AISM (2006) 58: 357–378
DOI 10.1007/s10463-005-0011-y

Elias Ould-Saïd · Mohamed Lemdani

Asymptotic properties of a nonparametric regression function estimator with randomly truncated data

Received: 19 April 2004 / Revised: 9 February 2005 / Published online: 30 May 2006

Abstract In this paper, we define a new kernel estimator of the regression function under a left truncation model. We establish the pointwise and uniform strong consistency over a compact set and give a rate of convergence of the estimate. The pointwise asymptotic normality of the estimate is also given. Some simulations are given to show the asymptotic behavior of the estimate in different cases. The distribution function and the covariable's density are also estimated.

Keywords Asymptotic normality · Kernel · Nonparametric regression · Rate of convergence · Strong consistency · Truncated data · V-C class

1 Introduction

Let Y be a real random variable (rv) with distribution function (df) F and $\mathbf{X}$ a random vector of covariates taking its values in $\mathbb{R}^s$ with df V and continuous density v. We want to estimate Y after observing $\mathbf{X}$. The regression function at a point x is the conditional expectation of Y given $\mathbf{X} = x$, that is

$$\mathbb{E}\,[Y \mid \mathbf{X} = x] =: m(x), \tag{1}$$

which can be written $m(x) = \psi(x)/v(x)$ with

$$\psi(x) = \int y\mathbf{F}(x, \mathrm{d}y) \tag{2}$$

E. Ould-Saïd
L.M.P.A. J. Liouville, Univ. du Littoral Côte d'Opale,
BP 699, 62228 Calais, France
E-mail: ouldsaid@lmpa.univ-littoral.fr

M. Lemdani (✉)
Lab. Biomaths-EA 3614. Univ. de Lille 2, Fac. Pharm.
3, rue Pr. Laguesse, 59006 Lille, France
E-mail: mlemdani@pharma.univ-lille2.fr

where $\mathbf{F}(\cdot,\cdot)$ is the joint df of the random vector $(\mathbf{X}, Y)$ with density $\mathbf{f}(\cdot,\cdot)$.

Now, based on an N-sample, $m(x)$ is often estimated by local averaging, i.e., by averaging the Y_i values for which $\mathbf{X}_i$ is close to x, that is

$$m_N(x) = \sum_{i=1}^{N} W_{i,N}(x) Y_i, \tag{3}$$

using weights $W_{i,N}(\cdot)$ that are measurable functions of x depending on $\mathbf{X}_1, \dots, \mathbf{X}_N$. The well-known kernel regression estimator (Nadaraya–Watson) is defined by

$$\begin{aligned} W_{i,N}(x) &= \frac{K\left\{(x-\mathbf{X}_i)/h_N\right\}}{\sum_{j=1}^{N} K\left\{(x-\mathbf{X}_j)/h_N\right\}} \\ &=: \frac{\left(Nh_N^s\right)^{-1} K\left\{(x-\mathbf{X}_i)/h_N\right\}}{v_N(x)} \end{aligned} \tag{4}$$

with the convention $0/0 = 0$. Here, $v_N(\cdot)$ is the kernel estimator of $v(\cdot)$, K is a nonnegative function defined on $\mathbb{R}^s$ and (h_N) a nonnegative sequence which goes to zero as N goes to infinity.

The nonparametric kernel estimator $m_N(\cdot)$ has first been considered in the case of complete data. As the literature is abundant, we mention only a few references, see e.g., Devroye, Györfi and Lugosi (1996), Györfi, Kohler and Walk (1998), Walk (2002) and the references therein.

In the case of censored data, the estimation of the regression function $m(\cdot)$ has been studied by many authors. The monograph of Fan and Gijbels (1996) and the recent work of Kohler, Máthé and Pintér (2002) can be consulted.

Now, let (Y_i, T_i), $1 \le i \le N$, be a sequence of iid random vectors such that (Y_i) is independent of (T_i). Let F and G denote the respective common dfs of the Y_i values and T_i values. In the random left-truncation model, the *rv of interest* Y_i is interferred by the *truncation rv* T_i in such a way that both Y_i and T_i are observable when $Y_i \ge T_i$, whereas neither is observed if $Y_i < T_i$. Such data occur in astronomy and economics (see, e.g., Woodroofe, 1985; Feigelson and Babu, 1986; Wang et al. 1986; Tsai et al. 1987 and also in epidemiology and biometry (see, e.g., He and Yang 1994).

As a consequence of truncation, n, the size of the actually observed sample, is random, with $n \le N$ and N is unknown. Let $\alpha := \mathbb{P}\{Y \ge T\}$ be the probability that we observe the rv of interest Y. It is clear that if $\alpha = 0$ no data can be observed. Therefore, we suppose throughout the paper that $\alpha > 0$.

From the strong law of large numbers (SLLN) we have, as $N \longrightarrow +\infty$

$$\hat{\alpha}_n := \frac{n}{N} \longrightarrow \alpha, \quad \mathbb{P} - a.s. \tag{5}$$

Without possible confusion we still denote $(Y_i, T_i), i = 1, \dots, n$ the observed sequence. Conditionally on the value of n, these observed random vectors are still iid. Following Stute (1993) the dfs of Y and T become:

$$F^*(y) = \alpha^{-1} \int_{-\infty}^{y} G(u) \mathrm{d}F(u) \qquad \text{and} \qquad G^*(y) = \alpha^{-1} \int_{-\infty}^{\infty} G(y \wedge u) \mathrm{d}F(u)$$

which are estimated by

$$F_n^*(y) = n^{-1}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \le y\}} \quad \text{and} \quad G_n^*(y) = n^{-1}\sum_{i=1}^{n} \mathbb{1}_{\{T_i \le y\}},$$

respectively.

Let $C(\cdot)$ be defined by

$$\begin{aligned} C(y) &:= G^*(y) - F^*(y) \\ &= \alpha^{-1} G(y)[1 - F(y)], \end{aligned}$$

with empirical estimator

$$C_n(y) = n^{-1}\sum_{i=1}^{n} \mathbb{1}_{\{T_i \le y \le Y_i\}} = G_n^*(y) - F_n^*(y^-).$$

Since N is unknown and n is known (although random), our results would not be stated with respect to the probability measure $\mathbb{P}$ (related to the N-sample) but will involve the conditional probability $\mathbf{P}$ related to the actually observed n-sample. Also $\mathbb{E}$ and $\mathbf{E}$ will denote the expectation operators related to $\mathbb{P}$ and $\mathbf{P}$, respectively.

Then, the nonparametric maximum likelihood estimators (MLE) of F and G are the product-limit estimators obtained by Lynden-Bell (1971) given by

$$\begin{aligned} 1 - F_n(y) &= \prod_{i; Y_i \le y} \left[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)}\right] \quad \text{and} \\ G_n(y) &= \prod_{i; T_i > y} \left[\frac{nC_n(T_i) - 1}{nC_n(T_i)}\right]. \end{aligned} \tag{6}$$

Now, for any df L, let $a_L = \inf\{y, L(y) > 0\}$ and $b_L = \sup\{y, L(y) < 1\}$ be its endpoints.

Asymptotic properties of Eq. (6) have been studied by Woodroofe (1985). In his Theorem 2, he establishes the uniform consistency results

$$\sup_y |F_n(y) - F_0(y)| \xrightarrow{\mathbf{P}-a.s.} 0 \quad \text{and} \quad \sup_y |G_n(y) - G_0(y)| \xrightarrow{\mathbf{P}-a.s.} 0, \tag{7}$$

where F_0 denotes the conditional distribution of Y given $Y \ge a_G$ and G_0 is the conditional distribution of T given $T \le b_F$. Therefore, F is identifiable ($F = F_0$) only when $a_G \le a_F$, whereas G is identifiable ($G = G_0$) only when $b_G \le b_F$. We point out that these are necessary but not sufficient identifiability conditions.

Consequently, α is identifiable only if $a_G \le a_F$ and $b_G \le b_F$. Note that the estimator $\hat{\alpha}_n$ defined in Eq. (5) cannot be calculated (since N is unknown). Another estimator, namely

$$\alpha_n = \frac{G_n(y)[1 - F_n(y^-)]}{C_n(y)} \tag{8}$$

is used. He and Yang (1998) proved that α_n does not depend on y and its value can then be obtained for any y such that $C_n(y) \neq 0$. Furthermore, they showed (Corollary 2.5) that

$$\alpha_n \overset{\mathbf{P}-a.s.}{\longrightarrow} \alpha \qquad \text{as} \quad n \to \infty. \tag{9}$$

Now, we come back to our main problem. We would like to estimate the regression function $m(x)$ under random left truncation.

Let $(\mathbf{X}_i, Y_i, T_i), i = 1, \ldots, N$, be a sequence of random vectors where Y is the variable of interest, T, the truncation variable and $\mathbf{X}$, the vector of covariates. The number of observed vectors (i.e., $Y_i \geq T_i$) is still denoted n. Based on this sample, the regression function that can be estimated is

$$m^*(x) = \mathbb{E}\left[Y|\mathbf{X} = x, Y \geq T\right]. \tag{10}$$

Since the marginal dfs of Y and T in the n-sample are different from their dfs in the original sample, we have that $m^*(x) \neq m(x)$. Therefore, an estimate of $m^*(\cdot)$ cannot be used as an estimate of $m(\cdot)$. Constructing an appropriate estimator is then obtained by adapting the weights (Eq. 4) in order to put more emphasis on small values of the interest variable Y which are more truncated than large values. This can be achieved, as we will show in Sect. 3, by dividing by G_n which yields our new estimator of $m(x)$:

$$m_n(x) = \frac{\sum_{i=1}^{n} Y_i G_n^{-1}(Y_i) K\left\{(x - \mathbf{X}_i)/h_n\right\}}{\sum_{i=1}^{n} G_n^{-1}(Y_i) K\left\{(x - \mathbf{X}_i)/h_n\right\}}.$$

Finally, let us quote some particular cases. For a linear model, i.e.,

$$Y = \delta'\mathbf{X} + \varepsilon, \tag{11}$$

where $\delta \in \mathbb{R}^s$ is an unknown parameter of interest and the error ε has a zero mean and is uncorrelated with $\mathbf{X}$, one of the earliest papers addressing the estimation of the regression parameter δ under left truncation is due to Bhattacharya et al. (1983) who built the estimator for the slope δ_1 in the simple linear regression $Y = \delta_0 + \delta_1 X + \varepsilon$, while the truncated variable T was assumed to be a known constant. The slope δ_1 corresponds to the Hubble constant in the study of luminosity function of galaxies. In the case of a fixed (and therefore nonrandom) truncation and with censored data, the nonparametric regression was studied by Lewbel and Linton (2002), by using the local polynomial method. We point out here that the truncation definition used in their work is not the same as the (usual) one we use. Indeed we stick for our part to one of the main consequences of truncation which is the randomness of n.

Other regression investigations can be found in Gross and Huber-Carol (1992), Gross and Lai (1996) and Kim and Lai (1999). Under the model (Eq. 11), He and Yang (2003) construted a weighted least square estimate $\hat{\delta}_n$ for the regression parameter δ, where the weights are random quantities depending on G_n. They established the strong consistency and asymptotic normality of the estimators. We point out that their scheme is slightly different than that of Woodroofe. Finally, Park (2004) gives the optimal rate of convergence of a B-spline regression estimator in the truncated and censored model.

The purpose of this paper is to give the asymptotic behavior of the estimator $m_n(x)$. As we are interested in the number n of observations (N is unknown), we give (unless otherwise specified) asymptotics as $n \to \infty$. Since $n \leq N$, these results also hold for $N \longrightarrow \infty$, $\mathbb{P}$—almost everywhere.

Throughout the paper $(\mathbf{X} \leq x)$ stands for $(X_1 \leq x_1, \ldots, X_s \leq x_s)$. The paper is organized as follows. In Sect. 2 we construct an estimator for the covariate's density. The nonparametric regression estimator is defined in Sect. 3. In Sect. 4 we give the assumptions and the main results. Some simulations are given in Sect. 5. Finally, the proofs are given in Sect. 6.

2 Estimation of covariate's density under random left truncation

As we did not find any literature about the estimation of the covariable distribution under random left truncation, we build here estimators of V and v. Note that we can no longer use the kernel estimator v_N defined in (Eq. 4) since only n observations are made. On the other hand,

$$v_n^*(x) = \frac{1}{nh_n^s} \sum_{i=1}^{n} K\left(\frac{x - \mathbf{X}_i}{h_n}\right)$$

is an estimator of the conditional density $v^*(x)$ (given $Y \geq T$). To overcome this difficulty, we first consider the conditional joint distribution of $(\mathbf{X}, Y, T)$

$$\begin{aligned} H^*(x, y, t) &= \mathbb{P}\left(\mathbf{X} \leq x, Y \leq y,\ T \leq t \mid Y \geq T\right) \\ &= \frac{1}{\alpha} \int_{u \leq x} \int_{a_G \leq w \leq y} G(w \wedge t)\, \mathbf{F}(\mathrm{d}u, \mathrm{d}w). \end{aligned}$$

Taking $t = +\infty$, we get the conditional joint df of $(\mathbf{X}, Y)$

$$\mathbf{F}^*(x, y) = \frac{1}{\alpha} \int_{u \leq x} \int_{a_G \leq w \leq y} G(w)\, \mathbf{F}(\mathrm{d}u, \mathrm{d}w) \tag{12}$$

which by differentiating gives

$$\mathbf{F}(\mathrm{d}x, \mathrm{d}y) = \frac{\alpha}{G(y)} \mathbf{F}^*(\mathrm{d}x, \mathrm{d}y) \ \text{ for } y > a_G.$$

Integrating over y we get the df of $\mathbf{X}$

$$V(x) = \alpha \int_{u \leq x} \int_{y \geq a_G} \frac{1}{G(y)}\, \mathbf{F}^*(\mathrm{d}u, \mathrm{d}y).$$

A natural estimator of V is then given by

$$V_n(x) = \frac{\alpha_n}{n} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)}\, \mathbb{I}_{\{\mathbf{X}_i \leq x\}}. \tag{13}$$

Note that in Eq. (13) and the forthcoming formulae, the sum is taken only for i such that $G_n(Y_i) \neq 0$.

Finally, Eq. (13) yields the density estimator

$$
\begin{aligned}
v_n(x) &:= \frac{1}{h_n^s}\int K\left(\frac{x-u}{h_n}\right)\mathrm{d}V_n(u)\\
&= \frac{\alpha_n}{nh_n^s}\sum_{i=1}^{n}\frac{1}{G_n(Y_i)}K\left(\frac{x-\mathbf{X}_i}{h_n}\right). \qquad (14)
\end{aligned}
$$

Note here that, without possible confusion, we use the same notation for the density estimator as in the case of complete data (v_N).

3 Construction of the new estimator of $m(\cdot)$

Throughout this paper, we assume that

$$
0 = a_G < a_F \qquad \text{and} \qquad b_G \le b_F. \qquad (15)
$$

$$
T \text{ and } (\mathbf{X}, Y) \text{ are independent.} \qquad (16)
$$

Remark 1 In view of Eq. (15) many authors (see Stute 1993) used the milder condition $a_G \le a_F$ with additional integrability conditions. As we need to prove some uniform results which imply a sufficient rate of convergence of G_n (see Lemma 6.3) we have to consider a set of values of Y_i which do not include a_G [a uniform rate for G_n is given in Woodroofe (1985) on $[a, b_G]$ with $a > a_G$] that is $a_F > a_G$.

Let $\xi : \mathbb{R}^s \times \mathbb{R} \longrightarrow \mathbb{R}$ be a measurable function and consider the problem of estimating the mean $\mathbb{E}[\xi(\mathbf{X}, Y)]$ from the sample $(\mathbf{X}_i, Y_i, T_i)$, $i = 1, \ldots, N$. Since only n observations (among N) are made, we try to construct an estimator of $\mathbb{E}[\xi(\mathbf{X}, Y)]$ which is calculated only if $Y \ge T$. Following the idea introduced by Carbonez, Györfi and van der Meulin (1995), one such unbiased estimate in the random-truncated model is given by

$$
\frac{1}{N}\sum_{i=1}^{N}\frac{\xi(\mathbf{X}_i, Y_i)}{G(Y_i)}\mathbb{I}_{\{Y_i \ge T_i\}}. \qquad (17)
$$

Indeed, using the properties of conditional expectation and Eq. (16) , we have

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\xi(\mathbf{X}_i, Y_i)}{G(Y_i)}\mathbb{I}_{\{Y_i \ge T_i\}}\right] &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\mathbb{E}\left[\frac{\xi(\mathbf{X}_i, Y_i)}{G(Y_i)}\mathbb{I}_{\{Y_i \ge T_i\}}|\mathbf{X}_i, Y_i\right]\right]\\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\frac{\xi(\mathbf{X}_i, Y_i)}{G(Y_i)}\mathbb{E}\left[\mathbb{I}_{\{Y_i \ge T_i\}}|\mathbf{X}_i, Y_i\right]\right]\\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\xi(\mathbf{X}_i, Y_i)\right]\\
&= \mathbb{E}\left[\xi(\mathbf{X}, Y)\right].
\end{aligned}
$$

In the particular case where $\xi(\mathbf{X}, Y) = (Y|\mathbf{X})$, we obtain an estimator of Eq. (1). Unfortunately, N and G being unknown, Eq. (17) cannot be used in practice and has to be adapted.

First $G(\cdot)$ is consistently estimated by $G_n(\cdot)$ and N by n/α_n. However, it is not possible to use the Nadaraya–Watson weights defined in Eq. (4) in order to estimate $\mathbb{E}[\xi(\cdot, \cdot)]$. Indeed, if we consider the weight $W_{i,n/\alpha_n}$, it is not possible to calculate its denominator since only n values of $\mathbf{X}_i$ (among N) are observed. On the other hand, if we take the weights

$$W_{i,n}(x) = \frac{K\{(x - \mathbf{X}_i)/h_n\}}{\sum_{j=1}^{n} K\left\{(x - \mathbf{X}_j)/h_n\right\}}$$

which are calculated from the observed sample, the estimated regression function will be $m^*(\cdot)$ defined in Eq. (10) which is not our purpose. Using Eq. (14) we adapt Eq. (4) to get the new weights

$$\overline{W}_{i,n}(x) = \frac{G_n^{-1}(Y_i)K\{(x - \mathbf{X}_i)/h_n\}}{\sum_{j=1}^{n} G_n^{-1}(Y_j)K\left\{(x - \mathbf{X}_j)/h_n\right\}}.$$

Since these weights lead to a complicated estimator of ψ (see Remark 2 below), they are modified into

$$\widetilde{W}_{i,n}(x) = \frac{\alpha_n^{-1}K\{(x - \mathbf{X}_i)/h_n\}}{\sum_{j=1}^{n} G_n^{-1}(Y_j)K\left\{(x - \mathbf{X}_j)/h_n\right\}} \tag{18}$$

in order to get a simpler and more natural estimator ψ_n of ψ [see (20) below] and therefore for m. This is motivated by the fact that both α_n^{-1} and $n^{-1}\sum_i G_n^{-1}(Y_i)$ are consistent estimators of α^{-1}. Note also that even if $\sum_i \widetilde{W}_{in} \neq 1$, the equality holds asymptotically, $\mathbf{P} - a.s.$

Now the estimator Eq. (17) can be written as:

$$\frac{1}{n} \times \frac{n}{N} \sum_{i=1}^{n} \frac{\xi(\mathbf{X}_i, Y_i)}{G(Y_i)},$$

which yields the new estimator of $m(x)$ by replacing the uniform weights $1/n$ by $\widetilde{W}_{i,n}(x)$. We get

$$\begin{aligned} m_n(x) &:= \alpha_n \sum_{i=1}^{n} \widetilde{W}_{i,n}(x) \frac{Y_i}{G_n(Y_i)} \\ &= \frac{\sum_{i=1}^{n} Y_i G_n^{-1}(Y_i)K\{(x - \mathbf{X}_i)/h_n\}}{\sum_{i=1}^{n} G_n^{-1}(Y_i)K\{(x - \mathbf{X}_i)/h_n\}} \\ &= \frac{\psi_n(x)}{v_n(x)} \end{aligned} \tag{19}$$

where

$$\psi_n(x) := \frac{\alpha_n}{nh_n^s} \sum_{i=1}^{n} \frac{Y_i}{G_n(Y_i)} K\left(\frac{x - \mathbf{X}_i}{h_n}\right) \tag{20}$$

and $v_n(x)$ is defined in Eq. (14). The estimator $m_n(\cdot)$ can be used in practice. Note here that $\psi_n(x)$ is an estimate of $\psi(x)$ defined in Eq. (2) and can be written as:

$$\psi_n(x) = \frac{1}{h_n^s} \int K\left(\frac{x-u}{h_n}\right) \mathrm{d}\Psi_n(u)$$

where

$$\Psi_n(x) = \frac{\alpha_n}{n} \sum_{i=1}^{n} \frac{Y_i}{G_n(Y_i)} \mathbb{I}_{\{\mathbf{X}_i \le x\}} \tag{21}$$

is an estimator of

$$\Psi(x) = \int_{u \le x} \psi(u)\, \mathrm{d}u. \tag{22}$$

Remark 2 Using the weights $\overline{W}_{in}$ instead of $\widetilde{W}_{in}$ yields the estimator

$$\overline{\psi}_n(x) = \frac{\alpha_n^2}{n h_n^s} \sum_{i=1}^{n} \frac{Y_i}{G_n^2(Y_i)} K\left(\frac{x - \mathbf{X}_i}{h_n}\right)$$

which is not a natural estimator of ψ. Moreover, the estimator α_n is not needed in the formula defining m_n, whereas it is needed if we choose the weights $\overline{W}_{in}$.

4 Assumptions and main results

In what follows, we focus our attention on the case of a univariate covariable ($s = 1$) and denote X for $\mathbf{X}$. Define $\Omega_0 = \{x \in \mathbb{R} \,|\, v(x) > 0\}$ and let $\Omega \subset \Omega_0$ be a compact set with $\eta = \inf_{x \in \Omega} v(x) > 0$.

We will make use of the following assumptions gathered here for easy reference:

(A1) There exists $\beta > 5/2$ such that

$$\int\int y^{\beta} \mathbf{F}(\mathrm{d}x,\, \mathrm{d}y) \ < \infty.$$

(A2) The bandwidth h_n satisfies

$$\frac{n^{1-2\gamma} h_n^2}{\log n} \longrightarrow \infty \qquad \text{as} \quad n \to \infty$$

for some $\gamma \in [0, \frac{3}{10})$ such that $\gamma > \dfrac{2}{\beta} - \dfrac{1}{2}$.

(A3) The kernel K is a $\mathcal{C}^1$-probability density with compact support.

(A4) v and ψ are locally Lipschitz continuous over Ω_0.

(B1) The joint density $\mathbf{f}(\cdot, \cdot)$ is twice differentiable with respect to the first variable. Moreover, $\partial \mathbf{f}/\partial x(\cdot, \cdot)$ and $\partial^2 \mathbf{f}/\partial x^2(\cdot, \cdot)$ are continuous over $\Omega_0 \times \mathbb{R}^+$.

(B2) The bandwidth h_n satisfies

$$n h_n^5 \longrightarrow 0 \qquad \text{as} \quad n \to \infty.$$

(B3) The kernel K satisfies

$$\int r K(r)\,\mathrm{d}r = 0.$$

Remark 3 If (A1) is satisfied for a $\beta > 4$ then we can take $\gamma = 0$ in (A2) which can then be stated without reference to γ. In that case we achieve the best possible rate in Theorem 4.1. More generally, hypotheses (A1)–(A3) are common in the nonparametric regression estimation, whereas (A4) is needed to obtain a rate of convergence in Theorem 4.1 (and Lemmas 6.4 and 6.6 below). (B1)–(B3) are additional assumptions to get asymptotic normality. Remark that under (B1), (A4) is also satisfied. Note finally that Theorem 4.1 is still valid with the milder conditions $\beta > 2$ (in A1) and $\gamma \in [0, 1/2)$ (in A2).

Now our first result deals with the pointwise and uniform strong consistency of $m_n(\cdot)$. The latter is derived with a rate over a fixed compact set.

Theorem 4.1 *Under assumptions* (A1)–(A3)*, for any x such that $v(x) > 0$, we have, as $n \to \infty$,*

$$m_n(x) \longrightarrow m(x), \quad \mathbf{P} - a.s.$$

Moreover, if (A4) *is satisfied then*

$$\sup_{x\in\Omega} |m_n(x) - m(x)| = O\left(\max\left\{\sqrt{\frac{\log n}{n^{1-2\gamma}h_n^2}}, h_n\right\}\right), \quad \mathbf{P} - a.s. \tag{23}$$

Remark 4 In the case where (A4) is not satisfied, it is still possible to get the first part of Theorem 4.1 (and Lemma 6.4). In that case, the rate in Eq. (23) (and Lemma 6.4) is modified according to the continuity moduli of v and ψ.

Our next result states the pointwise asymptotic normality of $m_n(\cdot)$. Let

$$\Sigma(x) = \begin{pmatrix} \Sigma_0(x) & \Sigma_1(x) \\ \Sigma_1(x) & \Sigma_2(x) \end{pmatrix}$$

with

$$\Sigma_j(x) = \int \frac{y^{2-j}\mathbf{f}(y,x)}{G(y)}\,\mathrm{d}y, \qquad \text{for } j = 0, 1, 2.$$

Then follows the Theorem 4.2.

Theorem 4.2 *Under assumptions* (A1)–(A3) *and* (B1)–(B3)*, for any x such that $v(x) > 0$ and*

$$\int \sup_{u\in\mathcal{U}(x)} \left| y \frac{\partial^j \mathbf{f}}{\partial x^j}(u, y)\right| \mathrm{d}y < +\infty, \qquad j = 1, 2 \tag{24}$$

for a neighborhood $\mathcal{U}(x)$ of x, we have

$$\sqrt{nh_n}\,(m_n(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(x)\right) \tag{25}$$

where $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution and*

$$\sigma^2(x) = \frac{\kappa\left[\Sigma_0(x)v^2(x) + \Sigma_2(x)\psi^2(x) - 2v(x)\psi(x)\Sigma_1(x)\right]}{\alpha v^4(x)}$$

with

$$\kappa = \int K^2(r)\,\mathrm{d}r.$$

Remark 5 Note on the one hand that $\Sigma_2(x) \geq v(x)$ and on the other hand, by Cauchy–Schwartz inequality, we have $\Sigma_1^2(x) < \Sigma_0(x)\Sigma_2(x)$ (the equality does not hold since the mapping $y \mapsto y$ is not constant). Therefore, $\Sigma(x)$ is positive definite as soon as $v(x) > 0$.

Remark 6 Under Eq. (24) ψ is twice differentiable on a neighborhood $\mathcal{U}'(x)$ ($\subset \mathcal{U}(x)$) of x, by the dominated convergence theorem. Moreover, ψ'' is continuous on $\mathcal{U}'(x)$ and we have

$$\psi''(x) = \int y \frac{\partial^2 \mathbf{f}}{\partial x^2}(x, y)\,\mathrm{d}y.$$

Remark 7 A plug-in-type estimate $\hat{\sigma}^2(x)$ for the asymptotic variance $\sigma^2(x)$ can easily be obtained using Eqs. (14), Eq. (20) and the estimators

$$\hat{\Sigma}_j(x) = \frac{1}{nh_n}\sum_{i=1}^{n} \frac{Y_i^{2-j}}{G_n^2(Y_i)} K\left(\frac{x - X_i}{h_n}\right)$$

of $\alpha^{-1}\Sigma_j(x)$, $j = 0, 1, 2$, respectively. This yields a confidence interval of asymptotic level $1 - \zeta$ for $m(x)$

$$\left[m_n(x) - \frac{u_{1-\zeta/2}\hat{\sigma}(x)}{\sqrt{nh_n}}, m_n(x) + \frac{u_{1-\zeta/2}\hat{\sigma}(x)}{\sqrt{nh_n}}\right]$$

where $u_{1-\zeta/2}$ denotes the $1 - \zeta/2$ quantile of the standard normal distribution.

5 Simulations

The first subsection deals with the consistency (Theorem 4.1), whereas the second looks at the asymptotic normality (Theorem 4.2).

5.1 Consistency

The aim of the following simulations is to study some particular regression functions $m(\cdot)$. First we consider a linear regression function with the model $Y_i = \delta_0 + \delta_1 X_i + \varepsilon_i$, $i = 1, \ldots, N$, where X_i and ε_i are two independent iid sequences distributed as $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, 0.2)$, respectively. We also simulate N iid rv $T_i \sim \mathcal{N}(\mu, 1)$ where μ is adapted in order to get different values of α. We then keep the data $(X_i, Y_i),\ i = 1, \dots, n$ such that $Y_i \geq T_i$. We do it in a way to obtain a given n (which means that in this case n is not random, whereas N is).

Then we compute our estimator with the observed data $(Y_i, T_i, X_i),\ i = 1, \dots, n$. We choose a Gaussian kernel and it is well known that, in nonparametric estimation, optimality (in the MSE sense) is not seriously swayed by the choice of the kernel (K) but is affected by the choice of the bandwidth h_n. We notice that the quality of fit increases with n (see Fig. 1). In all cases, we take either $h_n = 0.01$ or $h_n = 0.05$ (higher values give bad estimators).

We then try to see if the quality depends on the trunction proportion α. We take $n = 500$ and choose different values of $\alpha :\approx 25, 50$ and 100% (the latter is the case of complete data). The estimator's quality seems to be less affected by α as shown in Fig. 2 than it is by n (though higher values of N are needed for small α to achieve $n = 500$).

Finally, we consider the case of nonlinear m. Three models are studied:

$$\begin{aligned}
Y_i &= (X_i - 1)^2 + 0.5 + \varepsilon_i, && \text{parabolic case}\\
Y_i &= \arccos X_i + \varepsilon_i, && \text{inverse cosine case}\\
Y_i &= \sin X_i + \varepsilon_i && \text{sinus case.}
\end{aligned}$$

For each nonlinear model we find a quality of fit as good as for the linear case (see Figs. 3, 4).

5.2 Asymptotic normality

We now consider the problem of asymptotic normality. We show how good the normality is when dealing with samples of finite size which is the case in practice.

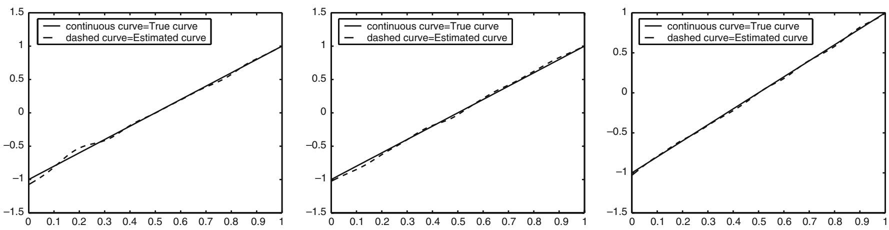

Fig. 1 Linear function $m(x) = 2x - 1$ with $\alpha \approx 30\%$ and $n = 100, 500$ and $1,000$, respectively

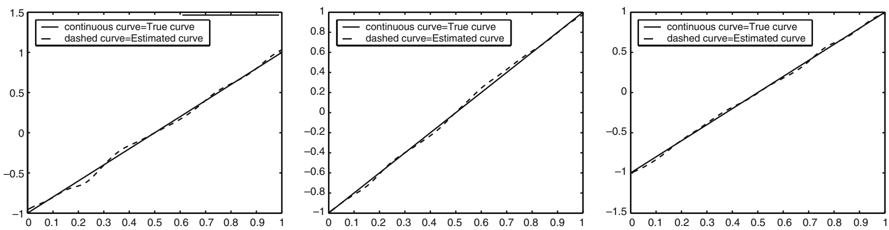

Fig. 2 Linear function $m(x) = 2x - 1$ with $n = 500$ and $\alpha \approx 25, 50$ and 100%, respectively

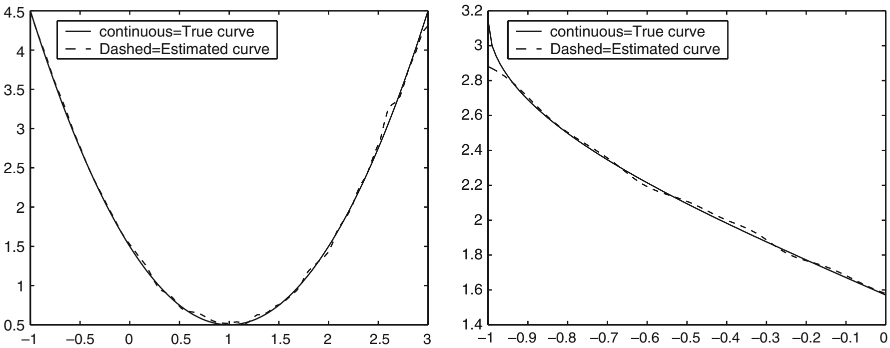

Fig. 3 Parabolic and inverse cosine functions with $\alpha \approx 30\%$ and $n = 500$

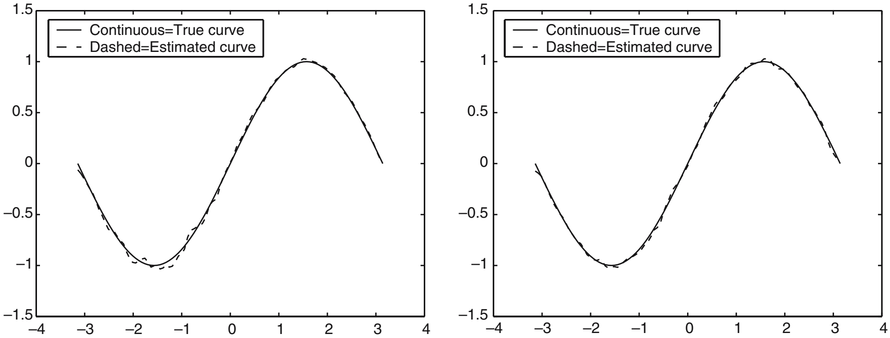

Fig. 4 Sinus function with $n = 500$ and $\alpha \approx 30$ and 90%, respectively

We compare the shape of the estimated density (of normalized deviations) to that of the standard normal density in the case of the linear regression model

$$Y_i = X_i + 2 + \varepsilon_i, \quad i = 1, \dots, n. \tag{26}$$

The data arise from the same distributions as previously. For a given sample size n, we estimate the regression function as before and calculate the normalized deviation between this estimate and the theoretical regression function given in Eq. (26) for $x = 0$ (i.e., $\overline{m}_n = \hat{\sigma}^{-1}(0)\sqrt{nh_n}\,\{m_n(0) - 2\}$). We draw, using this scheme, B independent n-samples. The bandwidth h_n is chosen according to hypotheses (A2) and (B2). In order to estimate the density function of $\overline{m}_n$ (by the kernel method), we make the classical bandwidth choice (see, e.g., Silverman 1986, p. 40) $h'_n = \mathcal{C}\, n^{-1/5}$, where the constant $\mathcal{C}$ is appropriately chosen.

We consider different values of n and B and give only the plots for $n = 200,\ 500$ and $B = 300$. In each case, we also give a histogram and the Q–Q-plot against a normal distribution. For $n = 200$ the results are considerably good but show a bit of skewness (Fig. 5). Moreover, a Shapiro–Wilk normality test suggests a small departure from normality (P value≈ 0.064). The results are more convincing with $n = 500$ (see Fig. 6). In that case the P value of the Shapiro–Wilk test is around 0.771.

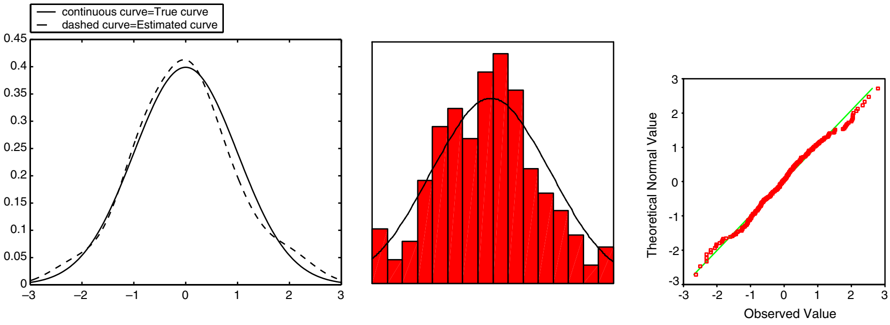

Fig. 5 $n = 200$, $B = 300$

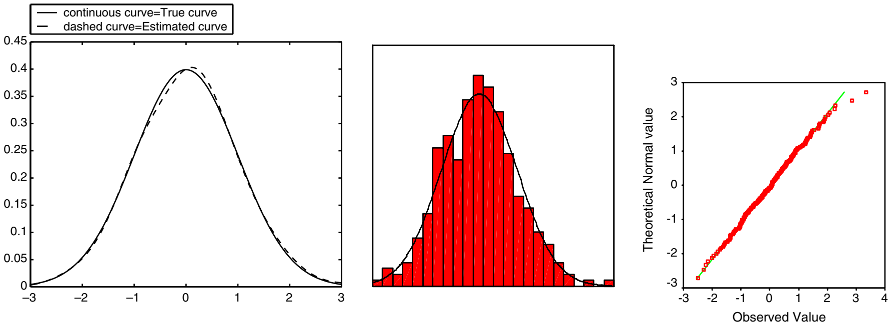

Fig. 6 $n = 500$, $B = 300$

6 Proofs

In order to prove Theorem 4.1, we first consider the estimators

$$\widetilde{V}_n(x) = \frac{\alpha}{n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} \mathbb{1}_{\{X_i \le x\}} \tag{27}$$

and

$$\widetilde{\Psi}_n(x) = \frac{\alpha}{n} \sum_{i=1}^{n} \frac{Y_i}{G(Y_i)} \mathbb{1}_{\{X_i \le x\}} \tag{28}$$

of $V(x)$ and $\Psi(x)$, respectively. Using Eqs. (12), (2) and Eq. (22) it can be shown that these are **E**-unbiased estimators (which cannot be calculated). Moreover, we have the following two lemmas.

Lemma 6.1 *As $n \longrightarrow +\infty$, we have*

$$\widetilde{V}_n(x) \longrightarrow V(x), \quad \mathbf{P} - a.s.$$

Furthermore, for any compact set $\Xi \subset \Omega_0$, we have

$$\sup_{x \in \Xi} \left| \widetilde{V}_n(x) - V(x) \right| = O\left[\left(\frac{\log n}{n} \right)^{1/2} \right] \ \mathbf{P} - a.s. \ \ as\ n \longrightarrow \infty.$$

Proof The first part of the Lemma 6.1 is a consequence of the strong law of large numbers (SLLN). For the second part, let us consider the iid sequence $(X_1, Y_1), \ldots, (X_n, Y_n)$ and define

$$\vartheta_n = \left\{\theta_x : \quad \Xi \times \mathbb{R}^+ \longrightarrow \mathbb{R}^+ / \theta_x(u, y) = \frac{\mathbb{I}_{\{u \le x\}}}{nG(y)}, \quad x \in \Xi\right\}.$$

By Lemma (3b) in Giné and Guillou (1999), ϑ_n is Vapnik–Červonenkis (V–C) class of nonnegative measurable functions which are uniformly bounded with respect to the envelope $\Theta = [nG(a_F)]^{-1}$. Moreover,

$$\mathbf{E}\left[\theta_x(X, Y)\right] = \mathbf{E}\left[\frac{\mathbb{I}_{\{X \le x\}}}{nG(Y)}\right] \le \frac{1}{nG(a_F)} =: U_n$$

and

$$\mathbf{E}\left[\theta_x^2(X, Y)\right] = \mathbf{E}\left[\left(\frac{\mathbb{I}_{\{X \le x\}}}{nG(Y)}\right)^2\right] \le \frac{1}{n^2 G^2(a_F)} =: \sigma_n^2.$$

While applying Talagrand's (1994) inequality [see Proposition 2.2 in Giné and Guillou (2001)] with $\lambda\sqrt{n^{-1}\log n}$, where D is a positive constant, we have

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\theta_x \in \vartheta_n}\left|\sum_{i=1}^n \{\theta_x(X_i, Y_i) - \mathbf{E}[\theta_x(X, Y)]\}\right| \ge D\sqrt{\frac{\log n}{n}}\right\} \\ &\le B_1 \exp\left\{-\frac{1}{B_1} D\sqrt{\frac{\log n}{n}} \times \; nG(a_F)\log\left[1 + \frac{\frac{D\sqrt{\log n}}{n^{3/2}G(a_F)}}{B_1\left(\frac{1}{n^{1/2}G(a_F)} + \frac{\sqrt{\log B_2}}{nG(a_F)}\right)^2}\right]\right\} \end{aligned} \tag{29}$$

where B_1 and B_2 are positive constants. For n large enough, the right-hand side of Eq. (29) becomes an order of $B_1 n^{-(\frac{DG(a_F)}{B_1})^2}$, which by an appropriate choice of the constant D can be made $O\left(n^{-3/2}\right)$ — which in turn is the general term of summable series. Hence, by Borel Cantelli's lemma, we have the second part of the result. □

Lemma 6.2 *Under* (A1)*, as $n \longrightarrow +\infty$, we have*

$$\widetilde{\Psi}_n(x) \longrightarrow \Psi(x), \quad \mathbf{P} - a.s.$$

Furthermore, for any compact set $\Xi \subset \Omega_0$

$$\sup_{x \in \Xi}\left|\widetilde{\Psi}_n(x) - \Psi(x)\right| = O\left[\left(\frac{\log n}{n^{1-2\gamma}}\right)^{1/2}\right] \quad \mathbf{P} - a.s. \;\; as\; n \longrightarrow \infty,$$

for γ as in (A2).

Proof As for Lemma 6.1, the first part is a consequence of the SLLN. For the second part and in the case of a bounded rv Y, the proof is similar to Lemma 6.1. In order to deal with the unbounded case ($b_F = +\infty$), we use a truncation method by a suitable sequence which diverges to infinity. Let

$$\widetilde{\Psi}_n(x) = \widetilde{\Psi}_{1n}(x) + \widetilde{\Psi}_{2n}(x)$$

with

$$\widetilde{\Psi}_{1n}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{G(Y_i)} \mathbb{I}_{\{X_i \le x,\ Y_i \le n^\gamma\}}$$

and $\widetilde{\Psi}_{2n}(x) = \widetilde{\Psi}_n(x) - \widetilde{\Psi}_{1n}(x)$.

It can be shown that $\widetilde{\Psi}_{1n}(x)$ and $\widetilde{\Psi}_{2n}(x)$ are $\mathbf{E}$-unbiased estimators of $\Psi_1(x) = \int_{u\le x}\int y\,\mathbb{I}_{\{y\le n^\gamma\}}\mathbf{F}(\mathrm{d}u,\ \mathrm{d}y)$ and $\Psi_2(x) = \Psi(x) - \Psi_1(x)$, respectively.

Now, we have

$$\begin{aligned}\sup_{x\in\Xi}|\widetilde{\Psi}_n(x) - \Psi(x)| &\le \sup_{x\in\Xi}|\widetilde{\Psi}_{1n}(x) - \Psi_1(x)| + \sup_{x\in\Xi}|\widetilde{\Psi}_{2n}(x) - \Psi_2(x)|\\ &:= I_n + II_n.\end{aligned}$$

I_n can be dealt analogously as in Lemma 6.1 by considering the V–C class

$$\Phi_n = \left\{\varphi_x\colon\ \Xi\times\mathbb{R}^+ \longrightarrow \mathbb{R}^+ \ /\ \varphi_x(u, y)\,\mathbb{I}_{\{u\le x;\ y\le n^\gamma\}}\frac{y}{nG(y)},\quad x\in\Xi\right\},$$

with envelope $\phi = \left[n^{1-\gamma}G(a_F)\right]^{-1}$. Furthermore, the first and second moments of the φ_x are bounded by

$$U_n = \frac{1}{n^{1-2\gamma}G(a_F)} \qquad \text{and} \qquad \sigma_n^2 = \frac{1}{n^{2-4\gamma}G^2(a_F)},$$

respectively. We then get

$$I_n = O\left[\left(\frac{\log n}{n^{1-2\gamma}}\right)^{1/2}\right]\ \mathbf{P}-\text{a.s. as } n\longrightarrow\infty.$$

For II_n, by Markov's inequality, we have

$$\mathbf{P}\left[II_n > n^{-(\frac{1}{2}-\gamma)}\right] \le \frac{\mathbf{E}[II_n^\beta]}{n^{-\beta(\frac{1}{2}-\gamma)}}. \tag{30}$$

By (A1) and Minkowski's inequality, the right-hand side of Eq. (30) can be bounded by $\dfrac{1}{n^{\beta(\frac{1}{2}+\gamma)-1}}$.

Since $\gamma > \dfrac{2}{\beta} - \dfrac{1}{2}$, the Borel–Cantelli lemma implies that

$$II_n = O\left[\left(\frac{1}{n^{1-2\gamma}}\right)^{1/2}\right]\ \mathbf{P}-\text{a.s. as } n\longrightarrow\infty.$$

□

The next two lemmas deal with the strong uniform consistency of $V_n(\cdot)$ and $v_n(\cdot)$ with rates.

Lemma 6.3 *As $n \to \infty$, we have for any x,*

$$V_n(x) \longrightarrow V(x), \quad \mathbf{P} - a.s.$$

Moreover, this convergence is achieved with a $O_{\mathbf{P}}\left((\frac{\log n}{n})^{1/2}\right)$ rate and is uniform on any compact set $\Xi \subset \Omega_0$.

Proof of Lemma 6.3. We have

$$\begin{aligned}
|V_n(x) - V(x)| &\leq \left|V_n(x) - \widetilde{V}_n(x)\right| + \left|\widetilde{V}_n(x) - V(x)\right| \\
&\leq \frac{|\alpha_n - \alpha|}{n} \sum_{j=1}^{n} \frac{1}{G_n(Y_j)} \mathbb{I}_{\{X_j \leq x\}} \\
&\quad + \frac{\alpha}{n} \sum_{j=1}^{n} \left| \frac{1}{G_n(Y_j)} - \frac{1}{G(Y_j)} \right| \mathbb{I}_{\{X_j \leq x\}} + \left|\widetilde{V}_n(x) - V(x)\right| \\
&\leq \frac{|\alpha_n - \alpha|}{G_n(a_F)} + \frac{\alpha \sup_{y \geq a_F} |G_n(y) - G(y)|}{G_n(a_F) G(a_F)} + \left|\widetilde{V}_n(x) - V(x)\right| \\
&=: (I) + (II) + (III).
\end{aligned} \tag{31}$$

Since $|\alpha_n - \alpha| = O_{\mathbf{P}}(n^{-1/2})$ [from Theorem 3.2 in He and Yang (1998)] and $G_n(a_F) \stackrel{\mathbf{P}-a.s.}{\longrightarrow} G(a_F) > 0$ we get, uniformly on x, $(I) = O_{\mathbf{P}}(n^{-1/2})$.

In the same way, and using Remark 6 in Woodroofe (1985) we get $(II) = O_{\mathbf{P}}(n^{-1/2})$. Finally, using Lemma 6.1 we get the result. □

Remark 8 Recall that, as noted in Eq. (13), all the sums involving the $G_n^{-1}(Y_j)$ are taken for the j such that $G_n(Y_j) \neq 0$. It follows that an additional term

$$(IV) = \frac{\alpha}{n} \sum_{j=1}^{n} \frac{1}{G(Y_j)} \mathbb{I}_{\{X_j \leq x\}} \mathbb{I}_{\{G_n(Y_j)=0\}}$$

should be added in Eq. (31). This term is clearly negligible by the law of large numbers. The same remark can be made for all similar quantities in what follows.

Lemma 6.4 *Under assumptions* (A2)–(A4) *we have, as $n \to \infty$,*

$$v_n(x) \longrightarrow v(x), \quad \mathbf{P} - a.s.$$

Moreover, we have for any compact set $\Xi \subset \Omega_0$

$$\sup_{x \in \Xi} |v_n(x) - v(x)| = O\left(\max\left\{\sqrt{\frac{\log n}{nh_n^2}};\ h_n\right\}\right), \quad \mathbf{P} - a.s.$$

Proof of Lemma 6.4

$$\begin{aligned}
v_n(x) - v(x) &= \left\{ \frac{1}{h_n} \int K\left(\frac{x-u}{h_n}\right) dV_n(u) - \frac{1}{h_n} \int K\left(\frac{x-u}{h_n}\right) dV(u) \right\} \\
&\quad + \left\{ \frac{1}{h_n} \int K\left(\frac{x-u}{h_n}\right) v(u)\, du - v(x) \right\} \\
&=: \mathcal{I}_1 + \mathcal{I}_2.
\end{aligned}$$

While integrating by parts, we have

$$\begin{aligned}
|\mathcal{I}_1| &\leq \frac{1}{h_n^2} \int |V_n(u) - V(u)| \left| K'\left(\frac{x-u}{h_n}\right) \right| \mathrm{d}u \\
&\leq \sup_u |V_n(u) - V(u)| \frac{1}{h_n} \int \left| K'(w) \right| \mathrm{d}w
\end{aligned}$$

which by Lemma 6.3 and (A3) is $O_{\mathbf{P}}\left(\sqrt{\frac{\log n}{nh_n^2}}\right)$ for any x. Moreover, this rate is uniform on any compact set $\Xi \subset \Omega_0$.

On the other hand, by (A3) and (A4) the bias term $\mathcal{I}_2$ is $O(h_n)$, which yields the result. □

Finally, we state consistency results with rates for $\Psi_n(\cdot)$ and $\psi_n(\cdot)$.

Lemma 6.5 *Under* (A1)*, we have for any x, as $n \to \infty$,*

$$\Psi_n(x) \longrightarrow \Psi(x), \quad \mathbf{P} - a.s.$$

Moreover, this convergence is achieved with a $O_{\mathbf{P}}\left((\frac{\log n}{n^{1-2\gamma}})^{1/2}\right)$ rate [where γ is as in (A2)*] and is uniform on any compact set $\Xi \subset \Omega_0$.*

Proof of Lemma 6.5 Analogous to Lemma 6.3, we have

$$\begin{aligned}
|\Psi_n(x) - \Psi(x)| &\leq \left[\frac{|\alpha_n - \alpha|}{G_n(a_F)} + \frac{\alpha \sup_{y \geq a_F} |G_n(y) - G(y)|}{G_n(y)G(y)} \right] \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{I}_{\{X_i \leq x\}} \\
&\quad + \left| \widetilde{\Psi}_n(x) - \Psi(x) \right|.
\end{aligned}$$

Note that, by the SLLN, as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{I}_{\{X_i \leq x\}} \longrightarrow \mathbf{E}\left[Y \mathbb{I}_{\{X \leq x\}} \right] = \frac{1}{\alpha} \int\int y G(y) \mathbb{I}_{\{u \leq x\}} \mathbf{F}(\mathrm{d}u, \mathrm{d}y),$$

the latter being finite, under (A1). Then, similar arguments to Lemma 6.3 and using Lemma 6.2 give the result. □

Lemma 6.6 *Under assumptions* (A1)–(A4) *we have for any x, as $n \to \infty$,*

$$\psi_n(x) \longrightarrow \psi(x), \quad \mathbf{P} - a.s.$$

Moreover, we have for any compact set $\Xi \subset \Omega_0$

$$\sup_{x \in \Xi} |\psi_n(x) - \psi(x)| = O\left(\max\left\{ \sqrt{\frac{\log n}{n^{1-2\gamma} h_n^2}};\ h_n \right\} \right), \quad \mathbf{P} - a.s.$$

Proof of Lemma 6.6 Analogous to the proof of Lemma 6.4, therefore, it is omitted. □

Proof of Theorem 4.1 Since

$$m_n(x) - m(x) = \frac{\psi_n(x) - \psi(x)}{v_n(x)} + m(x)\frac{v(x) - v_n(x)}{v_n(x)}, \tag{32}$$

the first parts of Lemmas 6.4 and 6.6 give the consistency result.

Now, from Eq. (32), we have $\mathbf{P} - a.s.$ and for n large enough

$$\sup_{x\in\Omega} |m_n(x) - m(x)| \le \frac{1}{\eta - \sup_{x\in\Omega} |v_n(x) - v(x)|} \left\{ \sup_{x\in\Omega} |\psi_n(x) - \psi(x)| \right.$$
$$\left. + \eta^{-1} \sup_{x\in\Omega} |\psi(x)| \sup_{x\in\Omega} |v_n(x) - v(x)| \right\},$$

which by the second parts of Lemmas 6.4 and 6.6 gives the rate of convergence. □

Now in order to prove Theorem 4.2, we write (using Eq. (19))

$$m_n(x) = \frac{\alpha_n^{-1}\psi_n(x)}{\alpha_n^{-1}v_n(x)}$$

where, from Eq. (20) and Eq. (14)

$$\frac{\psi_n(x)}{\alpha_n} = (nh_n)^{-1} \sum_{i=1}^{n} Y_i G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) \text{ and } \frac{v_n(x)}{\alpha_n}$$
$$= (nh_n)^{-1} \sum_{i=1}^{n} G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right).$$

Then we have

$$\begin{aligned}
\frac{\psi_n(x)}{\alpha_n} - \frac{\psi(x)}{\alpha} &= (nh_n)^{-1} \sum_{i=1}^{n} Y_i G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) \\
&\quad - (nh_n)^{-1} \sum_{i=1}^{n} Y_i G^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) \\
&\quad + (nh_n)^{-1} \sum_{i=1}^{n} Y_i G^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) \\
&\quad - \mathbb{E}\left[h_n^{-1} Y_1 G^{-1}(Y_1) K\left(\frac{x - X_1}{h_n}\right)\right] \\
&\quad + h_n^{-1}\mathbb{E}\left\{Y_1 G^{-1}(Y_1) K\left(\frac{x - X_1}{h_n}\right)\right\} - \frac{\psi(x)}{\alpha} \\
&=: \Lambda_{n1}(x) + \Lambda_{n2}(x) + \Lambda_{n3}(x).
\end{aligned} \tag{33}$$

In the same way,

$$\begin{aligned}
\frac{v_n(x)}{\alpha_n} - \frac{v(x)}{\alpha} &= (nh_n)^{-1} \sum_{i=1}^{n} G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) \\
&\quad - (nh_n)^{-1} \sum_{i=1}^{n} G^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) \\
&\quad + (nh_n)^{-1} \sum_{i=1}^{n} G^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) \\
&\quad - \mathbb{E}\left[h_n^{-1} G^{-1}(Y_1) K\left(\frac{x - X_1}{h_n}\right)\right] \\
&\quad + h_n^{-1} \mathbb{E}\left\{G^{-1}(Y_1) K\left(\frac{x - X_1}{h_n}\right)\right\} - \frac{v(x)}{\alpha} \\
&=: \Gamma_{n1}(x) + \Gamma_{n2}(x) + \Gamma_{n3}(x). \qquad (34)
\end{aligned}$$

We first consider the negligible terms in Eq. (33) and Eq. (34).

Lemma 6.7 *Under* (A1), (A2) *and for any x, both $\sqrt{nh_n}\Lambda_{n1}(x)$ and $\sqrt{nh_n}\Gamma_{n1}(x)$ are $o_P(1)$ as $n \to \infty$.*

Proof Using Remark 6 in Woodroofe (1985), we have

$$\begin{aligned}
\sqrt{nh_n}\Lambda_{n1}(x) &\leq \sqrt{nh_n}\, \frac{\sup_y |G_n(y) - G(y)|}{G(a_F) G_n(a_F)} \times \frac{1}{nh_n} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right) \\
&= O_P\left(\sqrt{h_n}\right). \qquad (35)
\end{aligned}$$

In the same way, we have

$$\sqrt{nh_n}\Gamma_{n1}(x) = O_P\left(\sqrt{h_n}\right). \qquad (36)$$

□

Lemma 6.8 *Under* (A1), (A3), (B1)–(B3) *and for any x, both $\sqrt{nh_n}\Lambda_{n3}(x)$ and $\sqrt{nh_n}\Gamma_{n3}(x)$ are $o_P(1)$ as $n \to \infty$.*

Proof Under (A1), (B1) and (B3) we have

$$\begin{aligned}
\sqrt{nh_n}\Lambda_{n3}(x) &= \sqrt{nh_n}\left[\frac{1}{h_n} \int\int \frac{y}{G(y)} K\left(\frac{x - u}{h_n}\right) \mathrm{d}\mathbf{F}^{\star}(u, y) - \frac{\psi(x)}{\alpha}\right] \\
&= \frac{\sqrt{nh_n}}{\alpha}\left[\int\int yK(r)\mathrm{d}\mathbf{F}(x - rh_n, y) - \psi(x)\right] \\
&= \frac{\sqrt{nh_n^5}}{2\alpha} \int\int yr^2 K(r) \frac{\partial^2 \mathbf{f}}{\partial x^2}(x_n, y)\, \mathrm{d}y\mathrm{d}r
\end{aligned}$$

(where x_n is between x and $x - rh_n$). Then under (A3), (B1) and (B2) we get (see our Remark 6 about ψ'' continuity)

$$\sqrt{nh_n}\Lambda_{n3}(x) = \frac{\sqrt{nh_n^5}}{2\alpha} \int r^2 K(r)\psi''(x_n)\,\mathrm{d}r = o_P(1) \tag{37}$$

and similarly

$$\sqrt{nh_n}\Gamma_{n3}(x) = o_P(1). \tag{38}$$

$\square$

Now we consider the dominant terms $\Lambda_{n2}(x)$ and $\Gamma_{n2}(x)$ and prove Lemma 6.9.

Lemma 6.9 *Under* (A1), (A3) *and* (B1) *we have*

$$(\Lambda_{n2}(x), \Gamma_{n2}(x))^T \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \alpha^{-1}\kappa\Sigma(x)\right)$$

Proof First we have

$$\begin{aligned}
\mathrm{Var}\left[\sqrt{nh_n}\Lambda_{n2}(x)\right] &= \frac{1}{nh_n} \times n\mathrm{Var}\left[Y_1 G^{-1}(Y_1) K\left(\frac{x - X_1}{h_n}\right)\right] \\
&= \frac{1}{\alpha h_n} \int\int \frac{y^2}{G(y)} K^2\left(\frac{x-u}{h_n}\right) \mathbf{f}(y,u)\,\mathrm{d}y\mathrm{d}u \\
&\quad - \frac{1}{\alpha^2 h_n^2}\left\{\int\int yK\left(\frac{x-u}{h_n}\right)\mathbf{f}(y,u)\,\mathrm{d}y\mathrm{d}u\right\}^2 \\
&= \alpha^{-1}\kappa\Sigma_0(x) + o(1)
\end{aligned} \tag{39}$$

by a simple Taylor expansion, under (A1), (A3) and (B1). In the same way, we easily get

$$\mathrm{Var}\left[\sqrt{nh_n}\Gamma_{n2}(x)\right] = \alpha^{-1}\kappa\Sigma_2(x) + o(1) \tag{40}$$

and

$$\mathrm{Cov}\left[\sqrt{nh_n}\Lambda_{n2}(x), \sqrt{nh_n}\Gamma_{n2}(x)\right] = \alpha^{-1}\kappa\Sigma_1(x) + o(1). \tag{41}$$

Now for a given pair of real numbers $c = (c_1, c_2)^{\mathrm{T}}$ put

$$\Delta_n(x) = \sqrt{nh_n}\,[c_1\Lambda_{n2}(x) + c_2\Gamma_{n2}(x)] =: \sum_{i=1}^{n} \Delta_{ni}(x)$$

where the $\Delta_{ni}(x)$ (readily obtained from Eq. (33) and Eq. (34)) are clearly iid. Let

$$\rho_{ni}^{\beta}(x) = \mathbb{E}\left[|\Delta_{ni}(x)|^{\beta}\right]$$

for the β in (A1). Then by Hölder's inequality we get from Eq. (33) and Eq. (34)

$$\rho_{ni}^{\beta}(x) \leq 2^{\beta-1}\left(\frac{h_n}{n}\right)^{\beta/2}\left\{c_1^{\beta}\mathbb{E}\left[\left|\frac{Y_1 G^{-1}(Y_1)}{h_n}K\left(\frac{x-X_1}{h_n}\right)\right|^{\beta}\right]\right.$$
$$\left. + c_2^{\beta}\mathbb{E}\left[\left|\frac{G^{-1}(Y_1)}{h_n}K\left(\frac{x-X_1}{h_n}\right)\right|^{\beta}\right]\right\}$$

which implies that, under (A1) (since $1-\beta/2<0$), we have

$$\rho_n^{\beta}(x) := \sum_{i=1}^{n}\rho_{ni}^{\beta}(x) = O\left(n^{1-\beta/2}h_n^{\beta/2}\right) = o(1). \tag{42}$$

On the other hand, we deduce from Eq. (39)–(41) that

$$s_n^2(x) := \mathrm{Var}\left\{\sqrt{nh_n}\,[c_1\Lambda_{n2}(x) + c_2\Gamma_{n2}(x)]\right\} \overset{n\to\infty}{\longrightarrow} \alpha^{-1}\kappa c^T\Sigma(x)c > 0 \tag{43}$$

for any $c \neq 0$ provided $v(x) > 0$ (see Remark 5). Then Eq. (42) and (43) give $\lim_{n\to\infty}\rho_n(x)/s_n(x) \to 0$. Hence, the result is a consequence of Berry–Esséen's Theorem (see Chow and Teicher, 1997, p. 322). □

Proof of Theorem 4.2 Consider the mapping θ from $\mathbb{R}^2$ to $\mathbb{R}$ defined by $\theta(x,y) = x/y$ for $y\neq 0$. Since $m_n(x)$ and $m(x)$ are the respective images of $\alpha_n^{-1}(\psi_n(x), v_n(x))$ and $\alpha^{-1}(\psi(x), v(x))$ by θ, we deduce from Lemmas 6.7 to 6.9 and from Mann–Wald's Theorem (see Rao 1965, p. 321) that $\sqrt{nh_n}\,(m_n(x) - m(x))$ converges in distribution to $\mathcal{N}\left(0, \alpha^{-1}\kappa\nabla\theta^T\Sigma(x)\nabla\theta\right)$, where the gradient $\nabla\theta$ is evaluated at $\alpha^{-1}(\psi(x), v(x))$. Simple algebra gives then the variance $\sigma^2(x)$.

Acknowledgements The authors are grateful to an anonymous referee whose careful reading gave them the opportunity to improve the quality of the paper and add both theoretical and practical results.

References

Bhattacharya, P.K., Chernoff, H., Yang, S.S. (1983). Nonparametric estimation of slope of a truncated regression. *The Annals of Statistics*, *11*, 505–514.

Carbonez, A., Györfi, L., van der Meulin, E.C. (1995). Partition-estimate of a regression function under random censoring. *Statistics and Decisions*, *13*, 21–37.

Chow, Y.S., Teicher, H. (1997). *Probability theory. independence, interchangeability, Martingales*. Berlin Heidelberg, New York: Springer.

Devroye, L., Györfi, L., Lugosi, G. (1996). *A probability theory of pattern recognition*. Berlin Heidelberg, New York: Springer.

Fan, J., Gijbels, I. (1996). *Local polynomial modelling and applications*. London: Chapman Hall.

Feigelson, E.D., Babu, G.J. (1992). *Statistical challenges in modern astronomy*. Berlin Heidelberg, New York: Springer.

Giné, E., Guillou, A. (1999). Laws of iterated logarithm for censored data. *The Annals of Probability*, *27*, 2042–2067.

Giné, E., Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Annales de l'Institut Henri Poincaré*, *37*, 503–522.

Gross, S.T., Huber-Carol, C. (1992). Regression models for truncated survival data. *Scandinavian Journal of Statistics*, *19*, 193–213.

Gross, S.T., Lai, T.L. (1996). Nonparametric estimation regression analysis with left truncated and censored data. *Journal of the American Statistical Association*, *91*, 1166–1180.

Györfi, L., Kohler, M., Walk, H. (1998). Weak and strong universal consistency semi-recursive partitioning and kernel regression estimates. *Statistics and Decisions*, *16*, 1–18.

He, S., Yang, G. (1994). Estimating a lifetime distribution under different sampling plan. In S.S. Gupta J.O. Berger (Eds.) *Statistical decision theory and related topics* (Vol. 5. pp. 73–85) Berlin Heidelberg, New York: Springer.

He, S., Yang, G. (1998). Estimation of the truncation probability in the random truncation model. *The Annals of Statistics*, *26*, 1011–1027.

He, S., Yang, G. (2003). Estimation of regression parameters with left truncated data. *Journal of Statistical Planning and Inference*, *117*, 99–122.

Kim, C.K., Lai, T.L. (1999). Regression smoothers and additive models for censored and truncated data. *Communications in Statistics: Theory Methods*, *28*(n^o 11), 2717–2747.

Kohler, M., Máthé, K., Pintér, M. (2002). Prediction from randomly right censored data. *Journal of Multivariate Analysis*, *80*, 73–100.

Lewbel, A., Linton, O. (2002). Nonparametric censored and truncated regression. *Econometrica*, *70*, 765–779.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices of the Royal Astronomical Society*, *155*, 95–118.

Park, J. (2004). Optimal global rate of convergence in nonparametric regression with left-truncated and right-censored data. *Journal of Multivariate Analysis*, *89*, 70–86.

Rao, C.R. (1965). *A linear statistical inference and its applications*, New-York: Wiley.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*, London: Chapman and Hall.

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

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.

Walk, H. (2002). Almost sure convergence of Nadaraya–Watson regression estimates. In M. Dror, P. l'Ecuyer F. Szidarovsky (Eds.) *Assays on uncertanily-S. Yakowitz memorial volume* (pp. 201–223) Dordrecht: Kluwer.

Wang, M.C., Jemell, N.P., Tsai, W.Y. (1986). Asymptotic properties of the product-limit estimate under random truncation. *The Annals of Statistics*, *14*, 1597–1605.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, *13*, 163–177.