under the truncation scheme. Assuming the identifiability of F and G , their nonparametric estimators, F_n and G_n , are given by

$$F_{Y,n}(y) = \prod_{i: Y_i > y} [1 - s(Y_i)/nC_n(Y_i)]$$

and

$$G_n(t) = 1 - \prod_{i: T_i \leq t} [1 - r(T_i)/nC_n(T_i)], \quad (1)$$

where for $u > 0, r(u) = \#\{i: T_i = u\}, s(u) = \#\{i: Y_i = u\}$, and

$$nC_n(u) = \#\{i: Y_i \leq u \leq T_i\}. \quad (2)$$

Lynden-Bell (1971) suggested $G_n(t)$ as the nonparametric maximum likelihood estimator (MLE) of $G(t)$ in a problem

Ülkü Gürler is Associate Professor, Industrial Engineering Department, Bilkent University, 06533 Bilkent, Ankara, Turkey. Part of this work was done while the author was visiting the Institute of Statistics, Catholic University of Louvain, Belgium, in the academic year 1993–1994. The author thanks an associate editor and the two referees, whose constructive comments led to a much-improved manuscript.

that arose in astronomy. Woodroffe (1985) and Wang, Jewell, and Tsai (1986) obtained consistency and weak convergence results over compact intervals; Chen, Chao, and Lo (1995) recently extended these to the whole real line. Chao and Lo (1988) presented an iid representation of $G_n(t)$, for which extensions and improvements were given by Stute (1993) and Gijbels and Wang (1993). Kernel estimators of the hazard function for Left Truncated Right Censored (LTRC) data were studied by Uzunogullari and Wang (1992). Gross and Huber-Carol (1992), Gürler, Stute, and Wang (1993), and Lai and Ying (1991) extended the results for truncated/censored data in various directions; Keiding and Gill (1990) presented a Markov process approach to the model. Recently, Gürler (1996) studied a nonparametric estimator for the bivariate distribution function when a component is subject to left truncation; this study presented nonparametric estimators for the bivariate distribution function and the diverse hazard vector and established their strong consistency and weak convergence via strong iid representations. In the present study, the results of Gürler (1996) are extended/generalized in the following directions:

- A nonparametric estimator for the bivariate “reverse-hazard” vector is proposed.
- An estimator for the bivariate survival function is proposed. An expression for the limiting covariance function and a nonparametric estimator for the limiting variance are provided.
- Alternative methods for estimating the bivariate distribution function are discussed.
- An application of the methods with a real data set is illustrated.

The rest of the article is organized as follows. In Section 2 the bivariate model is introduced and nonparametric estimators are provided. In Section 3 large-sample results are provided. In Section 4 simulation results and an application of these methods with TR-AIDS data are presented. Finally, concluding remarks are given in Section 5.

2. BIVARIATE MODEL AND THE ESTIMATORS

2.1 Preliminaries and Notation

Bivariate distribution, or survival function, is important in understanding the joint behavior of correlated lifetimes, as well as in assessing the strength of such association. Several recent studies have elaborated the subject (Akritas 1994; Lin and Ying 1993; Prentice and Cai 1993; and van der Laan 1993). However, all of these studies considered variations of the bivariate censored model. Despite the important applications in survival analysis, bivariate estimation with truncated data has not received much attention in the literature so far. In the present study nonparametric estimation procedures are considered for bivariate data when a component is subject to right truncation. The estimators are often compared to their counterparts in the censored case. Because the data structures and the sampling schemes in these cases are rather different, such analogies are based on the structure of the resulting estimators and the technical tools

involved therein. It turns out that similarities among censored and truncated models mostly hold between the singly (i.e., one component) truncated and doubly (i.e., both components) censored bivariate data. This may be attributed to the fact that the truncation effect introduces more complicated identifiability problems. Note also that for doubly truncated data, there do not exist estimators for the bivariate distribution/survival function, which is an open area for research.

Suppose that we are interested in the joint behavior of the random pair (Y, X) but, due to truncation effects, we can only observe the triplets (Y_i, X_i, T_i) , $i = 1, \dots, n$ for which $(Y_i \leq T_i)$. Here T is a random variable, which is assumed to be independent of (Y, X) , with distribution function G . Violation of this assumption could create additional bias in estimation problems. In the censored case, particularly in the estimation of regression coefficients, this violation creates significant problems. However, Tsai (1990) showed that the independence assumption could be tested for the truncated data, unlike the situation in censoring. The marginal distribution functions of Y and X are denoted by F_Y and F_X . For the identifiability of F , it is assumed that $a_{F_Y} \leq a_G$ and $b_{F_Y} \leq b_G$, as in the univariate model. Let $F(y, x) = P(Y \leq y, X \leq x)$ be the bivariate distribution function, and, with some abuse of notation, let $\bar{F}(y, x) = P(Y > y, X > x)$ be the bivariate survival function. For a univariate distribution function F , the survival function is $\bar{F} = 1 - F$. The observed samples have the transformed distribution H , given by

$$\begin{aligned} H_{Y,X,T}(y, x, t) &= P(Y \leq y, X \leq x, T \leq t | Y \leq T) \\ &= \alpha^{-1} \int_0^t F(y \wedge u, x) dG(u), \end{aligned}$$

where α is as defined before and $y \wedge u = \min(y, u)$. The problem is to reconstruct the latent $F(y, x)$ from the observable H and its marginals. Some of the bivariate and univariate marginals are as follows:

$$\begin{aligned} F_{Y,X}^*(y, x) &\equiv H_{Y,X,T}(y, x, \infty) \\ &= \alpha^{-1} \int_0^\infty F(y \wedge u, x) dG(u), \end{aligned}$$

$$H_{X,T}(x, t) = \alpha^{-1} \int_0^t F(u, x) dG(u),$$

$$H_{Y,T}(y, t) = \alpha^{-1} \int_0^t F_Y(y \wedge u) dG(u),$$

$$F_Y^*(y) = \alpha^{-1} \int_0^\infty F_Y(y \wedge u) dG(u),$$

and

$$G_T^*(t) = H_{Y,T}(\infty, t) = \alpha^{-1} \int_0^t F_Y(u) dG(u).$$

Assuming the existence of the densities (denoted in lower-case letters), we get

$$f_{Y,X}^*(y, x) = \alpha^{-1} [1 - G(y)] f_{Y,X}(y, x)$$

and

$$f_Y^*(y) = \alpha^{-1}[1 - G(y)]f_Y(y). \quad (3)$$

Another quantity of interest, which provides insight to the truncation model, is

$$C(z) = \alpha^{-1}\bar{G}(z-)F_Y(z) = F_Y^*(z) - G_T^*(z). \quad (4)$$

This quantity plays an important role in the identification and estimation of the model. The empirical counterpart, $C_n(z)$, is proportional to the size of the "risk set" at time z , which is given in (2). But unlike the usual risk sets used in survival analysis, (2) is not a monotone function. It tails off to zero at both ends, which introduces additional difficulty to the analysis. The bivariate functions considered next are assumed to be either discrete or differentiable at the continuity points, to avoid introducing more notation. For a bivariate function $\phi(u, v)$, let

$$\phi(\delta u, v) = \phi(u, v) - \phi(u-, v)$$

and

$$\phi(u, \delta v) = \phi(u, v) - \phi(u, v-).$$

Define

$$\phi_{1c} = \{u: \phi(\delta u, v) = 0\}, \quad \phi_{1d} = \{u: \phi(\delta u, v) \neq 0\},$$

$$\phi_{2c} = \{v: \phi(u, \delta v) = 0\}, \quad \phi_{2d} = \{v: \phi(u, \delta v) \neq 0\},$$

and

$$\phi(\partial u, v) = \begin{cases} \partial/\partial u \phi^1(u, v), & u \in \phi_{1c} \\ \phi(\delta u, v), & u \in \phi_{1d} \end{cases}$$

A similar definition holds for $\phi(u, \partial v)$, so that

$$\phi(\partial u, \partial v) = \begin{cases} (\partial^2/\partial u \partial v) \phi(u, v), & u \in \phi_{1c}, \quad v \in \phi_{2c} \\ (\partial/\partial u) \phi(u, \delta v), & u \in \phi_{1c}, \quad v \in \phi_{2d} \\ (\partial/\partial v) \phi(\delta u, v), & u \in \phi_{1d}, \quad v \in \phi_{2c} \\ \phi(\delta u, \delta v), & u \in \phi_{1d}, \quad v \in \phi_{2d} \end{cases}$$

Finally, the notation for the integration of the foregoing functions will be:

$$\begin{aligned} \int \phi(\partial u, v) &= \int \phi(\partial u v) du \\ &= \int_{\phi_{1c}} \phi(\partial u, v) du + \sum_{\phi_{1d}} \phi(\partial u, v). \end{aligned}$$

Defining $\int \phi(\partial u, v)$ similarly, we obtain for the double integral

$$\begin{aligned} \iint \phi(\partial u, \partial v) &= \int_{\phi_{2c}} \int_{\phi_{1c}} \phi(\partial u, \partial v) du dv \\ &+ \sum_{\phi_{2d}} \sum_{\phi_{1d}} \phi(\partial u, \partial v) + \int_{\phi_{2c}} \sum_{\phi_{1d}} \phi(\partial u, \partial v) dv \\ &+ \sum_{\phi_{2d}} \int_{\phi_{1c}} \phi(\partial u, \partial v) du. \end{aligned}$$

2.2 Bivariate "Reverse Hazard"

In the univariate models, it is well known that the hazard function and the distribution function determine each other in a unique way. This correspondence has proven useful in obtaining an estimator of the distribution function via that of the hazard, particularly with incomplete data. In the bivariate case, however, there have been several definitions of the hazard function or the failure rate. Dabrowska (1988) presented a nice representation of the bivariate distribution function in terms of the three-component bivariate hazard vector. We introduce here the bivariate reverse-hazard vector, which is analogous to her hazard vector and which turns out to be the natural quantity to consider in the right-truncation model.

Definition. For the bivariate distribution function $F(y, x)$, define the bivariate "reverse-hazard" vector $\lambda(u, v)$ as

$$\begin{aligned} \lambda(u, v) &= \{F(\partial u, \partial v)/F(u, v), F(\partial u, v)/F(u, v), \\ &\quad F(u, \partial v)/F(u, v)\} \\ &\equiv \{\Lambda_{12}(\partial u, \partial v), \Lambda_1(\partial u, v), \Lambda_2(u, \partial v)\}. \end{aligned} \quad (5)$$

The term "reverse hazard" is adopted as coined by Lagakos, Barraj, and DeGruttola (1988) in the univariate case, because in the usual hazard an immediate "future" failure is considered, whereas the foregoing vector relates to instantaneous "past" failure. Gross and Huber-Carol (1992) used the term "retro hazard" for the same quantity in the univariate setup. The general correspondence between the reverse-hazard vector and a bivariate survival function can be established following Dabrowska (1988) and is not presented here. However, when F is continuous and the density exists, we have the following relation: Let $R(y, x) = -\log F(y, x)$; then

$$F(y, x) = F_X(x)F_Y(y)\exp\{-\Lambda(y, x)\}, \quad (6)$$

where

$$\begin{aligned} \Lambda(y, x) &= \int_y^{b_{FY}} \int_x^{b_{FX}} R(du, dv) = \int_y^{b_{FY}} \int_x^{b_{FX}} \\ &\quad \times \{\Lambda_{12}(du, dv) - \Lambda_1(du, v)\Lambda_2(u, dv)\}. \end{aligned}$$

From the relations given in the previous section, it is not hard to verify the following:

$$\Lambda_{12}(\partial u, \partial v) = \frac{F_{Y,X}^*(\partial u, \partial v)}{C_2(u, v)},$$

$$\Lambda_1(\partial u, v) = \frac{F_{Y,X}^*(\partial u, v)}{C_2(u, v)},$$

and

$$\Lambda_2(u, \partial v) = \frac{C_2(u, \partial v)}{C_2(u, v)},$$

where

$$\begin{aligned} C_2(y, x) &= F_{Y,X}^*(y, x) - H_{T,X}(y-, x) \\ &= \alpha^{-1}[1 - G(y)]F(y, x). \end{aligned} \quad (7)$$

Let $F_{n,Y,X}^*(y, x)$ and $H_{n,T,X}(y-, x)$ be the empirical bivariate distribution functions of the observed pairs (Y, X) and (T, X) , and define the size of bivariate risk at time (u, v) as

$$\begin{aligned} nC_{2,n}(u, v) &= \#\{i: Y_i \leq u \leq T_i, X_i \leq v\} \\ &= n[F_{n,Y,X}^*(u, v) - H_{n,T,X}(u-, v)]. \end{aligned}$$

Then we have the following estimator for the reverse hazard:

$$\begin{aligned} \lambda_n(u, v) &= \left\{ \frac{F_{n,Y,X}^*(\partial u, \partial v)}{C_{2,n}(u, v)}, \right. \\ &\quad \left. \frac{F_{n,Y,X}^*(\partial u, v)}{C_{2,n}(u, v)}, \frac{C_{2,n}(u, \partial v)}{C_{2,n}(u, v)} \right\} \\ &\equiv \{\Lambda_{12,n}(\partial u, \partial v), \Lambda_{1,n}(\partial u, v), \Lambda_{2,n}(u, \partial v)\}. \end{aligned} \quad (8)$$

Note here that the reverse-hazard vector describes the identifiable components of the truncation model and is also of independent interest, because it describes the univariate and bivariate failure behavior in the immediate past. (See, e.g., Gross and Huber-Carol 1992 and Pons 1986 for hazard based inferences and further applications.)

2.3 Bivariate Survival and Distribution Function

Now we consider estimation of the joint survivor function, $\bar{F}(y, x) = P(Y > y, X > x)$, of (Y, X) . The following estimator is motivated by the relations given in (3) and (4), which lead to

$$\bar{F}_n(y, x) = \frac{1}{n} \sum_i \frac{F_{Y,n}(Y_i)}{C_n(Y_i)} I(Y_i > y, X_i > x). \quad (9)$$

In the next section we discuss the large-sample properties of \bar{F}_n and establish the strong consistency and weak convergence to a two-time parameter Gaussian process via an iid representation. We also provide the covariance function of the limiting process together with a nonparametric estimator of the asymptotic variance. These results are obtained when y is bounded away from the origin, which is inherited from similar properties of $F_{Y,n}(y)$ over the compact intervals away from a_{F_Y} . Estimation of the bivariate distribution function in the right-truncation model involves more technical difficulties, because in this case $F_Y(y)$ must be estimated at the lower tail. Next we present several approaches for estimating the bivariate distribution function. We first note that the relations (3) and (4), which lead to $\bar{F}_n(y, x)$, also allow us to estimate the bivariate distribution function $F(y, x)$ in the following manner:

$$F_{Y,n}^1(y, x) = \frac{1}{n} \sum_i \frac{F_{Y,n}(Y_i)}{C_n(Y_i)} I(Y_i \leq y, X_i \leq x). \quad (10)$$

The simple structure of this estimator is very appealing in comparison to the alternative estimators discussed later. It can easily be verified that $F_n^1(y, x)$ is a bivariate dis-

tribution function and that its marginal for Y reduces to $F_n(y)$ given in (1). On the other hand, as mentioned earlier, $F_n^1(y, x)$ involves integration of $F_{Y,n}(u)$ over the lower tail $[0, y]$. So far, we can obtain asymptotic orders for the convergence of $|F_n(u) - F(y)|$ only over intervals away from the left tail. Recently, Chen et al. (1995) proved that $\sup |F_n(u) - F(y)| = o(1)$ a.s. over the entire real line. Therefore, results similar to Theorem 1 in Section 3 are not immediate for $F_n^1(y, x)$. In the next section, however, the strong consistency of (10) is presented within more restricted intervals (cf. Thm. 2). Burke (1988) proposed an estimator with a similar structure for the bivariate data when both components are censored. His estimator could exceed the nominal bound of 1, which $F^1(y, x)$ does not suffer because it is dominated by $F_{Y,n}(y)$ for each x . This feature of Burke's estimator could be due to the additional complication created by the censoring in both components. Stute (1993b) studied a similar estimator for singly censored bivariate data, and Dabrowska (1995) used a variant of it in nonparametric regression context. Large-sample properties of $F_n^1(y, x)$ are discussed in Section 3.

2.4 Alternative Methods

Here we present some alternative approaches to estimating $F(y, x)$. Recall that if the bivariate density exists, then

$$\begin{aligned} F(y, x) &= F_X(x)F_Y(y)\exp\{-\Lambda(y, x)\} \\ &\equiv F_X(x)F_Y(y)\Theta(y, x). \end{aligned}$$

This relation suggests the following estimators for $i = 2, 3$:

$$F_n^i(y, x) = F_{X,n}(x)F_{Y,n}(y)\Theta_{i,n}(y, x),$$

where $F_{X,n}(x)$ and $F_{Y,n}(y)$ are the marginals of $F_{Y,X,n}^1(y, x)$ and

$$\Theta_{2,n}(y, x) = \exp\{-\Lambda_n(y, x)\},$$

$$\begin{aligned} \Lambda_n(y, x) &= \sum_{u>y} \sum_{v>x} [\Lambda_{1,n}(\partial u, v)\Lambda_{2,n}(u, \partial v) \\ &\quad - \Lambda_{12,n}(\partial u, \partial v)], \end{aligned}$$

and

$$\Theta_{3,n}(y, x) = \prod_{u>y; v>x} \frac{1 - \Lambda_{1,n}(\partial u, v) - \Lambda_{2,n}(u, \partial v) + \Lambda_{12,n}(\partial u, \partial v)}{[1 - \Lambda_{1,n}(\partial u, v)][1 - \Lambda_{2,n}(u, \partial v)]}.$$

Dabrowska (1988) used this estimator for bivariate data when both components are censored. The foregoing estimator $\Theta_{3,n}(y, x)$ is obtained by considering the discrete nature of empirical distributions. Note here that these estimators could assign negative mass to observed data points—an undesirable property. This problem also occurs with bivariate censored data (see, e.g., Dabrowska 1988 and Pruitt 1992). The difficulty in obtaining large-sample results for these estimators stems from the estimation of the X marginal.

Observe that the X marginal is obtained by integrating the bivariate estimator over an infinite region, which creates similar problems as discussed earlier for $F_n^1(y, x)$. Accordingly, we fail to establish the consistency of these estimators. Thus the relative performance of the estimators for the bivariate distribution function are studied via simulations. On the other hand, the estimators $\Theta_{i,n}(y, x)$ are of interest on their own, because they could provide a basis for other statistical methods, such as tests of independence, hazard-based regression, and model validation. They are consistent and Gaussian away from the lower tail of the Y distribution, which can be stated more precisely (as in Gürlér 1996) and is not further elaborated here. We next present two more approaches for estimating $F(y, x)$. The first approach simply follows from the relations given by (3) and (7), leading to

$$F_n(y, x) = F_{Y,n}(y)[C_{2,n}(y, x)/C_n(y)]. \quad (11)$$

This estimator is similar in spirit to the one suggested by Campbell (1981), which is based on a decomposition of the bivariate distribution function to a marginal and a conditional component. The other approach, presented in (12), is based on a path-dependent line integral of the univariate hazard function. This idea goes back to Campbell and Földes (1982) for the censored observations. Observe that

$$\begin{aligned} R(y, x) &= -\int_y^\infty R(du, \infty) + \int_x^\infty R(y, dv) \\ &= \int_y^\infty [F_Y(du)/F_Y(u)] + \int_x^\infty [F(y, dv)/F(y, v)] \\ &= \int_y^\infty [F_Y^*(du)/C(u)] + \int_x^\infty [C_{2,n}(y, dv)/C(y, v)] \end{aligned}$$

$$F_n(y, x) = \exp\{-R_n(y, x)\} \quad (12)$$

with

$$\begin{aligned} R_n(y, x) &= \sum_{i: Y_i > y} [nC_n(Y_i)]^{-1} \\ &\quad + \sum_{j: X_j > x} [nC_{2,n}(y, X_j)]^{-1} \cdot I(Y_j \leq y \leq T_j). \end{aligned}$$

The estimators (11) and (12) are nondecreasing in x , but not so in the y component. They could assign negative mass as the other estimators described earlier, and in the preliminary simulations they were dominated significantly by F_n^1 . The following comparison could give an idea. Consider the ratio of the estimated mean squared error (MSE) for F_n^1 to that of F_n^i , $i = 2, 3$. In the preliminary simulations of the next section, the minimum of such ratios was .996, indicating a comparable performance. However, for the alternative estimators presented earlier, the minimum for the better one was as low as .88 and the best was about .96 (Details of these simulations can be obtained from the author). Thus we did not consider these estimators further.

3. LARGE-SAMPLE PROPERTIES

In this section we present some asymptotic results for $\bar{F}_n(y, x)$ and $F_n^1(y, x)$. For $i = 1, \dots, n$, define

$$L_{i,n}(z) = \frac{I(Y_i > z)}{C(Y_i)} - \int_z^\infty \frac{I(Y_i \leq u \leq T_i)}{C^2(u)} F^*(du)$$

and

$$\bar{L}_n(z) = \frac{1}{n} \sum_i L_{i,n}(z) \quad A(z) = \frac{F(z)}{C(z)}.$$

Theorem 1. Suppose that $F(y, x)$ admits the density $f(y, x)$, and let $T_a = \{(y, x): y > a > a_F; 0 < x < \infty\}$. Suppose also that $\int_{a_F}^{b_F} F(du)/[1 - G(u)]^2 < \infty$. Then the following representation holds:

$$\begin{aligned} \bar{F}_n(y, x) - \bar{F}(y, x) &= [\bar{F}_n^*(y, x) - \bar{F}^*(y, x)]A(y) \\ &\quad + \int_y^\infty [\bar{F}_n^*(u, x) - \bar{F}^*(u, x)]A(du) \\ &\quad + \int_y^\infty \frac{A(u)}{C(u)} [C_n(u) - C(u)]\bar{F}^*(du, x) \\ &\quad + \int_y^\infty A(u)\bar{L}_n(u)\bar{F}^*(du, x) + R_n(y, x) \\ &\equiv \bar{\xi}_n(y, x) + R_n(y, x), \end{aligned}$$

where

$$\sup_{(y,x) \in T_a} |R_n(y, x)| = O\left(\frac{\log^3 n}{n}\right).$$

Proof. See Appendix A.

To establish the weak convergence properties, it is convenient to write this representation in the following form. Let

$$\begin{aligned} \theta_n(y, x) &= \sqrt{n}[\bar{F}_n(y, x) - \bar{F}(y, x)], \\ \tilde{F}_n(y, x) &= \sqrt{n}[\bar{F}_n^*(y, x) - \bar{F}^*(y, x)], \\ \tilde{C}_n(y) &= \sqrt{n}[C_n(y) - C(y)], \end{aligned}$$

and

$$\tilde{L}_n(y) = \sqrt{n}L_n(y) \quad \tilde{\xi}_n(y, x) = \sqrt{n}\bar{\xi}(y, x).$$

Then we have

$$\begin{aligned} \theta_n(y, x) &= \tilde{F}_n(y, x)A(y) + \int_y^\infty \tilde{F}_n(u, x)A(du) \\ &\quad - \int_y^\infty \frac{\tilde{C}_n(u)}{C(u)} F(du, x) + \int_y^\infty \tilde{L}_n(u)F(du, x) \\ &\quad + \sqrt{n}R_n(y, x) \equiv \tilde{\xi}_n(y, x) + \sqrt{n}R_n(y, x). \quad (13) \end{aligned}$$

Note here that $\tilde{L}_n(y)$ and $\tilde{C}_n(y)$ converge weakly to mean zero Gaussian processes on $D[0, b)$ with covariance structures given by standard results. The weak convergence of $\tilde{F}_n(y, x)$ to a mean zero two-time parameter Gaussian process on the complete separable metric space (D_2, d) defined

Table 1. Average Squared Bias and Variance, 1,000 Replications, $\times 10^{-5}$

<i>Model</i>	<i>n</i>	α		F_n^0	F_n^1	F_n^2	F_n^3
1	50	.20	squared bias	0	2,713	2,712	2,708
			variance	332	4,089	4,092	4,102
		.51	squared bias	0	4	4	4
			variance	338	2,453	2,454	2,456
	.77	squared bias	0	0	0	0	
		variance	343	520	520	521	
	75	.26	squared bias	0	770	768	768
			variance	150	1,333	1,336	1,337
		.53	squared bias	0	5	5	5
			variance	150	1,480	1,482	1,481
	.73	squared bias	0	2	2	2	
		variance	151	357	357	357	
2	50	.20	squared bias	0	2,421	2,420	2,417
			variance	221	3,708	3,712	3,717
		.50	squared bias	0	4	4	4
			variance	224	1,682	1,683	1,684
	.77	squared bias	0	0	0	0	
		variance	225	370	370	370	
	75	.25	squared bias	0	669	668	667
			variance	105	1,265	1,269	1,273
		.53	squared bias	0	1	1	1
			variance	100	883	885	884
	.73	squared bias	0	0	0	0	
		variance	103	296	296	296	

on $[0, 1] \times [0, 1]$ described by Neuhaus (1971) follows from the arguments in that article. We thus have the following result, which follows from the Theorem 1, the strong law of large numbers, and the functional law of iterated logarithm.

Corollary 1. Under the assumptions of Theorem 1, for $(y, x) \in T_a$,

- (a) $\bar{F}_n(y, x) \rightarrow \bar{F}(y, x)$ a.s.;
- (b) $\sup_{(y, x)} |\bar{F}_n(y, x) - \bar{F}(y, x)| = O((\log n/n)^{1/2})$; and
- (c) $\bar{F}_n(y, x)$ converges weakly to a mean zero, two-dimensional time Gaussian process on (D_2, d) , with the covariance structure $\sigma(\tilde{y}, \tilde{x})$ given later.

The proofs for the following lemmas are given in Appendix B.

Lemma 1. Suppose that $\int F(dz)/\bar{G}(z) < \infty$. Then for $a_{F_Y} < u, v < b_{F_Y}$, we have

$$\text{cov}(\tilde{F}_n(u, x), \tilde{F}_n(v, x)) = F^*(u \wedge v, x)[1 - F^*(u \vee v, x)],$$

$$\begin{aligned} \text{cov}(\tilde{F}_n(u, x), \tilde{C}_n(v)) &= (C(v)/F(v)) \\ &\times [\bar{F}(u, x) - \bar{F}(u \vee v, x)] - C(v)\bar{F}^*(u, v), \end{aligned}$$

$$\text{cov}(\tilde{C}_n(u), \tilde{C}_n(v)) = C(u \wedge v) \frac{F(u \wedge v)}{F(u \vee v)} - C(u)C(v),$$

$$\text{cov}(\tilde{L}_n(u), \tilde{L}_n(v)) = \int_{u \vee v} \frac{1}{C(z)F_Y(z)} F_Y(dz) \equiv b(u \vee v),$$

$$\text{cov}(\tilde{F}_n(u, x), \tilde{L}_n(v)) = \bar{F}(u, x)$$

$$- [\bar{F}(u, x) - \bar{F}(u \vee v, x)]/F(u \vee v)F_Y(v),$$

and

$$\text{cov}(\tilde{C}_n(u), \tilde{L}_n(v)) = -F_Y(u) \frac{\bar{F}_Y(u \wedge v)}{F_Y(u \wedge v)}.$$

Lemma 2. Suppose that $\int F(du)/[1 - G(u)] < \infty$. Then the covariance function of $\tilde{\xi}_n(y, x)$ is given as follows, where $\tilde{y} = (y_1, y_2)$, $\tilde{x} = (x_1, x_2)$ and $\sigma(\tilde{y}, \tilde{x}) \equiv \text{cov}(\tilde{\xi}_n(y_1, x_1), \tilde{\xi}_n(y_2, x_2))$:

$$\begin{aligned} \sigma(\tilde{y}, \tilde{x}) &= - \int_{y_1 \vee y_2} A(u) \bar{F}(du, x_1 \vee x_2) \\ &\quad - \int_{y_1} \int_{y_2} \left[\frac{1}{C(u \vee v)} - b(u \vee v) \right] \\ &\quad \times \bar{F}(du, x_1) \bar{F}(dv, x_2) \\ &= - \int_{y_1 \vee y_2} A^2(u) \bar{F}^*(du, x_1 \vee x_2) \\ &\quad - \int_{y_1} \int_{y_2} \left[\frac{1}{C(u \vee v)} - b(u \vee v) \right] \\ &\quad \times A(u)A(v) \bar{F}^*(du, x_1) \bar{F}^*(dv, x_2). \end{aligned}$$

It can be verified that this covariance function reduces to that of the bivariate empirical survival function in the absence of truncation, in which case $C(z) = F_Y(z)$ and $\bar{F}^*(y, x) = \bar{F}(y, x)$. The variance of the limiting process has practical importance, and it reduces to

$$\begin{aligned} \text{var}[\tilde{\xi}_n(y, x)] &= - \int_y A^2(u) \bar{F}^*(du, x) - 2 \int_y \left[\frac{1}{C(u)} - b(u) \right] \\ &\quad \times [\bar{F}(u, x) - \bar{F}(y, x)] A(u) \bar{F}^*(du, x). \quad (14) \end{aligned}$$

A natural estimator for this variance can be obtained as follows. First, note that $A(u)$ can be estimated by $A_n(u) = F_{Y,n}(u)/C_n(u)$ and $\bar{F}_{Y,X,n}^*(\partial Y_i, x) = -I(X_i > x)/n$. Let

$$V_{1,n}(y, x) = n^{-1} \sum_{i: Y_i > y, X_i = x} A_n^2(Y_i)$$

and

$$\begin{aligned} V_{2,n}(y, x) &= n^{-1} \sum_{i: Y_i > y, X_i > x} A_n(Y_i) [\bar{F}_{Y,X,n}(Y_i, x) \\ &\quad - \bar{F}_{Y,X,n}(y, x)] [1/C_n(Y_i) - b_n(Y_i)], \end{aligned}$$

where

$$b_n(u) = \sum_{i=1}^n I(Y_i \leq u)/nC_n^2(Y_i).$$

We then have the following nonparametric estimator for the limiting variance:

$$\sigma_n^2(y, x) = V_{1,n}(y, x) + 2V_{2,n}(y, x).$$

Table 2. Average $\sup |F_n^i(y, x) - F(y, x)|$ Over the Grid Points, 1,000 Replications, $\times 10^{-5}$

Model	<i>n</i>	α	F_n^1	F_n^2	F_n^3
1	30	.21	.49145	.49148	.49223
		.51	.24434	.24459	.24614
		.78	.16196	.16111	.16302
	60	.20	.43521	.43525	.43669
		.50	.18247	.18294	.18416
		.77	.11253	.11253	.11333
	100	.20	.39834	.39832	.39953
		.50	.14625	.14657	.14732
		.77	.08702	.08713	.08749
2	30	.26	.39630	.39845	.40277
		.54	.22839	.22882	.23255
		.74	.15556	.15485	.15654
	60	.25	.34936	.35087	.35669
		.53	.16902	.16966	.17132
		.73	.11015	.11014	.11063
	100	.25	.31530	.31732	.32384
		.53	.13514	.13549	.13702
		.73	.08661	.08677	.08700

As to the consistency of $F^1(y, x)$, we present the following result of uniform strong consistency over a restricted region. Let

$$R_b = \left\{ (y, x): \int^y F(du, x)/F(u) < \infty, y < b < b_F \right\}.$$

Theorem 2.

$$\sup_{(y,x) \in R_b} |F_n^1(y, x) - F(y, x)| \rightarrow 0 \quad \text{a.s.}$$

Proof. See Appendix A.

4. SIMULATION RESULTS AND APPLICATIONS

4.1 Simulations

This section compares the performance of the estimators for the bivariate distribution function with respect to average squared error (squared bias and variance) and average sup norm criteria. The estimators are evaluated at the Cartesian product of 40 grid points at each axis, over which 95% of the probability lies. Two models are considered for (Y, X) and T :

- Model 1. (Y, X) : Independent, where each is $\exp(1)$ and $T \sim \exp(\mu)$.
- Model 2. (Y, X) : Bivariate exponential; that is, $Y = \min(U_1, U_{12})$ and $X = \min(U_2, U_{12})$, where U_1, U_2 , and U_{12} are independent $\exp(1)$ and $T \sim \exp(\mu)$.

The parameter μ of the T variable is adjusted to obtain different values for α . Tables 1 and 2 summarize the results. The reported α 's in these tables are the average proportion of observable pairs in 1,000 replications, and F_n^0 corresponds to the bivariate empirical distribution function, applied to an untruncated sample for reference purposes. It is seen that F_n^3 has slightly better bias when there is heavy

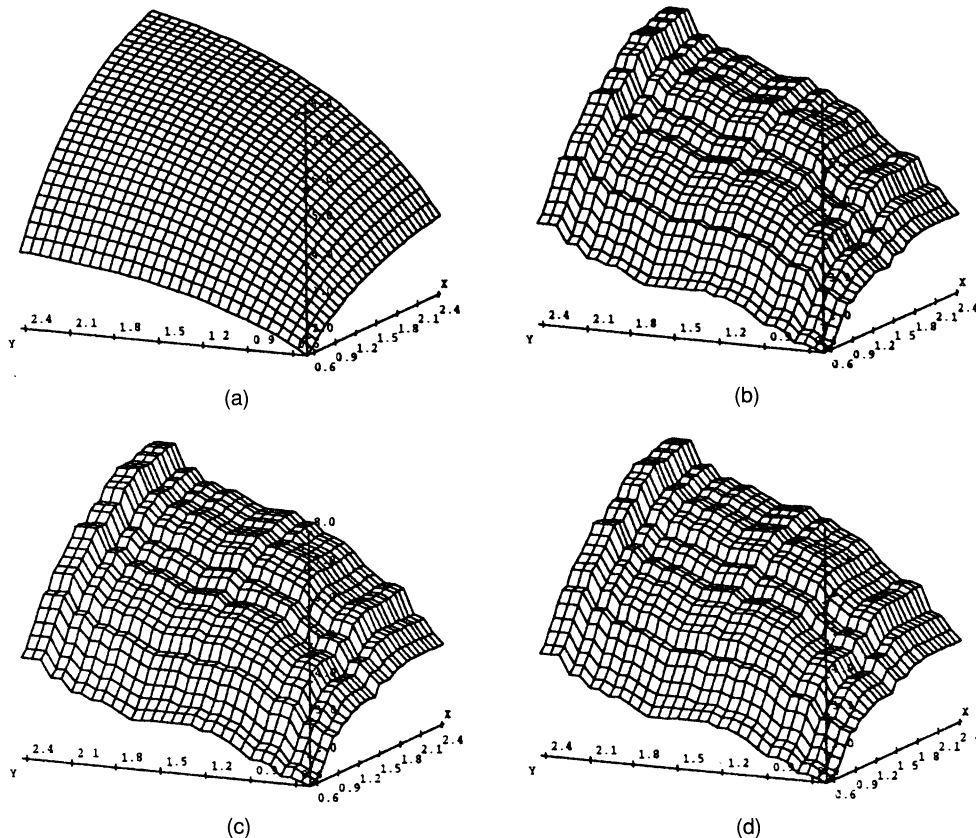


Figure 1. Plot of the True and the Estimated Bivariate Distribution Functions. $\alpha = .75$, $n = 100$; (Y, X) : Independent Exponential. (a) True distribution function $F(y, x)$; (b) estimator $F_n^1(y, x)$; (c) estimator $F_n^2(y, x)$; (d) estimator $F_n^3(y, x)$.

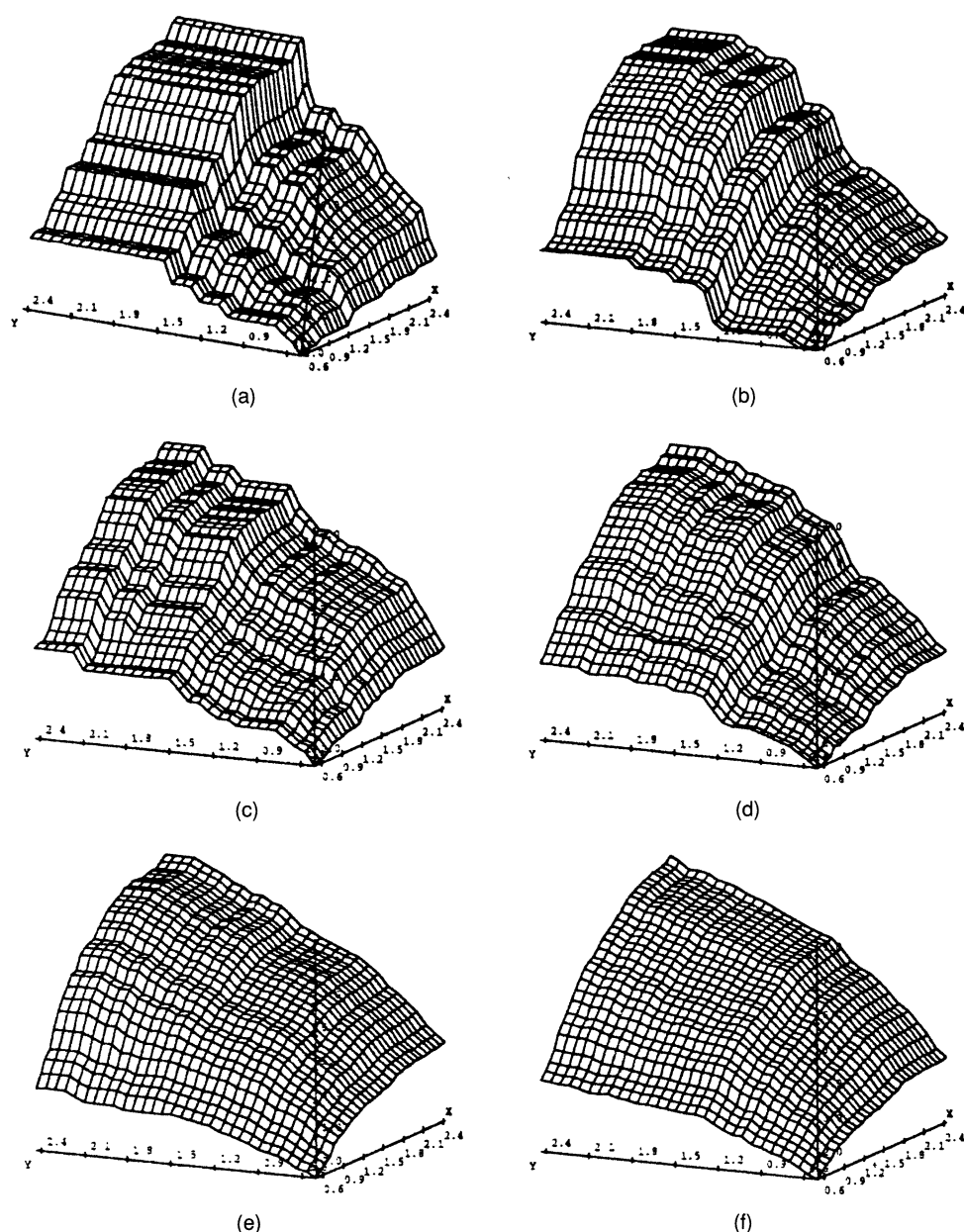


Figure 2. Plot of the Bivariate Distribution Function Estimated by $F_n^1(y, x)$, for Different Sample Sizes and Truncation Proportions: (Y, X), Independent Exponential; X, Incubation Time ($\times 10$); Y, Age ($\times 10$); Z, $F(y, x)$ ($\times .1$). (a) $\mu = 1.3$, $\alpha \simeq .45$, and $n = 50$; (b) $\mu = 1.0$, $\alpha \simeq .75$, and $n = 50$; (c) $\mu = 1.3$, $\alpha \simeq .45$, and $n = 100$; (d) $\mu = 1.0$, $\alpha \simeq .75$, and $n = 100$; (e) $\mu = 1.3$, $\alpha \simeq .45$, and $n = 400$; (f) $\mu = 1.0$, $\alpha \simeq .75$, and $n = 40$.

truncation, whereas F_n^1 has smaller variance. In terms of the MSE, F_n^1 is always dominating, but the difference is practically negligible. Observe, for instance, that the ratio of the MSE of F_n^1 to that of F_n^2 varies between .998 and .9995, and the ratio of the MSE of, to that of F_n^3 varies between .996 and .999. When sup norm is considered, there are cases when F_n^1 is dominated by F_n^2 . From Table 2, it is seen that in two-thirds of the cases corresponding to light truncation ($\alpha \simeq .75$), F_n^2 behaves at least as good. The ratio of the sup norms here changes between .9939 and 1.005 for F_n^2 and from .973 to .998 for F_n^3 . These figures also indicate that the estimators are practically equivalent with respect to MSE or sup norm criteria. This is also supported by Figure 1, which presents the result for a typical simulated sample. As to the relative behavior of F_n^2 and F_n^3 ,

the amount of negative mass and the proportion of points getting them are computed for these estimators. (Details can be obtained from the author.) In almost all cases, F_n^2 is significantly better. For this estimator, both the amount of negative mass and the proportion of the data receiving such mass tend to zero as the truncation proportion decreases, unlike the case for F_n^3 . This result agrees with the findings of Pruitt (1991) for the censored data and discourages the use of F_n^3 .

From the foregoing discussion, one can conclude that F_n^1 is the preferable estimator on the following grounds:

1. It is a proper bivariate distribution function.
2. In most of the simulated cases, its performance is dominating.
3. It has a simpler computational form.

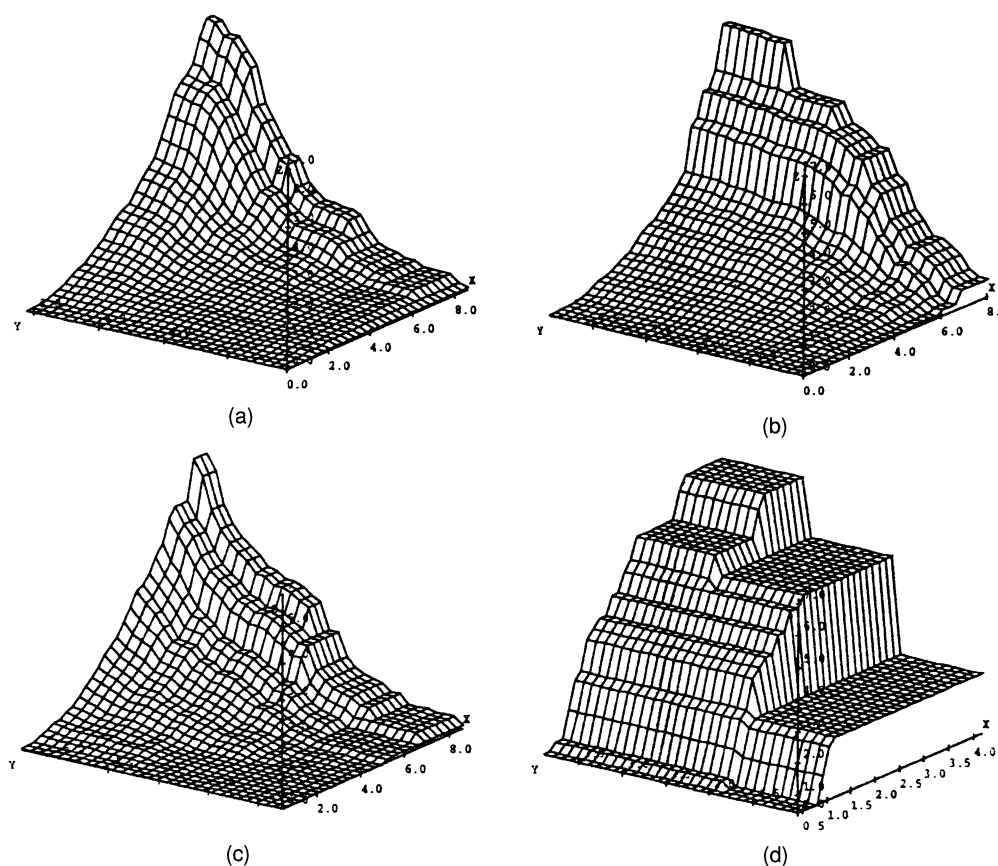


Figure 3. Plot of the Bivariate Distribution Function Estimated by $F_n^1(y, x)$, for the TR-AIDS Data. Combined and separate groups: X , incubation time ($\times 10$); Y , Age ($\times 10$); Z , $F(y, x)$ ($\times .1$). (a) Combined groups, $n = 295$; (b) elderly group, $n = 141$; (c) adult group, $n = 120$; (d) child group, $n = 34$.

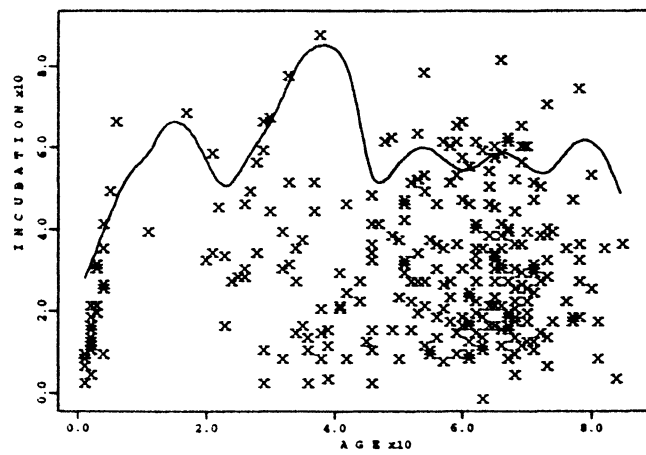
We thus used this estimator for the real data analysis of the next section. Finally, the impact of truncation, displayed in Table 1, shows that the efficiency loss is quite large for $\alpha \simeq .25$ or $.5$. For $\alpha \simeq .75$, however, the truncation effect seems to decrease significantly for all sample sizes. A typical sample from Model 1 (Fig. 1) is presented in Figure 2 with different α values. Here it can be seen that the performance is quite promising even for $n = 50$ and $\alpha \simeq .45$, indicating that less than one half of the population is observable.

4.2 An Application to AIDS Data

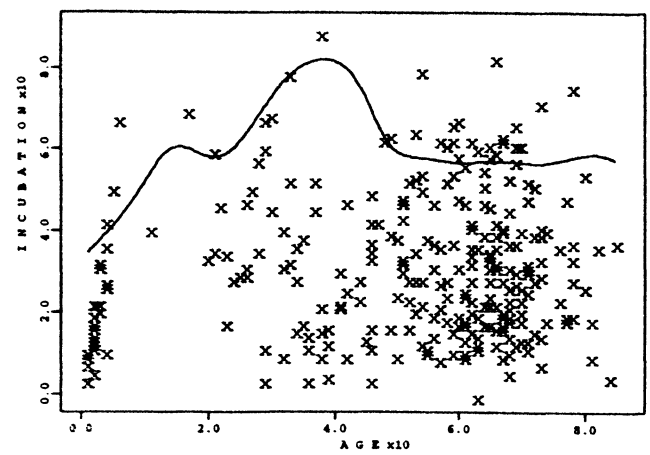
We now present a potential application of the proposed methods to TR-AIDS (transfusion-related AIDS) data given by Wang (1989). (Different versions of this data have been studied in Gross and Huber-Carol 1992, Kalbfleisch and Lawless 1989, Lagakos et al. 1988, Lui et al. 1986, and Medley, Billard, Cox, and Anderson 1987). The purpose here is to illustrate a possible application of the proposed methods, rather than provide a definitive analysis for AIDS data. In this example, Y is the incubation time (in months) as measured from the transfusion to the diagnosis of AIDS, X is the age (in years) of the individuals at the time of the study, and T is the time (in months) from the transfusion to the end of the study (July 1986). In most of the studies mentioned, the analysis is done separately for three age groups: "children," age 1–4; "adults," age 5–59; and "el-

derly," age 60 or older. Note that for this data set, it is not plausible to assume that $a_{F_Y} \leq a_G$, $b_{F_Y} \leq b_G$, and one can only estimate $F_c(y, x) \equiv F(y, x)/F_Y(T_0)$, where T_0 can be taken as $\max(T_1, \dots, T_n)$. Because it is not possible to estimate $F_Y(T_0) \equiv c$ from the available data set, the following results are based on $F_c(y, x)$ with $c = 1$, as done by Kalbfleisch and Lawless (1989). Figure 3 presents estimated $F(y, x)$ for the combined and the separate groups. A quantity of more interest is the regression of Y on X . Gross and Huber-Carol (1992) and Kalbfleisch and Lawless (1991) applied proportional hazard models to this data set, considering the age group as a covariate. Finkelstein and Moore (1992) also applied proportional hazards to an updated version of this data, where they considered the age and the gender as covariates. Consider the model $Y = m(x) + \varepsilon$, where $m(x) = E[Y|X = x]$ is the regression function and ε is the mean zero error term. We provide a rather informal approach to estimating $m(x)$ by nonparametric kernel regression methods based on the estimates of the bivariate distribution function that we have provided. In particular, we use the following:

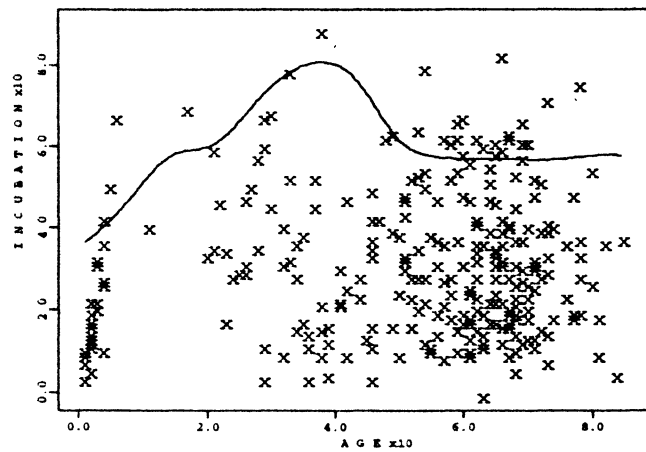
$$\hat{m}(x) = \sum_{i=1}^m Y_i F_n^1(\partial Y_i, \partial X_i) K_b(x - X_i) \div \sum_{i=1}^m F_{X,n}^1(\partial X_i) K_b(x - X_i),$$



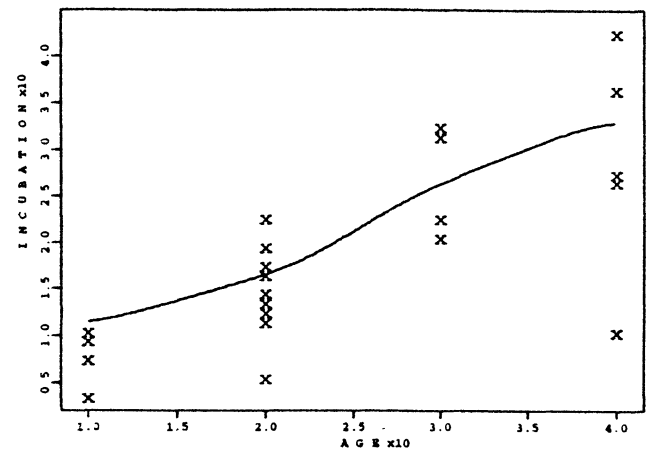
(a)



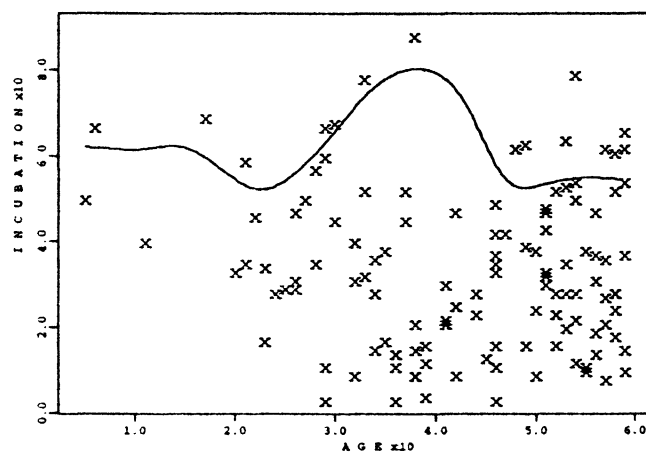
(b)



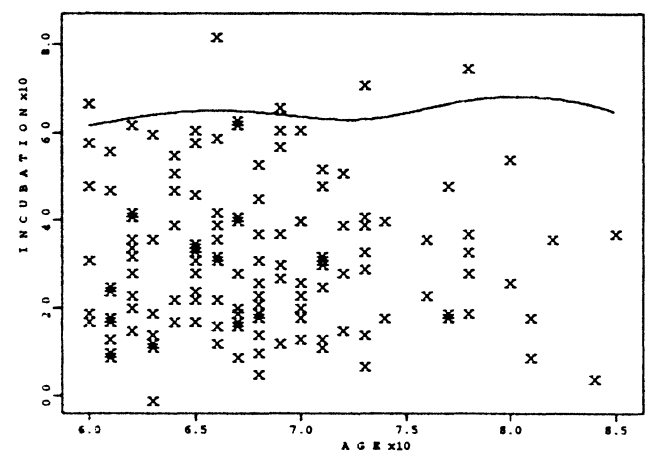
(c)



(d)



(e)



(f)

Figure 4. Plot of the Kernel Estimates of $E(Y|X=x)$ for TR-AIDS Data. Combined and separate groups with different bandwidth (bw) choices. (a) Combined groups, bw = 3; (b) combined groups, bw = 5; (c) combined groups, bw = 6; (d) child group, bw = 4; (e) adult group, bw = 1; (f) elderly group, bw = 4.

where K is the kernel function and $b = bw$ (in the figures) is the smoothing parameter, with $K_b(x - X_i) = K[(x - X_i)/b]/b$. This estimator is analogous to the Nadaraya-Watson regression estimators for the iid observations. For censored data, it is generally accepted that nonparametric regression methods based on Beran's (1980) conditional survival estimators perform better than kernel smoothing. For truncation, however, there are no results on either the estimation of conditional survival function or their applications for regression purposes. Thus such results and their comparison to the foregoing estimator is another open research area. Figure 4 illustrates the results for combined and separate groups, with several choices of the smoothing parameter. The results for the child and elderly groups suggest an increase in the former and an almost constant trend in the latter. For the adult group, there is not much indication of such trends. The early ages in the adult group behave more like the continuation of the child group; the later ones, like the beginning of the elderly group. This observation agrees with the findings of Finkelstein and Moore (1992), who augmented the child group to ages 1–12 and found that their latency is significantly different than that of the adults. The idea is also confirmed by the graph for the combined group with $b = 5$, which suggests an increasing trend until age 40, after which a decline is observed until age 60, followed by a stabilized curve.

5. CONCLUSIONS

This article has discussed nonparametric estimation methods for bivariate data when a component is randomly right truncated. A bivariate reverse hazard vector has been introduced and a nonparametric estimator proposed. An estimator for the bivariate survival function was presented, and its large-sample properties were established via a representation by iid variables. Several approaches were discussed for the bivariate distribution function; these are mostly motivated by their counterparts in the censored model. It turned out that difficulties are involved in obtaining asymptotic results for these estimators. These difficulties are inherited from the behavior of the $F_{Y,n}(y)$ on the left tail. Thus performances of these estimators were evaluated by a simulation study. Among the five nonparametric estimators of the bivariate distribution function discussed, only one— $F_n^1(y, x)$ —enjoys the properties of a bivariate distribution function that is also in a computationally simpler form. This estimator dominated the others in the simulations for most of the cases. The other estimators have the undesirable feature of allowing negative masses for observed data points. However, future research is needed to further elaborate and compare the large-sample behavior of these estimators. It is also worthwhile to mention that the technicalities in estimation with bivariate data when a single component is truncated correspond to those encountered in censoring when both components are censored. This may be attributed to the fact that truncation induces more complicated identifiability problems. It also explains the fact that the proposed methods cannot be easily extended for the data when both components are truncated.

APPENDIX A: PROOF OF THE THEOREMS

Let $\Lambda_Y(dy) \equiv \Lambda_1(\partial y, \infty)$ be the reverse-hazard rate of Y , and let $\Lambda_{Y,n}(y) = \int_y [1/C_n(u)] F_{Y,n}^*(du)$ denote the empirical counterpart of it. Then we can write

$$\begin{aligned} \Lambda_{Y,n}(y) - \Lambda_Y(y) &= \int_y^{b_{FY}} [\Lambda_{Y,n}(dz) - \Lambda_Y(dz)] \\ &= \int_y^{b_{FY}} [1/C(z)] [F_{Y,n}^*(dz) - F_Y^*(dz)] \\ &\quad - \int_y^{b_{FY}} [(C_n(z) - C(z))/C^2(z)] F_Y^*(dz) \\ &\quad + \varepsilon_n(y) \equiv \bar{L}_n(y) - \varepsilon_n(y), \end{aligned}$$

where

$$\begin{aligned} \varepsilon_n(y) &= \int_y^{a_{FY}} \{[(C_n(z) - C(z))/C^2(z)] [F_{Y,n}^*(dz) - F_Y^*(dz)] \\ &\quad + [(C_n(z) - C(z))^2/C^2(z) C_n(z)] F_{Y,n}^*(dz)\}. \end{aligned}$$

This representation and two-term Taylor expansion yield

$$F_{Y,n}(y) - F_Y(y) = -F_Y(y) \bar{L}_n(y) + \varepsilon_n(y),$$

where $O[\varepsilon_n(y)] = \log^3 n/n$ is obtained following Stute (1993a).

The results in the following lemma are utilized.

Lemma A.1

- (Lemma A2 of Chao and Lo 1988) $\sup_{y \leq b} [(C_n(y) - C(y))^2/C(y)] = O(\log n/n)$.
- (Corollary 1.3 of Stute 1991) $\sup_i [C(Y_i)/C_n(Y_i)] = O(\log n)$.
- For $(y, x) \in R_b$, $\int_y^x F_n^*(du, x)/C_n(u) = O(1)$ a.s.

Proof. Note that

$$\int_y^x F_n^*(du, x)/C_n(u) = \sum_{i=1}^n I[Y_i \leq y; X_i \leq x]/n C_n(Y_i).$$

Because $E[1/n C_n(Y_i) | Y_i = y] = (1/n C_n(y)) [1 - (1 - C(y))^n]$ (see, e.g., Woodroffe 1980), by applying conditional expectations we have

$$\begin{aligned} E \left[\int_y^x F_n^*(du, x)/C_n(u) \right] \\ = (\alpha/\bar{G}(y)) \int_y^x F(du, x)/F(u) < \infty. \end{aligned}$$

The strong law of large numbers then yields the result.

Proof of Theorem 1

To simplify the notation, the arguments of $F_Y(u)$, $F_{Y,n}(u)$, $C(u)$, and $C_n(u)$ are suppressed:

$$\begin{aligned} \bar{F}_n(y, x) - \bar{F}(y, x) &= - \int_y \{ [F_{Y,n}/C_n] \bar{F}_n^*(du, x) + [F_Y/C] \bar{F}^*(du, x) \} \\ &= - \int_y \{ [F_Y/C] (\bar{F}_n^*(du, x) - \bar{F}^*(du, x)) \\ &\quad + [F_Y(C_n - C)/C^2] \bar{F}^*(du, x) + \int_y [F_Y \bar{L}_n/C] \bar{F}^*(du, x) \\ &\quad - [(F_{Y,n} - F_Y)/C] (\bar{F}_n^*(du, x) - \bar{F}^*(du, x)) \} \end{aligned}$$

$$\begin{aligned}
& + [F_Y(C_n - C)/C^2](\bar{F}_n^*(du, x) - \bar{F}^*(du, x)) \\
& - [F_Y(C_n - C)^2/C^2 C_n] \bar{F}_n^*(du, x) \\
& + [(F_{Y,n} - F_Y)(C_n - C)/C C_n] \bar{F}_n^*(du, x) - \varepsilon'_n(y) \\
& \equiv I + II + III + R_{1,n} + R_{2,n} + R_{3,n} + R_{4,n} + \varepsilon'_n(y),
\end{aligned}$$

where

$$\varepsilon'_n(y) = \int_y \varepsilon_n(u)/C(u) \bar{F}^*(du, x) = O(\log^3 n/n).$$

The orders of $R_{i,n}$ ($i = 1, \dots, 4$) can now be obtained from Lemma A.1 (a) and (b) and the following facts:

- For $a_G \leq a_{F_Y}, b_G \leq b_{F_Y}, \sup_{0 \leq y < \infty} |C_n(y) - C(y)| = O((\log n/n)^{1/2})$, because C_n is a difference of empirical distribution functions.
- $\sup_{(y,x) \in [0,\infty) \times [0,\infty)} |F_n^*(y, x) - F^*(y, x)| = O((\log n/n)^{1/2})$.
- $\int_y [F/C^2] \bar{F}_n^*(du, x) \leq (\alpha/F(y)) \int_y [\bar{F}(du, x)/\bar{G}(u)] < \infty$, by the assumption of Theorem 1.

As an illustration, consider $R_{3,n}(y, x)$:

$$\begin{aligned}
\sup |R_{3,n}(y, x)| & \leq \sup [(C_n - C)^2/|C|] \int_y [F_Y/C \cdot C_n] F_n^*(du, x) \\
& \leq O(\log n/n) \sup_i [C(Y_i)/C_n(Y_i)] \int_y [F_Y/C^2] F_n^*(du, x) \\
& = O(\log n/n) O(\log n) O(1) = O(\log^2 n/n),
\end{aligned}$$

by Lemma A.1(a) and A.1(b).

Proof of Theorem 2

We have the following decomposition:

$$\begin{aligned}
& F_n^1(y, x) - F(y, x) \\
& = - \int_y \{ [F_{Y,n}/C_n] F_n^*(du, x) - [F_Y/C] F^*(du, x) \} \\
& = - \int_y \{ [F_Y/C] (F_n^*(du, x) - F^*(du, x)) \} \\
& \quad + [F_Y(C_n - C)/C] F_n^*(du, x) \\
& \quad + [(F_{Y,n} - F_Y)/C_n] F_n^*(du, x) \\
& \equiv \xi_n(y, x) + R_{1,n}(y, x) + R_{2,n}(y, x).
\end{aligned}$$

Using integration by parts, we can write

$$\begin{aligned}
\xi_n(y, x) & = [\alpha/\bar{G}(y)] [F_n^*(y, x) - F^*(y, x)] \\
& \quad - \int_y [F_n^*(u, x) - F^*(u, x)]/\bar{G}^2(u) G(du),
\end{aligned}$$

which entails that $\sup \xi(y, x) = o(1)$. By the result of Chen et al. (1995), $\sup_{0 < y < \infty} |F_n(y) - F_y| = o(1)$. By the Glivenko–Cantelli lemma we also have $\sup_{0 < y < \infty} |C_n(y) - C_y| = o(1)$. These facts, together with Lemma A.3, show that for $(y, x) \in R_b$, $\sup R_{n,i}(y, x) = o(1)$.

APPENDIX B: PROOFS OF THE LEMMAS

Proof of Lemma 1

We illustrate one of these; the others are obtained similarly:

$$\text{cov}(\tilde{C}_n(u), \tilde{L}_n(v))$$

$$\begin{aligned}
& = E\{I(v < Y \leq u \leq T)/C(Y)\} \\
& \quad - \int_v^\infty [1/C^2(z)] E[I(Y \leq u \wedge z, T \geq u \vee z)] F^*(dz) \\
& = \int_{u \wedge v}^u [\bar{G}(u)/\bar{G}(z) F(z)] F(dz) - I(u < v) \\
& \quad \times \int_v [F(u)/F^2(z)] F(dz) \int_{u \wedge v}^u [\bar{G}(u)/\bar{G}(z) F(z)] F(dz) \\
& \quad - I(v < u) \int_u [F(u)/F^2(z)] F(dz) \\
& = -F(u) \int_{u \vee v} [1/F^2(z)] F(dz) \\
& = -F(u) [\bar{F}(u \vee v)/F(u \vee v)].
\end{aligned}$$

Proof of Lemma 2

Because $E[\tilde{\xi}_n(y, x)] = 0$, we have

$$\text{cov}[\tilde{\xi}_n(y_1, x_1), \tilde{\xi}_n(y_2, x_2)] = E[\tilde{\xi}_n(y_1, x_1) \tilde{\xi}_n(y_2, x_2)].$$

We can write $E[\tilde{\xi}_n(y_1, x_1) \tilde{\xi}_n(y_2, x_2)] = \sum_{i=1}^{16} T_i(\tilde{y}, \tilde{x})$, where $\tilde{y} = (y_1, y_2)$, $\tilde{x} = (x_1, x_2)$, and

$$T_1(\tilde{y}, \tilde{x}) = \tilde{F}_n(y_1, x_1) \tilde{F}_n(y_2, x_2) A(y_1) A(y_2),$$

$$T_2(\tilde{y}, \tilde{x}) = A(y_1) \int_{y_1} \tilde{F}_n(y_1, x_1) \tilde{F}_n(u, x_2) A(du),$$

$$T_3(\tilde{y}, \tilde{x}) = A(y_1) \int_{y_1} \tilde{F}_n(y_1, x_1) [\tilde{C}_n(v)/C(v)] F(dv, x_2),$$

and

$$T_4(\tilde{y}, \tilde{x}) = A(y_1) \int_{y_1} \tilde{F}_n(y_1, x_1) \tilde{L}_n(v) F(dv, x_2)$$

The terms T_5, T_9 , and T_{13} are similar to the terms T_2, T_3 , and T_4 , except that \tilde{y} and \tilde{x} are interchanged:

$$T_6(\tilde{y}, \tilde{x}) = \int_{y_1} \int_{y_2} \tilde{F}_n(u, x_1) \tilde{F}_n(v, x_2) A(du) A(dv),$$

$$T_7(\tilde{y}, \tilde{x}) = \int_{y_1} \int_{y_2} \tilde{F}_n(u, x_1) \frac{\tilde{C}_n(v)}{C(v)} A(du) F(dv, x_2),$$

and

$$T_8(\tilde{y}, \tilde{x}) = \int_{y_1} \int_{y_2} \tilde{F}_n(u, x_1) \tilde{L}_n(v) A(du) F(dv, x_2).$$

The terms T_{10} and T_{14} are similar to the terms T_7 and T_8 , except that \tilde{y} and \tilde{x} are interchanged:

$$T_{11}(\tilde{y}, \tilde{x}) = \int_{y_1} \int_{y_2} \frac{\tilde{C}_n(u) \tilde{C}_n(v)}{C(u) C(v)} F(du, x_1) F(dv, x_2)$$

and

$$T_{12}(\tilde{y}, \tilde{x}) = \int_{y_1} \int_{y_2} \frac{\tilde{C}_n(u)}{C(u)} \tilde{L}_n(v) F(du, x_1) F(dv, x_2).$$

The term T_{15} is the same as T_{12} , except that \tilde{y} and \tilde{x} are interchanged:

$$T_{16}(\tilde{y}, \tilde{x}) = \int_{y_1} \int_{y_2} \tilde{L}_n(u) \tilde{L}_n(v) F(du, x_1) F(dv, x_2).$$

From the covariance structures given in Lemma 1, the expectations of the foregoing terms are found as

$$E[T_1(\tilde{y}, \tilde{x})] = A(y_1)A(y_2)[\bar{F}^*(y_1 \vee y_2, x_1 \wedge x_2) - \bar{F}^*(y_1, x_1)\bar{F}^*(y_2, x_2)] \quad (\text{B.1})$$

and

$$\begin{aligned} E[T_2(\tilde{y}, \tilde{x})] &= -A(y_1)A(y_2)\bar{F}^*(y_1 \vee y_2, x_1 \vee x_2) \\ &+ A(y_1)\bar{F}(y_1 \vee y_2, x_1 \vee x_2) \\ &+ A(y_1)A(y_2)\bar{F}^*(y_1, x_1)\bar{F}^*(y_2, x_2) \\ &- A(y_1)\bar{F}^*(y_1, x_1)F(y_2, x_2). \end{aligned} \quad (\text{B.2})$$

Observe that

$$E[\tilde{F}_n(u, x)\tilde{C}_n(v)/C(v) + \tilde{F}_n(u, x)\tilde{L}_n(v)] = [\bar{F}(u, x) - \bar{F}^*(u, x)] - [\bar{F}(u, x) - \bar{F}(u \vee v, x)][(F(u \vee v, x) - F(v, x))/F(u \vee v)],$$

and the second term above is always zero. Therefore, we have

$$E[T_3(\tilde{y}, \tilde{x}) + T_4(\tilde{y}, \tilde{x})] = -A(y_1)\bar{F}(y_2, x_2) \times [\bar{F}(y_1, x_1) - \bar{F}^*(y_1, x_1)], \quad (\text{B.3})$$

$$\begin{aligned} E[T_5(\tilde{y}, \tilde{x})] &= -A(y_1)A(y_2)\bar{F}^*(y_1 \vee y_2, x_1 \vee x_2) \\ &+ A(y_2)\bar{F}(y_1 \vee y_2, x_1 \vee x_2) \\ &+ A(y_1)A(y_2)\bar{F}^*(y_2, x_2)\bar{F}^*(y_1, x_1) \\ &- A(y_2)\bar{F}^*(y_2, x_2)\bar{F}(y_1, x_1), \end{aligned} \quad (\text{B.4})$$

$$\begin{aligned} E[T_6(\tilde{y}, \tilde{x})] &= - \int_{y_1 \vee y_2} A(u)\bar{F}(du, x_1 \vee x_2) \\ &+ A(y_1)A(y_2)\bar{F}^*(y_1 \vee y_2, x_1 \vee x_2) \\ &- [A(y_1) + A(y_2)](y_1 \vee y_2, x_1 \vee x_2) \\ &- A(y_1)A(y_2)\bar{F}^*(y_1, x_1)\bar{F}^*(y_2, x_2) \\ &+ A(y_1)\bar{F}^*(y_1, x_1)\bar{F}(y_2, x_2) \\ &+ A(y_2)F(y_1, x_1)F^*(y_2, x_2) \\ &- F(y_1, x_1)(y_2, x_2), \end{aligned} \quad (\text{B.5})$$

$$\begin{aligned} E[T_7(\tilde{y}, \tilde{x}) + T_8(\tilde{y}, \tilde{x})] &= -\bar{F}(y_2, x_2) \int_{y_1} [\bar{F}(u, x_1) - \bar{F}^*(u, x_1)]A(du), \end{aligned} \quad (\text{B.6})$$

$$\begin{aligned} E[T_9(\tilde{y}, \tilde{x}) + T_{13}(\tilde{y}, \tilde{x})] &= A(y_2)\bar{F}(y_1, x_1)[\bar{F}^*(y_2, x_2) - F(y_2, x_2)], \end{aligned} \quad (\text{B.7})$$

and

$$\begin{aligned} E[T_{10}(\tilde{y}, \tilde{x}) + T_{14}(\tilde{y}, \tilde{x})] &= -\bar{F}(y_1, x_1) \int_{y_2} [\bar{F}(u, x_2) - \bar{F}^*(u, x_2)]A(du). \end{aligned} \quad (\text{B.8})$$

Finally, we have

$$\begin{aligned} E[T_{11}(\tilde{y}, \tilde{x}) + T_{12}(\tilde{y}, \tilde{x}) + T_{15}(\tilde{y}, \tilde{x}) + T_{16}(\tilde{y}, \tilde{x})] &= \int_{y_1} \int_{y_2} \left\{ A(u) + A(v) - \left[\frac{1}{C(u \vee v)} - b(u \vee v) \right] \right\} \\ &\times \bar{F}(du, x_1)\bar{F}(dv, x_2) - \bar{F}(y_1, x_1)\bar{F}(y_2, x_2). \end{aligned} \quad (\text{B.9})$$

The result follows from adding up all of the terms (B.1) to (B.9).

[Received July 1994. Revised October 1995.]

REFERENCES

- Akritis, M. G. (1994), "Nearest Neighbor Estimation of a Bivariate Distribution Under Random Censoring," *The Annals of Statistics*, 22, 1299–1327.
- Beran, R. (1980), "Nonparametric Regression With Randomly Censored Survival Data," technical report, University of California, Berkeley.
- Burke, M. D. (1988), "Estimation of a Bivariate Distribution Function Under Random Censorship," *Biometrika*, 75, 379–382.
- Campbell, G. (1981), "Nonparametric Bivariate Estimation With Randomly Censored Data," *Biometrika*, 68, 417–442.
- Campbell, G., and Foldes, A. (1982), "Large-Sample Properties of Nonparametric Bivariate Estimators With Censored Data," in *Proceedings, International Colloquia on Nonparametric Statistical Inference, Budapest, 1980*, Amsterdam: North-Holland, pp. 23–28.
- Chao, M. T., and Lo, S.-H. (1988), "Some Representations of the Nonparametric Maximum Likelihood Estimators With Truncated Data," *The Annals of Statistics*, 16, 661–668.
- Chen, K., Chao, M.-T., and Lo, S.-H. (1995), "On Strong Uniform Consistency of the Lynden-Bell Estimator for Truncated Data," *The Annals of Statistics*, 23, 440–449.
- Dabrowska, D. M. (1988), "Kaplan-Meier Estimate on the Plane," *The Annals of Statistics*, 16, 1475–1489.
- (1995), "Nonparametric Regression With Censored Covariates," *Journal of Multivariate Analysis*, 54, 253–283.
- Finkelstein, D. M., Moore, D. F., and Schoenfeld, D. A. (1993), "A Proportional Hazards Model for Truncated AIDS Data," *Biometrics*, 49, 731–740.
- Gijbels, I., and Wang, J. L. (1993), "Strong Representations of the Survival Function Estimator for Truncated and Censored Data With Applications," *Journal of Multivariate Analysis*, 47, 210–229.
- Gross, S. T., and Huber-Carol, C. (1992), "Regression Models for Truncated Survival Data," *Scandinavian Journal of Statistics*, 19, 193–213.
- Gürler, Ü. (1996), "Bivariate Distribution and Hazard Functions When a Component is Randomly Truncated" submitted to *Journal of Multivariate Analysis*.
- Gürler, Ü., Stute, W., and Wang, J.-L. (1993), "Weak and Strong Quantile Representations for Randomly Truncated Data With Applications," *Statistics and Probability Letters*, 17, 139–148.
- Kalbfleisch, J. D., and Lawless, J. F. (1989), "Inference Based on Retrospective Ascertainment: An Analysis of the Data on Transfusion-Related AIDS," *Journal of the American Statistical Association*, 84, 360–372.
- (1991), "Regression Models for Right-Truncated Data With Application to AIDS Incubation Times and Reporting Lags," *Statistica Sinica*, 1, 19–32.
- Keiding, N., and Gill, R. D. (1990), "Random Truncation Models and Markov Processes," *The Annals of Statistics*, 18, 582–602.
- Lagakos, S. W., Barraj, L. M., and DeGruttola, V. (1988), "Nonparametric Analysis of Truncated Survival Data With Application to AIDS," *Biometrika*, 75, 515–523.
- Lai, T. L., and Ying, Z. (1991), "Estimating a Distribution Function With Truncated and Censored Data," *The Annals of Statistics*, 19, 417–442.
- Lin, D. Y., and Ying, Z. (1993), "A Simple Nonparametric Estimator of the Bivariate Survival Function Under Univariate Censoring," *Biometrika*, 80, 573–581.
- Lo, S. H., and Wang, J. L. (1989), "I.I.D. Representations for the Bivariate Product Limit Estimators and the Bootstrap Versions," *Journal of Multivariate Analysis*, 28, 211–226.
- Lui, K. J., Lawrence, D. N., Morgan, W. M., Peterman, T. A., Haverkos, H. W., and Bergman, D. J. (1986), "A Model-Based Approach for Estimating the Mean Incubation Period of Transfusion Associated Acquired Immunodeficiency Syndrome," in *Proceedings of the National Academy of Science, USA*, 83, 3051–3055.
- Lynden-Bell, D. (1971), "A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars," *Monthly Notices of the Royal Astronomy Society*, 155, 95–118.

- Medley, G. F., Billard, L., Cox, D. R., and Anderson, R. M. (1988), "The Distribution of the Incubation Period of the Acquired Immuno-Deficiency Syndrome (AIDS)," *Proceedings of the Royal Society of London*, Ser. B, 223, 267–277.
- Neuhaus, G. (1971), "On Weak Convergence of Stochastic Processes With Multidimensional Time Parameters," *Annals of Mathematical Statistics*, 42, 1285–1295.
- Pons, O. (1986), "A Test of Independence Between Two Censored Survival Times," *Scandinavian Journal of Statistics*, 13, 173–185.
- 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), "On Negative Mass Assigned by the Bivariate Kaplan–Meier Estimator," *The Annals of Statistics*, 19, 443–453.
- Stute, W. (1993a), "Almost Sure Representations of the Product-Limit Estimator for Truncated Data," *The Annals of Statistics*, 21, 146–156.
- (1993b), "Consistent Estimation Under Random Censorship When Covariables are Present," *Journal of Multivariate Analysis*, 45, 89–103.
- Tsai, W.-Y. (1990), "Testing the Assumption of Independence of Truncation Time and Failure Time," *Biometrika*, 77, 169–177.
- Uzunogullari, Ü., and Wang, J.-L. (1992), "A Comparison of the Hazard Rate Estimators for Left-Truncated and Right-Censored Data," *Biometrika*, 79, 297–310.
- Wang, M. C. (1989), "A Semiparametric Model for Randomly Truncated Data," *Journal of the American Statistical Association*, 84, 742–748.
- Wang, M. C., Jewell, N. P., and Tsai, W. Y. (1986), "Asymptotic Properties of the Product-Limit Estimate Under Random Truncation," *The Annals of Statistics*, 14, 1597–1605.
- Woodroffe, M. (1985), "Estimating a Distribution Function With Truncated Data," *The Annals of Statistics*, 13, 163–177.