**Supplementary Material for the paper entitled**:

# FDA: Theoretical and practical efficiency of the local linear estimation based on the $k$NN smoothing of the conditional distribution when there are missing data and functional covariables

## 1. The model and its estimation

Consider $n$ pairs of random variables $(X_i, Y_i)$, for $i = 1, \ldots, n$ which are drawn from the pair $(X, Y) \in \mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a Hilbert space equipped with the norm $\|.\|$. We study, for all $y \in \mathbb{R}$ and all $x \in \mathcal{F}$, the LLE of the CDF given by:

$$F(y|x) = \mathbb{P}(Y \leq y|X = x).$$

We assume, for a fixed $(y, x) \in \mathbb{R} \times \mathcal{F}$, that the CDF $F(y|x)$ is smoothed enough to be locally approximated by a linear function. That is, for all $x_0$ in a neighborhood of $x$, we have:

$$F(y|x_0) = a_{yx} + b_{yx}(x_0 - x) + \rho_{yx}(x_0 - x, x_0 - x) + o(\|x_0 - x\|^2), \tag{1.1}$$

where $b_{yx}$ (resp. $\rho_{yx}$) is a linear (resp. bilinear) continuous operator from $\mathcal{F}$ (resp. $\mathcal{F} \times \mathcal{F}$) to $\mathbb{R}$. The operators $a_{yx}$ and $b_{yx}$ are estimated by the $k$NN method as the minimizers of the following rule:

$$\min_{a,b \in \mathbb{R} \times \mathcal{F}} \sum_{i=1}^{n} \left(\mathbb{1}_{Y_i \leq y} - a - b(X_i - x)\right)^2 K\left(\frac{\|x - X_i\|}{h_k}\right), \tag{1.2}$$

where $K$ is a kernel and $h_k = \min\{h \in \mathbb{R}^+ \text{ such that } \sum_{i=1}^{n} \mathbb{1}_{B(x,h)}(X_i) = k\}$ with $B(x, r) = \{z \in \mathcal{F} : \|x - z\| \leq r\}$ denoting the topologically closed ball, in $\mathcal{F}$, centered at $x$ and with radius $r$ and $\mathbb{1}_A$ is the indicator function on the set $A$.

## 2. Asymptotic properties of the estimator $\widehat{F}(y|x)$

Let $(x, y)$ be a fixed point in $\mathcal{F} \times \mathbb{R})$, $\mathcal{N}_x$ (resp., $\mathcal{N}_y$) a fixed neighborhood of $x$ (resp., of $y$). Furthermore, we assume that our nonparametric model satisfies the following conditions:

(H1) For any $r > 0$, the function $\phi_x(r) := \mathbb{P}(X \in B(x, r)) > 0$ is an invertible function and there exist $0 < c < 1 < c^* < \infty$, such that

$$\lim_{r\to 0} \frac{\phi_x(rc)}{\phi_x(r)} < 1 < \lim_{r\to 0} \frac{\phi_x(rc^*)}{\phi_x(r)}.$$

(H2) The function $F(\cdot|\cdot)$ such that (1.1) with a continuous operator $\rho_{xy}$ and the coefficients $(c_j)_j$ such that

$$\sum_{j=J+1} c_j^2 = O_{a.co}(J^{-1}).$$

(H3) The function $P(\cdot)$ is continuous on $\mathcal{N}_x$ and such that $P(z) > 0$, for all $z \in \mathcal{N}_x$.

(H4) The kernel $K$ is a differentiable function which is supported within $(0, 1)$. Moreover, its first derivative $K'$ exists and is such that there exist two constants $C$ and $C^*$ satisfying $-\infty < C^* < K'(t) < C < 0$ for $0 \le t \le 1$.

(H5) The number of neighbors $k$ is such that

$$\frac{\ln n}{k} \to 0 \text{ as } n \to \infty.$$

**Theorem 1.** *Under Assumptions (H1)-(H5), we obtain*

$$|\widehat{F}(y|x) - F(y|x)| = O(J^{-1}) + O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\ln n}{k}}\right), \qquad (2.1)$$

*as* $\min(n, J) \to \infty$.

*Proof of Theorem 1.* The proof is based on similar ideas as those used by Chikr-Elmezouar et al. (2018). Indeed, we introduce the following notations, for $j, j' = 1, \ldots, J$:

$$\begin{aligned}
S_{n,j',j} &= \frac{1}{nh_k^2\phi_x(h_k)} \sum_{i=1}^{n} c_{ij'} c_{ij} \delta_i K_i(h_k), \\
T_{n,j}^0 &= \frac{1}{nh_k\phi_x(h_k)} \sum_{i=1}^{n} c_{ij} \delta_i K_i(h_k) 1\!\!1_{Y_i \le y}, \\
e_{n,j}^* &= S_{n,0,j} = \frac{1}{nh_k\phi_x(h_k)} \sum_{i=1}^{n} c_{ij} \delta_i K_i(h_k), \\
T_{n,j}^{0*} &= \frac{1}{nh_k\phi_x(h_k)} \sum_{i=1}^{n} c_{ij} \delta_i K_i(h_k)(1\!\!1_{Y_i \le y} - F(y|X_i)), \\
e_{n,j} &= \frac{1}{nh_k\phi_x(h_k)} \sum_{i=1}^{n} c_{ij} \delta_i K_i(h_k) \rho_{xy}(X_i - x, X_i - x),
\end{aligned}$$

where $K_i(h_k) = K(h_k^{-1}\|x - X_i\|)$. Under these notations, we can write:

$$\begin{pmatrix} \widehat{a}_{yx} \\ h_k\widehat{b}_1 \\ \vdots \\ h_k\widehat{b}_J \end{pmatrix} = (S_n)^{-1}(T_n^0),$$

where $S_n = (S_{n,j',j})_{j',j=0,\ldots,J}$ and $T_n^0 = (T_{n,j}^0)_{j=0,\ldots,J}$ So, by the regularity assumption (1.1) and the assumption (H2) imply that

$$F(y|X_i) = a_{yx} + \sum_{j}^{J} c_{ij} b_{yx}(v_j) + \rho_{yx}(X_i - x, X_i - x) + O(J^{-1}).$$

Further, we denote by $T_n^{0*} = (T_{n,j}^{0*})_{j=0,\ldots,J}$ and we put $e_n = (e_{n,j})_{j=0,\ldots,J}$ and $e_n^* = (e_{n,j}^*)_{j=0,\ldots,J}$. So, we have

$$\begin{aligned}
T_n^{0*} &= T_n^0 - (T_n^0 - T_n^{0*}) \\
&= S_n \begin{pmatrix} \widehat{a}_{yx} \\ h_k\widehat{b}_1 \\ \vdots \\ h_k\widehat{b}_J \end{pmatrix} - S_n \begin{pmatrix} a_{yx} \\ h_k b_1 \\ \vdots \\ h_k b_J \end{pmatrix} + e_n + O(J^{-1})e_n^*.
\end{aligned}$$

It follows that

$$\widehat{a}_{yx} - a_{yx} = e_1' \left( S_n^{-1} T_n^{0*} - S_n^{-1} e_n - O(J^{-1}) S_n^{-1} e_n^* \right),$$

Thus, Theorem 1's result will be a consequence of the following lemmas.

**Lemma 1.** *Under conditions (H1), (H4)and (H5), we get, for all $j, j' = 1, \ldots, J$,*

$$S_{n,j',j} = O_{a.co.}\left(1\right).$$

**Lemma 2.** *Under the conditions of the Theorem 1, we obtain, for all $j = 1, \ldots, J$,*

$$\mathbb{E}\left[T^{0*}_{n,j}\right] = 0 \quad \textit{and} \quad \mathbb{E}\left[e_{n,j}\right] = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right).$$

**Lemma 3.** *Under the conditions of Theorem 1, we obtain, for all $j = 1, \ldots, J$,*

$$T^{0*}_{n,j} - \mathbb{E}\left[T^{0*}_{n,j}(x)\right] = O_{a.co.}\left(\sqrt{\frac{\ln n}{k}}\right),$$

*and*

$$e_{n,j} - \mathbb{E}\left[e_{n,j}\right] = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\ln n}{k}}\right).$$

**Corollary 1.** *Under the conditions of Theorem 1, we have*

$$e_1' S_n^{-1} T_n^{0*} = O_{a.co.}\left(\sqrt{\frac{\ln n}{k}}\right),$$

*and*

$$e_1' S_n^{-1} e_n = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\ln n}{k}}\right).$$

■

Notice that the proofs of the intermediate results are given in short ways because they follow the same ideas as in Chikr-Elmezouar et al. (2019). The main challenge, here, is how to handle the additional variable $\delta$. In what follows, when no confusion is possible, we will denote by $C$ and $C^*$ some strictly positive generic constants. Similarly to Burba et al. (2009), we assume, in the proofs that, the random variables $c_{ij}c_{ij'}$ are nonnegative. The other cases can be deduced by taking

$$c_{ij}c_{ij'} = (c_{ij}c_{ij'})^+ - (c_{ij}c_{ij'})^-$$

where $(c_{ij}c_{ij'})^+ = \max(c_{ij}c_{ij'}, 0)$ and $(c_{ij}c_{ij'})^- = -\min(c_{ij}c_{ij'}, 0)$. Then, we adopt the same treatment for $c_{ij}\left(\mathbb{1}_{Y_i \le y} - F(y|X_i)\right)$ and $c_{ij}\ \rho_{xy}(X_i - x, X_i - x)$.

*Proof of Lemma 1.* Similarly to Chikr-Elmezouar et al. (2019), it suffices to show that

$$\ddot{S}_{n,j',j}(h) = O_{a.co.}(1) \quad \text{for} \quad h = h_k^{\pm},$$

with

$$\ddot{S}_{n,j',j}(h) = \frac{1}{nh^2\phi_x(h)} \sum_{i=1}^{n} c_{ij'}c_{ij}\delta_i K_i(h).$$

To do that we prove that

$$\ddot{S}_{n,j',j}(h) - \mathbb{E}\left[\ddot{S}_{n,j',j}(h)\right] = O_{a.co.}\left(\sqrt{\frac{\ln n}{n\phi_x(h)}}\right) \quad \text{and} \quad \mathbb{E}\left[\ddot{S}_{n,j',j}(h)\right] = O(1). \tag{2.2}$$

For the left one , we put

$$\widetilde{\Delta}_i = \frac{1}{h^2\phi_x(h)} c_{ij'}c_{ij}\delta_i K_i(h),$$

and we write

$$\left|\ddot{S}_{n,j',j}(h) - \mathbb{E}\left[\ddot{S}_{n,j',j}(h)\right]\right| = \left|\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\Delta}_i - \mathbb{E}[\widetilde{\Delta}_i]\right)\right|.$$

As $(v_j)_{j\geq 1}$ is an orthonormal basis, for all $j \leq J$, we obtain

$$|c_{1j}| \leq \|v_j\|\|x - X_1\| \leq \|x - X_1\|.$$

Hence,

$$\mathbb{E}\left[c_{1j'}c_{ij}\delta_1 K_1(h)\right] \leq \mathbb{E}\left[\|x - X_1\|^2 K_1(h)\right] \leq Ch^2\phi_x(h_k).$$

By using Assumptions (H1) and (H4), we obtain that

$$\mathbb{E}[\widetilde{\Delta}_i] < C^*, \quad \left|\widetilde{\Delta}_i\right| < C/\phi_x(h) \quad \text{and} \quad \mathbb{E}\left|\widetilde{\Delta}_i\right|^2 < C^*/\phi_x(h).$$

Then, by using the classical Bernstein's inequality (see Uspensky (1937), Page 205), we deduce, for $\eta > 0$, that

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\Delta}_i - \mathbb{E}[\widetilde{\Delta}_i]\right)\right| > \eta\sqrt{\frac{\ln n}{n\,\phi_x(h)}}\right\} \;\leq\; C^* n^{-C\eta^2}.$$

Now, for the right hand side expectation in (2.2), we have

$$\mathbb{E}[\ddot{S}_{n,j',j}(h)] = \frac{1}{h^2\phi_x(h)}\mathbb{E}\left[c_{1k}c_{ij}K_1(h)\right].$$

Once again we use the fact that the basis $(v_j)_{j\geq 1}$ is orthonormal to get

$$\mathbb{E}\left[c_{1k}c_{ij}K_1(h)\right] \leq \mathbb{E}\left[\|x - X_1\|^2 K_1(h)\right] \leq Ch^2\phi_x(h),$$

which allows to achieve the proof of this Lemma.

■

*Proof of Lemma 2.* Similarly to the Lemma 1's proof, he claimed results of this lemma are a consequences of the following statement

$$\mathbb{E}\left[\ddot{T}^{0*}_{n,j}\right] = 0 \qquad \text{and} \quad \mathbb{E}\left[\ddot{e}_{n,j}\right] = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right),$$

where

$$\ddot{T}^{0*}_{n,j} = \frac{1}{nh\phi_x(h)}\sum_{i=1}^{n} c_{ij}\delta_i K_i(h)(\mathbb{1}_{Y_i\leq y} - F(y|X_i)), \quad h = h_k^{\pm},$$

$$\ddot{e}_{n,j} = \frac{1}{nh\phi_x(h)}\sum_{i=1}^{n} c_{ij}\delta_i K_i(h)\rho_{xy}(X_i - x, X_i - x), \quad h = h_k^{\pm}.$$

Now, for the first term we have

$$\mathbb{E}[\ddot{T}^{0*}_{n,j}] = \frac{1}{h\phi_x(h)}\mathbb{E}\left[c_{1j}\delta_1 K_1(x)\left(\mathbb{1}_{Y_1\leq y} - F(y|X_1)\right)\right].$$

Therefore, we conditione by $X_1$, we show that

$$\mathbb{E}[\ddot{T}^{0*}_{n,j}] = \frac{1}{h\phi_x(h)}\mathbb{E}\left[c_{1j}K_1(h)P(X_1)(\mathbb{E}\left[\mathbb{1}_{Y_1\leq y}|X_1\right] - F(y|X_1))\right].$$

Thus

$$\mathbb{E}[\ddot{T}^{0*}_{n,j}] = 0.$$

The second term may be treated in the same manner. The proof of Lemma 2 is now finished.

■

*Proof of Lemma 3.* The proof of this lemma can be accomplished by the same manner as for the Lemma 1's proof.

■