

Bivariate Density Estimation with Randomly Truncated Data

Ülkü Gürler

Bilkent University, Turkey

and

Kathryn Prewitt

Arizona State University

Received September 25, 1998

In this study bivariate kernel density estimators are considered when a component is subject to random truncation. In bivariate truncation models one observes the i.i.d. samples from the triplets (T, Y, X) only if $T \leq Y$. In this set-up, Y is said to be left truncated by T and T is right truncated by Y . We consider the estimation of the bivariate density function of (Y, X) via nonparametric kernel methods where Y is the variable of interest and X a covariate. We establish an i.i.d. representation of the bivariate distribution function estimator and show that the remainder term achieves an improved order of $O(n^{-1} \ln n)$, which is desirable for density estimation purposes. Expressions are then provided for the bias and the variance of the estimators. Finally some simulation results are presented. © 2000 Academic Press

AMS 1990 subject classifications: 62G05, 62G20, 62G30.

Key words and phrases: bivariate distribution, truncation/censoring, kernel density estimators.

1 INTRODUCTION

Truncation is one of the common forms of incomplete data encountered in survival studies, as well as in insurance, economy and astronomy. When follow-up studies or life testing situations are considered, left truncation may arise if the time origin of the study is later than that of the individual events whereas a right truncation model may be in effect if the termination of the study takes place before the individuals experience the event of interest. One of the earliest applications of the left truncation model was given by Lynden-Bell (1971), where Y corresponds to the brightness of celestial objects which are only partially observable due to a preventing variable T . Other well known applications of right truncation model

include the transfusion related (TR) AIDS data and data arising from the reporting lags for insurance claims or epidemic surveillance. For the case of TR-AIDS data, let T_o be the time when the observation interval terminates and suppose that the interest is in the incubation period defined as the time from infection with HIV to the onset of AIDS. The infection times are determined retrospectively after an individual is diagnosed with AIDS by tracing back the transfusion dates. This type of data has a sampling bias however, since only those individuals who are infected and are diagnosed with AIDS before T_o can be included in the sample. Letting t and d denote the time of infection and that of the onset of AIDS respectively, only those individuals for whom $d - t \leq T_o - t$ can be observed, resulting in random right truncation. Analysis of TR-AIDS data can be found in Wang (1989), Kalbfleisch and Lawless (1989) and Tsai (1990). The USA Center for Disease Control (CDC) AIDS surveillance data is also right truncated due to reporting lags. In particular, the periodic surveillance reports of CDC are based on the cases that are diagnosed and reported to CDC before a certain deadline, excluding the cases for which the reporting lags are longer, which again results in right truncation. CDC publishes reports after correcting for the bias due to right truncation. Similar data with reporting lags also arise when the claims arriving to an insurance company are considered.

In bivariate left truncation models one observes the random triples (T, Y, X) where the main interest is in the Y variable, while T is a nuisance variable, preventing the complete observation of the main one and X is the covariate. Our interest in this study is the estimation of the bivariate density function of (Y, X) . In the AIDS incubation time example, an important covariate for the incubation time can be the age at infection with HIV. In reporting lags examples for AIDS surveillance and insurance claims, demographical factors and the amount of damage may as important covariates. Hence studying the joint behavior of these related variables would be important for understanding the nature of such phenomena. Bivariate models have received considerable attention recently. For random truncation models Gürlér (1996, 1997) has proposed non-parametric estimators for the bivariate distribution and hazard functions and established large sample properties via strong representations. The univariate left truncation model, first introduced by Lynden-Bell (1971) in the context of an application in astronomy is later studied by others including Woodroffe (1985), Chao and Lo (1988) and Stute (1993) as will be further discussed below. Recently, van der Laan (1996) introduced efficient nonparametric estimators for the bivariate distribution function when both components are truncated. Gürlér and Gijbels (1997) also proposed a nonparametric estimator for the bivariate d.f. $F(y, x)$ for left truncated and right censored (LTRC) data and discuss methods for

estimating the asymptotic variance. The kernel estimator of the bivariate density function studied in this paper is based on the bivariate distribution function estimator proposed by Gürler(1997), where the asymptotic analysis was based on an i.i.d. representation of the empirical distribution function. In order to allow for appropriate bandwidth selection, improved asymptotic results for the moments of the remainder term are obtained as presented in Theorem 1 below. The results of our study are presented for the left truncation model only; however as will become clear, they are applicable to the right truncation as well. This extension will be further discussed in Section 3.

The rest of the paper is organized as follows. In Section 2 preliminary results and a main theorem are presented. In Section 3 the kernel estimator for $f(y, x)$ is introduced and large sample properties are discussed via the results of the main theorem. Asymptotic bias and variance expressions are presented, and remarks concerning the independence assumption, modifications for the right truncation model and extension to the left truncation right censoring model are made. In Section 4 properties of suggested methods are illustrated by simulation results. Finally the proofs of the stated results are provided in the Appendix.

2. PRELIMINARIES AND THE MAIN THEOREM

Let $F(y, x)$ denote the joint distribution function (d.f.) of the random pair (Y, X) with the corresponding density $f(y, x)$ which we are aiming to estimate. In the model explained above where Y is subject to left truncation, one observes the random vector (Y_i, X_i, T_i) , $i = 1, \dots, n$, for which $T_i \leq Y_i$. A common assumption in univariate truncation literature is the independence of T and Y . However, as noted by Tsai (1990) the results for the univariate truncation model are valid under the assumption of only *quasi-independence* (see Remark 1). Analogously, the results of this paper are valid under the weaker assumption of quasi-independence of T from the bivariate random vector (Y, X) which allows the joint distribution of the observed (Y, X) to be written as given in (2.1). Consequently, the independence of T and (Y, X) is required only over the observable subspace where $T \leq Y$. Suppose T has d.f. G and the d.f.'s of the observable random variables are denoted by W with the subscript(s) indicating the particular variable(s) involved, so that W_Y stands for the d.f. of the observed Y . Let F_Y and F_X denote the marginal d.f.'s of Y and X respectively and for any d.f. F , let $a_F = \inf\{t: F(t) > 0\}$ and $b_F = \inf\{t: F(t) = 1\}$ denote the left and right endpoints of its support. The survivor function $1 - F(t)$ will be denoted as $\bar{F}(t)$. Similar to the univariate case (see Woodroffe (1985)), we assume that $F(y, x)$ satisfies the identifiability

condition $a_G \leq a_{W_Y}$ and $b_G \leq b_{W_Y}$. The observed variables have the following bivariate d.f.:

$$\begin{aligned} W_{Y,X}(y, x) &= P(Y \leq y, X \leq x \mid T \leq Y) \\ &= \alpha^{-1} \int_0^x \int_0^y G(u) F(du, dv) \end{aligned} \quad (2.1)$$

where $\alpha = P(T \leq Y)$, and $t \wedge u = \min(t, u)$, $t \vee u = \max(t, u)$. We present the following functions which will be of interest:

$$W_{Y,T}(y, t) = \alpha^{-1} \int_0^y G(t \wedge u) F_Y(du) \quad (2.2)$$

$$W_Y(y) = \alpha^{-1} \int_0^y G(u) F_Y(du)$$

$$W_{Y,X}(dy, dx) = \alpha^{-1} G(y) F(dy, dx). \quad (2.3)$$

$$W_Y(dy) = \alpha^{-1} G(y) F_Y(dy). \quad (2.4)$$

$$C(u) = \alpha^{-1} G(u) \bar{F}_Y(u-) \quad (2.5)$$

$$a(u) = \frac{\bar{F}(u-)}{C(u)} \quad (2.6)$$

The above relations then motivate the following estimator for $F(y, x)$:

$$F_n(y, x) = \frac{1}{n} \sum_i a_n(Y_i) I(Y_i \leq y, X_i \leq x), \quad (2.7)$$

where

$$\bar{F}_{Y,n}(y) = \prod_{i: Y_i \leq y} \left[1 - \frac{s(Y_i)}{nC_n(Y_i)} \right] \quad a_n(u) = \frac{\bar{F}_{Y,n}(u-)}{C_n(u)}$$

and

$$nC_n(u) = \# \{i: T_i \leq u \leq Y_i\},$$

with $s(u) = \{i: Y_i = u\}$ for $u > 0$.

Consistency properties of the Lynden-Bell (1971) estimator $F_{y,n}(y)$ above are investigated in Woodroffe (1985), Wang *et al.* (1986), Chao and Lo (1988) and Stute (1993) among others. An extension to both truncated and censored data is presented in Tsai *et al.* (1987), and asymptotic results are improved in Gijbels and Wang (1993). The bivariate estimator (2.7) is introduced in Gürlér (1997) together with a representation by an i.i.d.

process. Improved asymptotic results for this estimator are obtained as stated in Theorem 1. Apart from being of theoretical interest, this improvement is beneficial for the choice of a bandwidth sequence with a desirable order as further discussed in Section 3.2.

Define

$$L_i(Y) = \frac{I(Y_i \leq y)}{C(Y_i)} - \int_0^y \frac{I(T_i \leq u \leq Y_i)}{C^2(u)} W_Y(du)$$

Then, as in Gürler (1997), we have the following representation of $F_n(y, x)$

$$\begin{aligned} & F_n(y, x) - F(y, x) \\ &= \int_0^y a(u) [W_{Y, X, n}(du, x) - W_{Y, X}(du, x)] \\ &\quad - \int_0^y \frac{a(u)}{C(u)} [C_n(u) - C(u)] - \bar{L}_n(u) a(u) W_{Y, X}(du, x) + R_n(y, x) \\ &\equiv \bar{\xi}_n(y, x) + R_n(y, x) \end{aligned} \quad (2.8)$$

The remainder term $R_n(y, x)$ above can be written as

$$R_n(y, x) = \sum_{i=1}^5 R_{i, n}(y, x)$$

where

$$R_{1, n}(y, x) = \int_0^y \frac{\bar{F}_{Y, n}(u) - \bar{F}_Y(u)}{C(u)} [W_{Z, X, n}(du, x) - W_{Z, X}(du, x)] \quad (2.9)$$

$$\begin{aligned} R_{2, n}(y, x) &= \int_0^y \frac{\bar{F}_{Y, n}(u) [C_n(u) - C(u)]}{C^2(u)} \\ &\quad \times [W_{Z, X, n}(du, x) - W_{Z, X}(du, x)] \end{aligned} \quad (2.10)$$

$$R_{3, n}(y, x) = \int_0^y \frac{\bar{F}_{Y, n}(u) [C_n(u) - C(u)]^2}{C^2(u) C_n(u)} W_{Z, X, n}(du, x) \quad (2.11)$$

$$R_{4, n}(y, x) = \int_0^y \frac{[\bar{F}_{Y, n}(u) - \bar{F}_Y(u)] [C(u) - C_n(u)]}{C(u) C_n(u)} W_{Z, X, n}(du, x) \quad (2.12)$$

$$R_{5, n}(y, x) = \int_0^y \frac{R_n(u)}{C(u)} W_{Z, X, n}(du, x) \quad (2.13)$$

and $R_n(u)$ is the remainder term of the following representation of Chao and Lo (1988):

$$\bar{F}_{Y,n}(u) - \bar{F}_Y(u) = \bar{L}_n(u) + R_n(u)$$

The following Lemmas provide the asymptotic behavior of these remainder terms, which in turn determine that of $R_n(y, x)$. For the discussion below, it is assumed that $a_G < a_{W_Y}$ which implies that for $u \in (0, b)$ $C(u) > \varepsilon$ for some $\varepsilon > 0$. If $a_G = a_{W_Y}$, then an order $\sup_{(y,x) \in T_b} |R_n(y, x)| = O(\ln^3 n/n)$ can be obtained as in Gúrler (1997), however this rate would not be good enough for the bandwidth choice purposes as discussed in the next section. In the results provided below K refers to a generic constant that will take different values in different statements. Let

$$\begin{aligned} H(\lambda, \alpha, n, t, p) = & n \exp(-\lambda t) + (t/50)^{-2} + \exp(-\lambda t^3) \\ & + \exp(-\lambda t) + (n-1) t^{-2p} + \exp(\lambda n t^2) \\ & + t^{-2} n^\alpha \exp(n^{\alpha/2} \lambda/t) \end{aligned}$$

where λ , and α , are nonnegative constants independent of t and $p > 0$ is an integer. Lemma 1 below is due to Gijbels and Wang (1993) for LTRC models which also applies to the left truncation only when the censoring variable is taken to be infinity:

LEMMA 1. *If $a_G < a_W$ then*

- (i) $\sup_{0 \leq y \leq b} |R_n(y)| = O(n^{-1} \ln n)$
- (ii) $E[\sup_{0 \leq y \leq b} |R_n(y)|^\tau] = O(n^{-\tau})$

Lemma 2 below presents one of the main results of this paper, the proof of which involves methods other than those of Gijbels and Wang (1993). In particular, the results of Stute (1982) on the oscillation behavior of the empirical processes are utilized. Let $t^* = t + \varepsilon^{-2}$, then

LEMMA 2. *If $a_G < a_{W_Y}$ then*

$$P[n \sup_{(y,x) \in T_b} |R_{1,n}(y, x)| > t] \leq KH(\lambda, \alpha, n, t, p)$$

Proof. See Appendix.

LEMMA 3. *If $a_G < a_{W_Y}$*

$$P[n \sup_{(y, x) \in T_b} |R_{2, n}(y, x)| > t] \leq KH(\lambda, \alpha, n, t, p)$$

Proof. See Appendix.

LEMMA 4. *If $a_G < a_{W_Y}$ then*

$$(i) \quad P[n \sup_{(y, x) \in T_b} |R_{3, n}(y, x)| > t] \leq K[\exp(-n\alpha) + \exp(-\lambda t)]$$

$$(ii) \quad P[n \sup_{(y, x) \in T_b} |R_{4, n}(y, x)| > t] \leq K[\exp(-n\alpha) + \exp(-\lambda t)]$$

Proof. See Appendix.

We can now state the following theorem, regarding the asymptotic behavior of the remainder term:

THEOREM 1. *Assume $F(y, x)$ is continuous in both components, $b < b_{W_Y}$ and let $T_b = \{(y, x): 0 < y < b; 0 < x < \infty\}$. Then $F_n(y, x)$ admits the following representation:*

$$F_n(y, x) - F(y, x) \equiv \bar{\xi}_n(y, x) + R_n(y, x)$$

If $a_G < a_{W_Y}$, then for any $\gamma > 0$

$$(i) \quad \sup_{(y, x) \in T_b} |R_n(y, x)| = O(\log n/n) \quad a.s.$$

$$(ii) \quad E[\sup_{(y, x) \in T_b} |R_n(y, x)|^\gamma] = O(n^{-\gamma})$$

Proof. First observe that the order of $R_{5, n}(y, x)$ is obtained from an application of Lemma 1 above. The results then follow from Lemma's 2–4.

3. KERNEL ESTIMATION OF THE BIVARIATE DENSITY

The strong i.i.d. representation of $F_n(y, x)$ in Theorem 1 can now be used to obtain a smooth nonparametric estimator of the density $f(y, x)$ by a convolution of a kernel function with $F_n(y, x)$. The convolution-type estimator we propose uses a bandwidth choice which allows for a different bandwidth choice in the directions of y and x , denoted by b_y and b_x . It is

well-known that the choice of this smoothing parameter is crucial for the quality of the resulting estimators and a discussion on bandwidth choice issues will be provided in the following subsection. We assume that the kernels employed belong to a certain class of functions. In particular, $K: S \rightarrow R$ is a kernel function with compact support $S \subset R^2$. We also require the following moment conditions:

$$\iint K(u, v) u^i v^j du dv = \begin{cases} 1 & i + j = 0 \\ 0 & i + j < k \\ \beta(i, j) < \infty & \neq 0 \text{ for some } (i, j): i + j = k \end{cases}$$

The bandwidth sequences $b_y = b_y(n)$ and $b_x = b_x(n)$ depend on the sample size but for brevity in notation we drop the argument n . It is assumed that $b_x, b_y \rightarrow 0$ and $nb_x b_y \rightarrow \infty$ which are the standard assumptions regarding the bandwidth sequences. The bivariate density estimator we suggest is:

$$\begin{aligned} f_n(y, x) &= \iint \frac{1}{b_y b_x} K\left(\frac{y-u}{b_y}, \frac{x-u}{b_x}\right) F_n(du, dv) \\ &= \frac{1}{nb_x b_y} \sum_i \frac{\bar{F}_{Y,n}(Y_i)}{C_n(Y_i)} K\left(\frac{y-Y_i}{b_y}, \frac{x-X_i}{b_x}\right) \end{aligned}$$

Applying integration by parts we write

$$\begin{aligned} f_n(y, x) &= \int \frac{1}{b_x b_y} \left[K\left(\frac{y-u}{b_y}, \frac{x-v}{b_x}\right) F_n(u, dv) \right]_{u=-\infty}^{\infty} \\ &\quad - \int F_n(u, dv) K\left(d\left(\frac{y-u}{b_y}\right), \frac{x-v}{b_x}\right) \Bigg] \\ &= -\frac{1}{b_x b_y} \iint F_n(u, dv) K\left(d\left(\frac{y-u}{b_y}\right), \frac{x-v}{b_x}\right) \end{aligned} \quad (3.14)$$

Another application of integration by parts and a change of variable results in:

$$f_n(y, x) = \iint \frac{1}{b_x b_y} F_n(y - b_y u, x - b_x v) K(du, dv)$$

We therefore have:

THEOREM 2. *Under the assumptions of Theorem 1,*

$$\begin{aligned} f_n(y, x) - f(y, x) &= \frac{1}{b_x b_y} \iint \bar{\xi}_n(y - b_y u, x - b_x v) K(du, dv) \\ &\quad + \frac{1}{b_x b_y} \iint F(y - b_y u, x - b_x v) K(du, dv) \end{aligned} \quad (3.15)$$

$$- f(y, x) + r_n(y, x) \quad (3.16)$$

$$\equiv S_n(y, x) + B_n(y, x) + r_n(y, x) \quad (3.17)$$

where

$$\sup_{(y, x) \in T_b} |r_n(y, x)| = O(\log n / nb_x b_y)$$

Proof. Follows from the representation of Theorem 1 after applying Taylor expansion.

In the above decomposition $B_n(y, x)$ corresponds the bias part and S_n captures the main stochastic component. From these terms we obtain below the asymptotic bias and the variance of our estimators. Let

$$f^{ij}(y, x) = \frac{\partial^{i+j} f(u, v)}{\partial u^i \partial v^j} \Big|_{(u, v) = (y, x)}$$

Let $K(u, v) = K(u) K(v)$ be a product kernel and let $K(\cdot)$ be constructed as in Müller (1988) p.28 by selecting a $\tilde{K}(\cdot)$ of order $(1, 3)$ with the relationship: $\tilde{K} = K^{(1)}$ on $[-1, 1]$. Then

THEOREM 3. *Under the assumptions of Section 1*

$$\begin{aligned} \text{BIAS}(f_n(y, x)) &= (-1)^k \sum_{i+j=k} b_y^i b_x^j \frac{\binom{k}{i}}{k!} f^{ij}(y, x) \beta(i, j) + o((b_x b_y)^k) \\ &\quad + O\left(\frac{1}{nb_x b_y}\right) \\ \text{VAR}(f_n(y, x)) &= \frac{1}{nb_x b_y} \left[A(y)^2 \frac{\partial^2}{\partial y \partial x} W(y, x) \right] \left[\int_{-1}^1 K^2(u) du \right]^2 \\ &\quad + o\left(\frac{1}{nb_x b_y}\right) \\ &= \frac{1}{nb_x b_y} \frac{F_Y(y)}{C(y)} f(y, x) \left[\int_{-1}^1 K^2(u) du \right]^2 + o\left(\frac{1}{nb_x b_y}\right) \end{aligned}$$

Proof. See Appendix.

Prewitt and Gürlér (1999) discuss the bandwidth choice issues and provide the form of the optimal bandwidths that minimize the asymptotic MSE derived from the expressions given above. We refer the interested readers to this paper for further discussion on bandwidth choice.

Remarks. (1) *Independence of truncating variable:* For the univariate truncation model, Y and T are said to be quasi-independent (Tsai, 1990) if the joint distribution of the observed (Y, T) can be written as in (2.2). Tsai (1990) introduces a method to test this assumption from the observed data and Chen, Tsai and Chao (1996) generalize the product-moment correlation coefficient to measure the association between Y and T over the observed region. Our results require that the joint distribution of the observed (Y, X, T) can be written in the form:

$$\begin{aligned} W_{Y, X, T}(y, x, t) &= P(Y \leq y, X \leq x, T \leq t | T \leq Y) \\ &= \alpha^{-1} \int_0^y \int_0^x G(t \wedge u) F(du, dv) \end{aligned} \quad (3.18)$$

which we consider as the quasi-independence in the bivariate set-up and is less restrictive than requiring independence in the entire space. The empirical d.f. of the observed (Y, X, T) is an estimator for the LHS of (3.18) and the RHS can be estimated by incorporating the nonparametric estimators available for G and the bivariate F . Hence it is possible to develop a test statistic for testing the quasi-independence assumption. A complete pursuit of this idea however is beyond the scope of this work and will not be further discussed.

(2) *Right truncation model:* The methods provided in this paper for the left truncation model are directly applicable to the right truncation model, where one observes the pairs (Y_i, T_i, X_i) only if $(Y_i \leq T_i)$, $i = 1, \dots, n$. Then the foregoing results are still valid with the following modifications of $a(u)$, $a_n(u)$ and T_b

$$a(u) = \frac{F_Y(u)}{C(u)} \quad \text{and} \quad a_n(u) = \frac{F_{Y,n}(u)}{C_n(u)}$$

with

$$\begin{aligned} F_{Y,n}(u) &= \prod_{i: Y_i > u} [1 - s(Y_i)/nC(Y_i)] \\ nC_n(u) &= \# \{i: Y_i \leq u \leq T_i\} \end{aligned}$$

and

$$T_b = \{(y, x): 0 < b < y; 0 < x < \infty\}; \quad b > a_{w_Y}$$

(3) *Extension to LTRC*: In the models where Y is subject to LTRC, the observed data is $(Z_i, X_i, T_i, \delta_i)$, $i = 1, \dots, n$, for which $T_i \leq Z_i$; $Z_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$ where C is the censoring variable with d.f. H . It is usually assumed that T and C are independent and they are also independent of both Y and X . The proposed estimator in (2.7) can be extended for this model with slight modifications to incorporate the censoring effect. In particular, let $\bar{W}_Z = (1 - F_Y)(1 - H)$ stand for the d.f. of Z . Then the observed uncensored variables have the following bivariate sub-distribution function:

$$\begin{aligned} W_{Z, X}^1(y, x) &= P(Z \leq y, X \leq x, \delta = 1 \mid T \leq Z) \\ &= \alpha^{-1} \int_0^\infty \int_0^x \int_0^{y \wedge c} G(u) F(du, dv) H(dx) \end{aligned}$$

where $\alpha = P(T \leq Z)$. The estimator for the bivariate d.f. can be modified as follows

$$F_n(y, x) = \frac{1}{n} \sum_i a_n(Z_i) I(Z_i \leq y, X_i \leq x, \delta_i = 1), \quad (3.19)$$

where

$$\bar{F}_{Y, n}(y) = \prod_{i: Z_i \leq y} \left[1 - \frac{s(Y_i)}{nC_n(Y_i)} \right]^{\delta_i}$$

with the previous definitions of $a_n(u)$ and $nC_n(u)$. This estimator can be used to develop a bivariate kernel estimator for the LTRC model. However, the resulting estimator may not achieve the best performance for censored observations, since the information on the X variable is not fully utilized when the corresponding Y is censored. Some of the simulation experiments are extended for this estimator for comparison purposes.

(4) *Alternative estimators*: An efficient estimator for the bivariate distribution function is provided in van der Laan (1996) when both components are truncated. Since our model is a special case of the model introduced there, an alternative density estimator can be obtained by using the empirical d.f. of van der Laan. The simulation results of Gürler and Keleş (1998) indicates that the two estimators of the bivariate distribution function proposed by van der Laan and Gürler are equally efficient when a single component is subject to truncation. We therefore do not expect a superior performance for the density estimator based on this alternative

empirical d.f.. On the other hand, since our proposed estimator yields explicit expressions for the asymptotic bias and the variance, it is more suitable for bandwidth choice considerations.

4. SIMULATIONS

We investigated the behavior of the suggested density estimator under varying parameters. The performance measure we used is the average squared deviation of $f_n(y, x)$ from $f(y, x)$ over a set of grid points on the support of the joint density $f(y, x)$. That is

$$MSE = \frac{1}{M^2} \sum_{j=1}^M \sum_{i=1}^M [f_n(y_i, x_j) - f(y_i, x_j)]^2$$

where $M = 30$ is the number of grid points on each axis and (y_i, x_j) are the points in the Cartesian product. Recall that $\alpha = P(T < Y)$ is the probability that an observation is not truncated. A generalization of the proposed estimator for the LTRC model was introduced in (3.19) when both truncation and censoring are in effect. Let $\beta = P(Y < C)$ be the probability that Y is not censored for the LTRC model. In the simulations various truncation proportions are considered when approximately 25 percent of the observations are censored. The censoring only case corresponds to $\alpha = 1.0$ and the case of complete i.i.d. observations corresponding to $\alpha = 1.0$ $\beta = 1.0$, which are also included. The MSE 's are evaluated for the sample sizes $n = 100, 300$ for fixed bandwidths. For the bivariate kernel function, the product of the following univariate $(0, 2)$ order kernel (see Müller (1988)) is used

$$K(u) = 15/16(1 - 2u^2 + u^4) I(-1 < u < 1)$$

The (Y, X) samples are taken from two bivariate models. In Model 1, (Y, X) are independent exponential variables with parameter one. In Model 2, $F(y, x)$ is taken as a bivariate normal d.f. In both models $G(t)$ and $H(c)$ are taken to be Exponential with parameters τ_T and τ_C respectively, where the values of τ_T and τ_C are adjusted to achieve the desired levels of α and β . For illustration purposes, bandwidths in both directions are taken the same and we used the notation $bw = b_x = b_y$ in the tables.

The simulation results for the independent exponential model are presented in Table 1 and those for the bivariate normal model are displayed in Table 2. From these results we observe once again that the bandwidth choice has a significant impact on the resulting MSE in both models and hence further investigations for the optimal choice of these parameters are desirable. Note also that the estimation for bivariate normal model yields

TABLE 1
MSE Values for Independent Exponential Model,
 $N = 500, (\times 10^3)$.

β	α	$n = 100$ bw = 0.2	$n = 300$ bw = 0.15	$n = 300$ bw = 0.25
0.75	0.26	1.54	1.01	1.11
0.75	0.52	1.32	0.86	1.04
0.75	0.75	1.16	0.74	1.00
0.75	1.00	1.05	0.70	0.97
1.00	0.26	1.67	0.89	0.75
1.00	0.50	1.30	0.75	0.69
1.00	0.76	0.98	0.50	0.59
1.00	1.00	0.80	0.44	0.56

generally higher MSE's than the exponential one, which may be attributed to the fact that the density for the exponential case displays a smoother behavior.

Since truncation and censoring are two different forms of incomplete data, it is interesting to note their relative impacts on estimation and to this end we compare the MSE's of the censored only and truncated only cases. For the exponential model this corresponds to the comparison of the results for $\alpha = 1.0, \beta = 0.75$ vs. $\alpha = 0.76, \beta = 1.0$. For all three (n, bw) choices, we observe that censoring has worse effect than truncation and the improvement as the sample size increases is faster in the truncation only case. Similar observations are made for the bivariate normal model from Table 2. Another point worth noting is that relatively larger bandwidths seems to perform better in heavy truncation. For exponential model, from

TABLE 2
MSE Values for Bivariate Normal Model,

β	α	$n = 100$ bw = 0.2	$n = 100$ bw = 0.3	$n = 300$ bw = 0.25
0.72	0.25	4.58	2.75	1.80
0.72	0.48	2.60	1.82	1.48
0.73	0.72	2.16	1.60	1.33
0.73	1.00	2.08	1.58	1.32
1.00	0.28	6.86	3.15	1.59
1.00	0.50	2.66	1.40	0.77
1.00	0.71	1.83	0.99	0.62
1.00	1.00	1.62	0.85	0.57

Table 1 we note that a larger bandwidth reduces the MSE for $\alpha = 0.26, 0.50$ but increases it for $\alpha = 0.76, 1.0$ increases the MSE from 0.5 to 0.59 for $\alpha = 0.76$. In the bivariate normal model, larger bandwidth results in lower MSE values for all cases, however this reduction is more significant for the heavy truncation case of $\alpha = 0.28$.

In order to illustrate the behavior of the estimators for particular realizations, we present below some examples for $n = 300$. For the exponential case in Fig. 1, (a) is the true density, (b) is the estimated density from an untruncated i.i.d. sample. The remaining figures display the estimated densities with decreasing truncation proportions. The last graph illustrates the impact of choosing a different bandwidth when $\alpha = 0.75$. Similar results are provided for the bivariate normal case in Fig. 2.

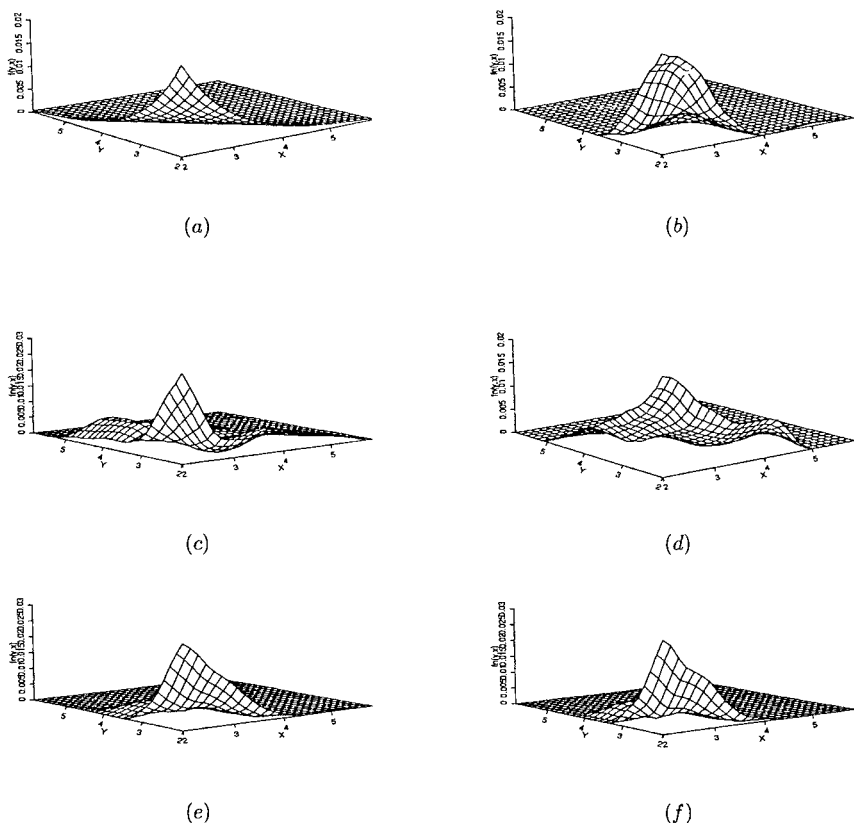


FIG. 1. Independent exponential model, real density and kernel estimators; $n = 300$, $bw = 0.8$ (except in (f)), (a) Real density, (b) No truncation, $\alpha = 1.0$, (c) $\alpha = 0.25$, (d) $\alpha = 0.5$, (e) $\alpha = 0.75$, (f) $\alpha = 0.75$, $bw = 0.7$.

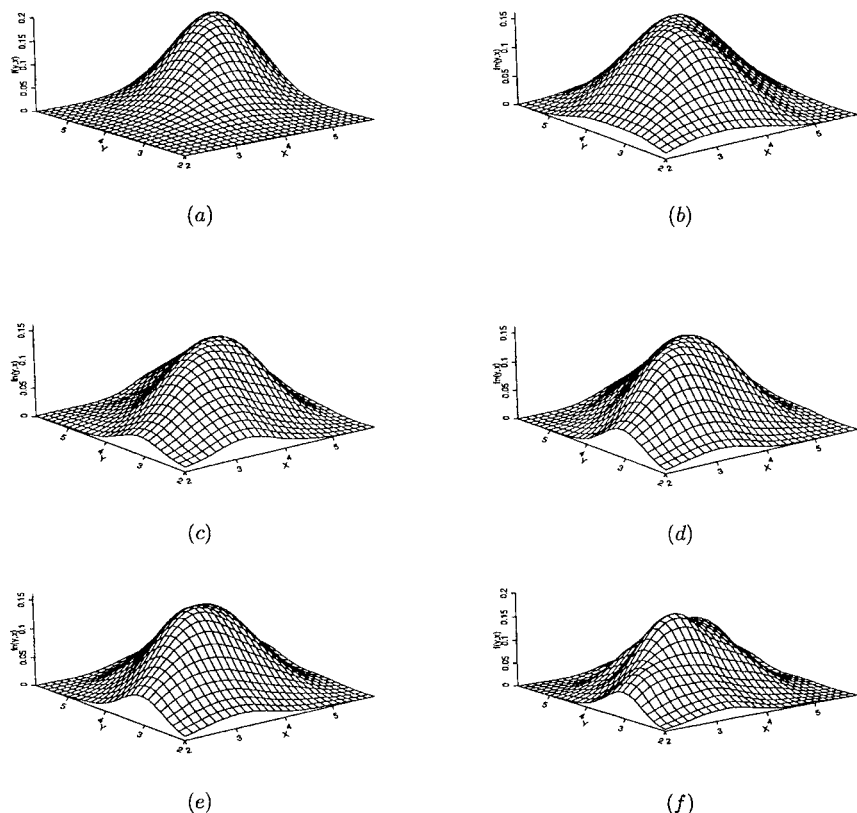


FIG. 2. Bivariate normal model, real density and kernel estimators; $n = 300$, $bw = 1.0$ (except in (f)), (a) Real density, (b) No truncation, $\alpha = 1.0$, (c) $\alpha = 0.25$, (d) $\alpha = 0.5$, (e) $\alpha = 0.75$, (f) $\alpha = 0.75$, $bw = 0.8$.

In two dimensions, since the super-imposition of graphs is not convenient, it is hard to assess visually the impact of the truncation proportion on the deviations of the estimated density from the true one. The observed average (over the grid points) squared errors (ASE) of the estimated density could be indicative for this purpose, which are found to be as below. For Fig. 1, the ASE values were ($\times 10^{-5}$) 5.79, 3.14, 3.09, 2.49 and 2.57 for parts (b) to (e) respectively. For Fig. 2, the corresponding ASE values were ($\times 10^{-3}$) 50.62, 2.73, 2.54, and 2.78. These figures agree with the expectation that the estimation errors increase with the unobserved proportion of the population, although it is hard to visualize it from the figures. We also observe that truncation up to 50 percent is reasonably tolerated, while there is a drastic increase in the ASE for $\alpha = 0.25$.

5. APPENDIX

We define certain terms which will be used throughout.

$$A_n(b) = \bigcap_{i=1}^n \{ |C_n(Y_i) - C(Y_i)| \leq \frac{1}{2} C(Y_i) \text{ or } Y_i \notin (a, b] \} \quad (5.20)$$

$R_n(y)$ is the remainder term of the representation of Gijbels and Wang (1993, p. 214 (1.14)).

Proof of Lemma 2. Recall from (2.9) of Section 2

$$R_{1,n}(y, x) = \int_0^y \frac{\bar{F}_{Y,n}(u) - \bar{F}_Y(u)}{C(u)} [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \quad (5.21)$$

Since

$$\bar{F}_{Y,n}(u) - \bar{F}_Y(u) = \bar{L}_n(u) + R_n(u) \quad (5.22)$$

$$n \sup_{y \leq b} |R_{1n}(y, x)|$$

$$\begin{aligned} &\leq n \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} \bar{L}_n(u) [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \right| \\ &\quad + n \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} R_n(u) [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \right| \end{aligned} \quad (5.23)$$

Then

$$\begin{aligned} &P \left[n \sup_{y \leq b} |R_{1,n}(y, x)| \geq \frac{t^*}{5} \right] \\ &\leq P[A_n(b)^c] + P \left[n \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} \bar{L}_n(u) [W_{Y,X,n}(du, x) \right. \right. \\ &\quad \left. \left. - W_{Y,X}(du, x)] \right| > \frac{t}{10}, A_n(b) \right] \\ &\quad + P \left[n \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} R_n(u) [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \right| \right. \\ &\quad \left. > \frac{t^*}{10}, a_n(b) \right] \\ &= I + II + III \end{aligned} \quad (5.24)$$

$$I \leq 2n \exp[-\alpha n] \quad (5.25)$$

by A.1, p. 223 (Gijbels and Wang (1993)), and

$$\begin{aligned} III &\leq P[\sup_{y \leq b} |R_n(u)| > Kt^*] \\ &\leq K \left[\exp[-\lambda t] + \left(\frac{t}{50}\right)^{-2n} + \exp[-\lambda t^3] \right] \end{aligned} \quad (5.26)$$

follow from Gijbels and Wang (1993, p. 214 (1.15)), with $K, \lambda > 0$ and n large enough.

Next we consider II. Note

$$\begin{aligned} &n \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} \bar{L}_n(u) [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \right| \\ &\leq \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} \left[\int_0^u \frac{C(w) - C_n(w)}{C^2(w)} W_{Y,n}(dw) \right] \right. \\ &\quad \times [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \left. \right| \\ &\quad + n \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} \left[\int_0^u \frac{C_n(w)}{C^2(w)} [W_{Y,n}(dw) - W_Y(dw)] \right] \right. \\ &\quad \times [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \left. \right| \end{aligned} \quad (5.27)$$

Using integration by parts and

$$\left| \int u \, dv \right| \leq 2 \max \left\{ |uv|, \left| \int v \, du \right| \right\} \quad (5.28)$$

with

$$u = \frac{1}{C(u)} \left[\int_0^u \frac{C(w) - C_n(w)}{C^2(w)} W_{Y,n}(dw) \right] \quad (5.29)$$

$$dv = [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \quad (5.30)$$

$$\begin{aligned} du &= -\frac{C'(u)}{C(u)^2} \left[\int_0^u \frac{C(w) - C_n(w)}{C^2(w)} W_{Y,n}(dw) \right] \\ &\quad + \frac{1}{C(u)} \frac{C(u) - C_n(u)}{C^2(u)} W_{Y,n}(du) \end{aligned} \quad (5.31)$$

$$v = [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] \quad (5.32)$$

$$\left| \int v \, du \right| \leq K \sup_{y \leq b} |C(y) - C_n(y)| |W_{Y,X,n}(y, x) - W_{Y,X}(y, x)| \quad (5.33)$$

Applying (5.28) to $B1$,

$$\begin{aligned}
 B1 &\leq 2n \max \left\{ \left| \frac{1}{C(y)} \int_0^y \frac{C(w) - C_n(w)}{C^2(w)} W_{Y,n}(dw) \right| \right. \\
 &\quad \times |W_{Y,X,n}(y, x) - W_{Y,X}(y, x)|, \\
 &\quad \left. K \sup_{y \leq b} |C(y) - C_n(y)| |W_{Y,X,n}(y, x) - W_{Y,X}(y, x)| \right\} \\
 &\leq nK \sup_{y \leq b} [|W_{Y,X,n}(y, x) - W_{Y,X}(y, x)| |C(y) - C_n(y)|]
 \end{aligned}$$

Consider now $B2$ where

$$\begin{aligned}
 B2 &= n \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} \left[\int_0^u \frac{C_n(w)}{C^2(w)} [W_{Y,n}(dw) - W_Y(dw)] \right] \right. \\
 &\quad \times [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \Big| \\
 &\leq n \sup_{y \leq b} \left| \frac{1}{C(u)} \int_0^y \left[\int_0^u \frac{C_n(w) - C(w)}{C^2(w)} [W_{Y,n}(dw) - W_Y(dw)] \right] \right. \\
 &\quad \times [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \Big| \\
 &\quad + n \sup_{y \leq b} \left| \int_0^y \frac{1}{C(u)} \left[\int_0^u \frac{C(w)}{C^2(w)} [W_{Y,n}(dw) - W_Y(dw)] \right] \right. \\
 &\quad \times [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \Big| \tag{5.34}
 \end{aligned}$$

$$= B21 + B22 \tag{5.35}$$

Using the same techniques as in the treatment of $B1$,

$$B21 \leq \sup_{y \leq b} Kn [|W_{Y,X,n}(y, x) - W_{Y,X}(y, x)| |C_n(y) - C(y)|] \tag{5.36}$$

So, $B21$ and $B1$ are bounded above by the same term. We therefore look at one of those probabilities.

$$\begin{aligned}
P\left(B_{21} > \frac{t}{20}\right) &\leq P\left(n^{1/2} \sup_{y \leq b} |[W_{Y,X,n}(y, x) - W_{Y,X}(y, x)]| > \left(\frac{t}{20}\right)^{1/2}\right) \\
&\quad + P\left(n^{1/2} \sup_{y \leq b} K |C_n(u) - C(u)| > \left(\frac{t}{20}\right)^{1/2}\right) \\
&\leq K \exp(-\lambda t)
\end{aligned} \tag{5.37}$$

which follows from the DKW result for some constant K , $\lambda > 0$. We can similarly show

$$P\left(B_1 > \frac{t}{10}\right) \leq K \exp(-\lambda t)$$

Regarding the term B_{22} , we have for some K ,

$$\begin{aligned}
P\left(B_{22} > \frac{t}{20}\right) &\leq P\left(Kn \sup_{y \leq b} |[W_{Y,n}(y) - W_Y(y)]| \right. \\
&\quad \times |W_{Y,X,n}(y, x) - W_{Y,X}(y, x)| > \frac{t}{40}\Big) \\
&\quad + P\left(Kn \sup_{y \leq b} \left| \int_0^y \frac{1}{C^2(u)} [W_{Y,X,n}(y, x) - W_{Y,X}(u, x)] \right. \right. \\
&\quad \times [W_{Y,n}(du) - W_Y(du)] \Big| > \frac{t}{40}\Big) \\
&\leq K \exp[-\lambda t] \\
&\quad + P\left(n \sup_{y \leq b} \left| \int_0^y [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] \right. \right. \\
&\quad \times [W_{Y,n}(du) - W_Y(du)] \Big| > Kt\Big)
\end{aligned} \tag{5.38}$$

We now examine the remaining probability in (5.38). The arguments of Gijbels and Wang (1993) regarding term $B_3(z)$ on pp. 225–226 may be applied (including Lemma 1). By continuity of $W_{Y,X}$ it is possible to partition $[0, b]$ as $0 = t_1 < t_2 < \dots < t_n = b$ such that

$$W_{Y,X}(t_{i+1}, x) - W_{Y,X}(t_i, x) \leq n^{-1}$$

and

$$\begin{aligned}
& P \left(n \sup_{y \leq b} \left| \int_0^y \frac{1}{C^2(u)} [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] \right. \right. \\
& \quad \left. \left. \times [W_{Y,n}(du) - W_Y(du)] \right| > Kt \right) \\
& \leq P \left(\max_{1 \leq i \leq n-1} n \left| \int_0^{t_i} \frac{1}{C^2(u)} [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] \right. \right. \\
& \quad \left. \left. \times [W_{Y,n}(du) - W_Y(du)] \right| > Kt \right) \\
& \quad + P \left(\max_{1 \leq i \leq n-1} \sup_{t_i \leq t \leq t_{i+1}} n \left| \int_{t_i}^t \frac{1}{C^2(u)} [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] \right. \right. \\
& \quad \left. \left. \times [W_{Y,n}(du) - W_Y(du)] \right| > Kt \right) \\
& \leq \sum_{i=1}^{n-1} P \left(n \left| \int_0^{t_i} \frac{1}{C^2(u)} [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] \right. \right. \\
& \quad \left. \left. \times [W_{Y,n}(du) - W_Y(du)] \right| > Kt \right) + P(nV_n > Kt) \\
& \leq (n-1)(Kt)^{-2p} + P(nV_n > Kt) \tag{5.39}
\end{aligned}$$

The first bound of (5.39) above results from Lemma 1 (Gijbels and Wang (1993)) after observing that

$$\begin{aligned}
& \int_0^s \frac{1}{C^2(u)} [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] [W_{Y,n}(du) - W_Y(du)] \\
& = \int_0^s \frac{1}{C^2(u)} \int_0^u [W_{Y,X,n}(dw, x) - W_{Y,X}(dw, x)] \\
& \quad \times [W_{Y,n}(du) - W_Y(du)] \\
& = \int_0^b \int_0^b \frac{1}{C^2(u)} I(0 \leq u \leq s) I(0 \leq w \leq u) \\
& \quad \times [W_{Y,X,n}(dw, x) - W_{Y,X}(dw, x)] [W_{Y,n}(du) - W_Y(du)] \tag{5.40}
\end{aligned}$$

Considering the second term on the RHS of (5.39), we have

$$\begin{aligned}
V_n &\leq \max_{1 \leq i \leq n-1} \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \frac{1}{C^2(u)} [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] W_{Y,n}(du) \right| \\
&\quad + \max_{1 \leq i \leq n-1} \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \frac{1}{C^2(u)} [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] W_Y(du) \right| \\
&\leq \sup_{y \leq b} \left| \frac{1}{C^2(u)} [W_{Y,X,n}(u, x) - W_{Y,X}(u, x)] \right| \\
&\quad \times \left\{ \max_{1 \leq i \leq n-1} [W_{Y,X,n}(t_{i+1}, x) - W_{Y,X,n}(t_i, x)] \right. \\
&\quad \left. + \max_{1 \leq i \leq n-1} [W_{Y,X}(t_{i+1}, x) - W_{Y,X}(t_i, x)] \right\} \\
&\leq \sup_{y \leq b} \left| \frac{1}{C^2(u)} \sup_{y \leq b} |[W_{Y,X,n}(u, x) - W_{Y,X}(u, x)]| \right| \\
&\quad \times \left\{ \max_{1 \leq i \leq n-1} |[W_{Y,X,n}(t_{i+1}, x) - W_{Y,X}(t_{i+1}, x)] \right. \\
&\quad \left. - [W_{Y,X,n}(t_i, x) - W_{Y,X}(t_i, x)]| + \frac{2}{n} \right\} \tag{5.41}
\end{aligned}$$

Then,

$$\begin{aligned}
P(nV_n > Kt) &\leq P \left(\frac{n}{\varepsilon^2} \sup_{y \leq b} |[W_{Y,X,n}(u, x) - W_{Y,X}(u, x)]| \right. \\
&\quad \times \max_{1 \leq i \leq n-1} |[W_{Y,X,n}(t_{i+1}, x) - W_{Y,X}(t_{i+1}, x)] \\
&\quad \left. - [W_{Y,X,n}(t_i, x) - W_{Y,X}(t_i, x)]| \right. \\
&\quad \left. + \frac{n}{\varepsilon^2} \sup_{y \leq b} |[W_{Y,X,n}(u, x) - W_{Y,X}(u, x)]| \frac{2}{n} > Kt \right) \\
&\leq P \left(\frac{n^{1/2}}{\varepsilon} \sup_{y \leq b} |[W_{Y,X,n}(u, x) - W_{Y,X}(u, x)]| > (Kt)^{1/2} \right) \\
&\quad + P \left(\frac{n^{1/2}}{\varepsilon} \max_{1 \leq i \leq n-1} |[W_{Y,X,n}(t_{i+1}, x) - W_{Y,X}(t_{i+1}, x)] \right. \\
&\quad \left. - [W_{Y,X,n}(t_i, x) - W_{Y,X}(t_i, x)]| > (Kt)^{1/2} \right) \\
&\quad + P(n^{1/2} \sup_{y \leq b} |[W_{Y,X,n}(u, x) - W_{Y,X}(u, x)]| > \tilde{K}tn^{1/2})
\end{aligned}$$

$$\begin{aligned}
&\leq K \exp[-\lambda t] \\
&\quad + P(n^{1/2} \max_{1 \leq i \leq n-1} |[W_{Y,X,n}(t_{i+1}, x) - W_{Y,X}(t_{i+1}, x)] \\
&\quad - [W_{Y,X,n}(t_i, x) - W_{Y,X}(t_i, x)]| > K^* t^{1/2}) \\
&\quad + K \exp[-\lambda n t^2]
\end{aligned} \tag{5.42}$$

With respect to (5.42) above, write

$$W_{Y,X}^{-1}(s_{i+1}, x) = t_{i+1} \quad s_{i+1} = W_{Y,X}(t_{i+1}, x)$$

$$\begin{aligned}
&\max_{1 \leq i \leq n-1} |[W_{Y,X,n}(t_{i+1}, x) - W_{Y,X}(t_{i+1}, x)] \\
&\quad - [W_{Y,X,n}(t_i, x) - W_{Y,X}(t_i, x)]| \\
&= \max_{1 \leq i \leq n-1} |[W_{Y,X,n}(W_{Y,X}^{-1}(s_{i+1}, x)) - W_{Y,X}(W_{Y,X}^{-1}(s_{i+1}, x))] \\
&\quad - [W_{Y,X,n}(W_{Y,X}^{-1}(s_i, x)) - W_{Y,X}(W_{Y,X}^{-1}(s_i, x))]| \\
&\leq \sup_{|u-v| \leq 1/n} |[W_{Y,X,n}(W_{Y,X}^{-1}(u, x)) - W_{Y,X}(W_{Y,X}^{-1}(u, x))] \\
&\quad - [W_{Y,X,n}(W_{Y,X}^{-1}(v, x)) - W_{Y,X}(W_{Y,X}^{-1}(v, x))]|
\end{aligned}$$

and recall that the t_i were chosen so that $W_{Y,X}(t_{i+1}, x) - W_{Y,X}(t_i, x) \leq 1/n$. We now apply a result by Stute (1982). For n large enough, we make the following choices which satisfy the conditions for Lemma 2.4 of Stute (1982)

$$s = a^{1/4} \tag{5.43}$$

$$\frac{1}{n} < a < (K^* t^{1/2})^4 \tag{5.44}$$

K^* from (5.42).

$$a < \frac{1}{8} \tag{5.45}$$

$$8 \leq \frac{1}{9} a^{-1/2} \tag{5.46}$$

$$a^{-3/4} \leq (1/4)(1/2) x_{1/2} n^{1/2} \tag{5.47}$$

where

$$x_{1/2} \text{ is } \ni \frac{\ln(1 + x_{1/2})}{x_{1/2}} = \frac{1}{2} \text{ and } C_{1/2} \text{ is a constant depending only on } 1/2 \tag{5.48}$$

We choose

$$a^{-1} = t^{-2} K^{*-4} n^\beta \quad \beta < \frac{2}{3} \quad (5.49)$$

Notice that for n large enough, conditions (5.44) through (5.48) are satisfied (i.e. conditions of Stute's lemma 2.4) and his lemma can be applied with the following result.

$$\begin{aligned} & P(n^{1/2} \max_{1 \leq i \leq n-1} |[W_{Y,X,n}(t_{i+1}, x) - W_{Y,X}(t_{i+1}, x)] \\ & \quad - [W_{Y,X,n}(t_i, x) - W_{Y,X}(t_i, x)]| > K^* t^{1/2}) \\ & \leq P(n^{1/2} \sup_{|u-v| \leq 1/n} |[W_{Y,X,n}(W_{Y,X}^{-1}(u, x)) - W_{Y,X}(W_{Y,X}^{-1}(u, x))] \\ & \quad - [W_{Y,X,n}(W_{Y,X}^{-1}(v, x)) - W_{Y,X}(W_{Y,X}^{-1}(v, x))]| > K^* t^{1/2}) \\ & < C_{1/2} a^{-1} \exp[-a^{-1/2}/64] = K t^{-2} n^\beta \exp[-t n^{\beta/2} \lambda] \end{aligned} \quad (5.50)$$

$$\begin{aligned} II & \leq P \left[B1 > \frac{t^*}{20}, A_n(b) \right] + P \left[B2 > \frac{t^*}{20}, A_n(b) \right] \\ & \leq P \left[B1 > \frac{t^*}{20}, A_n(b) \right] + P \left[B21 > \frac{t^*}{40}, A_n(b) \right] + P \left[B22 > \frac{t^*}{40}, A_n(b) \right] \\ & \leq K \exp[-\lambda t] + K \exp[-\lambda t] + \{ K \exp[-\lambda t] + (n-1)(Kt)^{-2p} \\ & \quad + K \exp[-\lambda t] + K t^{-2} n^\alpha \exp[-t n^{\beta/2} \lambda] + K \exp[-\lambda n t^2] \} \end{aligned} \quad (5.51)$$

Proof of Lemma 3. Note that we can write

$$\begin{aligned} |R_{2n}(y, x)| &= \int_0^y \frac{\bar{L}_n(u)[C_n(u) - C(u)]}{C^2(u)} [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \\ & \quad + \int_0^y \frac{R_n(u)[C_n(u) - C(u)]}{C^2(u)} [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \\ & \quad + \int_0^y \frac{\bar{F}_Y(u)[C_n(u) - C(u)]}{C^2(u)} [W_{Y,X,n}(du, x) - W_{Y,X}(du, x)] \end{aligned} \quad (5.52)$$

Recognizing

$$\begin{aligned}
& n \sup_{(y, x) \in T_b} |R_{2n}(y, x)| \\
& \leq Kn \sup_{(y, x) \in T_b} |\bar{L}_n(y)| |C_n(y) - C(y)| + \sup_{(y, x) \in T_b} Kn |R_n(y)| \\
& \quad + \sup_{(y, x) \in T_b} n \left| \int_0^y \frac{\bar{F}_Y(u) [C_n(u) - C(u)]}{C^2(u)} \right. \\
& \quad \left. \times [W_{Y, X, n}(du, x) - W_{Y, X}(du, x)] \right|
\end{aligned} \tag{5.53}$$

and following Lemma 2,

$$\begin{aligned}
& P \left(\sup_{(y, x) \in T_b} n |R_{2n}(y, x)| > \frac{t^*}{5} \right) \\
& \leq P \left(\sup_{y \leq b} n^{1/2} |\bar{L}_n(y)| n^{1/2} |C_n(y) - C(y)| > \frac{t}{15K} \right) \\
& \quad + P \left(\sup_{y \leq b} n |R_n(y)| > \frac{t^*}{15K} \right) \\
& \quad + P \left(\sup_{y \leq b} n \left| \int_0^y \frac{\bar{F}_Y(u) [C_n(u) - C(u)]}{C^2(u)} \right. \right. \\
& \quad \left. \left. \times [W_{Y, X, n}(du, x) - W_{Y, X}(du, x)] \right| > \frac{t}{15} \right) \\
& \leq K \exp[-\lambda t] + K \left[\exp[-\lambda t + \left(\frac{t}{50}\right)^{-2n}] + \exp[\lambda t^3] \right] \\
& \quad + (n-1)(Kt)^{-2p} + K \exp(-\lambda t) + Kt^{-2} n^\alpha \exp[-\tilde{K}t^{-1} n^{\alpha/2}] \\
& \quad + K \exp[-\lambda^2 t^2 n]
\end{aligned} \tag{5.54}$$

and the result obtains.

Proof of Lemma 4.

$$\begin{aligned}
& P \left[n \sup_{(y, x) \in T_b} |R_{3n}(y, x)| > \frac{t^*}{5} \right] \\
& \leq P(A_n(b)^c) + P \left[n \sup_{(y, x) \in T_b} |C_n(y) - C(y)|^2 > \tilde{K}t, A_n(b) \right]
\end{aligned}$$

$$\begin{aligned}
& 2n \exp^{-\alpha n} + 2P[n^{1/2} \sup_{(y, x) \in T_b} |C_n(y) - C(y)| > (\tilde{K}t)^{1/2}] \\
& \leq 2n \exp^{-\alpha n} + K \exp[-\lambda t]
\end{aligned} \tag{5.56}$$

Since the integral below is integrated with respect to the empirical measure $W_{Y, X, n}(du, x)$, $1/C_n(u)$ is bounded on $A_n(b)$ and

$$\begin{aligned}
& |n \sup_{(y, x) \in T_b} R_{4, n}(y, x)| \\
& = n \left| \int_0^y \frac{[\bar{F}_{Y, n}(u) - \bar{F}_Y(u)][C_n(u) - C(u)]}{C(u) C_n(u)} W_{Y, X, n}(du, x) \right| \\
& \leq nK \sup_{(y, x) \in T_b} |[\bar{F}_{Y, n}(u) - \bar{F}_Y(u)][C_n(u) - C(u)]|
\end{aligned} \tag{5.57}$$

and the result now follows as in Lemma 3.

Proof of Theorem 1(ii). The result follows from applying

$$E(|W|^\beta) = \beta \int_0^\infty u^{\beta-1} P(|X| \geq u) du \tag{5.58}$$

We alter the probability bound for the remainder term in Theorem 1(i) by replacing the bound for one of the terms with another which is obtained by a different application of Lemma 2.4 Stute (1982). The probability which must be bounded is:

$$\begin{aligned}
& P(n^{1/2} \sup_{|u-v| \leq 1/n} |[W_{Y, X, n}(W_{Y, X}^{-1}(u, x)) - W_{Y, X}(W_{Y, X}^{-1}(u, x))] \\
& - [W_{Y, X, n}(W_{Y, X}^{-1}(v, x)) - W_{Y, X}(W_{Y, X}^{-1}(v, x))]| > K^* t^{1/2})
\end{aligned} \tag{5.59}$$

Consider two cases (1) $K^* > 3$ and (2) $K^* < 3$ for large enough n . In the first case, for $t \geq n$, the probability of (5.59) is zero. For $\gamma < t < n$ where $\gamma = (1/K^{*4})^{1/3}$ and $a = 1/t$

$$(5.59) \leq C_{1/2} t, \exp[-t^{1/2}/64] \tag{5.60}$$

In the second case with $K^* < 3$, (5.59) is zero for $t \geq 4(1/K^{*2})n$. For $\gamma < t \leq 4(1/K^{*2})n$ with $a = 4/tK^{*2}$, a similar bound is obtained as in (5.60) with the appropriate resulting expectation order.

Proof of Theorem 3. Bias

$$Ef_n(y, x) - f(y, x) = ES_n(y, x) - f(y, x) + EB_n(y, x) + Er_n(y, x)$$

$$EB_n(y, x) + Er_n(y, x) = EB_n(y, x) + O\left(\frac{1}{nb_x b_y}\right) \quad (5.61)$$

$$\begin{aligned} EB_n(y, x) &= \frac{1}{b_x b_y} \iint F(y - b_y u, x - b_x v) K(du, dv) \\ &= \frac{1}{b_x b_y} \iint K\left(\frac{y-u}{b_y}, \frac{x-v}{b_x}\right) F(du, dv) \end{aligned}$$

$$\begin{aligned} &\frac{1}{b_x b_y} \iint K\left(\frac{y-u}{b_y}, \frac{x-v}{b_x}\right) f(u, v) du dv \\ &= \iint K(u, v) f(y - ub_y, x - vb_x) dy dx \\ &= \iint K(u, v) \sum_i \sum_{+j=l} \sum_{l=0}^k \frac{\binom{l}{i}}{l!} f^{ij}(y, x) (-ub_y)^i (-vb_x)^j du dv \\ &\quad + O((b_x \vee b_y)^{k+1}) \\ &= (-1)^k \sum_i \sum_{+j=k} \frac{\binom{k}{i}}{k!} f^{ij}(y, x) \beta(i, j) + O((b_x \vee b_y)^{k+1}) \quad (5.62) \end{aligned}$$

Variance

$$\begin{aligned} V(f_n(y, x)) &= V(S_n(y, x) + r_n(y, x)) \\ &= V(s_n(y, x)) + V(r_n(y, x)) + 2 \operatorname{Cov}(S_n(y, x), r_n(y, x)) \quad (5.63) \end{aligned}$$

$$\begin{aligned} V(r_n(y, x)) &= E[(r_n(y, x))]^2 - (E[r_n(y, x)])^2 \\ &= O\left(\frac{1}{(bb_y b_x)^2}\right) \end{aligned}$$

$$|\operatorname{Cov}(S_n(y, x), r_n(y, x))| = O\left(\frac{1}{(nb_y b_x)^2}\right)$$

So, by the results of Prewitt and Gürler (1999), the variance of the density estimator reduces to

$$(5.63) = \left(\frac{1}{nb_x b_y} \right) \left[A(y)^2 \frac{\partial^2}{\partial y \partial x} W_{Y, X}(y, x) \right] \left[\int_{-1}^1 K^2(x) ds \right]^2 \\ + o \left(\frac{1}{nb_x b_y} \right) + O \left(\frac{1}{(nb_y b_x)^2} \right)$$

REFERENCES

1. C. H. Chen, W. Y. Tsai, and W. H. Chao, The product-moment correlation coefficient and liner regression for truncated data, *J. Amer. Statist. Assoc.* **91** (1996), 1181–1186.
2. M. Chao and S. H. Lo, Some representations of the nonparametric maximum likelihood estimators with truncated data, *Ann. Statist.* **16** (1988), 661–668.
3. I. Gijbels and J. L. Wang, Strong representations of the S survival function estimator for truncated and censored data with applications, *J. Multivariate Anal.* **47** (1993), 210–229.
4. Ü. Gürler, Bivariate distribution and hazard functions when a component is randomly truncated, *J. Multivariate Anal.* **60** (1997), 20–47.
5. Ü. Gürler, Bivariate estimation with right truncated data, *J. Amer. Statist. Assoc.* **91** (1996), 1152–1165.
6. Ü. Gürler and I. Gijbels, A bivariate distribution function estimator and its variance under left truncation and right censoring, Discussion Paper 9702, Institute de Statistique, Université Catholique de Louvain, Belgium, 1997.
7. Ü. Gürler and S. Keleş, Comparison of Independence Tests for Truncated Data, Tech. Rep. IEOR-9816, Industrial Engineering Department, Bilkent University, Ankara, Turkey, 1998.
8. J. D. Kalbfleisch and J. F. Lawless, Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS, *J. Amer. Statist. Assoc.* **84** (1989), 360–372.
9. D. Lynden-Bell, A method of allowing for known observational selection in small samples applied to 3CR quasars, *Monthly Notices Roy. Astronom. Soc.* **155** (1971), 95–118.
10. H.-G. Müller, “Nonparametric Regression Analysis of Longitudinal,” Lecture Notes in Statist., Vol. 46, Springer-Verlag, Berlin, 1988.
11. H.-G. Müller and K. Prewitt, Multivariate bandwidth processes and adaptive surface smoothing, *J. Multivariate Anal.* **47** (1993), 1–21.
12. K. Prewitt and Ü. Gürler, Variance of the bivariate density estimator for left truncated and right censored data, *Statist. Probab. Let.*, to appear.
13. W. Stute, The oscillation behavior of empirical processes, *Ann. Probab.* **10** (1982), 86–107.
14. W. Stute, Almost sure representations of the product-limit estimator for truncated data, *Ann. Statist.* **21** (1993), 146–156.
15. W. Y. Tsai, Testing the assumption of independence of truncation time and failure time, *Biometrika* **77** (1990), 169–177.
16. W. Y. Tsai, N. P. Jewell, and M. C. Wang, A note on the product limit estimator under right censoring and left truncation, *Biometrika* **74** (1987), 883–886.
17. M. J. van der Laan, Nonparametric estimation of the bivariate survival function with truncated data, *J. Multivariate Anal.* **58** (1996), 107–131.

18. M. C. Wang, A semiparametric model for randomly truncated data, *J. Amer. Statist. Assoc.* **84** (1989), 742–748.
19. M. C. Wang, N. P. Jewell, and W. Y. Tsai, Asymptotic properties of the product limit estimate under random truncation, *Ann. Statist.* **14** (1986), 1597–1605.
20. M. Woodroffe, Estimating a distribution function with truncated data, *Ann. Statist.* **13** (1985), 163–177.