Asymptotic properties of a conditional quantile estimator with randomly truncated data

Mohamed Lemdani a,∗, Elias Ould-Saïd b, Nicolas Poulin b

a Lab. de Biomathématiques, Univ. de Lille 2. Fac. de Pharmacie, 3, rue du Pr. Laguesse, 59006 Lille, France
b L.M.P.A J. Liouville, Univ. du Littoral Côte d’Opale, BP 699, 62228 Calais, France

Article history:
Received 1 October 2007
Available online 17 June 2008

AMS 1991 subject classifications:
62G05
62G20

Keywords:
Asymptotic normality
Consistency
Kernel
Quantile function
Truncated data

1. Introduction

In this paper we consider the classical conditional quantile estimation problem in the case where the response variable is left-truncated.

Let (Y i) 1≤i≤N be a sample of independent and identically distributed (iid) real random variables (rv) with common unknown distribution function (df) F and (X i) 1≤i≤N a corresponding sample of random covariate vectors taking their values in Rd with df V and continuous density v.

The Y i’s are regarded as the lifetimes of the items under study and are supposed to be subject to left-truncation which may occur if the time origin of the lifetime precedes the time origin of the study. Only subjects that fail after the start of the study are being followed, the others being truncated. We denote by (T i) 1≤i≤N the iid sample of truncation rv with continuous df G. The T i’s are assumed to be independent of the Y i’s. Then (Y i, T i) is observed only when Y i ≥ T i. Therefore the original sample is not completely observed and only n observations (among N) are obtained. This model arises in various fields such as astronomy, economics and medical studies (see e.g. [9,25]).

Now consider the joint df function F(·, ·) of (X 1, Y 1) and suppose it is of class C 1(Rd+1). Then the conditional df of Y 1 given X 1 = x = (x 1, . . . , x d) is expressed as

\[ F(· | x) = \frac{F_1(x, ·)}{v(x)} \] (1)
with
\[
F_1(\mathbf{x}, \cdot) = \frac{\partial F(\mathbf{x}, \cdot)}{\partial \mathbf{x}} := \frac{\partial^d F(\mathbf{x}, \cdot)}{\partial x_1 \cdots \partial x_d}.
\]

For a fixed \( p \in (0, 1) \), let \( \xi_p(\mathbf{x}) \) be the \( p \)th conditional quantile of \( Y_1 \) given \( \mathbf{X}_1 = \mathbf{x} \) defined by
\[
\xi_p(\mathbf{x}) = \inf \{ y : F(y|\mathbf{x}) \geq p \}.
\]

It is well known that the conditional quantile function gives a good description of the data (see, e.g. [2]), such as robustness to heavy-tailed error distributions and outliers, especially the conditional median function. In the iid framework without truncation, several authors considered this problem. Let us quote for example Samanta [22] who studied a class of nonparametric estimators of conditional quantiles. Bhattacharya and Gangopadhyay [1] gave a Bahadur-type representation of the conditional quantile and asymptotic models; Mehra et al. [18] and Xiang [26] gave the a.s. convergence of a kernel-type conditional quantile estimator and its asymptotic normality. Furthermore, Qin and Wu [20] obtained the asymptotic normality of an estimator for a conditional quantile using the empirical likelihood method and a linear fitting when some auxiliary information is available. Finally Honda [12] dealt with \( \alpha \)-mixing processes and proved the uniform convergence and asymptotic normality of an estimate of \( \xi_p(.) \) using the local polynomial fitting method.

In the left-truncation model, on the iid framework, Gürler et al. [8] gave a Bahadur-type representation for the quantile function and asymptotic normality. The extension to a time series case was obtained by Lemdani et al. [15]. A nonparametric regression function estimator with randomly truncated data is considered in [7,11,19]. To our knowledge no result is available in the literature about the conditional distribution or the conditional quantile estimators under random truncation.

Our main goal is to establish the asymptotic behavior of a kernel conditional quantile estimator for a truncated model. The paper is organized as follows: in Section 2 we recall the truncation framework before introducing in Section 3, the different notations and defining our conditional quantile estimator. The assumptions and main results are detailed in Section 4. Simulations are given in Section 5 whereas Section 6 is devoted to the proofs.

## 2. Background for truncation models

In this section we give the main definitions and results related to truncation models.

Recall that our original sample is \((\mathbf{X}_i, Y_i, T_i)_{1 \leq i \leq N}\). Taking into account the truncation’s effect we denote by \((\mathbf{X}_i, Y_i, T_i), \ldots, (\mathbf{X}_n, Y_n, T_n)\) the actually observed sample (i.e. \( Y_i \geq T_i, 1 \leq i \leq n \)) and suppose that \( \alpha \equiv \mathbb{P}(Y_1 \geq T_1) > 0 \). Conditionally on the value of \( n \), these observed random vectors are still iid (see [14]). Note here that \( n \) is a rv itself and that from the strong law of large numbers (SLLN) we have, as \( N \to \infty \):
\[
\hat{\alpha}_n := \frac{n}{N} \to \alpha, \quad \mathbb{P}\text{-a.s.}
\tag{2}
\]

For any real df \( L \) denote the left and right endpoints of its support by \( a_L = \inf \{ t : L(t) > 0 \} \) and \( b_L = \sup \{ t : L(t) < 1 \} \), respectively. Following [23] the dfs of \( Y_1 \) and \( T_1 \) are
\[
F^\ast(y) = \alpha^{-1} \int_{-\infty}^y G(t) \, dF(t) \quad \text{and} \quad G^\ast(y) = \alpha^{-1} \int_{-\infty}^\infty G(y \wedge t) \, dF(t)
\]
respectively and are estimated by
\[
F_n^\ast(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq y\}} \quad \text{and} \quad G_n^\ast(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{T_i \leq y\}}
\]
respectively, where \( \mathbf{1}_A \) is the indicator of the set \( A \). Note that, in what follows, the star notation (\( \ast \)) relates to any characteristic of the actually observed data (that is, conditionally on \( n \)).

Define
\[
C(y) = G^\ast(y) - F^\ast(y) = \alpha^{-1} G(y) \left( 1 - F(y) \right), \quad y \in [a_F, +\infty)
\tag{3}
\]
and consider its empirical estimate
\[
C_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{T_i \leq y \leq Y_i\}} = G_n^\ast(y) - F_n^\ast(y^+).
\]

Since \( N \) is unknown and \( n \) is known (although random), our results will not be stated with respect to the probability measure \( \mathbb{P} \) (related to the \( N \)-sample). Let \( \mathbb{P}_n(\cdot) = \mathbb{P}(\cdot|n) \) be the conditional probability. Since independence is preserved we can write \( \mathbb{P}_n(\cdot) = \mathbb{P}^{\otimes n} \) where \( \mathbb{P}(\cdot) = \mathbb{P}(\cdot|Y \geq T) \). Estimation results are then established considering \( n \to \infty \) and so are expressed with respect to the probability \( \mathbb{P} \). Finally let \( \mathbb{E} \) and \( \mathbb{E} \) denote the respective expectation operators of \( \mathbb{P} \) and \( \mathbb{P} \).
It is well known that the respective nonparametric maximum likelihood estimators of $F$ and $G$ are the product-limit estimators given by

$$
F_n(y) = 1 - \prod_{Y_i \leq y} \left( \frac{n C_n(Y_i) - 1}{n C_n(Y_i)} \right)
$$
and

$$
G_n(y) = \prod_{Y_i > y} \left( \frac{n C_n(Y_i) - 1}{n C_n(Y_i)} \right)
$$

which were obtained by Lynden-Bell [17]. Their asymptotic properties were studied by Woodroofe [25] who showed that

$$
\sup_y |F_n(y) - F(y)| \xrightarrow{\text{P- a.s.}} 0 \quad \text{and} \quad \sup_y |G_n(y) - G(y)| \xrightarrow{\text{P- a.s.}} 0.
$$

(4)

provided $a_G \leq a_F$, $b_G \leq b_F$ and $\int DF/G < \infty$. Additional results can be found in [13].

Consequently $\alpha$ is identifiable only if $a_G \leq a_F$ and $b_G \leq b_F$. Note that the estimator $\hat{\alpha}_n$ defined in (2) cannot be calculated (since $N$ is unknown). Another estimator, namely

$$
\alpha_n = \frac{G_n(y) \left( 1 - F_n(y^{-}) \right)}{C_n(y)}
$$

(5)
is used. He and Yang [10] proved that $\alpha_n$ 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 (in their Corollary 2.5) its $P - a.s.$ consistency.

3. Quantile and distribution functions’ estimators

In this section we recall some results and then define our quantile estimator. Our estimation of the conditional df is based on the choice of weights. These are obtained in [19].

Recall that, in the case of complete data, a well-known kernel estimator of the regression function is based on the Nadaraya–Watson weights

$$
W_{i,N}(x) := \frac{1}{N} \sum_{j=1}^{N} k_d \left( \frac{x - x_i}{h_N} \right)
$$

(6)

associated to the $N$-sample (with the convention $0/0 = 0$). Here $k_d$ is a nonnegative function on $\mathbb{R}^d$ and $(h_N)$ is a nonnegative sequence which goes to zero as $N$ goes to infinity. Considering the density $v$, the corresponding estimator $v_N$ is based on the complete data and cannot therefore be calculated. On the other hand

$$
v^*_N(x) = \frac{1}{nh^d} \sum_{i=1}^{n} k_d \left( \frac{x - x_i}{h_n} \right)
$$

(7)
is an estimator of the conditional density $v^*(x)$ (given $y \geq y_1$).

In order to estimate $v$ we have to take into account the truncation and the estimator

$$
v_N(x) := \frac{\alpha_n}{nh^d} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} k_d \left( \frac{x - x_i}{h_n} \right)
$$

(8)
is considered in [19]. Note that, in this formula and the forthcoming, the sum is taken only for $i$ such that $G_n(Y_i) \neq 0$.

Then the adapted weights

$$
\tilde{W}_{i,n}(x) = \frac{\alpha_n^{-1} k_d \left( \frac{x - x_i}{h_n} \right)}{\sum_{j=1}^{n} G_n^{-1}(Y_j) k_d \left( \frac{x - x_j}{h_n} \right)}
$$

(9)

are considered in order to derive the estimator of the conditional df of $y$ given $x = x$:

$$
F_n(y|x) = \frac{\alpha_n}{n} \sum_{i=1}^{n} \tilde{W}_{i,n}(x) \frac{1}{G_n(Y_i)} K_0 \left( \frac{y - Y_i}{h_n} \right)
$$

$$
= \frac{\sum_{i=1}^{n} G_n^{-1}(Y_i) k_d \left( \frac{x - x_i}{h_n} \right) K_0 \left( \frac{y - Y_i}{h_n} \right)}{\sum_{i=1}^{n} G_n^{-1}(Y_i) k_d \left( \frac{x - x_i}{h_n} \right)}
$$
which, in view of (1) and (8), can be written as

\[ F_n(y|x) = \frac{F_{1,n}(x, y)}{v_n(x)} \]

where \( K_0 \) is a smooth df and

\[ F_{1,n}(x, y) = \frac{\alpha_n}{nh_n} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} k_d \left( \frac{x - X_i}{h_n} \right) K_0 \left( \frac{y - Y_i}{h_n} \right) \]

is an estimator of \( F_1(x, y) \).

Then a natural estimator of \( \xi(x) \) is given by

\[ \xi_{\rho,n}(x) = \inf \{ y : F_n(y|x) \geq \rho \} \]

**Remark 1.** Considering the density \( k_0 = K_0 \) and (10), we easily get an estimator of the conditional density of \( y \) given \( x \) (that is \( f(y|x) = \frac{\partial}{\partial y} F(y|x) \)) defined by:

\[ f_n(y|x) = \frac{\partial}{\partial y} F_n(y|x) = f_n(x, y) \]

where

\[ f_n(x, y) = \frac{\alpha_n}{nh_n^{d+1}} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} k_d \left( \frac{x - X_i}{h_n} \right) k_0 \left( \frac{y - Y_i}{h_n} \right) \]

is an estimator of the joint density \( f(x, y) = \frac{\partial}{\partial y} f_n(x, y) \).

### 4. Assumptions and main results

Throughout this paper we assume that \( 0 = a_c < a_r \leq b_c \leq b_r \) and suppose that \( T_i \) and \( (X_i, Y_i), 1 \leq i \leq N \) are independent. We consider two real numbers \( a \) and \( b \) such that \( a_c < a < b < b_r \).

Define \( \Omega_0 = \{ x \in \mathbb{R}^d : v(x) > 0 \} \) and let \( \Omega \) be a compact subset of \( \Omega_0 \). Then

\[ y := \inf_{x \in \Omega} v(x) > 0. \]

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

A1 The bandwidth \( h_n \) satisfies \( h_n \downarrow 0 \) and \( nh_n^d \log n \to \infty \) as \( n \to \infty \).

A2 The kernel \( k_d \) is a \( C^1 \)-probability density with compact support.

A3 \( K_0 \) is a df with \( C^1 \)-probability density \( k_0 \) and compact support.

A4 The kernels \( k_d \) and \( k_0 \) satisfy

\[ \int t k_0(t) dt = 0 \quad \text{and} \quad \sum_{i=1}^{d} \int r_i k_d(r) dr = 0 \]

with \( r = (r_1, \ldots, r_d) \).

A5 The joint density \( f(\cdot, \cdot) \) is bounded and twice continuously differentiable.

The last hypotheses intervene in the asymptotic normality.

A6 The bandwidth \( h_n \) satisfies \( h_n \downarrow 0 \) and \( nh_n^{d+1} \log n \to \infty \) as \( n \to \infty \).

A7 The bandwidth \( h_n \) satisfies \( h_n \downarrow 0 \) and \( nh_n^{d+4} \to 0 \).

**Remark 2.**

1. Assumption A3 implies that the kernel \( k_0 \) is bounded by a constant \( M_0 > 0 \). In the same way, under A2, we put \( M_d = \| k_d \|_{\infty} \).

2. Assumption A2 implies condition \( (K_1) \) in [6] under which

\[ \mathcal{F} = \left\{ k_d \left( \frac{x - h}{h} \right) : x \in \mathbb{R}^d, h \in \mathbb{R} \setminus \{ 0 \} \right\} \]

is a bounded VC-class of measurable functions. This is a consequence of Theorems 4.2.1 and 4.2.4 in [4]. This assumption is needed in order to use Talagrand’s inequality. Note also that A1 is a consequence of A6.

3. Assumption A5 is a classical smoothness assumption which permits studying bias terms by means of Taylor expansions. Moreover, this assumption implies the continuity of the covariate’s density \( v(\cdot) \). Getting rid of A5 may be possible but different and more tedious conditions have then to be assumed.

Next, our first result is the uniform almost sure convergence with a rate of the conditional df estimator defined in (10).
Proposition 1. Under Assumptions A1–A5, we have

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |F_n(y|x) - F(y|x)| = O \left( \max \left\{ \sqrt{\frac{\log n}{nh_n^d}}, h_n^2 \right\} \right) \text{P-a.s. as } n \to \infty.$$  

Theorem 1. Under the assumptions of Proposition 1 and for each fixed $p \in (0, 1)$, if the conditional density satisfies

$$\inf_{x \in \Omega} f(\xi_p(x)|x) > 0$$

then

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = O \left( \max \left\{ \sqrt{\frac{\log n}{nh_n^d}}, h_n^2 \right\} \right) \text{P-a.s. as } n \to \infty.$$  

Remark 3. The uniform positiveness assumption on the conditional density (in Theorem 1) implies the uniform unicity of the conditional quantile that is

$$\forall \varepsilon > 0, \exists \beta > 0, \forall \eta_p : \Omega \to \mathbb{R}, \sup_{x \in \Omega} |\xi_p(x) - \eta_p(x)| \geq \varepsilon \Rightarrow \sup_{x \in \Omega} |F(\xi_p(x)|x) - F(\eta_p(x)|x)| \geq \beta.$$  

On the other hand, assuming (16) with no additional condition, guarantees the consistency of the conditional quantile but does not give a rate of convergence.

The following result is interesting in itself and intervenes in the asymptotic normality.

Proposition 2. Under Assumptions A2–A6, we have

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |f_n(y|x) - f(y|x)| = O \left( \max \left\{ \sqrt{\frac{\log n}{nh_n^d}}, h_n^2 \right\} \right) \text{P-a.s. as } n \to \infty.$$  

Now, in order to state the asymptotic normality we need some additional notations.

Consider the matrix

$$\Sigma(x, y) = \left( \begin{array}{cc} \Sigma_1(x, y) & \Sigma_1(x, y) \\ \Sigma_2(x) & \Sigma_2(x) \end{array} \right)$$

where

$$\Sigma_1(x, y) = \int_{-\infty}^y \frac{f(x, s)}{G(s)} \, ds \quad \text{and} \quad \Sigma_2(x) = \int_{-\infty}^\infty \frac{f(x, s)}{G(s)} \, ds.$$  

Proposition 3. Under Assumptions A2–A7, for any $x \in \Omega_0$ and $y < b_F$, we have

$$\sqrt{nh_n^d} (F_n(y|x) - F(y|x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(x, y))$$

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

$$\sigma^2(x, y) = \kappa \left[ \Sigma_1(x, y) \nu^2(x) + \Sigma_2(x) F_1^2(x, y) - 2 \Sigma_1(x, y) F_1(x, y) \nu(x) \right] \quad \alpha \nu^4(x)$$

and $\kappa = \int k_3^2(r) \, dr$.

Theorem 2. Under the assumptions of Proposition 3 we have, for each $p \in (0, 1)$ and any $x \in \Omega_0$ such that $f(\xi_p(x)|x) \neq 0$,

$$\sqrt{nh_n^d} (\xi_{p,n}(x) - \xi_p(x)) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \frac{\sigma^2(x, \xi_p(x))}{f^2(\xi_p(x)|x)} \right).$$

Remark 4. Note, on the one hand, that $\Sigma_2(x) \geq \nu(x)$ and on the other hand, we have $0 < \Sigma_1(x, y) < \Sigma_2(x)$ if $y < b_F$. Therefore $\Sigma(x, y)$ is positive definite for any $x \in \Omega_0$, as soon as $y < b_F$. 


Now using Theorem 2 it is possible to construct confidence intervals for \( \xi_p(x) \). For that purpose a plug-in estimate \( \sigma_n(x, y) \) for \( \sigma(x, y) \) can be used. Having already defined the estimators \( \alpha_n, \nu_n(x) \) and \( F_{1,n}(x, y) \) (see formulae (5), (8) and (11)) we have to consider the estimators

\[
\Sigma_{1,n}(x, y) = \frac{\alpha_n}{nh_1} \sum_{i=1}^n \frac{\xi(Y_i, x)}{G_n(Y_i)} k_d \left( \frac{x - X_i}{h_n} \right) \quad \text{and} \quad \Sigma_{2,n}(x) = \frac{\alpha_n}{nh_2} \sum_{i=1}^n \frac{1}{G_n(Y_i)} k_d \left( \frac{x - X_i}{h_n} \right)
\]

for \( \Sigma_1(x, y) \) and \( \Sigma_2(x) \), respectively. We then have:

**Corollary 1.** Under the assumptions of Theorem 2, a confidence interval of asymptotic level \( 1 - \zeta \) for \( \xi_p(x) \) is given by

\[
\left[ \frac{\hat{\xi}_{p,n}(x) - u_{1-\zeta/2} \sigma_n(x, \xi_{p,n}(x))}{\sqrt{nh_n^2 \sigma_n^2(x, \xi_{p,n}(x))}} , \frac{\hat{\xi}_{p,n}(x) + u_{1-\zeta/2} \sigma_n(x, \xi_{p,n}(x))}{\sqrt{nh_n^2 \sigma_n^2(x, \xi_{p,n}(x))}} \right]
\]

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

### 5. Simulations

The main purpose of this section is to show the behavior of our estimator for some particular conditional quantile functions. A lot of issues can be considered: (i) the value of \( p \); (ii) the sample size; (iii) the percentage of truncation \( \alpha \); (iv) the type of model; (v) the dimension of covariate’s space. In order to limit the number of examples, we first address Issues (i), (ii) and (iii) in the unidimensional linear case. In the second step, we fix \( p \) and \( \alpha \) and consider the bidimensional case in the linear and nonlinear (quadratic) cases which answers Issue (iv).

#### 5.1. Unidimensional case

We consider the linear model \( Y_i = X_i + \sigma \epsilon_i, i = 1, \ldots, N \), where \( X_i \) and \( \epsilon_i \) are two independent iid sequences distributed as \( \mathcal{N}(0, 1) \) and \( \sigma \) is a constant appropriately chosen (the choice of \( \sigma \) is calibrated in order to get different values of the truncation probability \( \alpha \)).

We also simulate \( N \) iid rv \( T_i \sim \mathcal{E}(1) - 2 \), where \( \mathcal{E}(\lambda) \) is the exponential distribution with parameter \( \lambda \). We then keep the data \( (X_i, Y_i, T_i), i = 1, \ldots, n \) such that \( Y_i \geq T_i \). We do it in a way to obtain a given number \( n \) of observed triplets (which means that in this case \( n \) is not random whereas \( N \) is). We then compute our estimator based on these observed data by choosing Gaussian kernels \( k_1 \) and \( k_2 \). Recall that, for nonparametric estimation, optimality (in the MSE sense) is not seriously swayed by the choice of the kernel but is affected by that of the bandwidth \( h_n \).

Adapting \( \sigma \) so as to obtain \( \alpha \approx 90\% \), we noticed that the estimator was bad for small \( n \), but had a good behavior for \( n \) large enough (see Figs. 1 and 2). We recall that in our model, for \( p = 0.5 \), the conditional median is the identity function and, for \( p = 0.25 \), the conditional lower quartile is the line \( y = x - 0.4384 \).

We then tried to see if the quality depended on the truncation proportion \( \alpha \). We took \( n = 500 \) and chose different values of the percentage of truncated data (by adapting \( \sigma \): \( \alpha \approx 30\% \), \( \approx 50\% \) and \( \approx 70\% \)). The estimator’s quality does not seem to be affected by \( \alpha \) as shown in Fig. 3 (though higher values of \( N \) are needed for small \( \alpha \) to achieve \( n = 500 \)).

#### 5.2. Bidimensional case

In this subsection, we restrict ourselves to \( p = 0.5 \) (median). The linear and nonlinear bidimensional cases are studied. For the linear case, we consider the model given by \( Y = 0.25 + U + 2V + \sigma \epsilon \) with \( X = (U, V)^t \) where \( U \sim \mathcal{E}(1), V \sim \mathcal{E}(2) \) and \( \epsilon \) is uniformly distributed on \([-1, 1]\). The truncation variable is distributed as \( \mathcal{N}(2, 1) \). The choice of \( \sigma \) is still related to \( \alpha \) but satisfies \( \alpha \leq 2 \) so that \( Y \geq 0 \) (in fact we take \( \sigma = 0.12 \)). In Fig. 4 we give the theoretical plane \( Y = 0.25 + U + 2V \) and the estimated surfaces for \( n = 200 \) and \( n = 500 \). For both cases \( \alpha \approx 50\% \).
Fig. 2. Conditional quartile function $p = 0.25$ with $\alpha \approx 90\%$ and $n = 100, 500$ and 1000, respectively.

Fig. 3. Conditional median function $p = 0.5$ with $n = 500, \alpha \approx 30\%, \alpha \approx 50\%$ and $\alpha \approx 70\%$, respectively.

For the nonlinear case our model is given by $Y = 0.25 + U^2 + V^2 + \sigma \epsilon$. The estimated surfaces for $n = 200$ and $n = 500$ are shown and compared (in Fig. 5) to the theoretical mean response.

As far as we can read from the last figures, the estimation quality looks as good in the bidimensional case as is for a univariate covariate.

Fig. 4. Conditional median surface $p = 0.5$: $n = 200$, $n = 500$ and theoretical linear function, respectively.

Fig. 5. Conditional median surface $p = 0.5$: $n = 200$, $n = 500$ and theoretical quadratic function, respectively.
6. Auxiliary results and proofs

In order to make the proofs easier, we need some auxiliary results and notations. The first lemma gives the uniform consistency with rate of the estimator \(v_n^*(x)\) defined in (7).

**Lemma 1.** Under Assumptions A1, A2, A4 and A5, we have

\[
\sup_{x \in \mathbb{R}^d} |v_n^*(x) - v^*(x)| = O \left( \max \left\{ \frac{\log n}{nh_n^2}, h_n^2 \right\} \right), \quad \mathbb{P}\text{-a.s.} \tag{19}
\]

**Proof.** First, simple algebra gives

\[
v^*(x) = \alpha^{-1} \int G(y)f(x, y) \, dy
\]

and therefore, by A5, \(v^*\) is bounded and twice continuously differentiable.

Using the triangle inequality, we have

\[
\sup_{x \in \mathbb{R}^d} |v_n^*(x) - v^*(x)| \leq \sup_{x \in \mathbb{R}^d} |v_n^*(x) - \mathbb{E}[v_n^*(x)]| + \sup_{x \in \mathbb{R}^d} |\mathbb{E}[v_n^*(x)] - v^*(x)| =: I_1 + I_2. \tag{20}
\]

Under A2, the class of functions

\[
\mathcal{F}_1 = \left\{ \theta_k(\cdot) = \frac{1}{nh_n} k_d \left( \frac{x - \cdot}{h_n} \right) : x \in \mathbb{R}^d \right\}
\]

is a bounded VC-class of measurable functions (see Remark 2.2) which are uniformly bounded with envelope \(\Theta = \frac{M_d}{nh_n^d}\).

Moreover

\[
\mathbb{E}\left[ \theta_k(X_1) \right] \leq \frac{\|v^*\|_\infty}{n} \quad \text{and} \quad \mathbb{E}\left[ \theta_k^2(X_1) \right] \leq \frac{M_d\|v^*\|_\infty}{n^2h_n^d}.
\]

Applying Talagrand’s inequality (see Proposition 2.2 in [5]) with \(t = D\sqrt{\frac{\log n}{nh_n^2}}\), where \(D\) is a positive constant there exist two positive constants \(c_1\) and \(c_2\) such that

\[
P \left\{ \sup_{\theta_k \in \mathcal{F}_1} \left| \sum_{i=1}^n \{\theta_k(X_i) - \mathbb{E}[\theta_k(X_1)]\} \right| \geq D \sqrt{\frac{\log n}{nh_n^d}} \right\} \leq c_1 \exp \left\{ - \frac{1}{c_1} D \frac{1}{\|v^*\|_\infty} \sqrt{\frac{n \log n}{nh_n^d}} \log \left( 1 + \frac{D \sqrt{\frac{\log n}{nh_n^d}}}{c_1\|v^*\|_\infty} \right) \right\}
\]

\[
\leq c_1 \exp \left\{ - \frac{1}{c_1} D \frac{1}{\|v^*\|_\infty} \sqrt{\frac{n \log n}{h_n^d}} \frac{1}{n} \log \left( \frac{n \log n}{nh_n^d}\right) \right\}
\]

\[
\leq c_1 n^{-\frac{D^2}{c_1^2\|v^*\|_\infty^2}}
\]

(by \(\log(1 + t) \leq t\) which, for \(n\) large enough and by an appropriate choice of \(D\), can be made \(O\left(n^{-3/2}\right)\). The latter being a general term of a summable series, we then have by the Borel–Cantelli lemma under A1

\[
I_1 = O \left( \sqrt{\frac{\log n}{nh_n^d}} \right), \quad \mathbb{P}\text{-a.s.} \tag{22}
\]

On the other hand, using a change of variable and a Taylor expansion, we get, under A4 and A5

\[
I_2 = O \left( h_n^2 \right) \tag{23}
\]

Hence, replacing (22) and (23) in (20), we get the result. □
Under Assumptions A1–A5, we have

\[ \tilde{F}_{1,n}(x, y) = \frac{\alpha}{nh_n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} k_d \left( \frac{x - X_i}{h_n} \right) K_0 \left( \frac{y - Y_i}{h_n} \right). \]  

We have

**Lemma 2.** Under Assumptions A1–A5, we have

\[ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |F_{1,n}(x, y) - F_1(x, y)| = O \left( \max \left\{ \log n, \frac{1}{h_n^d} \right\} \right) , \quad \mathbb{P}\text{-a.s. as } n \to \infty. \]  

**Proof.** We have

\[
\begin{align*}
\sup_{x \in \Omega} \sup_{a \leq y \leq b} |F_{1,n}(x, y) - F_1(x, y)| & \leq \sup_{x \in \Omega} \sup_{a \leq y \leq b} |F_{1,n}(x, y) - \tilde{F}_{1,n}(x, y)| \\
& + \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\tilde{F}_{1,n}(x, y) - \mathbb{E}[\tilde{F}_{1,n}(x, y)]| + \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbb{E}[\tilde{F}_{1,n}(x, y)] - F_1(x, y)| \\
& =: \mathcal{g}_1 + \mathcal{g}_2 + \mathcal{g}_3.
\end{align*}
\]

As \( K_0 \) is a df under A3, it is bounded by 1 so

\[ |F_{1,n}(x, y) - \tilde{F}_{1,n}(x, y)| \leq \left\{ \begin{array}{ll} |\alpha_n - \alpha| & \text{if } |\alpha_n - \alpha| \leq \frac{\alpha}{G_n(a_F)} \\
\frac{\alpha}{G_n(a_F)} \sup_{a \leq y \leq b} |G_n(y) - G(y)| & \text{if } |\alpha_n - \alpha| > \frac{\alpha}{G_n(a_F)} \end{array} \right\} |\nu_n^*(x)|. \]

From Theorem 3.2 in [10] we have \( |\alpha_n - \alpha| = O( n^{-1/2} ) , \mathbb{P} - a.s. \). Moreover, \( G_n(a_F) \xrightarrow{P-a.s.} G(a_F) \). On the other hand, \( \sup_{a \leq y \leq b} |G_n(y) - G(y)| = O( n^{-1/2} ) , \mathbb{P} - a.s. \) (see Remark 6 in [25]). As an immediate consequence of **Lemma 1** we obtain

\[ \mathcal{g}_1 = O( n^{-1/2} ) , \quad \mathbb{P}\text{-a.s.} \]

Then using Lemma 2.6.18(vi) in [24] (p. 147) with \( g(y) = 1/G(y) \), the class of functions

\[ \mathcal{F}_2 = \left\{ \eta_y(\cdot) = \frac{1}{G(y)} K_0 \left( \frac{y - \cdot}{h_n} \right) : y \in \mathbb{R} \right\} \]

is a bounded VC-class of bounded measurable functions under A3. Now recalling that \( \mathcal{F}_1 \) defined in (21) is a VC-class, using Lemma 2.6.20 in [24] (p. 148) with \( \mathcal{F}_1 \) and \( \mathcal{F}_2 \), the class of functions

\[ \mathcal{F}_1 = \left\{ \theta_{x,y} (r, w) = \frac{1}{nh_n^d G(y)} k_d \left( \frac{x - r}{h_n} \right) K_0 \left( \frac{y - w}{h_n} \right) : x \in \mathbb{R}^d, y \in \mathbb{R} \right\} \]

is a bounded VC-class of measurable functions with envelope \( \Theta = \frac{M_d}{nh_n^d G(y)} \), under A2.

In the same way as for **Lemma 1**, the result is then obtained under A1 by applying Talagrand’s inequality, which gives

\[ \mathcal{g}_2 = O \left( \frac{\log n}{nh_n^d} \right) , \quad \mathbb{P}\text{-a.s.} \]

Finally, from (24) we have

\[ \mathbb{E}[\tilde{F}_{1,n}(x, y)] = \mathbb{E} \left[ \frac{1}{nh_n^d} k_d \left( \frac{x - X_1}{h_n} \right) \mathbb{E} \left[ \frac{\alpha}{G(Y_1)} K_0 \left( \frac{y - Y_1}{h_n} \right) \left| X_1 \right. \right] \right]. \]

Remark that

\[ \mathbb{E} \left[ \frac{\alpha}{G(Y_1)} K_0 \left( \frac{y - Y_1}{h_n} \right) \left| X_1 \right. \right] = \int \frac{\alpha}{G(w)} K_0 \left( \frac{y - w}{h_n} \right) f^*(w|X_1) dw \]

\[ = \int K_0 \left( \frac{y - w}{h_n} \right) f(w|X_1) dw \]

\[ = \int k_0(s) F(y - sh_n|X_1) ds \]
by integration by parts and change of variable. Then using (1) and a Taylor expansion, under A4 and A5, we get, uniformly over \( x \) and \( y \)

\[
\begin{align*}
\mathbb{E} \left[ \tilde{F}_{1,n}(x,y) - F_1(x,y) \right] &= \int \int k_d(r)k_0(s) \left[ F_1(x - rh_n, y - sh_n) - F_1(x, y) \right] \, dr \, ds \\
&= O \left( h_n^2 \right). 
\end{align*}
\]

Combining (26)–(29) permits concluding the proof of Lemma 2. \( \blacksquare \)

Using the same framework as in Lemma 2, we can show:

**Lemma 3.** Under A1, A2, A4 and A5, we have

\[
\sup_{x \in \Omega} \sup_{a \leq y \leq b} |F_n(y|x) - F(y|x)| \leq \frac{1}{\gamma - \sup_{x \in \Omega} |v_n(x) - v(x)|} \left\{ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |F_{1,n}(x,y) - F_1(x,y)| \right. \\
+ \left. \sup_{x \in \Omega} \sup_{a \leq y \leq b} |F(y|x)| \sup_{x \in \Omega} |v_n(x) - v(x)| \right\}. 
\]

The result follows straightforwardly from Lemmas 2 and 3. \( \blacksquare \)

**Proof of Proposition 1.** In view of (10), from (14), we have

\[
\sup_{x \in \Omega} \sup_{a \leq y \leq b} |F_n(y|x) - F(y|x)| \leq \frac{1}{\gamma - \sup_{x \in \Omega} |v_n(x) - v(x)|} \left\{ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |F_{1,n}(x,y) - F_1(x,y)| \right. \\
+ \left. \sup_{x \in \Omega} \sup_{a \leq y \leq b} |F(y|x)| \sup_{x \in \Omega} |v_n(x) - v(x)| \right\}.
\]

(30)

The result follows straightforwardly from Lemmas 2 and 3. \( \blacksquare \)

**Proof of Theorem 1.** Let \( x \in \Omega \). As \( F(\cdot|x) \) and \( F_n(\cdot|x) \) are continuous, we have \( F(\xi_p(x)|x) = F_n(\xi_{p,n}(x)|x) = p \). Then

\[
|F(\xi_{p,n}(x)|x) - F(\xi_p(x)|x)| = |F(\xi_{p,n}(x)|x) - F_n(\xi_{p,n}(x)|x)| \\
\leq \sup_{a \leq y \leq b} |F_n(y|x) - F(y|x)|.
\]

Then the consistency of \( \xi_{p,n}(x) \) follows immediately from Proposition 1 and the continuity of \( F(\cdot|x) \). Now

\[
F(\xi_{p,n}(x)|x) - F(\xi_p(x)|x) = (\xi_{p,n}(x) - \xi_p(x)) f(\xi_p(x)|x)
\]

where \( \xi_p(x) \) is between \( \xi_{p,n}(x) \) and \( \xi_{n,p}(x) \). Then, by (31), we have

\[
\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| |f(\xi_p(x)|x)| \leq \sup_{x \in \Omega} |F_n(y|x) - F(y|x)|.
\]

The result is then a consequence of Proposition 1 and the assumption of \( f(\xi_p(\cdot)|\cdot) \) being uniformly lower-bounded away from zero. \( \blacksquare \)

Here we point out that, if \( f(\xi_p(x)|x) = 0 \) for some \( x \in \Omega \), the consistency of \( \xi_{p,n}(x) \) (with an adapted rate) may be obtained by a higher order Taylor expansion.

Now to prove Proposition 2, we need to define:

\[
\tilde{f}_n(x,y) = \frac{\alpha}{nh_n^{d+1}} \sum_{i=1}^{n} \frac{1}{G(Y_i)} k_d \left( \frac{x - X_i}{h_n} \right) k_0 \left( \frac{y - Y_i}{h_n} \right). 
\]

(32)

**Proof of Proposition 2.** We have, as in (30)

\[
\sup_{x \in \Omega} \sup_{a \leq y \leq b} |f_n(y|x) - f(y|x)| \leq \frac{1}{\gamma - \sup_{x \in \Omega} |v_n(x) - v(x)|} \left\{ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |f_n(x,y) - \tilde{f}_n(x,y)| \\
+ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\tilde{f}_n(x,y) - \mathbb{E} \left[ \tilde{f}_n(x,y) \right]| + \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbb{E} \left[ \tilde{f}_n(x,y) \right] - f(x,y)| + \sup_{x \in \Omega} \sup_{a \leq y \leq b} |f(y|x)| \sup_{x \in \Omega} |v_n(x) - v(x)| \right\}.
\]

(33)

= \frac{1}{\gamma - \sup_{x \in \Omega} |v_n(x) - v(x)|} \left\{ \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + \mathcal{L}_4 \right\}. 

We deal with each term of the right–hand side of (33).

For $\mathcal{L}_1$, as in (27), we get (see also Remark 2.1)

$$\mathcal{L}_1 = 0 \left( \left( \frac{1}{n^{2/d}} \right)^{-1/2} \right), P\text{-a.s. as } n \to \infty. \quad (34)$$

Then, under A2 and A3, in the same way as for (28), it can be shown that the class of functions

$$\tilde{F} = \left\{ \tilde{\theta}_{x,y}(r, w) = \frac{1}{nh_n^d + 1} k_d \left( \frac{x-r}{h_n} \right) k_0 \left( \frac{y-w}{h_n} \right) : x \in \mathbb{R}^d, y \in \mathbb{R} \right\}$$

is a bounded VC-class of measurable functions with respect to the envelope $\Theta = \frac{M_0M_0}{n^{d+1}G(a_F)}$. Moreover

$$\mathbb{E} \left[ \tilde{\theta}_{x,y}(X_1, Y_1) \right] \leq \frac{\|f\|_\infty}{nG(a_F)} \quad \text{and} \quad \mathbb{E} \left[ \tilde{\theta}_{x,y}^2(X_1, Y_1) \right] \leq \frac{M_0M_0\|f\|_\infty}{n^2h_n^{d+1}G^2(a_F)}.$$  

Applying Talagrand’s inequality in the same way as in Lemma 1 with $t = D \sqrt{\frac{\log n}{n^{d+1}}} \log n$ and by the Borel–Cantelli lemma and A6 we get

$$\mathcal{L}_2 = 0 \left( \frac{\log n}{n^{d+1}} \right), P\text{-a.s. as } n \to \infty. \quad (35)$$

Then, in view of $\mathcal{L}_3$, we write

$$\mathbb{E} \left[ \tilde{f}_n(x, y) \right] = \mathbb{E} \left[ \frac{1}{h_n^{d+1}} k_d \left( \frac{x-X_1}{h_n} \right) \mathbb{E} \left[ \frac{\alpha}{G(Y_1)} k_0 \left( \frac{y-Y_1}{h_n} \right) | X_1 \right] \right]. \quad (36)$$

As in (28) we get

$$\mathbb{E} \left[ \frac{\alpha}{G(Y_1)} k_0 \left( \frac{y-Y_1}{h_n} \right) | X_1 \right] = \int \frac{1}{h_n^d} k_0' \left( \frac{y-w}{h_n} \right) F(w | X_1) \, dw$$

which replaced in (36) yields

$$\mathbb{E} \left[ \tilde{f}_n(x, y) \right] = \int \frac{1}{h_n^{d+2}} k_d \left( \frac{x-r}{h_n} \right) \int k_0' \left( \frac{y-w}{h_n} \right) F(w | r) \, dw \, v(r) \, dr$$

$$= \int \frac{1}{h_n^{d+2}} k_d \left( \frac{x-r}{h_n} \right) \int k_0' \left( \frac{y-w}{h_n} \right) F_1(r, w) \, dw \, dr.$$  

Simple algebra gives

$$\int k_0' \left( \frac{y-w}{h_n} \right) F_1(r, w) \, dw = \int h_n^d k_0(s) f(r, y-sh_n) \, ds$$

and then

$$\mathbb{E} \left[ \tilde{f}_n(x, y) \right] = \int \int k_d(u) k_0(s) f(x-u h_n, y-sh_n) \, du \, ds.$$  

Then, under the lemma’s assumptions, a Taylor expansion yields

$$\mathcal{L}_3 = 0 \left( \frac{1}{h_n^2} \right), P\text{-a.s. as } n \to \infty. \quad (37)$$

Finally, combining Lemma 3, (33)–(35) and (37) we get the result (recall that A6 implies A1).

In order to prove Theorem 2 and Proposition 3 we write, from (10),

$$F_n(y | x) = \frac{\alpha_n^{-1} F_{1,n}(x, y)}{\alpha_n^{-1} \nu_n(x)}.$$  

A three-term decomposition of $\alpha_n^{-1} \nu_n(x) - \alpha^{-1} \nu(x)$ is given in [19]. Under A2 and A4–A7 it is proved in Lemmas 6.7 and 6.8 of that paper that two of the terms are $o_P(1)$ and in Lemma 6.9 that the dominant term

$$\Gamma_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n G^{-1}(Y_1) k_d \left( \frac{x-X_i}{h_n} \right) - \mathbb{E} \left[ \frac{1}{nh_n^d} G^{-1}(Y_1) k_d \left( \frac{x-X_1}{h_n} \right) \right]$$  

(38)
converges in distribution to a centred Gaussian law under assumptions A2, A4, A5 and A7. Moreover, they gave
\[
\Var \left[ \sqrt{n\eta_n} \Gamma_{n2}(x) \right] = \alpha^{-1} \kappa \Sigma_2(x) + o(1)
\] (39)

where \( \kappa \) and \( \Sigma_2(x) \) are given in Proposition 3.

Now we have to give an analogous result for the difference \( \alpha_n^{-1}F_{1,n}(x, y) - \alpha^{-1}F_1(x, y) \). We write
\[
\frac{F_{1,n}(x, y)}{\alpha_n} \alpha - \frac{F_1(x, y)}{\alpha} = \frac{F_{1,n}(x, y)}{\alpha_n} - \frac{\bar{F}_{1,n}(x, y)}{\alpha} + \frac{\bar{F}_{1,n}(x, y)}{\alpha} \frac{\mathbb{E}[\bar{F}_{1,n}(x, y)]}{\alpha} \leq \frac{\mathbb{E}[\bar{F}_{1,n}(x, y)]}{\alpha} - \frac{F_1(x, y)}{\alpha} =: A_{n1}(x, y) + A_{n2}(x, y) + A_{n3}(x, y).
\] (40)

We first consider the negligible terms in (40).

**Lemma 4.** Under A1, A2, A4 and A5, for any \( x \) and \( y \), \( \sqrt{n\eta_n} A_{n1}(x, y) \) is \( o_p(1) \) as \( n \to \infty \).

**Proof.** We have, from (11) and (24)
\[
\sqrt{n\eta_n} A_{n1}(x, y) = \sqrt{n\eta_n} \left[ \frac{1}{n\eta_n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} \left( \frac{x - X_i}{h_n} \right) K_0 \left( \frac{y - Y_i}{h_n} \right) \right] \leq \sqrt{n\eta_n} \sup_{y} \left| G_n(y) - G(y) \right| \leq \sqrt{n\eta_n} \frac{1}{G(a_f) G_n(a_f)} \nu_n^{*}(x).
\]

The result is then a consequence of Remark 6 in [25] and Lemma 1. \( \blacksquare \)

**Lemma 5.** Under A1–A5 and A7, for any \( x \) and \( y \), \( \sqrt{n\eta_n} A_{n3}(x, y) \) is \( o_p(1) \) as \( n \to \infty \).

**Proof.** Using (29) we get
\[
\sqrt{n\eta_n} A_{n3}(x, y) = \alpha^{-1} \sqrt{n\eta_n} \left[ \mathbb{E}[\bar{F}_{1,n}(x, y)] - F_1(x, y) \right] = O \left( \sqrt{n\eta_n}^{d+\theta} \right)
\]
and the result is a direct consequence of A7. \( \blacksquare \)

Now we consider the dominant terms \( A_{n2}(x, y) \) and \( \Gamma_{n2}(x, y) \).

**Lemma 6.** Under A1–A3 and A5 we have, for any \( x \) such that \( v(x) > 0 \) and \( y < b_F \)
\[
\sqrt{n\eta_n} \left( A_{n2}(x, y), \Gamma_{n2}(x, y) \right)^T \overset{d}{\rightarrow} \mathcal{N} \left( 0, \alpha^{-1} \kappa \Sigma(x, y) \right)
\]
where \( \Sigma(x, y) \) is defined in (18).

**Proof.** Using a change of variable and a Taylor expansion we can write
\[
\Var \left[ \sqrt{n\eta_n} A_{n2}(x, y) \right] = \frac{1}{n\eta_n} \times \Var \left[ G^{-1}(Y_1) k_d \left( \frac{x - X_1}{h_n} \right) K_0 \left( \frac{y - Y_1}{h_n} \right) \right] = \frac{1}{\alpha h_n^d} \int \int \frac{1}{G(s)} k_d^2 \left( \frac{x - r}{h_n} \right) K_0 \left( \frac{y - s}{h_n} \right) f(r, s) \, dr \, ds - \frac{1}{\alpha^2 h_n^d} \left[ \int \int k_d \left( \frac{x - s}{h_n} \right) K_0 \left( \frac{y - s}{h_n} \right) f(r, s) \, dr \, ds \right]^2 = \frac{1}{\alpha} \int \int \frac{1}{G(s)} k_d^2 (u) K_0 \left( \frac{y - s}{h_n} \right) f(x, s) \, du \, ds + O(h_n) = \frac{\kappa}{\alpha} \Sigma_1(x, y) + o(1).
\] (41)
In the same way,
\[
\text{Cov} \left[ \sqrt{nh_n^d} A_{n2}(x, y), \sqrt{nh_n^d} F_{n2}(x, y) \right] = \frac{1}{\alpha h_n} \int \int \frac{1}{G(s)} k^2_d \left( \frac{x - r}{h_n} \right) K_0 \left( \frac{y - s}{h_n} \right) f(r, s) \, dr ds + o(1)
= \frac{1}{\alpha} \int \int \frac{1}{G(s)} k^2_d (u) K_0 \left( \frac{y - s}{h_n} \right) f(x, s) \, du ds + o(1)
= \frac{\kappa}{\alpha} \Sigma_1(x, y) + o(1).
\] (42)

Now for a given pair of real numbers \(c = (c_1, c_2)^T\) put
\[
\Delta_n(x, y) = \sqrt{nh_n^d} [c_1 A_{n2}(x, y) + c_2 F_{n2}(x, y)] = \sum_{i=1}^{n} \Delta_i(x, y)
\]
where the \(\Delta_i(x, y)\) (readily obtained from (38) and (40)) are clearly iid. Let
\[
\rho_n^3(x, y) = \mathbb{E} [|\Delta_n(x, y)|^3].
\]
Then by the \(C_r\)-inequality (see [16], p. 156) we get from (38) and (40)
\[
\rho_n^3(x, y) \leq 4 \left( \frac{h_n^l}{h_n^d} \right)^{3/2} \left\{ c^3_1 \mathbb{E} \left[ \left( \frac{1}{h_n^d G(Y_1)} k_d \left( \frac{x - X_1}{h_n} \right) K_0 \left( \frac{y - Y_1}{h_n} \right) \right)^3 \right] + c^3_2 \mathbb{E} \left[ \left( \frac{1}{h_n G(Y_1)} k_d \left( \frac{x - X_1}{h_n} \right) \right)^3 \right] \right\}
\]
which implies
\[
\rho_n^3(x, y) : = \sum_{i=1}^{n} \rho_i^3(x, y) = O \left( n^{-1/2} h_n^{3d/2} \right) = o(1).
\] (43)

On the other hand, we deduce from (39), (41) and (42) that
\[
s_n^2(x, y) : = \text{Var} \left\{ \sqrt{nh_n^d} [c_1 A_{n2}(x, y) + c_2 F_{n2}(x, y)] \right\} \xrightarrow{n \to \infty} \alpha^{-1} \kappa c^T \Sigma(x, y) c > 0
\] (44)
for any \(c \neq 0\) provided \(v(x) > 0\) (see Remark 4). Then (39), (43) and (44) give \(\lim_{n \to \infty} \rho_n(x, y)/s_n(x, y) \to 0\). Hence the result is a consequence of the Berry–Essèen Theorem (see [3], p. 322).

Proof of Proposition 3. Consider the mapping \(\theta\) from \(\mathbb{R}^2\) to \(\mathbb{R}\) defined by \(\theta(u, v) = u/v\) for \(v \neq 0\). Since \(F_n(y|x)\) and \(F(y|x)\) are the respective images of \(\alpha^{-1}(F_{n}(y|x), v_{n}(x))\) and \(\alpha^{-1}(F_{1}(x, y), v(x))\) by \(\theta\) we deduce from Lemmata 4–6, from Lemmata 6.7–6.9 in [19] and from the \(\delta\)-method Theorem (see [21], p. 321) that \(\sqrt{nh_n^d} (F_{n}(y|x) - F(y|x))\) converges in distribution to \(\mathcal{N}(0, \alpha^{-1} \kappa \nabla \theta^T \Sigma (x, y) \nabla \theta)\) where the gradient \(\nabla \theta\) is evaluated at \(\alpha^{-1}(F_{1}(x, y), v(x))\). Simple algebra gives then the variance \(\sigma^2(x, y)\).

Proof of Theorem 2. We make use of the property \(F_{p}(\xi_{p}(x)|x) = p = F_{n}(\xi_{p,n}(x)|x)\). Using a Taylor expansion we have
\[
\xi_{p,n}(x) - \xi_{p}(x) = \frac{F_{n}(\xi_{p,n}(x)|x) - F_{n}(\xi_{p}(x)|x)}{f_{n}(\xi_{p,n}(x)|x)} - \frac{\xi_{p,n}^{*}(x)|x)}{f_{n}(\xi_{p,n}^{*}(x)|x)}
\]
where \(\xi_{p,n}^{*}(x)\) lies between \(\xi_{p,n}(x)\) and \(\xi_{p,n}(x)\).

The continuity of \(f(\cdot|x)\), Theorem 1 and Proposition 2 imply the convergence in probability of the above denominator to \(f(\xi_{p}(x)|x)\). Proposition 3 is used to finish the proof.

Proof of Corollary 1. First the consistency of \(\Sigma_{1,n}(x, y)\) and \(\Sigma_{2,n}(x)\) can be proved in the same way as for Lemma 3. Then using Theorem 1, wherever \(\xi_{p,n}(x)\) appears, it can be replaced by \(\xi_{p}(x)\) since all the involved quantities are uniformly continuous with respect to the \(\xi_{p}\) argument. The result is then a direct consequence of Theorem 2.

Acknowledgments

The authors are grateful to two anonymous referees for their careful reading and appropriate remarks. One of the referee’s queries also gave them the opportunity to improve the quality of the paper and enrich the simulation part.
References