Ann. Inst. Statist. Math.
Vol. 51, No. 1, 51–70 (1999)

# BANDWIDTH SELECTION IN DENSITY ESTIMATION WITH TRUNCATED AND CENSORED DATA*

C. SÁNCHEZ-SELLERO[1], W. GONZÁLEZ-MANTEIGA[1] AND R. CAO[2]

[1]*Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Santiago de Compostela, 15706 Santiago de Compostela, Spain*
[2]*Departamento de Matemáticas-Facultad de Informática, Universidad de La Coruña, Campus de Elviña, 15071 La Coruña, Spain*

(Received November 15, 1996; revised November 25, 1997)

**Abstract.** A plug-in type bandwidth selector is presented for density estimation with truncated and censored data. It is based on a representation of the MISE function obtained in the paper. Rate of convergence and limit distribution are derived for this selector. A bootstrap method is introduced to estimate the MISE whose minimizer is an alternative bandwidth selector. A simulation study was carried out to assess the behavior with small samples. This methodology is applied to a real-data problem consisting of reporting delay of AIDS cases. The almost sure representation of the product-limit estimator is a key tool in our proofs.

*Key words and phrases*: LTRC model, almost sure representation, kernel estimator.

## 1. Introduction

In many situations, as for example in medical or in engineering life studies, one may not be able to observe the variable of interest, $X$. Left truncation and right censoring are frequent reasons for which incomplete data may appear. Left truncation may occur if the time origin, $X^0$, of the lifetime precedes the time origin of the study, $X^1$. To be precise, when $T = X^1 - X^0 > X$ the case is not observed at all (we do not even know its existence). Right censoring appears when the lifetime of interest is only partially observed due to the previous occurrence of censoring (death from a cause unrelated to the study, withdrawal of the patient during the study, ...). Left truncation and right censoring may happen simultaneously.

Censoring has been receiving considerable attention for a long time, whereas truncation is relatively new and only in recent years has appealed interest, mainly

* Research supported by a DAAD Grant for the first author, the Grant XUGA 20701B96 for the second author, the DGICYT Grant PB94-0494 for the third author and the DGICYT Grant PB95-0826 for the three authors.

because of the AIDS epidemic. A very commented example is the AIDS study by Struthers and Farewell (1989) where the lifetime is the incubation period and the truncating variable is the time from infection until the entry to the study. In this case, the study was carried out with people in which AIDS was not yet present. Another example of truncation comes from reporting delay of AIDS cases to national agencies (see Sánchez Sellero *et al.* (1995)). In this problem, the time of delay is subject to right truncation by the time elapsed from diagnosis until reporting. Woodroofe (1985) also reviews some examples from astronomy and economy where left truncation may occur.

Of course, other types of truncation and censoring are possible, but in this paper the random left truncation and right censoring model (hereafter abreviated as LTRC) has been considered because it is the most common in the literature. More specifically, let $(X,T,C)$ be random variables, where $X$ is the variable of interest, $T$ represents the random left truncation and $C$ the random right censoring. These variables $X$, $T$ and $C$ are often assumed to be positive, although this assumption is not neccessary for our results. It is assumed that $X$ is independent of $(T,C)$ but $T$ and $C$ may be dependent. One observes $(Y,T,\delta)$ if $T \leq Y$ where $Y = X \wedge C = \min(X,C)$ and $\delta = I(X \leq C)$. When $T > Y$ nothing is observed.

Let $\alpha = P(T \leq Y)$, and $F$, $G$, $M$ and $W$ denote the distribution functions of $X$, $T$, $C$ and $Y$, respectively. It is clear that, denoting $\bar{F} = 1 - F$ the survival function, then $\bar{W} = \bar{F}\bar{M}$. Let $(Y_i, T_i, \delta_i)$, $i = 1, \ldots, n$ be an independent and identically distributed sample of $(Y,T,\delta)$ which one observes (i.e., $T_i \leq Y_i$). The well-known product-limit estimator (PLE) $\hat{F}_n$ of $F$, defined in Tsai *et al.* (1987), can be obtained by an empirical estimate of the function $C(z) = P(T \leq z \leq Y/T \leq Y)$ as follows:

$$C_n(z) = n^{-1} \sum_{i=1}^{n} I(T_i \leq z \leq Y_i)$$

$$1 - \hat{F}_n(x) = \prod_{Y_i \leq x} (1 - [nC_n(Y_i)]^{-1})^{\delta_i}.$$

One of the most important properties of the PLE, extensively employed in our proofs, is the strong representation of this estimator as a sum of iid random variables plus a remainder term of order $n^{-1} \log n$ given by Gijbels and Wang (1993),

$$\hat{F}_n(z) - F(z) = \bar{F}(z) n^{-1} \sum_{i=1}^{n} \eta(Y_i, T_i, \delta_i, z) + R_n(z)$$

where

$$\eta(y,t,\delta,z) = I(y \leq z, \delta = 1)/C(y) - \int_{-\infty}^{z} [I(t \leq u \leq y)/C^2(u)] dW_1(u)$$

with $W_1(y) = P(Y \leq y, \delta = 1/T \leq Y)$ and $W_2(y) = P(Y \leq y, \delta = 0/T \leq Y)$.

It is worth noting that $E(\eta(Y,T,\delta,z)) = 0$ and

$$\mathrm{Cov}(\eta(Y,T,\delta,z_1), \eta(Y,T,\delta,z_2)) = q(z_1 \wedge z_2)$$

where we denote $q(z) = \int_{-\infty}^{z} [C(u)]^{-2} dW_1(u)$. Other almost sure representations of this type can be found in Stute (1993) or in Arcones and Giné (1995) only for truncated data and in Lo *et al.* (1989) only for censored data. Zhou (1996) has recently given an almost sure representation for truncated and censored data extending the theorems of Stute (1993).

Our study is confined to kernel density estimation with truncated and censored data, so we assume $F$ is absolutely continuous with density function $f$. Then, the kernel density estimate for $f$ can be defined as the convolution of a kernel function with the PLE of the distribution function $F$,

$$\hat{f}_h(z) = \int K_h(z-u)d\hat{F}_n(u)$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$ is the rescaled kernel function $K$ according to the bandwidth $h$ (i.e., smoothing parameter). Consistency and asymptotic normality of this estimator were obtained by Gijbels and Wang (1993) as an application of their strong representation of the PLE.

However, one of the most important aspects, extensively studied in the last years in the untruncated and uncensored case, is the choice of the smoothing parameter (see Cao *et al.* (1994) and Jones *et al.* (1996) for reviews about the current situation).

Very frequently, the bandwidth is chosen as the minimizer of some measure of the distance between the true density and its estimate. The mean integrated squared error is a distance of that kind, and maybe the most broadly applied among the global criteria based on a deterministic function of the bandwidth.

$$MISE(h) = E\left[\int (\hat{f}_h(x) - f(x))^2 \omega(x)dx\right]. \tag{1.1}$$

The bandwidth minimizing (1.1) is the so-called MISE bandwidth, denoted by $h_{MISE}$. In our definition of the mean integrated squared error, a non-negative weighting function $\omega$ is introduced. This function should be carefully chosen, verifying at least the assumptions imposed below. Other properties of $\omega$ will become advisable in light of the results.

The early study of the MISE for the kernel density estimator without censoring or truncation suggested many different "plug-in" bandwidth selectors (as can be seen in the above-mentioned papers) based on the asymptotic form of $h_{MISE}$,

$$h_{MISE} = c_0 n^{-1/5} + o(n^{-1/5}) = h_{AMISE} + o(n^{-1/5})$$

where

$$c_0 = \left[d_K^{-2}\left(\int f''(x)^2\omega(x)dx\right)^{-1} c_K\right]^{1/5}.$$

We use the functional notation $d_K = \int u^2 K(u)du$, $c_K = \int K(u)^2 du$, which will be applied to other functions as well.

In the next section, a representation is given for the MISE function, which suggests a plug-in bandwidth selector. New and very interesting properties are exhibited by this selector. By arguments similar to those employed in this paper, our results could also be derived for hazard rate estimation. Some simulations are supplied in Section 3 to assess the small-sample behavior of our bandwidth selector. Section 4 contains an application to real-data consisting of reporting delay of AIDS cases. Finally, the proofs are given in the Appendix.

## 2. Main results

Here are the assumptions made on the kernel function, the weighting function and the probability density:

(K1) The kernel $K$ is a symmetric probability density.

(K2) $K$ is three times continuously differentiable, its first derivative is integrable, and

$$\lim_{|x|\to\infty} x^j K^{(j)}(x) = 0, \qquad j = 0,1,2,3.$$

($\omega$1) $\omega$ is compactly supported by a set of points $x$ satisfying $C(x) \geq \epsilon$ for some constant $\epsilon > 0$.

($\omega$2) $\omega$ is three times continuously differentiable.

(D) The density function $f$ is six times differentiable and its sixth derivative is bounded. The first, second, third and fourth derivatives of $f$ are integrable, and the limits of $f$ and any of its first five derivatives at $-\infty$ or $+\infty$ are zero.

(q) The function $q$ is twice continuously differentiable.

Assumption ($\omega$1) relates to the important problem of identifiability in the LTRC model, and makes possible to use the almost sure representation of the PLE given by Gijbels and Wang (1993). Assumption ($\omega$2), like (K2) and (D), is needed to apply Taylor expansions.

Assumption (q) requires regularity for the distribution function of the truncating and censoring variables. Yet, only the second derivative of $q$ means an actual requirement, because the first is given by $q'(x) = f(x)/(C(x)\bar{F}(x))$ and always exists, assuming the existence of $f$.

We obtain an asymptotic expansion of the mean integrated squared error for the kernel estimator.

THEOREM 2.1. *Assume that the bandwidth $h$ satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then the function MISE admits the following representation*

$$\begin{aligned} MISE(h) = \frac{1}{4} d_K^2 h^4 \int f''(x)^2 \omega(x)dx + n^{-1}h^{-1}c_K \int \bar{F}(x)^2 C(x)^{-2}\omega(x)dW_1(x) \\ + n^{-1} \int (q(x) - C(x)^{-1}) f(x)^2 \omega(x)dx \\ + O(h^6) + O(n^{-1}h) + O((nh)^{-3/2}). \end{aligned}$$

This extends Theorem 1 of Cao (1993), which is applicable to untruncated and uncensored data. In this sense, the first term (representing the integrated squared bias) is identical to that of Cao (1993). Whereas, the most relevant novelty in our representation is the second term, related to the variance, which depends on the underlying distributions. Note that, in the untruncated and uncensored case, $q(x) = 1/\bar{F}(x) - 1$, $C(x) = \bar{F}(x)$ and $W_1(x) = F(x)$, so the second and third terms in the representation reduce to those in Theorem 1 of Cao (1993). The new expression is reflecting an increase in the variance of the kernel density estimate for the LTRC model.

A plug-in bandwidth selector is now defined replacing the integrals by estimates of them in the minimizer of the dominant terms of the representation,

$$\hat{h} = \left[ c_K \left( \int \hat{\bar{F}}(x)^2 C_n(x)^{-2} \omega(x) dW_{1n}(x) \right) \left( d_K^2 \int \hat{f}_g''(x)^2 \omega(x) dx \right)^{-1} \right]^{1/5} n^{-1/5}$$

where $W_{1n}(x) = n^{-1} \sum_{i=1}^n I(Y_i \leq x, \delta_i = 1)$ is the empirical estimate of $W_1(x)$, $\hat{f}_g$ denotes a kernel density estimate with bandwidth $g$ and kernel $L$, allowed to be different from $h$ and $K$, and $\hat{f}_g''$ is the second derivative of $\hat{f}_g$. The same Assumptions (K1) and (K2) will be made on the kernel $L$.

Observe that a pilot bandwidth $g$ is needed to give an estimate of the curvature (integrated squared second derivative) of the true density, whereas the integral appearing in the second term of the MISE and, subsequently, in the numerator of the plug-in bandwidth, can be estimated without smoothing. Theorems 2.2 and 2.3 study the accuracy of the estimates of each of the two integrals.

THEOREM 2.2. *The estimate of the integral $\int \bar{F}(x)^2 C(x)^{-2} \omega(x) dW_1(x)$ is root-n consistent, that is,*

$$\int \hat{\bar{F}}(x)^2 C_n(x)^{-2} \omega(x) dW_{1n}(x) - \int \bar{F}(x)^2 C(x)^{-2} \omega(x) dW_1(x) = O_p(n^{-1/2}).$$

This result is quite remarkable, because implies that the new integral appearing in the representation of MISE can be optimally estimated without adding new pilot bandwidths, and so, no further complications are incorporated to the bandwidth selection problem.

Following the ideas in Sheather and Jones (1991), we consider a diagonals-in estimate of the curvature. The next theorem extends the good properties of such an estimate to the LTRC model.

THEOREM 2.3. *Assume that the pilot bandwidth $g$ goes to zero as $n$ tends to infinity, then*

$$(2.1) \qquad E\left( \int \hat{f}_g''(x)^2 \omega(x) dx - \int f''(x)^2 \omega(x) dx \right)^2$$
$$= n^{-2} g^{-10} \left( \int L''(u)^2 du \int \bar{F}^2 C^{-2} \omega dW_1 \right)^2$$
$$+ g^4 d_L^2 \left( \int f'' f^{(4)} \omega \right)^2 + 2n^{-1} g^{-3} d_L$$

$$\cdot \int L''(u)^2 du \int \bar{F}^2 C^{-2} \omega dW_1 \int f'' f^{(4)} \omega$$
$$+ O(n^{-1}g^{-1}) + O(n^{-2}g^{-9}) + O(g^6) + O(n^{-3/2}g^{-9/2}).$$

*Furthermore, if* $ng^3 \to \infty$*, then the dominant part of the bandwidth that minimizes the mean squared error in* (2.1)*, is given by*

$$\left[ \left( \int L''(u)^2 du \right) \left( \int \bar{F}^2 C^{-2} \omega dW_1 \right) \left( -d_L \int f'' f^{(4)} \omega \right)^{-1} n^{-1} \right]^{1/7} \tag{2.2}$$

*if the integral* $\int f'' f^{(4)} \omega$ *is negative, and*

$$\left[ (-5/2) \left( \int L''(u)^2 du \right) \left( \int \bar{F}^2 C^{-2} \omega dW_1 \right) \left( -d_L \int f'' f^{(4)} \omega \right)^{-1} n^{-1} \right]^{1/7} \tag{2.3}$$

*if it is positive.*

To reduce notation we use $\int \bar{F}^2 C^{-2} \omega dW_1 \equiv \int \bar{F}(x)^2 C(x)^{-2} \omega(x) dW_1(x)$, $\int f'' f^{(4)} \omega \equiv \int f''(x) f^{(4)}(x) \omega(x) dx$ and so on.

As in the untruncated and uncensored case, all the dominant terms come from the bias of the estimation and, if the integral $\int f'' f^{(4)} \omega$ is negative, any pilot bandwidth asymptotically equal to (2.2) produces a sum of the three dominant terms in (2.1) that goes to zero. As a consequence, the mean squared error for such bandwidths is of the order $o(n^{-4/7})$. Whereas, if $\int f'' f^{(4)} \omega$ is positive, the optimal pilot bandwidth (2.3) does not produce that effect of cancellation and the mean squared error for this bandwidth is of the order $O(n^{-4/7})$.

The next theorem shows the good rate of approximation of our plug-in bandwidth selector to the optimal bandwidth, $h_{MISE}$, and establishes its normal limit distribution.

THEOREM 2.4. *Let us define the constants*

$$\beta := \left[ 2 \left( \int L'' * L''(u)^2 du \right) \int \left( \frac{f(x)}{\alpha^{-1} P\{T \le x \le C\}} \omega(x) \right)^2 dx \right]^{1/2}$$
$$\gamma := 5\beta^{-1} \left[ c_K \left( \int \bar{F}^2 C^{-2} \omega dW_1 \right) d_K^{-2} \left( \int f''(x)^2 \omega(x) dx \right)^{-6} \right]^{-1/5},$$

*assume* $\int f'' f^{(4)} \omega$ *is negative, and consider a pilot bandwidth asymptotically equal to* (2.2)*, up to a second order term of the type* $o(n^{-1/14})$*. Then the plug-in bandwidth satisfies*

$$\hat{h} - h_{MISE} = O_p(n^{-39/70})$$

*and its asymptotic distribution is given by*

$$n^{6/5} g^{9/2} \gamma (\hat{h} - h_{MISE}) \xrightarrow{d} N(0,1).$$

The symbol $*$ denotes convolution.

*Remark* 1. It is clear that last theorem could equally well be established in terms of relative rate of convergence. In that case, it would be obtained

$$\frac{\hat{h} - h_{MISE}}{h_{MISE}} = O_p(n^{-5/14}).$$

The results show that the truncation and/or censoring do not affect the rates of convergence but do change the coefficients coming up in the dominant terms in the asymptotic representations. We would assert that other procedures for bandwidth selection will also maintain their properties when adapting to the random LTRC model. In this sense, we would like to mention the paper of Patil (1993), where the least squares cross-validation bandwidth selector was studied in the framework of hazard rate estimation with right censored data. In that paper, the same properties as in the iid case were obtained for this kind of selector. However, at this point it should be remembered the extremely slow rate of convergence of this selector to its optimum, proved to be $n^{-1/10}$ by Hall and Marron (1987). Nothing should be expected to overcome this drawback in the models with truncated and/or censored data.

*Remark* 2. In Theorems 3 and 4 it was needed that the integral $\int f'' f^{(4)} \omega$ were negative in order to get the best rates of convergence. We emphasize that this requirement means no restriction, since the weighting function $\omega$ can be suitably chosen to verify this property. Now, it should be remembered that the function $C(x)$ must be bounded away from zero in the support of $\omega$ (Assumption ($\omega$1)). To guarantee this, the identifiability conditions $a_T \leq a_F$ and $b_F \leq b_C$ (where we use the functional notation $(a_H, b_H)$ for the convex support of a distribution function $H$) together with the independence between $T$ and $C$ are assumed. This allows taking $\omega$ close enough to an indicator function on the support of $X$, and thus the integral $\int f'' f^{(4)} \omega$ will be approximately equal to $-\int (f^{(3)})^2$ which is clearly negative.

*Remark* 3. The bootstrap methods, that exhibited a quite good behaviour for bandwidth selection in the common case (see Cao (1993)), could also be considered under the LTRC model. When the data are subject to censoring or truncation, the bootstrap methodology gives rise to new questions about the resampling scheme to be employed. Now, with the aim of estimating the MISE function, the following procedure is presented:

(a) Select pilot bandwidths $g_1$, $g_2$ and $g_3$, one for each variable $X$, $T$ and $C$.

(b) Draw independent random values of $X$, $T$ and $C$ following the distributions $\hat{F}_n * K_{g_1}$, $\hat{G}_n * K_{g_2}$ and $\hat{M}_n * K_{g_1}$, respectively. ($\hat{F}_n$, $\hat{G}_n$ and $\hat{M}_n$ are the product-limit estimates). The bootstrap sample $(Y_i^*, T_i^*, \delta_i^*)$, $i = 1, \ldots, n$ will be supplied by those values $(X, T, C)$ verifying $T \leq Y = X \wedge C$, and as many values of $(X, T, C)$ will be drawn as needed to obtain exactly $n$ of them verifying the inequality.

From the bootstrap sample and for each value of $h$, the density estimate $\hat{f}_h^*(x)$ is constructed and its integrated squared distance to the density estimate $\hat{f}_{g_1}(x)$ (based on the original sample) is calculated. Finally, the mean of these values over all possible bootstrap samples is taken as the bootstrap estimate of the $MISE$ function, that is,

$$MISE^*(h) = E^* \left[ \int (\hat{f}_h^*(x) - \hat{f}_{g_1}(x))^2 \omega(x) dx \right].$$

The minimizer of the previous function is our bootstrap bandwidth selector. Although $MISE^*(h)$ can be written in terms of the original sample and, therefore, from a theoretical point of view no resampling is needed, an explicit expression is quite hard to obtain. Thus, in practice Montecarlo methods are proposed to calculate the values of $MISE^*(h)$.

By the same kind of techniques as in previous results, it can be shown that

$$\begin{aligned} MISE^*(h) = & \frac{1}{4} d_K^2 h^4 \int \hat{f}_{g_1}''(x)^2 \omega(x) dx \\ & + n^{-1} h^{-1} c_K \int \hat{\bar{F}}(x)^2 \hat{C}(x)^{-2} \omega(x) dW_{1n}(x) \\ & + o_p(h^4) + o_p(n^{-1}h^{-1}). \end{aligned}$$

As a consequence, the best choice for the pilot bandwidths $g_2$ and $g_3$ (devised for the truncating and censoring variables, respectively) is taking them equal to zero. This relates to the already-mentioned fact that the integral $\int \bar{F}(x)^2 C(x)^{-2} \omega(x) dW_1(x)$ can be estimated without smoothing.

## 3. Simulations

A simulation study has been carried out to assess the behavior of our bandwidth selector. As in Uzunogullari and Wang (1992), the lifetime variable is simulated from a distribution $F$ with hazard rate $\lambda(x) = (x-1)^2 + 1$ for $x \geq 0$. The corresponding density is given by

$$f(x) = ((x-1)^2 + 1) \cdot e^{-(x-1)^3/3 - x - 1/3} \quad \text{for} \quad x \geq 0.$$

Both the censoring and truncation distributions are simulated from exponential distributions with means 4 and 0.1 respectively. The triples $(X, T, C)$ were drawn independently until one hundred of them satisfied the condition $T \leq Y = \min(X, C)$. In this way, censored and truncated samples $(Y_1, T_1, \delta_1), \ldots, (Y_n, T_n, \delta_n)$ of size $n = 100$ were obtained.

One very relevant feature of this model is the high relative values of the density function $f$ in the neighborhood of zero. This gives rise to some boundary effects. This is a typical situation where the weighting function $\omega$ is required. We have chosen a function $\omega$ which discards 25% of the distribution in the lower tail and 10% in the upper tail. The function $\omega$ was taken uniform in most of this reduced support and it was smoothened at the edges to satisfy the regularity

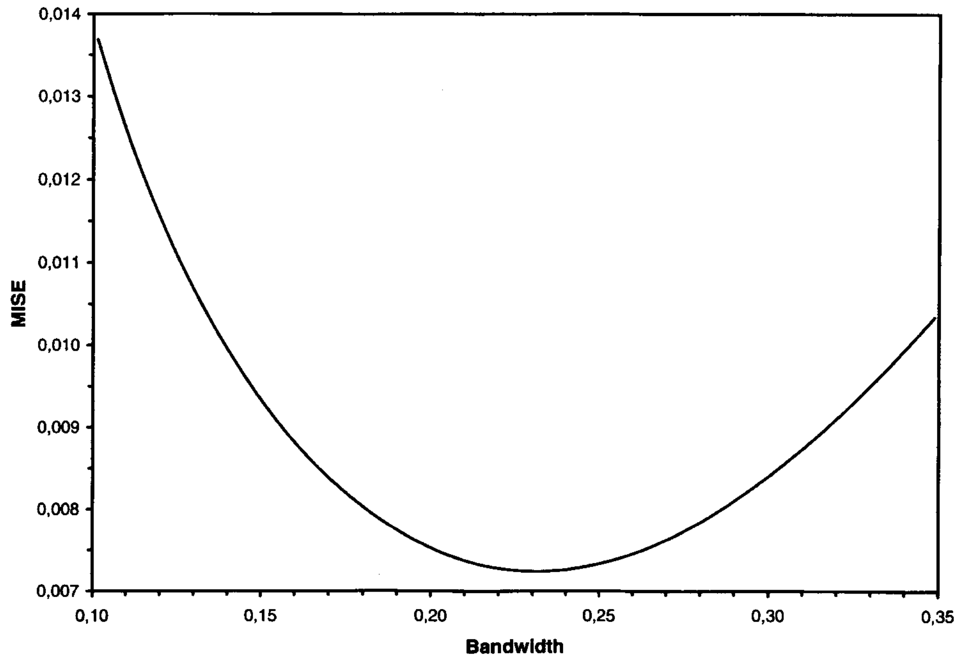


Fig. 1. MISE function for the simulated model.

conditions. The Gaussian kernel was employed both for the density estimate and for the estimate of the curvature.

The MISE function was approximated by the sample mean of the integrated squared error calculated for one thousand samples. The optimal bandwidth 0.2315 was obtained by numerical minimization of the MISE function. Figure 1 shows the MISE function. It should be observed that the MISE function is not far away from its optimum in a wide neighborhood around the optimal bandwidth.

The distribution of our bandwidth selector was approximated by its values computed over one thousand samples. Figure 2 contains a kernel density estimate based on these one thousand values. For this density estimate, the plug-in bandwidth for complete data was taken.

In the light of Fig. 2, the most remarkable property of the bandwidth selector is the slight bias towards values smaller than the optimal. Despite this bias, the approximation to the optimal is quite acceptable given the shape of the MISE function, near to its optimum in an interval containing most of the distribution of our bandwidth selector.

## 4. An application to real data

The methodology developed in this paper has been applied to a real life example where truncation is present. The problem under consideration is that of delay in reporting AIDS cases to the National Commission on AIDS (Spain). The same problem has also been observed in many other countries (for instance, in USA

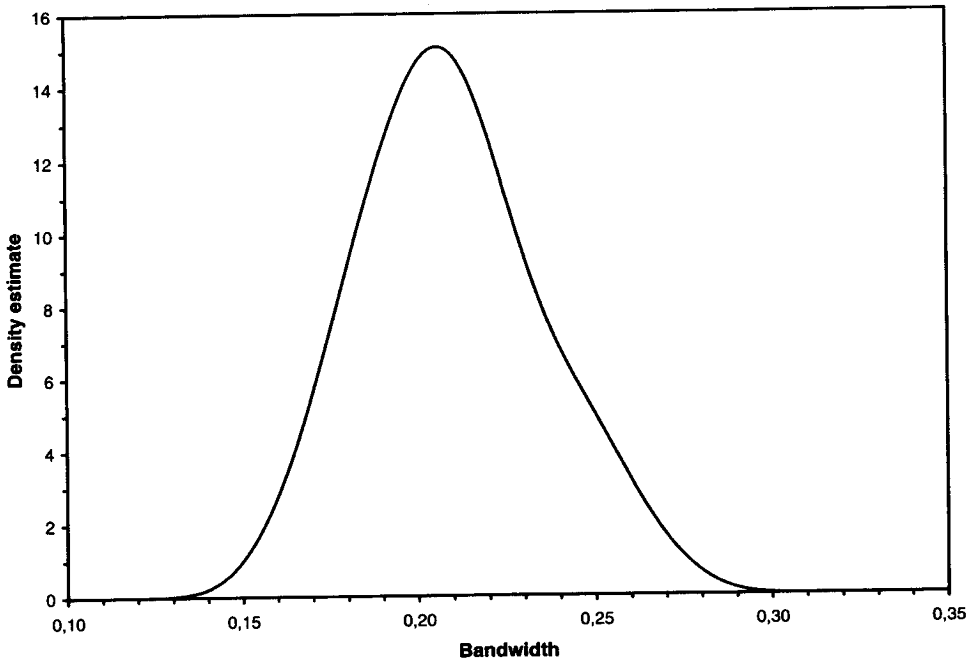


Fig. 2. Kernel density estimate of the bandwidth selector's distribution for the simulated model.

by the Centers for Disease Control). Reporting delay gives place to an underestimation of the size of the AIDS epidemic, due to the lack of those cases not yet reported. Knowing the actual incidence and providing accurate projections of this incidence is crucial for health care planning. Thus, methods to estimate the actual incidence are of great epidemiological interest. For a more detailed description, with particular reference to Spain, see Sánchez Sellero *et al.* (1995).

At the same time, the reporting delay density itself is studied to assess the efficiency of the surveillance system for AIDS. Reporting delay is subject to right truncation by the time elapsed since the date of diagnosis to the current date. The results given in this paper for the LTRC model have been adapted to right truncated data.

The Spanish incidence reported to the National Commission on AIDS since July 1, 1993 until June 30, 1995 has risen to 8736 cases, from which the two curves shown in Fig. 3 have been constructed. Each of them represents the density estimate for reporting delay obtained by the convolution of a distribution function estimate with the kernel. The solid line takes the product-limit estimate and the dashed line the empirical distribution function of the observed reporting delays. The bandwidth has been chosen following the plug-in rule presented in this paper for the solid line, and the common plug-in selector with iid observations for the dashed line.

The effect of truncation becomes apparent. Only the shorter delays are observable and this fact makes the dashed line give unduly high probability to short

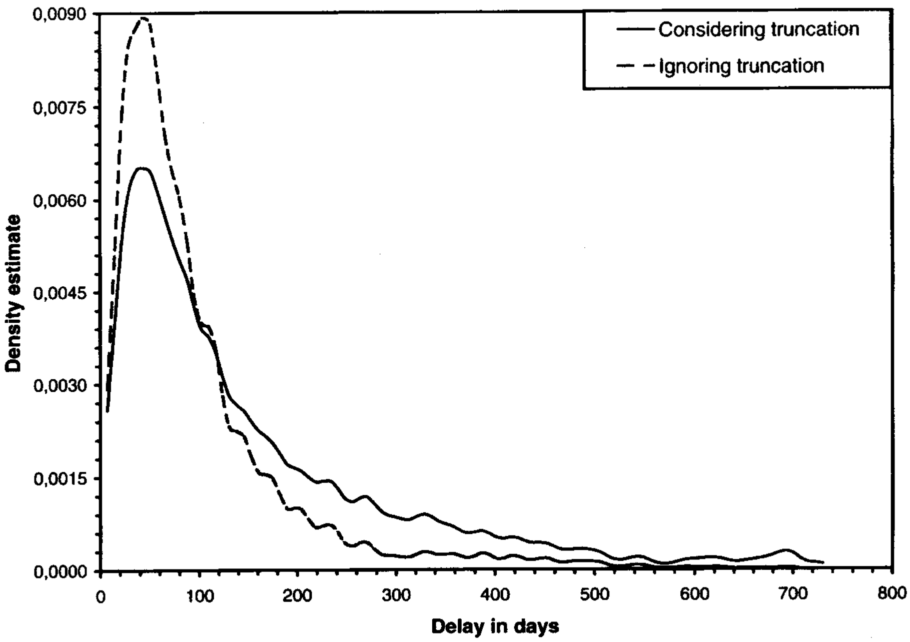


Fig. 3. Density estimates of reporting delay in Spanish AIDS data.

delays at the expense of that of long delays.

It is also stressed that the plug-in bandwidth was considerably larger for the estimate that considered the truncation, $\hat{h} = 13.59$, than for the estimate ignoring it, $\hat{h} = 8.56$.

## Acknowledgements

The authors greatly appreciate the kind comments coming from Professors W. Stute and I. Gijbels, and from two anonymous referees. The work of the first author has been partially carried out during a visit at the University of Giessen, Germany. Many thanks to Dr. Jesús Castilla from the National Commission on AIDS (Spain), for making the database of AIDS cases available.

## Appendix: Proofs

Not every assumption stated in Section 2 is needed for all the different proofs. Some of them are only used in one or two results. Since it often becomes clear along the proofs, the minimal conditions will not be mentioned. Furthermore, these minimal conditions may be weakened versions of our assumptions.

We begin with a lemma that gives an almost sure representation of the $r$-th derivative of the kernel density estimator.

LEMMA A.1. *Denote* $\xi_i(z) = \bar{F}(z)\eta(Y_i, T_i, \delta_i, z)$, $i = 1, \ldots, n$, $z \in R$, *then* *for* $r = 1, \ldots, 6$

$$\hat{f}_h^{(r)}(x) = f^{(r)}(x) + \beta_n^{(r)}(x) + \sigma_n^{(r)}(x) + e_n^{(r)}(x)$$

*where*

$$\beta_n^{(r)}(x) = \int f^{(r)}(x - vg)K(v)dv - f^{(r)}(x)$$

$$\sigma_n^{(r)}(x) = n^{-1}g^{-r-1}\sum_{i=1}^{n}\int \xi_i(x - vg)dK^{(r)}(v)$$

$$e_n^{(r)}(x) = g^{-r-1}\int R_n(x - vg)dK^{(r)}(v).$$

*The error* $e_n^{(r)}(x)$ *satisfies*

(A.1) $$\sup_x |e_n^{(r)}(x)| = O(\log n/(nh^{r+1})) \qquad a.s.$$

$$E\left(\sup_x |e_n^{(r)}(x)|^\rho\right) = O((nh^{r+1})^{-\rho}) \qquad \textit{for any} \quad \rho \geq 1.$$

This result gives a decomposition of $\hat{f}^{(r)}(x)$ in terms of the true derivative, $\beta_n^{(r)}(x)$ representing the bias part, $\sigma_n^{(r)}(x)$ representing the variance part, and $e_n^{(r)}(x)$ the error of the approximation. It should be observed that the bias part is not random and the variance part is a sum of i.i.d. random variables.

The proof of this lemma is an immediate application of the almost sure representation of the distribution function. This strategy was already employed by Lo *et al.* (1989) in their Proposition 3.1 with censored data and more recently by Gijbels and Wang (1993) under the LTRC model.

PROOF OF THEOREM 2.1. Recall Lemma A.1 for $r = 0$. The classical decomposition of MISE into the integrated squared bias and variance can be obtained with a very practicable form

$$\begin{aligned} MISE(h) &= E\left[\int (\hat{f}_h(x) - f(x))^2\omega(x)dx\right] \\ &= E\left[\int (\beta_n(x) + \sigma_n(x) + e_n(x))^2\omega(x)dx\right] \\ &= S_1 + S_2 + S_3 + 2(S_4 + S_5) \end{aligned}$$

where

$$S_1 = \int \beta_n(x)^2\omega(x)dx$$

$$S_2 = \int E(\sigma_n(x)^2)\omega(x)dx$$

$$S_3 = \int E(e_n(x)^2)\omega(x)dx$$

$$S_4 = \int \beta_n(x)E(e_n(x))\omega(x)dx$$

$$S_5 = \int E(\sigma_n(x)e_n(x))\omega(x)dx.$$

The term $S_1$ represents the integrated squared bias and $S_2$ the integrated variance. The other three terms $S_3$, $S_4$ and $S_5$ will be shown to be negligible with respect to the dominant terms in the representation, coming from $S_1$ and $S_2$.

A Taylor expansion and some algebra lead to

$$S_1 = \frac{1}{4} d_K^2 h^4 \int f''(x)^2 \omega(x) dx + O(h^6).$$

For $S_2$, we employ the same arguments, and so, straightforward, albeit more long-winded, calculations supply

$$S_2 = n^{-1}h^{-1}c_K \int \bar{F}^2 C^{-2} \omega dW_1 + n^{-1} \int (q - C^{-1}) f^2 \omega + O(n^{-1}h).$$

Now, using the rate for the error term $e_n(x) \equiv e_n^{(0)}(x)$ given in (A.1), an asymptotic bound for $S_3$ is easily obtained.

$$|S_3| \le \int E\left(\sup_x |e_n(x)|^2\right) = O((nh)^{-2}).$$

The orders of $S_4$ and $S_5$ are finally derived by application of Cauchy-Schwarz inequality.

$$|S_4| \le S_1^{1/2} S_3^{1/2} = O(h^2 (nh)^{-1}) = O(n^{-1}h)$$
$$|S_5| \le S_2^{1/2} S_3^{1/2} = O((n^{-1}h^{-1})^{1/2} (nh)^{-1}) = O((nh)^{-3/2}).$$

PROOF OF THEOREM 2.2. The first step is the following decomposition

$$\int \hat{\bar{F}}^2 C_n^{-2} \omega dW_{1n} - \int \bar{F}^2 C^{-2} \omega dW_1 = A_1 + A_2$$

where

$$A_1 = \int (\hat{\bar{F}}^2 C_n^{-2} - \bar{F}^2 C^{-2}) \omega dW_{1n}$$
$$A_2 = \int \bar{F}^2 C^{-2} \omega d[W_{1n} - W_1].$$

The idea is to prove that $A_1 = O_p(n^{-1/2})$ and $A_2 = O_p(n^{-1/2})$. Let us start with $A_1$.

$$|A_1| \le \left( \sup_{x:\omega(x)>0} |\hat{\bar{F}}(x)^2 C_n(x)^{-2} - \bar{F}(x)^2 C(x)^{-2}| \right) \left| \int \omega dW_{1n} \right|.$$

The supremum can be expressed in terms of the supremum distances of $\hat{\bar{F}}(x)$ and $C_n(x)$ to the true functions, by the following arguments:

$$\begin{aligned} &|\hat{\bar{F}}(x)^2 C_n(x)^{-2} - \bar{F}(x)^2 C(x)^{-2}| \\ &\le |C_n(x)|^{-2} |\hat{\bar{F}}(x)^2 - \bar{F}(x)^2| + |\bar{F}(x)|^2 |C_n(x) C(x)|^{-2} |C_n(x)^2 - C(x)^2| \\ &\le 2|C_n(x)|^{-2} |\hat{\bar{F}}(x) - \bar{F}(x)| + 2|\bar{F}(x)|^2 |C_n(x) C(x)|^{-2} |C_n(x) - C(x)| \ \forall x. \end{aligned}$$

Now, from the almost sure representation of the PLE given by Gijbels and Wang (1993) it is immediately derived that

$$\sup_{x:\omega(x)>0} |\hat{\bar{F}}(x) - \bar{F}(x)| = O_p(n^{-1/2}).$$

Since $C_n$ is the difference of two empirical distribution functions, Dvoretzki *et al.* (1956) inequality suffices to show that

$$\sup_{x:\omega(x)>0} |C_n(x) - C(x)| = O_p(n^{-1/2}).$$

From this and the fact that $C(x) \geq \epsilon$ whenever $\omega(x) > 0$, we can also obtain

$$\sup_{x:\omega(x)>0} |C_n(x)|^{-2} = O_p(1).$$

Now, $\int \omega dW_{1n} = O_p(1)$ provides the desired result $A_1 = O_p(n^{-1/2})$.

To deal with $A_2$, we seek to replace the subdistribution $W_1$ by a distribution function (consequently, $W_{1n}$ by the empirical estimate of that cdf). This can be done in a similar way as in the proof of Lemma 3 of Gijbels and Wang (1993), so we omit the details. At the end, we have

$$A_2 = \int h(t)d[U_n(t) - U(t)]$$

where

$$h(t) = \bar{F}(W_1^{-1}(t))^2 C(W_1^{-1}(t))^{-2}\omega(W_1^{-1}(t))I(t \leq W_1(\infty)),$$
$$W_1^{-1}(t) = \inf\{x : W_1(x) \geq t\},$$

$U(t)$ is the uniform distribution function on the unit interval and $U_n(t)$ is the empirical distribution function of the iid uniform $[0,1]$ random variables $V_1, \ldots, V_n$ given by

$$V_i = \begin{cases} W_1(Y_i) & \text{if } \delta_i = 1 \\ 1 - W_2(Y_i) & \text{if } \delta_i = 0. \end{cases}$$

Since $h(t) \leq \epsilon^{-2} \sup_x |\omega(x)| \; \forall t$, Lemma B in Serfling ((1980), p. 223) and Markov inequality lead to $A_2 = O_p(n^{-1/2})$.

PROOF OF THEOREM 2.3. From Lemma A.1 for $r = 2$, we can obtain a representation for the curvature of the density estimator

$$\int \hat{f}_g''(x)^2\omega(x)dx - \int f''(x)^2\omega(x)dx = B_1+B_2+B_3+2(B_4+B_5+B_6+B_7+B_8+B_9)$$

where

$$B_1 = \int \beta_n^{(2)}(x)^2\omega(x)dx \quad B_2 = \int \sigma_n^{(2)}(x)^2\omega(x)dx \quad B_3 = \int e_n^{(2)}(x)^2\omega(x)dx$$
$$B_4 = \int f''(x)\beta_n^{(2)}(x)\omega(x)dx \quad B_5 = \int f''(x)\sigma_n^{(2)}(x)\omega(x)dx$$
$$B_6 = \int f''(x)e_n^{(2)}(x)\omega(x)dx \quad B_7 = \int \beta_n^{(2)}(x)\sigma_n^{(2)}(x)\omega(x)dx$$
$$B_8 = \int \beta_n^{(2)}(x)e_n^{(2)}(x)\omega(x)dx \quad B_9 = \int \sigma_n^{(2)}(x)e_n^{(2)}(x)\omega(x)dx.$$

The only relevant summands, as regards the representation (2.1) and the asymptotic normality of $\int \hat{f}_g''(x)^2\omega(x)dx$ established below, are $B_2$ and $B_4$. The term $B_4$ is not random and can be expanded as

$$B_4 = \frac{1}{2}g^2 d_L \int f''(x)f^{(4)}(x)\omega(x)dx + O(g^4). \tag{A.2}$$

However $B_2$ is a double sum that can be decomposed into the diagonal terms (which give place to a sum of iid random variables) and a $U$-statistic, denoted by $Q_n$ and $U_n$, respectively

$$B_2 = Q_n + U_n, \qquad Q_n = n^{-2}g^{-6}\sum_{i=1}^{n} H_n(Z_i, Z_i), \qquad U_n = n^{-2}g^{-6}\sum_{i\neq j} H_n(Z_i, Z_j)$$

where $Z_i = (Y_i, T_i, \delta_i)$, $i = 1, \ldots, n$ and

$$H_n(Z_i, Z_j) = \int \lambda(Z_i, x, g)\lambda(Z_j, x, g)\omega(x)dx, \qquad \lambda(Z_i, x, g) = \int \xi_i(x - vg)dL''(v).$$

The term $B_1$ is not random and is easily shown to be $B_1 = O(g^4)$. The term $B_5$ satisfies $E(B_5) = 0$ and $\operatorname{Var} B_5 = O(n^{-1})$ and the same happens with $B_7$. The terms involving the error $e_n^{(2)}(x)$ will be negligible. Finally, it only remains to obtain the mean and variance of $B_2$. The next two lemmas will be helpful to this aim. They are derived by straightforward calculations, so we omit the details in their proofs.

LEMMA A.2. *The following expression holds for the expectation of* $Q_n$

$$E(Q_n) = n^{-1}g^{-5}\left(\int L''(u)^2du\right)\left(\int \bar{F}^2C^{-2}\omega dW_1\right) + O(n^{-1}g^{-3}). \tag{A.3}$$

LEMMA A.3. *The expectation of each summand appearing in* $U_n$ *is zero and its second moment satisfies*

$$E(H_n(Z_1, Z_2)^2) = g^3 \int L'' * L''(u)^2du \int \left(\frac{f(x)}{\alpha^{-1}P\{T \le x \le C\}}\omega(x)\right)^2 dx + o(g^3)$$

*which implies*

$$\operatorname{Var} U_n = 2n^{-2}g^{-9}\int L''*L''(u)^2du\int\left(\frac{f(x)}{\alpha^{-1}P\{T\le x\le C\}}\omega(x)\right)^2dx + o(n^{-2}g^{-9}).$$

From Lemma A.3 and some further derivations, it can be shown that

$$\operatorname{Var} B_2 = O(n^{-2}g^{-9}).$$

Finally, the squared bias provides the first three summands in the right-hand side of (2.1) and these terms are obtained from expressions (A.2) and (A.3).

To prove (2.2) and (2.3), first note that if $ng^3 \to 0$, then the first three summands in the right-hand side of (2.1) are the dominant terms.

On the other hand, it is not difficult, but tedious, to prove that the asymptotic form of the pilot bandwidth $g$, minimizing (2.1), coincides with the minimizer of the dominant part. From this point, the last paragraph in Theorem 2.3 becomes evident. □

PROOF OF THEOREM 2.4. Under the assumptions on the kernel $K$, the weighting function and the density, the function $MISE$ is three times differentiable with respect to $h$. Defining

$$\widehat{MISE}(h) = h^4\frac{d_K^2}{4}\int \hat{f}_g''(x)^2\omega(x)dx + n^{-1}h^{-1}c_K\int \hat{\bar{F}}^2 C_n^{-2}\omega dW_{1n}$$

this function $\widehat{MISE}$, which is nothing else but an estimate of $MISE$, is obviously three times differentiable with respect to $h$. Then, as a consequence of a Taylor expansion,

$$\begin{aligned}MISE'(h_{MISE}) = 0 = \widehat{MISE}'(\hat{h}) &= \widehat{MISE}'(h_{MISE})\\ &\quad + \widehat{MISE}''(h_{MISE})(\hat{h}-h_{MISE})\\ &\quad + \frac{1}{2}\widehat{MISE}'''(\tilde{h})(\hat{h}-h_{MISE})^2\end{aligned}$$

where $\tilde{h}$ is a point in between $h_{MISE}$ and $\hat{h}$.

Rearranging terms in the previous equation,

$$\text{(A.4)}\quad \begin{aligned}&MISE'(h_{MISE}) - \widehat{MISE}'(h_{MISE})\\ &= \widehat{MISE}''(h_{MISE})(\hat{h}-h_{MISE}) + \frac{1}{2}\widehat{MISE}'''(\tilde{h})(\hat{h}-h_{MISE})^2.\end{aligned}$$

At this point we use two lemmas. Lemma A.4 gives a rough bound for the rate of convergence of the plug-in bandwidth. It is an immediate consequence of Theorems 2.2 and 2.3 and the definition of the plug-in bandwidth. Lemma A.5 is easily derived by the same arguments as in the proof of Theorem 2.1 (almost sure representation, Taylor expansion, ...) and some elementary analysis (mean value theorem, dominated convergence theorem, ...).

LEMMA A.4. *In addition to the conditions stated in Section* 1, *assume that the pilot bandwidth* $g$ *is of precise order* $n^{-1/7}$, *then*

$$\hat{h} - h_{MISE} = O_p(n^{-2/5}).$$

LEMMA A.5. *Under the current assumptions*,

$$MISE'(h) = h^3 d_K^2 \int f''(x)^2 \omega(x)dx - n^{-1}h^{-2}c_K \int \bar{F}^2 C^{-2}\omega dW_1 + O(h^5) + O(n^{-1}).$$

Applying Theorem 2.3 and Lemmas A.4 and A.5 to (A.4), we obtain

$$\begin{aligned} &MISE'(h_{MISE}) - \widehat{MISE}'(h_{MISE}) \\ &= \left(3d_K^2 h_{MISE}^2 \int \hat{f}_g''(x)^2\omega(x)dx + 2n^{-1}h_{MISE}^{-3}c_K \int \hat{\bar{F}}^2 C_n^{-2}\omega dW_{1n}\right) \\ &\quad \cdot (\hat{h} - h_{MISE}) + O_p(n^{-1}). \end{aligned}$$

The estimated integrals appearing in this formula can be replaced by their limits by means of Theorems 2.2 and 2.3 and Lemma A.4 to obtain

$$\begin{aligned} \text{(A.5)}\quad &MISE'(h_{MISE}) - \widehat{MISE}'(h_{MISE}) \\ &= \left(3d_K^2 h_{MISE}^2 \int f''(x)^2\omega(x)dx + 2n^{-1}h_{MISE}^{-3}c_K \int \bar{F}^2 C^{-2}\omega dW_1\right) \\ &\quad \cdot (\hat{h} - h_{MISE}) + O_p(n^{-1}). \end{aligned}$$

On the other hand, making use of Lemma A.5, a different expression is found for the left-hand side of (A.5),

$$\begin{aligned} \text{(A.6)}\quad &\widehat{MISE}'(h_{MISE}) - MISE'(h_{MISE}) \\ &= d_K^2 h_{MISE}^3 \left[\int \hat{f}_g''(x)^2\omega(x)dx - \int f''(x)^2\omega(x)dx\right] \\ &\quad - n^{-1}h_{MISE}^{-2}c_K \left[\int \hat{\bar{F}}^2 C_n^{-2}\omega dW_{1n} - \int \bar{F}^2 C^{-2}\omega dW_1\right] \\ &\quad + O_p(n^{-1}) \\ &= d_K^2 h_{MISE}^3 \left[\int \hat{f}_g''(x)^2\omega(x)dx - \int f''(x)^2\omega(x)dx\right] + O_p(n^{-1}) \end{aligned}$$

where the last equality was derived from Theorem 2.2. It is then clear that the next step will consist of investigating the limit distribution of the curvature of the density estimator. This follows as a consequence of the next lemma.

LEMMA A.6. *If the pilot bandwidth $g$ goes to zero and $ng \to \infty$, the $U$-statistic $U_n$, defined above, is asymptotically normal with zero mean and variance given in Lemma* A.3.

PROOF OF LEMMA A.6. The proof will be based on the central limit theorem by Hall (1984) for degenerate $U$-statistics. First, note that the $U$-statistic may be written in the form

$$U_n = 2n^{-2}g^{-6}\sum_{i<j} H_n(Z_i, Z_j).$$

The quantity $U_n$ is a degenerate $U$-statistic and its kernel function, $H_n$, is symmetric, with finite second moment (already proved in Lemma A.3). Since

$$\sum_{i<j} H_n(Z_i, Z_j)$$

has zero mean and variance given by

$$\frac{1}{2}n(n-1)E(H_n(Z_1, Z_2)^2),$$

the only requirements needed (as stated in Hall's result) are the existence of the fourth moment of the term $H_n(Z_1, Z_2)$ and the limit condition

$$[E(H_n(Z_1, Z_2)^2)]^{-2}[E(G_n(Z_1, Z_2)^2) + n^{-1}E(H_n(Z_1, Z_2)^4)] \to 0 \tag{A.7}$$

where

$$G_n(x, y) = E(H_n(Z_1, x)H_n(Z_1, y)), \qquad x, y \in R.$$

A representation for the first factor in (A.7) was already derived in Lemma A.3. Now, we deal with the fourth moment. It may be shown that

$$\sup_{x,y} |H_n(x, y)| = O(g)$$

and this, together with Lemma A.3, leads to

$$\begin{aligned} 0 \le E(H_n(Z_1, Z_2)^4) &\le E\left[\left(\sup_{x,y}|H_n(x,y)|\right)^2 H_n(Z_1, Z_2)^2\right] \\ &\le \left(\sup_{x,y}|H_n(x,y)|\right)^2 E[H_n(Z_1, Z_2)^2] = O(g^5). \end{aligned} \tag{A.8}$$

As regards $E(G_n(Z_1, Z_2)^2)$, a direct substitution of its components leads to

$$\begin{aligned}E(G_n(Z_1, Z_2)^2) = \int \overset{12)}{\cdots} \int & \bar{F}(x_1 - v_1 g)\bar{F}(x_3 - v_2 g)\bar{F}(x_2 - v_3 g)\bar{F}(x_4 - v_4 g) \\ & \cdot \bar{F}(x_1 - v_5 g)\bar{F}(x_2 - v_6 g)\bar{F}(x_3 - v_7 g)\bar{F}(x_4 - v_8 g) \\ & \cdot q((x_1 - v_1 g) \wedge (x_3 - v_2 g))q((x_2 - v_3 g) \wedge (x_4 - v_4 g)) \\ & \cdot q((x_1 - v_5 g) \wedge (x_2 - v_6 g))q((x_3 - v_7 g) \wedge (x_4 - v_8 g)) \\ & \cdot \prod_{i=1}^{8} K^{(3)}(v_i)dv_i \prod_{j=1}^{4} \omega(x_j)dx_j.\end{aligned}$$

By successive calculation of each integral with changes of variable and application of assumptions like the symmetry of the kernel, this expression is progressively bounded. In this way by straightforward, although long-winded, calculation the following rate is achieved

$$E(G_n(Z_1, Z_2)^2) = O(g^7).$$

The rate just found, Lemma A.3 and (A.8) imply condition (A.7). Hence, the whole proof of this lemma is finished. □

As a consequence of the previous lemma, the limit distribution of

$$\begin{aligned}ng^{9/2}\beta^{-1}\Big[\int \hat{f}_g''(x)^2\omega(x)dx - \int f''(x)^2\omega(x)dx - g^2 d_L \int f'' f^{(4)}\omega \\ -n^{-1}g^{-5}\left(\int L''(u)^2 du\right)\left(\int \bar{F}^2 C^{-2}\omega dW_1\right)\Big]\end{aligned}$$

may be proved to be normal with zero mean and unit variance, whenever the pilot bandwidth $g$ is of precise rate $n^{-1/7}$. Furthermore, if the ratio with numerator $g$ and denominator equal to (2.2) goes to one at the rate $o(n^{-1/14})$, then we may delete the last two summands. Hence the same standard normal limit distribution holds for

$$ng^{9/2}\beta^{-1}\left[\int \hat{f}_g''(x)^2\omega(x)dx - \int f''(x)^2\omega(x)dx\right].$$

This fact, together with expression (A.6), proves that

$$ng^{9/2}\beta^{-1}d_K^{-2}h_{MISE}^{-3}(\widehat{MISE}'(h_{MISE}) - MISE'(h_{MISE}))$$

is of order $O_p(1)$ and has a standard normal limit distribution. Now, using (A.5), the same statement remains true for

$$\begin{aligned}ng^{9/2}\beta^{-1}d_K^{-2}h_{MISE}^{-3}\Big(3d_K^2 h_{MISE}^2 \int f''(x)^2\omega(x)dx \\ + 2n^{-1}h_{MISE}^{-3}c_K \int \bar{F}^2 C^{-2}\omega dW_1\Big)(\hat{h} - h_{MISE}).\end{aligned}$$

Using the asymptotic form of $h_{MISE}$, the theses in Theorem 2.4 are easily derived. □

## References

Arcones, M. A. and Giné, E. (1995). On the law of iterated logarithm for canonical $U$-statistics and processes, *Stochastic Process. Appl.*, **58**, 217–245.

Cao, R. (1993). Bootstrapping the mean integrated squared error, *J. Multivariate Anal.*, **45**, 137–160.

Cao, R., Cuevas, A. and González-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation, *Comput. Statist. Data Anal.*, **17**, 153–176.

Dvoretzki, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *Ann. Math. Statist.*, **27**, 642–669.

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

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators, *J. Multivariate Anal.*, **14**, 1–16.

Hall, P. and Marron, J. S. (1987). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation, *Probab. Theory Related Fields*, **74**, 567–581.

Jones, M. C., Marron, J. S. and Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation, *J. Amer. Statist. Assoc.*, **91**, 401–407.

Lo, S. H., Mack, Y. P. and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator, *Probab. Theory Related Fields*, **80**, 461–473.

Patil, P. N. (1993). Bandwidth choice for nonparametric hazard rate estimation, *J. Statist. Plann. Inference*, **35**, 15–30.

Sánchez Sellero, C., Vázquez, E., González, W., Otero, X. L., Hervada, X., Fernández, E. and Taboada, X. A. (1995). Reporting delay: A review with a simulation study and application to Spanish AIDS data, *Statistics in Medicine*, **15**, 305–321.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *J. Roy. Statist. Soc. Ser. B*, **53**, 683–690.

Struthers, C. A. and Farewell, V. T. (1989). A mixture model for time to AIDS data with left truncation and an uncertain origin, *Biometrika*, **76**, 814–817.

Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data, *Ann. Statist.*, **21**, 146–156.

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

Uzunogullari, U. and Wang, J. L. (1992). A comparison of hazard rate estimators for left truncated and right censored data, *Biometrika*, **79**, 297–310.

Woodroofe, M. (1985). Estimating a distribution function with truncated data, *Ann. Statist.*, **13**, 163–177.

Zhou, Y. (1996) A note on the TJW product-limit estimator for truncated and censored data, *Statist. Probab. Lett.*, **26**, 381–387.