**STATISTICS &
PROBABILITY
LETTERS**

Variance of the bivariate density estimator
for left truncated right censored data

Kathryn Prewitt[a,*], Ulkü Gürler[b]

[a] *Department of Mathematics, Arizona State University, P.O. Box 871804, Tempe, AZ 85287-1804, USA*
[b] *Bilkent University, Turkey*

Abstract

In this study the variance of the bivariate kernel density estimators for the left truncated and right censored (LTRC) observations are considered. In LTRC models, the complete observation of the variable Y is prevented by the truncating variable T and the censoring variable C. Consequently, one observes the i.i.d. samples from the triplets (T, Z, δ) only if $T \leqslant Z$, $Z = \min(Y, C)$ and δ is one if $Z = Y$ and zero otherwise. Gürler and Prewitt (1997, submitted for publication) consider the estimation of the bivariate density function via nonparametric kernel methods and establish an i.i.d. representation of their estimators. Asymptotic variance of the i.i.d. part of their representation is developed in this paper. Application of the results are also discussed for the data-driven and the least-squares cross validation bandwidth choice procedures. ⓒ 1999 Published by Elsevier Science B.V. All rights reserved

1. Introduction

Our main purpose in this paper is to establish the variance of the bivariate density estimate when one component is left truncated and right censored (LTRC). LTRC data naturally occurs if the time origin of the study is later than the time origin of the individual events, which leads to the truncation of certain items/individuals. Moreover, the truncated sample can also become subject to right censoring during the course of study, due to the drop outs or failure to follow-up which is common to cohort follow-up studies. Studies analyzing univariate data arising from such models include Tsai et al. (1987), Uzunoğullarí and Wang (1992) and Gijbels and Wang (1993), among others. Recently, Gürler and Gijbels (1996) proposed a nonparametric bivariate distribution function estimator when a component is subject to LTRC. Gürler and Prewitt (1997) considered a similar data structure and introduced the bivariate kernel density estimates. Due to the truncation

---

* Corresponding author.

and censoring effects, the resulting estimates are in the form of the sums of dependent random variables which complicates the large sample analysis. Gürler and Prewitt (1997) present a strong representation of their estimator as a sum of mean zero i.i.d. variables with an asymptotically negligible remainder term. Although the variance of the i.i.d. term of this representation is very complicated, knowing the variance is important for several purposes such as the analysis of the small and large sample behavior, construction of confidence intervals, hypothesis testing, comparison of the alternative estimators, etc. Moreover, evaluating and estimating the variance of estimators gain particular importance in the context of kernel estimation. As is well known, kernel estimators require a bandwidth choice which is a crucial parameter of such methods, and often optimal bandwidth selection criteria depend on the leading terms of the variance. For example, an optimal bandwidth might be the one which minimizes the leading terms of the MSE with a data-driven bandwidth choice, defined by replacing the theoretical terms with their associated estimates. The techniques involved in evaluating the variance of the kernel density estimate make use of several properties of the kernels which are not straightforward, particularly in the bivariate case. As mentioned before, the expression for the variance of the considered estimator displays a quite complex structure due to both the higher dimension considered and the involvement of nuisance functionals produced by truncation/censoring. We provide in this paper the key aspects of decomposing the variance expression to a leading and the negligible terms and finding their corresponding orders.

The rest of the paper is organized as follows: In Section 2, we present the bivariate density estimator, the essential definitions and the preliminary results. In Section 3, we provide the main result regarding the variance and the bias of the density estimator and provide a brief discussion on the bandwidth choice. Finally in the appendix we present the proofs of the presented results.

2. Preliminaries

Let $F(y,x)$ denote the joint distribution function (d.f.) of the random pair (Y,X) with the corresponding density $f(y,x)$. In the model where Y is subject to LTRC, the observed data has the following features: Let T be the truncating variable with d.f. G, and let C be the censoring variable with d.f. H; then we observe $(Z_i, X_i, T_i, \delta_i)$, $i = 1, \ldots, n$, for which $T_i \leqslant Z_i$, where $Z_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \leqslant C_i)$. Here T and C are assumed to be independent, and are also independent of both Y and X. The d.f.'s of the observed random variables will be denoted by W with the subscript(s) indicating the particular variable(s) involved. Hence $W_Z = (1 - F_Y)(1 - H)$ will stand for the d.f. of Z. Let F_Y and F_X denote the marginal d.f.'s of Y and X, respectively. Let $W_{Z,X}^1(y,x)$ stand for the sub-distribution function of the observed uncensored variables and let $\alpha = P(T \leqslant Z)$, $t \wedge u = \min(t,u)$ and $t \vee u = \max(t,u)$. Also, for any d.f. F, let $\bar{F} = 1 - F$ and $F(z-) = \lim_{h \to 0} F(z - h)$. Then

$$W_{Z,X}^1(y,x) = P(Z \leqslant y, X \leqslant x, \delta = 1 | T \leqslant Z)$$

$$= \alpha^{-1} \int_0^\infty \int_0^x \int_0^{y \wedge c} G(u) F(\mathrm{d}u, \mathrm{d}v) H(\mathrm{d}c).$$

From which, one can also derive

$$W_Z^1(z) = \alpha^{-1} \int_0^z G(u) \bar{H}(u-) F_Y(\mathrm{d}u),$$

$$W_{Z,X}^1(\mathrm{d}z, \mathrm{d}x) = \alpha^{-1} G(z)[1 - H(z-)] F(\mathrm{d}z, \mathrm{d}x),$$

$$W_Z^1(\mathrm{d}z) = \alpha^{-1} G(z)[1 - H(z-)] F_Y(\mathrm{d}z).$$

We also define $C(u) = \alpha^{-1}G(u)\bar{F}_Y(u-)\bar{H}(u-)$ and $A(u) = \bar{F}(u-)/C(u)$ and assume that $\inf_u C(u) \geqslant \varepsilon$ for some $\varepsilon > 0$. Exploiting the foregoing relations, Gürler and Gijbels (1996) suggest the following estimator for the joint d.f. of (Y, X), where $s(u) = \#\{i : Z_i = u\}$, for $u > 0$:

$$F_n(y, x) = \frac{1}{n}\sum_i A_n(Z_i)I(Z_i \leqslant y, X_i \leqslant x, \delta_i = 1), \tag{2.1}$$

where

$$A_n(u) = \frac{\bar{F}_{Y,n}(u-)}{C_n(u)}, \quad \bar{F}_{Y,n}(y) = \prod_{i\,:\,Z_i \leqslant y}\left[1 - \frac{s(Y_i)}{nC_n(Y_i)}\right]^{\delta_i} \tag{2.2}$$

and

$$nC_n(u) = \#\{i : T_i \leqslant u \leqslant Z_i\}. \tag{2.3}$$

For the estimator of $F(y, x)$ given above, Gürler and Gijbels (1996) provide the following representation with an asymptotically negligible remainder term:

$$\tilde{F}_n(y, x) = \tilde{W}_n(y, x)a(y) - \int_0^y \tilde{W}_n(s, x)A(\mathrm{d}s) - \int_0^y \frac{A(s)}{C(s)}\tilde{C}_n(s)W_{Z,X}^1(\mathrm{d}s, x)$$

$$- \int_0^y A(s)\tilde{L}_n(y)W_{Z,X}^1(\mathrm{d}s, x) + \sqrt{n}R_n(y, x) \tag{2.4}$$

$$\equiv A_1(y, x) - A_2(y, x) - A_3(y, x) + A_4(y, x) + R_n^*(y, x) \equiv \tilde{\tilde{\xi}}_n(y, x) + \tilde{R}_n(y, x), \tag{2.5}$$

where

$$\tilde{F}/_n(y, x) = \sqrt{n}\{F_n(y, x) - F(y, x)\}\tilde{C}_n(y) = \sqrt{n}\{C_n(y) - C(y)\},$$
$$\tilde{W}_n(y, x) = \sqrt{n}\{W_{Z,X,n}^1(y, x) - W_{Z,X}^1(y, x)\}\tilde{L}_n(y) = \sqrt{n}\bar{L}_n(y) \tag{2.6}$$

and

$$L_i(z) = \frac{I(Z_i \leqslant z, \delta_i = 1)}{C(Z_i)} - \int_0^z \frac{I(T_i \leqslant u \leqslant Z_i)}{C^2(u)}W_Z^1(\mathrm{d}u) \quad \text{and} \quad \bar{L}_n(y) = \sum_{i=1}^n L_i(y)/n. \tag{2.7}$$

For the purpose of density estimation, the order of the remainder term in the foregoing representation is further improved in Gürler and Prewitt (1997) and the following result is obtained, where T_b is a compact set:

$$E\left[\sup_{(y,x)\in T_b}|R_n(y, x)|^2\right] = \mathrm{O}(n^{-2}). \tag{2.8}$$

The covariance functions of the processes $\tilde{C}_n(y)$, $\tilde{W}_n(y, x)$ and $\tilde{L}_n(y)$ (see Gijbels and Gürler, 1998) are used to derive the variance of the bivariate density estimator.

2.1. Bivariate density estimator

Gürler and Prewitt (1997) suggest the following bivariate density estimator $f_n(y, x)$ for the LTRC data, by convolving the bivariate d.f. estimator $F_n(y, x)$ with an appropriately chosen kernel function. In particular, they consider the following estimator:

$$f_n(y, x) = \iint \frac{1}{b_x b_y}F_n(y - b_y u, x - b_x v)K(\mathrm{d}u, \mathrm{d}v), \tag{2.9}$$

where $K(u,v)$ is a bivariate kernel function satisfying

$$\iint K(u,v)u^i v^j \,\mathrm{d}u\,\mathrm{d}v = \begin{cases} 1, & i+j=0, \\ 0, & i+j<k, \\ \beta(i,j) < \infty & (\neq 0 \text{ for some } (i,j): i+j=k). \end{cases} \tag{2.10}$$

As mentioned earlier, the choice of the kernel function is important for the tractability of the variance terms, particularly for the bivariate case. We adopted the following properties for the construction of these kernel functions. For the kernel function and for any bivariate function, define

$$K^{ij}(u,v) = \frac{\partial^{i+j}}{\partial u^i \partial v^j} K(u,v) \tag{2.11}$$

with

$$K(y,x) = \int_{-1}^{y} \int_{-1}^{x} K^{11}(u,v)\,\mathrm{d}u\,\mathrm{d}v. \tag{2.12}$$

Then we construct $K(u,v)$ by using a product kernel $K(u,v) = K(u)K(v)$, from which it is obvious that $K^{11}(u,v) = K^1(u)K^1(v)$. The kernel $K(\cdot)$ is constructed by choosing $K^1(\cdot) \in M_{v,k}$ with $v=1$ and $k=3$, where $M_{v,k}$ is as defined in Müller (1988, p. 28), satisfying $K(-1)=K(1)=0$, and $K \in \mathcal{M}_{0,2}$.

3. Main results

In this section we present the leading terms of the asymptotic mean and the variance of the bivariate density estimator. These expressions are important since the quality of the resulting estimator depends critically on the bandwidth choice, and most of the suggested methods for choosing the bandwidth utilize estimates of the mean squared error. A brief discussion about the possible approaches for the bandwidth choice is provided below. First note that we can write

$$f_n(y,x) - f(y,x) = \frac{1}{b_x b_y} \iint \bar{\xi}_n(y - b_y u, x - b_x v) K(\mathrm{d}u, \mathrm{d}v)$$

$$+ \frac{1}{b_x b_y} \iint F(y - b_y u, x - b_x v) K(\mathrm{d}u, \mathrm{d}v) - f(y,x) + r_n(y,x),$$

$$\equiv S_n(y,x) + B_n(y,x) + r_n(y,x), \tag{3.13}$$

where

$$r_n(y,x) = \frac{1}{b_x b_y} \iint R_n(y - ub_y, x - vb_x) K(\mathrm{d}u, \mathrm{d}v). \tag{3.14}$$

The following lemma which is a consequence of the result given in (2.8) indicates that the variance of the remainder term in the foregoing representation (3.13) is asymptotically negligible:

Lemma 1.

$$\mathrm{Var}(r_n(y,x)) = \mathrm{O}\left(\frac{1}{(nb_y b_x)^2}\right).$$

Since the $B_n(y,x)$ term of (3.13) which corresponds to the bias of the kernel estimator is not stochastic, the leading term for the variance of the density estimator is contributed by the term $S_n(y,x)$. We present below this main result regarding the asymptotic variance, the proof of which is given in the appendix. For completeness and reference purposes the asymptotic bias expression is also given in Theorem 1; the proof of

which involves standard applications of Taylor expansion. See Gürler and Prewitt (1997) for more details. Let $V = \int K^2(u)\,du$.

Theorem 1. *Suppose* $\int F_Y(du)/G(u) < \infty$. *Then*

$$\text{Bias}(f_n(y,x)) = (-1)^k \sum_{i+j\ =k} \sum \frac{b_y^i b_x^j}{i!(k-i)!} f^{ij}(y,x)\beta(i,j) + o((b_x b_y)^k) + O\left(\frac{1}{nb_x b_y}\right),$$

$$\text{Var}(f_n(y,x)) = \frac{1}{nb_x b_y}\left[A(y)^2 \frac{\partial^2}{\partial y \partial x} W(y,x)\right] V^2 + o\left(\frac{1}{nb_x b_y}\right)$$

$$= \frac{1}{nb_x b_y} \frac{F_Y(y)}{C(y)} f(y,x)V^2 + o\left(\frac{1}{nb_x b_y}\right).$$

3.1. Special cases

First note that the variance expression given above reduces to the variance of the bivariate kernel estimator for the case of i.i.d. observations since no truncation and no censoring implies that $C(y) = \bar{F}_Y(y)$. For the LTRC model we observe that as a consequence of the incomplete data structure, this variance is magnified by the component $a(y) = \bar{F}_Y(y)/C(y)$, which reflects the noise introduced in the model by truncation and censoring. Apart from the trivial i.i.d. model, we can also elaborate the following cases which corresponds to the truncated only and the censored only type of data:

(a) *Right censored data.* In this case $\alpha = 1$, $G(y) = 1$, $\forall y$ and $C(y) = \bar{F}_Y(y)\bar{H}(y)$, so that $a(y) = 1/\bar{H}(y)$. This implies that the estimation becomes particularly difficult, indicating large variances for large y values. This is a result consistent with the complications of estimation on the right tail with the right censored data.

(b) *Left truncated data.* In this case $C(y) = \alpha^{-1}\bar{F}_Y(y)G(y)$, so that $a(y) = 1/\alpha^{-1}G(y)$. We then confront a magnified variance in the left tail, which is an expected problem for the left truncated data.

3.2. Bandwidth choice

As mentioned before, the most important choice in kernel smoothing is the choice of the bandwidth parameter. There is a vast literature on different perspectives such as local, global and adaptive choices and numerous approaches within each perspective. Most of these results however are directly applicable to the univariate data with i.i.d. observations. A detailed discussion of most of the available methods and their applicability in the case of truncated/censored data is beyond the scope of this study. Therefore, we briefly present below a possible approach, namely a data-driven *local* bandwidth procedure which minimizies the asymptotic MSE (AMSE) at the point (y,x) which is written as below if a product kernel is used

$$\text{AMSE}(y,x) = \frac{1}{4}[b_y^2 f^{20}(y,x)\beta + b_x^2 f^{02}(y,x)\beta]^2 + \frac{1}{nb_x b_y} \frac{F_Y(y)}{C(y)} f(y,x)V^2,$$

where

$$\beta = \int u^2 K(u)\,du.$$

The optimal choices would then be the b_x and b_y which minimize $\text{AMSE}(y,x)$. A solution is guaranteed for the case $b_x \neq b_y$ if $f^{02}(y,x)$ and $f^{20}(y,x)$ have the same sign and are non-zero. It is given by

$$b_x = \left[\frac{F_Y(y)f(y,x)V^2(f^{20}(y,x)/f^{02}(y,x))^{1/2}}{2C(y)f^{02}(y,x)^2\beta^2}\right]^{1/6} n^{-1/6} \quad \text{and} \quad b_y = b_x[f^{02}(y,x)/f^{20}(y,x)]^{1/2}. \quad (3.15)$$

For the simple case of $b = b_x = b_y$ with at least one of $f^{20}(y, x)$ or $f^{02}(y, x)$ non-zero the minimizing value for b is given by

$$b(y, x) = \left(\frac{[F_Y(y)/C(y)]f(y, x)V}{[\frac{1}{2}\beta^2[f^{20}(y, x) + f^{02}(y, x)]^2]} \right)^{1/6} n^{-1/6}.$$

A consistent estimator of this bandwidth can be obtained by replacing the unknown quantities by their consistent estimators. In particular, (2.2) and (2.3) provide consistent estimators for F_Y and $C(y)$, respectively. The estimator given in (2.9) with a pilot bandwidth is consistent for $f(y, x)$. Consistent derivative estimators $\hat{f}^{20}(y, x)(=(\partial/\partial y^2)\hat{f}(y, x))$ and $\hat{f}^{02}(y, x)(=(\partial/\partial x^2)f_n(y, x))$ can be obtained from (2.9).

Appendix. Proof of the asymptotic variance

Using the notation of (3.13) and noting that B_n is not stochastic,

$$\text{Var}(f_n(y, x)) = \text{Var}(S_n(y, x)) + \text{Var}(r_n(y, x)) + 2\,\text{Cov}(S_n(y, x), r_n(y, x)). \tag{A.1}$$

We will show that the leading term in the expression $\text{Var}(S_n) = \text{O}(1/nb_x b_y)$, which together with Lemma 1 will imply that $|\text{Cov}(S_n(y, x), r_n(y, x))| \leqslant [\text{Var}(S_n(y, x))]^{1/2}\text{Var}(r_n(y, x))]^{1/2} = \text{O}(1/(nb_y b_x)^2) = \text{o}(1/nb_y b_x)$. Letting $\tilde{S}_n = \sqrt{n}S_n$, we write $\text{Var}(S_n(y, x)) = n^{-1}E[\tilde{S}_n(y, x)^2]$, and from (3.13) we have

$$E[\tilde{S}_n(y, x)^2] = \left(\frac{1}{b_x b_y} \right)^4 \int_{x-b_x}^{x+b_x} \int_{y-b_y}^{y+b_y} \int_{x-b_x}^{x+b_x} \int_{y-b_y}^{y+b_y} E\tilde{\tilde{\xi}}_n(u_1, u_2)\tilde{\tilde{\xi}}_n(v_1, v_2)$$

$$\times K^{11}\left(\frac{y - u_1}{b_y}, \frac{x - u_2}{b_x} \right) K^{11}\left(\frac{y - v_1}{b_y}, \frac{x - v_2}{b_x} \right) \, du_1 \, du_2 \, dv_1 \, dv_2. \tag{A.2}$$

From the expressions given in Section 2, $\tilde{\tilde{\xi}}_n(u_1, u_2)\tilde{\tilde{\xi}}_n(v_1, v_2)$ can be written as the sum of 16 terms which are the squares and cross-products of $A_i(y, x)$'s, $i = 1, \ldots, 4$. Let $T_1 = A1(u_1, u_2)A1(v_1, v_2)$ be the first of these. It is shown in Lemma 3 below that $T1$ contributes the leading term for the variance and all the others have negligible orders. Proofs of the remaining terms use similar techniques and can be found in Prewitt and Gürler (1998) with further details. The following result is used in the lemmas below:

$$\left[\int_{-1}^{1} s_1 K(s_1)K^1(s_1) \, ds_1 \right]^2 = \frac{1}{4} \left[\int_{-1}^{1} K^2(s_1) \, ds_1 \right]^2. \tag{A.3}$$

Lemma 2. *For* $i + j + k + l \leqslant 2$

$$\int_{-1}^{1} \int_{t_2}^{1} \int_{-1}^{1} \int_{t_1}^{1} s_1^i s_2^j t_1^k t_2^l K^{11}(s_1, s_2)K^{11}(t_1, t_2) \, dt_1 \, ds_1 \, ds_2 \, dt_2$$

$$= \begin{cases} -\frac{1}{4}\left[\int_{-1}^{1} K^2(s_1) \, ds_1 \right]^2 & \text{for } j = 1, \ k = 1, \ i, l = 0 \text{ or } j, k = 0, \ i = 1, \ l = 1, \\ \frac{1}{4}\left[\int_{-1}^{1} K^2(s_1) \, ds_1 \right]^2 & \text{for } k, l = 0, \ i = 1, \ j = 1, \text{ or } i, j = 0, \ k = 1, \ l = 1, \\ 0 \end{cases} \tag{A.4}$$

Lemma 3.

$$T1^* \equiv E\left(\frac{1}{b_x b_y}\right)^4 \int_{x-b_x}^{x+b_x} \int_{y-b_y}^{y+b_y} \int_{x-b_x}^{x+b_x} \int_{y-b_y}^{y+b_y} T1$$

$$\times K^{11}\left(\frac{y-u_1}{b_y}, \frac{x-u_2}{b_x}\right) K^{11}\left(\frac{y-v_1}{b_y}, \frac{x-v_2}{b_x}\right) \, du_1 \, du_2 \, dv_1 \, dv_2$$

$$= \left(\frac{1}{nb_x b_y}\right) \left[A(y)^2 \frac{\partial^2}{\partial y \partial x} W^1_{Z,X}(y,x)\right] \left[\int_{-1}^1 K^2(u) \, du\right]^2 + o\left(\frac{1}{nb_x b_y}\right). \tag{A.5}$$

Proof. Using the covariance result of Gürler and Gijbels (1996) we write

$$T1^* = \left(\frac{1}{b_x b_y}\right)^4 \int_{x-b_x}^{x+b_x} \int_{y-b_y}^{y+b_y} \int_{x-b_x}^{x+b_x} \int_{y-b_y}^{y+b_y} W^1_{Z,X}(u_1 \wedge v_1, u_2 \wedge v_2) \tag{A.6}$$

$$\times A(u_1)A(v_1) K^{11}\left(\frac{y-u_1}{b_y}, \frac{x-u_2}{b_x}\right) K^{11}\left(\frac{y-v_1}{b_y}, \frac{x-v_2}{b_x}\right) \, du_1 \, du_2 \, dv_1 \, dv_2$$

$$- \left(\frac{1}{b_x b_y}\right)^4 \int_{x-b_x}^{x+b_x} \int_{y-b_y}^{y+b_y} \int_{x-b_x}^{x+b_x} \int_{y-b_y}^{y+b_y} W^1_{Z,X}(u_1,u_2) W^1_{Z,X}(v_1,v_2) A(u_1)A(v_1) \tag{A.7}$$

$$K^{11}\left(\frac{y-u_1}{b_y}, \frac{x-u_2}{b_x}\right) K^{11}\left(\frac{y-v_1}{b_y}, \frac{x-v_2}{b_x}\right) \, du_1 \, du_2 \, dv_1 \, dv_2 = I1 + I2. \tag{A.8}$$

Splitting the area of integration first with respect to (u_1, v_1), then (u_2, v_2) and making the appropriate change of variables, we have after some algebra

$$I1 = 4\left(\frac{1}{b_x b_y}\right)^2 \int_{-1}^1 \int_{t_2}^1 \int_{-1}^1 \int_{t_1}^1 W^1_{Z,X}(y-b_y s_1, x-b_x s_2) A(y-b_y s_1) A(y-b_y t_1)$$

$$\times K^{11}(s_1,s_2) K^{11}(t_1,t_2) \, ds_1 \, dt_1 \, ds_2 \, dt_2. \tag{A.9}$$

Let $g_{T1}(y-b_y s_1, x-b_x s_2) = W^1_{Z,X}(y-b_y s_1, x-b_x s_2) A(y-b_y s_1)$, $g^{01}_{T1}(y,x) = \partial/\partial x g_{T1}(y,x)$ and $g^{11}_{T1}(y,x) = \partial^2/\partial y \partial x$. Applying a Taylor expansion yields

$$g_{T1}(y-b_y s_1, x-b_x s_2) = W^1_{Z,X}(y,x) A(y) + g^{10}_{T1}(y,x)(-b_y s_1) + g^{01}_{T1}(y,x)(-b_x s_2)$$

$$+ 1/2 g^{20}_{T1}(y,x)(-b_y s_1)^2 + 1/2 g^{02}_{T1}(y,x)(-b_x s_2)^2$$

$$+ g^{11}_{T1}(y,x)(-b_y s_1)(-b_x s_2) + O((b_x \vee b_y)^3), \tag{A.10}$$

$$A(y-b_y t_1) = A(y) + A^1(y)(-b_y t_1) + 1/2 A^{(2)}(y)(-b_y t_1)^2 + O(b_y^3). \tag{A.11}$$

When the product of (A.10) and (A.11) is taken, any term producing a product of three bandwidths will be of smaller order than the leading term. Also, by Lemmas 2 and 3, many of the other terms will vanish, and

the only remaining non-zero intergral produces the following after application of Lemma 2 and (4.18):

$$I1 = \left[\int_{-1}^{1} K^2(s_1)\,\mathrm{d}s_1\right]^2 \left\{\frac{A(y)}{b_x b_y}\left[A(y)\frac{\partial^2}{\partial y \partial x}W_{Z,X}^1(y,x) + A^1(y)\frac{\partial}{\partial x}W_{Z,X}^1(y,x)\right] + \mathrm{o}\left(\frac{1}{nb_x b_y}\right)\right\}$$

$$- \left[\int_{-1}^{1} K^2(s_1)\,\mathrm{d}s_1\right]^2 \left[\frac{1}{b_x b_y}\frac{\partial}{\partial x}W_{Z,X}^1(y,x)A(y)A^1(y) + \mathrm{o}\left(\frac{1}{nb_x b_y}\right)\right] \tag{A.12}$$

$$= \left(\frac{1}{b_x b_y}\right)\left[A(y)^2\frac{\partial^2}{\partial y \partial x}W_{Z,X}^1(y,x)\right]\left[\int_{-1}^{1} K^2(s_1)\,\mathrm{d}s_1\right]^2 + \mathrm{o}\left(\frac{1}{b_x b_y}\right). \tag{A.13}$$

For the second term in (A.8) we obtain $I_2 = \mathrm{o}(1/b_x b_y)$, which follows since after Taylor expansions, the integral is zero for s_1^m or $s_2^m, m \leqslant 2$, it is of order $1/n$ for terms including $s_1 s_2$ and the terms including $s_1^i s_2^j$, $i + j \geqslant 3$ procedure integrals of order b^2.

References

Gijbels, I., Gürler, Ü., 1998. Covariance function of a bivariate distribution function estimator for left truncated and right censored data. Statist. Sinica 8, 1219–1232.

Gijbels, I., Wang, J.L., 1993. Strong representations of the survivor function estimator for truncated and censored data with applications. J. Multivariate Anal. 47, 210–229.

Gürler, Ü., Gijbels, I., 1996. A bivariate distribution function estimator and its variance under left truncation and right censoring, Discussion Paper 9702, Institut de Statistique, Université Catholique de Louvain.

Gürler, Ü., Prewitt, K., 1997. Bivariate density estimator for left truncated right censored data, submitted for publication.

Müller, H.-G., 1988. Nonparametric Regression Analysis of Longitudinal Data. Lecture Notes in Statistics, Vol. 46. Springer, Berlin.

Prewitt, K.A., Gürler, Ü., 1998. Variance function of the bivariate kernel density estimator for left truncated right censored observations. Technical Report 140, Department of Mathematics, Arizona State University.

Tsai, W.Y., Jewell, N.P., Wang, M.C., 1987. A note on the product limit estimator under right censoring and left truncation. Biometrika 74, 883–886.

Uzunoğullari, Ü., Wang, J.-L., 1992. A comparison of the hazard rate estimators for left truncated and right censored data. Biometrika 79, 297–310.