Uniform law of the logarithm for the conditional distribution function and application to certainty bands

Sandie Ferrigno, Myriam Maumy-Bertrand, Aurélie Muller

UNIFORM LAW OF THE LOGARITHM FOR THE CONDITIONAL DISTRIBUTION FUNCTION AND APPLICATION TO CERTAINTY BANDS

S. FERRIGNO[1], M. MAUMY-BERTRAND[2], AND A. MULLER-GUEUDIN[1]

Abstract. In this paper, we establish exact rate of strong uniform consistency for the local linear estimator of the conditional distribution function. As a consequence of our main result, we give asymptotic uniform certainty bands. We illustrate our results with simulations.

[1]Équipe Probabilités et Statistiques, IECN, Nancy-Université, CNRS, INRIA
Bd des Aiguillettes, B.P. 239, 54506 Vandoeuvre-lès-Nancy, France

[2] Équipe de Statistique, IRMA, Université de Strasbourg,
7 rue René Descartes, 67084 Strasbourg, France

ferrigno@iecn.u-nancy.fr, mmaumy@math.unistra.fr, muller@iecn.u-nancy.fr

Keywords: Conditional distribution function, local polynomial estimator, asymptotic uniform certainty bands

1. Introduction

1.1. **Motivations.** Consider (X, Y), a random vector defined in $\mathbb{R}\times\mathbb{R}$. Here Y is the variable of interest and X the concomitant variable. Throughout, we work with a sample $\{(X_i, Y_i)_{1\leqslant i\leqslant n}\}$ of independent and identically replica of (X, Y). In the sequel, we impose the following assumptions upon the distribution of (X, Y). We will assume that (X, Y) [resp. X] has a density function $f_{X,Y}$ [resp. f_X] with respect to the Lebesgue measure. In this paper, we will mostly focus on the regression function of $\psi(Y)$ evaluate at $X = x$ defined by

$$m_\psi(x) = \mathbb{E}\left(\psi(Y)|X = x\right) = \frac{1}{f_X(x)}\int_{\mathbb{R}} \psi(y) f_{X,Y}(x, y)dy, \tag{1}$$

whenever this regression function is meaningful. Here and elsewhere, ψ denotes a specified measurable function, which is assumed to be bounded on each compact subinterval of $\mathbb{R}$. Because of numerous applications, the problem of estimating the function m_ψ, the density function f_X and the regression function $m_{\psi=\mathrm{Id}}$ has been the subject of considerable interest during the last decades. We can cite for example Nadaraya [14], Watson [17], Devroye [7], Collomb [3], Härdle [12] and specially mention two articles, Einmahl and Mason [8] and Deheuvels and Mason [6] for two reasons. The first is that these articles study the function m_ψ and its properties. The second is that we use the tools which are developped in this article in order to establish our proofs. We now choose $\psi(y) = \mathbb{1}_{\{y\leqslant t\}}$ with $t \in \mathbb{R}$ arbitrairy but fixed, and $\mathbb{1}$ the indicator function, so we obtain the conditional distribution function, for all $t \in \mathbb{R}$, defined by

$$F(t|x) = \mathbb{E}\left(\mathbb{1}_{\{Y\leqslant t\}}|X = x\right) = \mathbb{P}\left(Y \leqslant t|X = x\right), \quad \text{for all } x \in \mathbb{R}. \tag{2}$$

In this article, we study the conditional distribution function and a nonparametric estimator associated to this function. The distribution function has the advantages of completely characterizing the law of the random considered variable, allowing to obtain the regression function, the density function, the moments and the quantile function. For example, the conditional distribution function is used, in medicine domain, for the estimation of the references curves (see Gannoun *et al.* [11]).

Introduce the Nadaraya-Watson estimator (see Nadaraya [14] and Watson [17]) of the conditional distribution function $F(t|x)$, defined by

$$\widehat{F}_n^{(0)}(t|x) = \frac{\displaystyle\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \leqslant t\}} K\left(\frac{x - X_i}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)}, \tag{3}$$

where $K(\cdot)$ is a real-valued kernel function defined on $\mathbb{R}$ and $(h_n)_{n\geqslant 1}$ is the bandwidth, and denotes a non-random sequence of positive constants satisfying some assumptions. The Nadaraya-Watson estimator and its properties have been first studied by Stute [15].

Einmahl and Mason [8] have determined, under mild regularity conditions on the joint and marginal density functions and under hypotheses on the bandwidth h_n, exact rates of uniform strong consistency of kernel-type function estimators and specially for the conditional distribution function when the random vector (X, Y) is in $\mathbb{R} \times \mathbb{R}$. We recall here this result:

Corollary 1.1. *(see Corollary 2 in Einmahl and Mason* [8]*.) Let $I = [a, b]$ be a compact interval. Assume that $f_{X,Y}$ and f_X satisfy some regularity conditions and moreover that h_n satisfies $h_n \searrow 0$, $nh_n \nearrow +\infty, \log h_n^{-1}/\log\log n \to +\infty$ and $nh_n/\log n \to +\infty$ as $n \to +\infty$. Then we have for any kernel defined in* [8]*, with probability one*

$$\lim_{n\to+\infty} \sup_{t\in\mathbb{R}} \sup_{x\in I} \sqrt{\frac{nh_n}{2\log(h_n^{-1})}} \left| \widehat{F}_n^{(0)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(0)}(t|x)\right)\right| = \sigma_F(I) \tag{4}$$

where the centering term is

$$\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(0)}(t|x)\right) = \frac{\mathbb{E}\left(\mathbb{1}_{\{Y\leqslant t\}} K\left(\frac{x-X}{h_n}\right)\right)}{h_n \mathbb{E}\left(\widehat{f}_n(x)\right)},$$

and

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \tag{5}$$

is the kernel density estimator of $f_X(x)$, and

$$\sigma_F^2(I) = \frac{||K||_2^2}{4 \inf\limits_{x\in I} f_X(x)}$$

and $||K||_2^2 = \int_{\mathbb{R}} K^2(u)du$.

Remark 1.2. The limit $\sigma_F(I)$ in Equation (4) does not depend on the distribution of Y.

In a more recent article, Einmahl and Mason [9] have given an uniform in bandwidth consistency of kernel-type function estimators, in the case where (X, Y) is in $\mathbb{R}^r \times \mathbb{R}$, $r \in \mathbb{N}^*$, and specially for the estimator $\widehat{F}_n^{(0)}(t|x)$. Moreover, Blondin establishes in [2] a similar result of the Corollary 1.1 in the multivariate case, *i.e.* (X,Y) is in $\mathbb{R}^r \times \mathbb{R}^d$, $r,d \in \mathbb{N}^*$. The major difficulty is that the approach developed by Einmahl and Mason [8] and Deheuvels and Mason [6], based on the empirical processes, has not been used in the multivariate case. The results on the empirical processes indexed by classes of functions are established only for classes of real-valued functions.

It is a well-known fact the asymptotic bias of the Nadaraya-Watson estimator has a bad form. To overcome this problem, there exists an alternative: the local polynomial techniques described in Fan and Gijbels [10]. Mint El Mouvid [13] study the local linear estimator of the conditional distribution function using U-statistics. But this method implies heavy calculations. Here, we do not use U-statistics in our proofs but the theory of the empirical processes.

The present paper is organized as follows. First, we introduce the main notations and hypotheses needed for our task. Then we establish a uniform law of the logarithm for the local linear estimator of the conditional distribution function (see Theorem 2.1 in Section 2). In Section 3, we show that limit laws of the logarithm are useful in the construction of asymptotic uniform certainty bands for the true considered function. Such certainty bands are obtained from simulations and displayed in Section 4. Finally, Section 5 is devoted to the proofs of our results.

1.2. **Notations and assumptions.** Let $(X_1,Y_1),(X_2,Y_2),\ldots,$ be independent and identically distributed replica of (X,Y) in $\mathbb{R}\times\mathbb{R}$. Let $I=[a,b]$, $J=[a',b']\supset I$, two fixed compacts of $\mathbb{R}$.

First, we impose the following set of assumptions upon the distribution of (X,Y):

(F.1) $f_{X,Y}$ is continuous on $J\times\mathbb{R}$ and f_X is continuous and strictly positive on J;
(F.2) $Y\mathbb{1}_{\{X\in J\}}$ is bounded on $\mathbb{R}$.

K denotes a real-valued kernel function defined on $\mathbb{R}$, fulfilling the conditions:

(K.1) K is right-continuous function with bounded variation on $\mathbb{R}$;
(K.2) K is compactly supported and $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$;
(K.3) $\int_{\mathbb{R}} uK(u)\mathrm{d}u = 0$ and $\int_{\mathbb{R}} u^2K(u)\mathrm{d}u \neq 0$.

For such a kernel K, we set for any integer $r \geqslant 0$, $\mu_r(K) = \int_{\mathbb{R}} u^rK(u)du$, and $||K||_2^2 = \int_{\mathbb{R}} K^2(u)du$.

Further, introduce the following growth conditions on the sequence $(h_n)_{n\geqslant 1}$:

(H.1) $h_n \to 0$, as $n\to+\infty$;
(H.2) $nh_n/\log n \to +\infty$, as $n\to+\infty$;
(H.3) $h_n \searrow 0$ and $nh_n \nearrow +\infty$, as $n\to+\infty$;
(H.4) $\log(h_n^{-1})/\log\log n \to +\infty$, as $n\to+\infty$.

Remark 1.3. The hypotheses (H.1-2) are necessary and sufficient for our uniform probability convergence result (see Theorem 2.1). In order to have almost surely convergence results, we should add the hypotheses (H.3-4) (see Blondin [2]).

Our aim will be to establish the strong uniform consistency of the local linear estimator of the conditional distribution function, defined by

$$\widehat{F}_n^{(1)}(t|x) = \frac{\widehat{f}_{n,2}(x)\widehat{r}_{n,0}(x,t) - \widehat{f}_{n,1}(x)\widehat{r}_{n,1}(x,t)}{\widehat{f}_{n,0}(x)\widehat{f}_{n,2}(x) - \left(\widehat{f}_{n,1}(x)\right)^2} \tag{6}$$

where $^{(1)}$ denotes the order 1 of the local polynomial estimator, and

$$\widehat{f}_{n,j}(x) = \frac{1}{nh_n}\sum_{i=1}^{n}\left(\frac{x-X_i}{h_n}\right)^j K\left(\frac{x-X_i}{h_n}\right), \text{ for } j=0,1,2, \tag{7}$$

$$\widehat{r}_{n,j}(x,t) = \frac{1}{nh_n}\sum_{i=1}^{n}\mathbb{1}_{\{Y_i\leqslant t\}}\left(\frac{x-X_i}{h_n}\right)^j K\left(\frac{x-X_i}{h_n}\right), \text{ for } j=0,1. \tag{8}$$

Remark 1.4.

(1) The Nadaraya-Watson estimator $\widehat{F}_n^{(0)}(t|x)$ can be also written with the functions $\widehat{f}_{n,j}$ and $\widehat{r}_{n,j}$ as
$$\widehat{F}_n^{(0)}(t|x) = \frac{\widehat{r}_{n,0}(x,t)}{\widehat{f}_{n,0}(x)}\cdot$$
It is the local polynomial estimator of order 0 of the conditional distribution function. Uniform consistency result for this estimator have been given by Einmahl and Mason [8].

(2) The estimator $\widehat{F}_n^{(1)}(t|x)$ is better than the Nadaraya-Watson estimator when the design is random and has the favorable property to reproduce polynomial of degree 1. Precisely, the local linear estimator has a high minimax efficiency among all possible estimators, including nonlinear smoothers (see Fan and Gijbels [10]).

(3) We have state in the beginning of this Section that we restrict ourselves to the local polynomial estimator of degree 1. The local polynomial estimator can be generalized to the degrees $p \geqslant 2$, but the equations become more complicated. The general results will find elsewhere. We show briefly the form of the local polynomial estimator of order 2:
$$\widehat{F}_n^{(2)}(t|x) = \frac{a_1\widehat{r}_{n,0}(x,t)+a_2\widehat{r}_{n,1}(x,t)+a_3\widehat{r}_{n,2}(x,t)}{a_1\widehat{f}_{n,0}(x)+a_2\widehat{f}_{n,1}(x)+a_3\widehat{f}_{n,2}(x)}$$
where $\begin{cases} a_1 &= \widehat{f}_{n,2}(x)\widehat{f}_{n,4}(x) - \left(\widehat{f}_{n,3}(x)\right)^2 \\ a_2 &= \widehat{f}_{n,2}(x)\widehat{f}_{n,3}(x) - \widehat{f}_{n,1}(x)\widehat{f}_{n,4}(x) \\ a_3 &= \widehat{f}_{n,1}(x)\widehat{f}_{n,3}(x) - \left(\widehat{f}_{n,2}(x)\right)^2 \end{cases}$

and $\widehat{f}_{n,3}$, $\widehat{f}_{n,4}$ and $\widehat{r}_{n,2}$ are the direct extensions of the definitions given in Equations (7) and (8).

Now, we study the consistency of the estimator $\widehat{F}_n^{(1)}(t|x)$ via the following decomposition:
$$\widehat{F}_n^{(1)}(t|x) - F(t|x) = \underbrace{\widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)}_{(1)} + \underbrace{\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) - F(t|x)}_{(2)}$$
where, following the ideas of Deheuvels and Mason [6], the centering term is defined by
$$\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) = \frac{f_{n,2}(x)r_{n,0}(x,t) - f_{n,1}(x)r_{n,1}(x,t)}{f_{n,0}(x)f_{n,2}(x) - f_{n,1}^2(x)},$$
where $f_{n,j}(x) = \mathbb{E}\left(\widehat{f}_{n,j}(x)\right)$ for $j=0,1,2$ and $r_{n,j}(x,t) = \mathbb{E}\left(\widehat{r}_{n,j}(x)\right)$ for $j=0,1$.

The term (1) is the *random part* and is the object of our main theorem given in Section 2. Under (F.1-2), (H.1) and (K.1-3), the term (2) converges uniformly to 0 over $(x,t)\in I\times\mathbb{R}$. The argument to proof this is the Bochner's Lemma (see for instance [8], or our Equations (25) in Section 5).

Remark 1.5. If X and Y are independent variables, then $\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) = \mathbb{P}(Y \leqslant t)$.

2. Main result

We have now all the ingredients to state our main result, the following uniform law of the logarithm concerning the local linear estimator of the conditional distribution function, captured in Theorem 2.1 below. We give both probability convergence results, and almost surely convergence results.

Theorem 2.1. *Under (F.1-2), (H.1-2) and (K.1-3), we have probability convergence results:*

$$\sup_{x\in I}\sqrt{\frac{nh_n}{2\log(h_n^{-1})}}\left|\widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma_{F,t}(I), \tag{9}$$

where

$$\sigma_{F,t}^2(I) = \sup_{x\in I}\left(\frac{F(t|x)(1-F(t|x))}{f_X(x)}\right)\int_{\mathbb{R}} K^2(u)du = ||K||_2^2 \sup_{x\in I}\left(\frac{F(t|x)(1-F(t|x))}{f_X(x)}\right).$$

Moreover, we have

$$\sup_{t\in\mathbb{R}}\sup_{x\in I}\sqrt{\frac{nh_n}{2\log(h_n^{-1})}}\left|\widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma_F(I), \tag{10}$$

where

$$\sigma_F^2(I) = \sup_{t\in\mathbb{R}}\sup_{x\in I}\left(\frac{F(t|x)(1-F(t|x))}{f_X(x)}\right)\int_{\mathbb{R}} K^2(u)du = \frac{||K||_2^2}{4\inf_{x\in I} f_X(x)}.$$

Under (F.1-2), (H.2-4) and (K.1-3) we have almost surely convergence results:

$$\sup_{x\in I}\sqrt{\frac{nh_n}{2\log(h_n^{-1})}}\left|\widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)\right| \xrightarrow[n\to+\infty]{} \sigma_{F,t}(I), \quad \text{almost surely.} \tag{11}$$

Moreover, we have

$$\sup_{t\in\mathbb{R}}\sup_{x\in I}\sqrt{\frac{nh_n}{2\log(h_n^{-1})}}\left|\widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)\right| \xrightarrow[n\to+\infty]{} \sigma_F(I), \quad \text{almost surely.} \tag{12}$$

The proof of Theorem 2.1 is postponed to Section 5. In the following section, we present some direct applications of Theorem 2.1.

Remark 2.2.

(1) The hypothesis (F.2) says that we can found a real $M > 0$ such that if $X \in J$ then $|Y| \leqslant M$. This implies that for $x \in I$ and for enough hight n, $\widehat{F}_n^{(1)}(t|x) = 0$ if $t < -M$ and $\widehat{F}_n^{(1)}(t|x) = 1$ if $t > M$. Our results are then only interesting for $|t| \leqslant M$. But this is not very restrictive, since no upper bound is imposed for the choice of M.

(2) In a next paper, we will give an uniform in bandwidth result. For our applications in the Section 4, a reference choice for h_n is given by minimizing the MISE criteria (see for instance Berlinet [1] or Deheuvels [6]).

$$h_n \propto n^{-1/5}.$$

(3) The choice of the kernel K is not important in practice. The most common used kernels are the Gaussian, the indicator function over $[-\frac{1}{2}, \frac{1}{2}]$, and the Epanechnikov kernels (see for instance Deheuvels [4]). Note that the Gaussian kernel is not compactly supported, but our results can be extended to this case.

3. Uniform asymptotic certainty bands

We show now how our Theorem 2.1 can be used to construct uniform asymptotic certainty bands for $F(t|x)$.

Let remember the kernel density estimator $\widehat{f}_n(x)$ of $f_X(x)$ (see Equation (5)). It is easy to deduce from the Proof in Section 5 (see more precisely Equations (23) and (25)), that under (F.1-2), (H.1-2), (K.1-3) we have,

$$\sup_{x \in I} \left| \frac{\widehat{f}_n(x)}{f_X(x)} - 1 \right| \xrightarrow[n \to +\infty]{\mathbb{P}} 0. \tag{13}$$

Now, we need the following theorem:

Theorem 3.1. *Under (F.1-2), (H.1-2), (K.1-3) we have,*

$$\sup_{t \in \mathbb{R}} \sup_{x \in I} \sqrt{\frac{nh_n}{2\log(h_n^{-1})} \widehat{f}_n(x)} \left| \widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) \right| \xrightarrow[n \to +\infty]{\mathbb{P}} \frac{\|K\|_2}{2}. \tag{14}$$

Remark 3.2. The matching almost surely convergence result can also be obtained, by assuming (H.2-4) instead of (H.1-2).

The proof of this theorem is similar to the proof of Theorem 2.1, in Section 5. To resume, it is a direct application of our Theorem 5.1, with $c(x) = 1/\sqrt{f_X(x)}$ and $d(x) = -F(t|x)/\sqrt{f_X(x)}$ for $x \in I$, combined with Equation (13).

The interest of this result is that the right part does not depend on the unknown density f_X. We built now the certainty bands from this uniform law of the logarithm. We introduce

$$L_n(x) := \frac{\|K\|_2}{2} \sqrt{\frac{nh_n}{2\log(h_n^{-1})} \widehat{f}_n(x)}^{\,-1}$$

such that

$$\sup_{t \in \mathbb{R}} \sup_{x \in I} \{L_n(x)\}^{-1} \left| \widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) \right| \xrightarrow[n \to +\infty]{\mathbb{P}} 1.$$

Now, this probability convergence result enables us to obtain uniform asymptotic certainty bands for $\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)$ in the following sense. For each $0 < \varepsilon < 1$, we have, as $n \to +\infty$:

$$\mathbb{P}\left\{ \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) \in \left[\widehat{F}_n^{(1)}(t|x) \pm (1+\varepsilon)L_n(x)\right],\ \forall (x,t) \in I \times \mathbb{R} \right\} \to 1, \tag{15}$$

and

$$\mathbb{P}\left\{ \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) \in \left[\widehat{F}_n^{(1)}(t|x) \pm (1-\varepsilon)L_n(x)\right],\ \forall (x,t) \in I \times \mathbb{R} \right\} \to 0. \tag{16}$$

Remark 3.3. Then, for all $0 < \varepsilon < 1$ and $0 < \delta < 1$, there exists $n_0 = n_0(\varepsilon, \delta)$ such that for all $n \geqslant n_0$ and with probability greater than $1 - \delta$:

$$\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) \in \left[\widehat{F}_n^{(1)}(t|x) \pm (1+\varepsilon)L_n(x)\right], \quad \text{uniformly in } (x,t) \in I \times \mathbb{R}, \tag{17}$$

and

$$\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) \notin \left[\widehat{F}_n^{(1)}(t|x) \pm (1-\varepsilon)L_n(x)\right], \quad \text{for some } (x,t) \in I \times \mathbb{R}. \tag{18}$$

Whenever (15) and (16) hold jointly for each $0 < \varepsilon < 1$, we will say that the intervals

$$\left[\widehat{F}_n^{(1)}(t|x) \pm L_n(x)\right] \tag{19}$$

provide asymptotic certainty bands (at an asymptotic confidence level of 100%) for $\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)$, uniformly in $(x,t) \in I \times \mathbb{R}$.

We have noted in the Section 1 that the bias part can be neglected. Then, certainty bands given in Equation (19) provide also asymptotic certainty bands for the attempted function $F(t|x)$ uniformly in $(x,t) \in I \times \mathbb{R}$.

Remark 3.4.

(1) Probability convergence is sufficient for forming certainty bands, and requires less restrictive hypotheses on the bandwidth h_n than the almost surely convergence results. That is why we use only the probability convergence result of our main theorem.
(2) Following a suggestion of Deheuvels and Derzko [5], we use, for these upper and lower bounds for $F(t|x)$, the qualification of certainty bands, rather that of confidence bands, because there is no preassigned confidence level $\alpha \in (0,1)$. Some authors (see for instance Deheuvels and Mason [6], or Blondin [2]) have used the term confidence bands.

4. A simulation study

In this paragraph, the certainty bands introduced above are constructed on simulated data. We worked with sample size $n = 1000$ and considered the case: $X \sim \mathcal{N}(0,1)$ where $\mathcal{N}(0,1)$ denotes the Gaussian distribution with mean 0 and standard deviation 1. We present simulations for two models:

(1) We introduce $\lambda(x) = 1 + x^2$, and suppose in a first model that conditionally to $X = x$, for every $x \in \mathbb{R}$, Y follows an exponential distribution with mean $(\lambda(x))^{-1}$, that is $F(t|x) = 1 - \exp\{-\lambda(x)t\}$ for every $(x,t) \in \mathbb{R} \times \mathbb{R}$.
(2) We suppose in a second model that conditionally to $X = x$, for every $x \in \mathbb{R}$, Y follows a Gaussian distribution with mean $(\lambda(x))^{-1}$ and variance $(\lambda(x))^{-2}$.

For the kernel K, we opted for the Epanechnikov kernel. For the bandwidth, we selected $h_n = n^{-1/5}$. The results are presented in Figures 1 and 2, for $x = 0$ and $x = 1$. In the two cases, $\mathbb{E}(Y|X = x) = (\lambda(x))^{-1}$, then for practical reasons in the graphic representation, we have chosen $t \in [0,3]$.

The certainty bands are adequate because they contained the true conditional distribution function, for every $t \in [0,3]$. The fact that the true function did not belong to our bands for some points was expected: it is due to the ε term in Equations (17) and (18).

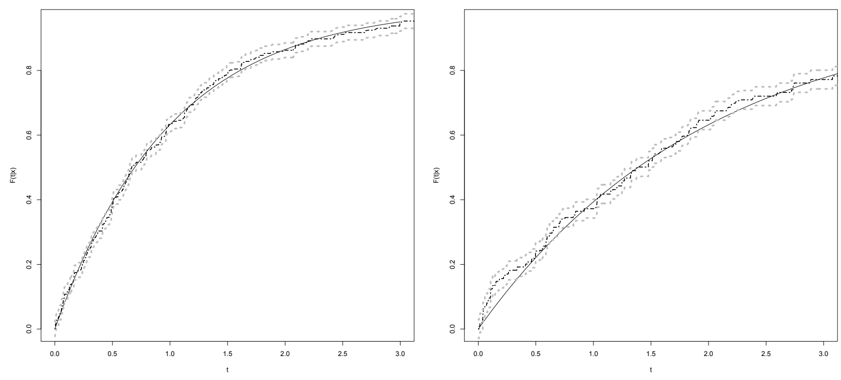

FIGURE 1. First Model: true conditional distribution function $F(t|x)$ in full line, estimated conditional distribution $\widehat{F}_n^{(1)}(t|x)$ in black dashed line, and certainty bands in grey dotted line, for $x = 0$ (left), and $x = 1$ (right).

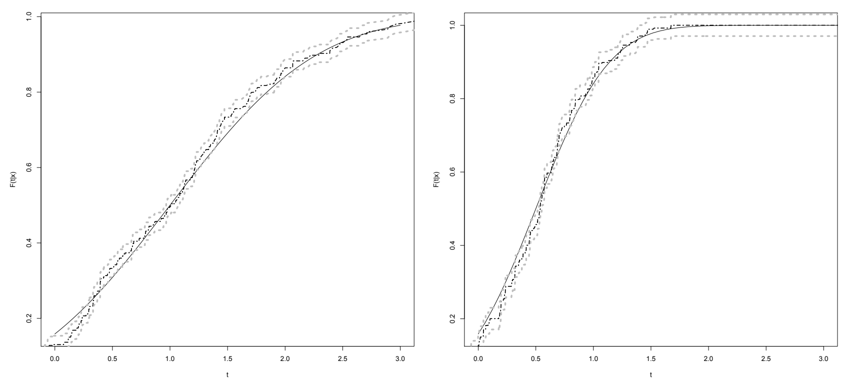

FIGURE 2. Second Model: true conditional distribution function $F(t|x)$ in full line, estimated conditional distribution $\widehat{F}_n^{(1)}(t|x)$ in black dashed line, and certainty bands in grey dotted line, for $x = 0$ (left), and $x = 1$ (right).

5. Proof of Theorem 2.1

We prove the probability convergence result, uniformly in $(x,t) \in I \times \mathbb{R}$. The uniform in $x \in I$ result (less difficult) is left to the reader. The almost surely convergence result needs some additional arguments, based on the Borel-Cantelli's Lemma (see for instance Blondin [2]).

Step 1:

In a first step, we introduce a general local empirical process. For any $j = 0, 1, 2$ and continuous real valued functions $c(\cdot)$ and $d(\cdot)$ on J, set for $x \in J$, $t \in \mathbb{R}$,

$$W_{n,j}(x,t) = \sum_{i=1}^{n} \left(c(x)\mathbb{1}_{\{Y_i \leqslant t\}} + d(x)\right) K_j\left(\frac{x - X_i}{h_n}\right) - n\mathbb{E}\left(\left(c(x)\mathbb{1}_{\{Y \leqslant t\}} + d(x)\right) K_j\left(\frac{x - X}{h_n}\right)\right) \tag{20}$$

where $K_j(u) = u^j K(u)$, for $j = 0, 1, 2$ and $u \in \mathbb{R}$.

For every fixed $t \in \mathbb{R}$, and $j = 0, 1, 2$, the process $W_{n,j}(\cdot, t)$ can be represented as a bivariate empirical process indexed by a class of functions. More precisely,

$$W_{n,j}(\cdot, t) = \sqrt{n}\alpha_n(g) = \sum_{i=1}^{n} \{g(X_i, Y_i) - \mathbb{E}\,(g(X, Y))\},$$

where α_n is the empirical process based upon $(X_1, Y_1), \ldots, (X_n, Y_n)$ and indexed by a suitable subclass $\mathcal{F}_{n,j}$ of the class of functions defined on $J \times \mathbb{R}$

$$\mathcal{F}_j = \left\{ (x, y) \mapsto \{c(z)1\!\!1_{y \leqslant t} + d(z)\}\, K_j \left(\frac{z - x}{h} \right) : t \in \mathbb{R}, z \in I, 0 < h < 1 \right\} \cdot$$

We give now the result from which our main Theorem 2.1 follows.

Theorem 5.1. *Under (F.1-2), (H.1-2), (K.1-3) we have,*

$$\sqrt{\frac{1}{2nh_n \log(h_n^{-1})}} \sup_{t\in\mathbb{R}} \sup_{x\in I} |W_{n,j}(x, t)| \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma_{W,j}(I)\cdot \tag{21}$$

Under (F.1-2), (H.2-4), (K.1-3) we have,

$$\sqrt{\frac{1}{2nh_n \log(h_n^{-1})}} \sup_{t\in\mathbb{R}} \sup_{x\in I} |W_{n,j}(x, t)| \xrightarrow[n\to+\infty]{} \sigma_{W,j}(I), \quad \textit{a.s. (almost surely)} \tag{22}$$

where

$$\sigma^2_{W,j}(I) = \sup_{t\in\mathbb{R}} \sup_{x\in I} \mathbb{E} \left(\left| c(x)1\!\!1_{\{Y \leqslant t\}} + d(x) \right|^2 |X = x \right) f_X(x) \int_{\mathbb{R}} K_j^2(u)du.$$

Proof: The proof is divided into an upper bound result, and a lower bound result.

Upper bound part: The proof of the upper bound result is divided into two steps. The hypothesis (F.2) is important in the upper bound part: we can found a real $M > 0$ such that if $X \in J$ then $|Y| \leqslant M$.

Step A: Discretization in $x \in I$ and $t \in [-M, M]$. First, we examine the behavior of our process $(x, t) \mapsto W_{n,j}(x, t)$ on an appropriate chosen grid of $I \times [-M, M]$, with increment δh_n for I and increment δ for $[-M, M]$, for fixed $0 < \delta < 1$:

$$\begin{cases} z_{i,n} = a + i\delta h_n, & i = 1, \ldots, i_n = \left[\frac{b-a}{\delta h_n}\right], \\ \\ t_l = -M + l\delta, & l = 1, \ldots, L = \left[\frac{2M}{\delta}\right], \end{cases}$$

where $[u] \leqslant u < [u] + 1$ represents the integer part of u.

The study of the supremum on $I \times [-M, M]$ is then reduced to the study of the maximum on a finite number of points. The empirical process is then indexed on the finite class of functions

$$\mathcal{F}_{n,j} = \left\{ (x, y) \mapsto \{c(z_{i,n})1\!\!1_{y \leqslant t_l} + d(z_{i,n})\}\, K_j \left(\frac{z_{i,n} - x}{h_n} \right) : i = 1, \ldots, i_n,\ l = 1, \ldots, L \right\} \cdot$$

The useful tool in this Step is the Bernstein inequality (see for instance Deheuvels and Mason [6]).

Step B: Oscillation. Next we study the behavior of our process between the grid points $z_{i,n}$ and t_l for $1 \leqslant i \leqslant i_n$ and $1 \leqslant l \leqslant L$. The objective of this Step is to study the maximal oscillation between the grid points. The useful tool in this Step is the Talagrand inequality for VC Classes (see for instance Talagrand [16], Einmahl and Mason [8] or Blondin [2]).

Few words about the almost surely convergence (blocking argument). To prove the almost surely convergence, we use the precedent Steps and the Borel-Cantelli's Lemma. The problem is simplified in a finite dimensional problem, by considering the empirical process indexed by finite subclasses $\mathcal{F}_{n_k,j}$, and for a subsequence $(n_k)_{k\geqslant 1}$. The almost sure convergence of the empirical processes is extended for processes $\sqrt{n}\alpha_n$ indexed by the whole $\mathcal{F}$, thanks to relative compacity technics.

Lower bound part: The lower bound result is proved with technical results based on Poisson processes, and needs the Bochner's Lemma. In this part, the hypothesis (F.1) is particularly important.

Few words about the almost surely convergence. The almost surely convergence is obtained with the same arguments, associated to the Borel-Cantelli's Lemma (see Einmahl and Mason [8]).

Step 2: We give now useful corollaries of Theorem 5.1. We give only the probability convergence results. The almost surely convergence results are similar. It suffice to take hypotheses (H.2-4) instead of hypotheses (H.1-2).

Corollary 5.2. *Under (F.1-2), (H.1-2) and (K.1-3), we have, by application of Theorem 5.1 with $c(x) = 0, d(x) = 1, j = 0, 1, 2$:*

$$\sqrt{\frac{nh_n}{2\log(h_n^{-1})}}\sup_{x\in I}|\widehat{f}_{n,j}(x) - f_{n,j}(x)| \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma_{f,j}(I), \tag{23}$$

where

$$\sigma^2_{f,j}(I) = \sup_{x\in I}\{f_X(x)\}\int_{\mathbb{R}} K_j(u)^2du = ||K_j||_2^2 \sup_{x\in I}\{f_X(x)\}\cdot$$

Corollary 5.3. *Under (F.1-2), (H.1-2) and (K.1-3), we have, by application of Theorem 5.1 with $c(x) = 1, d(x) = 0, j = 0, 1$:*

$$\sqrt{\frac{nh_n}{2\log(h_n^{-1})}}\sup_{t\in\mathbb{R}}\sup_{x\in I}|\widehat{r}_{n,j}(x,t) - r_{n,j}(x,t)| \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma_{r,j}(I) \tag{24}$$

where

$$\sigma^2_{r,j}(I) \quad = \quad ||K_j||_2^2 \sup_{t\in\mathbb{R}}\sup_{x\in I}\left\{\mathbb{E}\left(1\!\!1_{\{Y\leqslant t\}}|X=x\right)f_X(x)\right\}.$$

Moreover, under (F.1-2), (H.1) and (K.1-3), the Bochner's Lemma (cf for instance [8]) implies, uniformly in $(x,t)\in I\times\mathbb{R}$:

$$\begin{array}{ll} f_{n,0}(x) & = f_X(x)\mu_0(K) + o(1), \\ f_{n,1}(x) & = f_X(x)\mu_1(K) + o(1), \\ f_{n,2}(x) & = f_X(x)\mu_2(K) + o(1), \\ r_{n,0}(x,t) & = f_X(x)F(t|x) + o(1), \\ r_{n,1}(x,t) & = f_X(x)F(t|x)\mu_1(K) + o(1), \end{array} \tag{25}$$

where $\mu_j(K) = \int_{\mathbb{R}} K_j(u)du$, for $j = 0,1,2$. The hypotheses (K.2-3) imply that $\mu_0(K) = 1$, $\mu_1(K) = 0$ and $\mu_2(K) \neq 0$.

Step 3: In a third step, the deviation $\widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)$ can be asymptotically expressed as a linear function of the bivariate empirical process.

$$
\begin{aligned}
&\widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right) \\
&= \frac{\widehat{r}_{n,0}(x,t)\widehat{f}_{n,2}(x) - \widehat{f}_{n,1}(x)\widehat{r}_{n,1}(x,t)}{\widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}^2_{n,1}(x)} - \frac{r_{n,0}(x,t)f_{n,2}(x) - f_{n,1}(x)r_{n,1}(x,t)}{f_{n,2}(x)f_{n,0}(x) - f^2_{n,1}(x)}, \qquad (26) \\
&= \frac{\widehat{r}_{n,0}(x,t)\widehat{f}_{n,2}(x) - r_{n,0}(x,t)f_{n,2}(x) + f_{n,1}(x)r_{n,1}(x,t) - \widehat{f}_{n,1}(x)\widehat{r}_{n,1}(x,t)}{\widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}^2_{n,1}(x)} \qquad (27) \\
&\quad + \frac{\left(r_{n,0}(x,t)f_{n,2}(x) - f_{n,1}(x)r_{n,1}(x,t)\right)\left(f_{n,2}(x)f_{n,0}(x) - f^2_{n,1}(x) - \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) + \widehat{f}^2_{n,1}(x)\right)}{\left(\widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}^2_{n,1}(x)\right)\left(f_{n,2}(x)f_{n,0}(x) - f^2_{n,1}(x)\right)} \qquad (28)
\end{aligned}
$$

First, studying the numerator of the expression (27), we have

$$
\begin{aligned}
\mathrm{Num}(27) \;=\; & \left(\left(\widehat{f}_{n,2}(x) - f_{n,2}(x)\right) + \left(f_{n,2}(x) - f_X(x)\mu_2(K)\right) + f_X(x)\mu_2(K)\right)\left(\widehat{r}_{n,0}(x,t) - r_{n,0}(x,t)\right) \\
& + \left(\left(r_{n,0}(x,t) - f_X(x)F(t|x)\right) + f_X(x)F(t|x)\right)\left(\widehat{f}_{n,2}(x) - f_{n,2}(x)\right) \\
& + f_{n,1}(x)\left(r_{n,1}(x,t) - \widehat{r}_{n,1}(x,t)\right) \\
& + \left(\left(\widehat{r}_{n,1}(x,t) - r_{n,1}(x,t)\right) + r_{n,1}(x,t)\right)\left(f_{n,1}(x) - \widehat{f}_{n,1}(x)\right).
\end{aligned}
$$

Let $\alpha_n = \sqrt{\dfrac{nh_n}{2\log(h_n^{-1})}} \longrightarrow +\infty$ as $n \longrightarrow +\infty$. Combining with previous corollaries and results (25), we see that

$$
\begin{aligned}
&\alpha_n \sup_{t\in\mathbb{R}} \sup_{x\in I} \left|\mathrm{Num}(27) - \widehat{g}_n(x,t)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0 \\
&\text{where } \widehat{g}_n(x,t) = f_X(x)\mu_2(K)\left(\widehat{r}_{n,0}(x,t) - r_{n,0}(x,t)\right) - f_X(x)F(t|x)\left(\widehat{f}_{n,2}(x) - f_{n,2}(x)\right)
\end{aligned} \qquad (29)
$$

Now, we show that the denominator of the expression (27), now denoted Den(27), satisfies

$$
\sup_{x\in I}\left|\mathrm{Den}(27) - \frac{1}{f_X(x)^2\mu_2(K)}\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0 \qquad (30)
$$

Indeed, for all $\epsilon > 0$, let the event

$$
\mathcal{A}_\epsilon = \left\{\sup_{x\in I}\left|\mathrm{Den}(27) - \frac{1}{f_X(x)^2\mu_2(K)}\right| > \epsilon\right\},
$$

and for all $B > 0$, let the event

$$\mathcal{B} = \left\{ \sup_{x \in I} \left| \left(\widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 \right) f_X(x)^2 \mu_2(K) \right| > B \right\} \cdot$$

Then,

$$\mathbb{P}(A_\epsilon) = \mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}) + \mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}^c)$$

and

$$\mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}^c) \leqslant \mathbb{P}(\mathcal{B}^c) \leqslant \mathbb{P}\left(\sup_{x \in I} \left| \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 \right| \leqslant \frac{B}{\inf_{x \in I} f_X(x)^2 \mu_2(K)} \right),$$

taking $B = \frac{1}{2}\left(\inf_{x \in I} f_X(x)^2 \mu_2(K)\right)^2$, we obtain

$$\mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}^c) \leqslant \mathbb{P}\left(\sup_{x \in I} \left| \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 - f_X(x)^2 \mu_2(K) \right| \geqslant \frac{1}{2} \inf_{x \in I} f_X(x)^2 \mu_2(K) \right) \xrightarrow[n \to +\infty]{} 0 \cdot$$

This last limit is obtained by the following trivial decomposition

$$\begin{aligned}
&\widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 - f_X(x)^2 \mu_2(K) = \\
&\Big(\big(\widehat{f}_{n,2}(x) - f_{n,2}(x)\big) + \big(f_{n,2}(x) - f_X(x)\mu_2(K)\big) + f_X(x)\mu_2(K) \Big) \\
&\cdot \Big(\big(\widehat{f}_{n,0}(x) - f_{n,0}(x)\big) + \big(f_{n,0}(x) - f_X(x)\big) \Big) \\
&+ \Big(\big(\widehat{f}_{n,2}(x) - f_{n,2}(x)\big) + \big(f_{n,2}(x) - f_X(x)\mu_2(K)\big) \Big) f_X(x),
\end{aligned}$$

and by applying the corollary 5.2 and results (25), combined with the boundedness property of f_X on I.

Now,

$$\mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}) \leqslant \mathbb{P}\left(\sup_{x \in I} \left| \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 - f_X(x)^2 \mu_2(K) \right| > \epsilon B \right) \xrightarrow[n \to +\infty]{} 0$$

for the same reason as before. We have then proved (30).

Combining (29) and (30), we have:

$$\alpha_n \sup_{t \in \mathbb{R}} \sup_{x \in I} \left| (27) - \frac{\widehat{r}_{n,0}(x,t) - r_{n,0}(x,t)}{f_X(x)} - \frac{F(t|x)}{f_X(x)\mu_2(K)} \left(\widehat{f}_{n,2}(x) - f_{n,2}(x) \right) \right| \xrightarrow[n \to +\infty]{\mathbb{P}} 0 \cdot \qquad (31)$$

This last limit is due to the following lemma:

Lemma 5.4. *Let $\mathcal{I} \subset \mathbb{R}^d$, with $d \in \mathbb{N}^*$, and X_n, Z_n, Y_n and Y random functions defined on $\mathcal{I}$ such that*

$$\sup_{w \in \mathcal{I}} |X_n(w) - Z_n(w)| \xrightarrow[n \to +\infty]{\mathbb{P}} 0, \qquad \sup_{w \in \mathcal{I}} |Y_n(w) - Y(w)| \xrightarrow[n \to +\infty]{\mathbb{P}} 0,$$

and for enough high $B > 0$, $\lim_{n \to \infty} \mathbb{P}(\sup_{w \in \mathcal{I}} |X_n(w)| > B) = 0$ and Y is bounded on $\mathcal{I}$. Then,

$$\sup_{w \in \mathcal{I}} |X_n(w)Y_n(w) - Z_n(w)Y(w)| \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

Proof:

$$\sup_{\mathcal{I}} |X_n Y_n - Z_n Y| \leqslant \sup_{\mathcal{I}} |X_n| \sup_{\mathcal{I}} |Y_n - Y| + \sup_{\mathcal{I}} |Y| \sup_{\mathcal{I}} |X_n - Z_n|,$$

then for all $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{\mathcal{I}} |X_n Y_n - Z_n Y| > \epsilon\right) \leqslant \mathbb{P}\left(\sup_{\mathcal{I}} |X_n| \sup_{\mathcal{I}} |Y_n - Y| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\sup_{\mathcal{I}} |Y| \sup_{\mathcal{I}} |X_n - Z_n| > \frac{\epsilon}{2}\right).$$

Introducing the event $\mathcal{B} = \{\sup_{\mathcal{I}} |Y| > B\}$ with $B > 0$, we bound the second term

$$\mathbb{P}\left(\sup_{\mathcal{I}} |Y| \sup_{\mathcal{I}} |X_n - Z_n| > \frac{\epsilon}{2}\right) \leqslant \mathbb{P}(\mathcal{B}) + \mathbb{P}\left(\sup_{\mathcal{I}} |X_n - Z_n| > \frac{\epsilon}{2B}\right) \xrightarrow[n\to+\infty]{} 0$$

for enough high B. Now, we bound the first term

$$\mathbb{P}\left(\sup_{\mathcal{I}} |X_n| \sup_{\mathcal{I}} |Y_n - Y| > \frac{\epsilon}{2}\right) \leqslant \mathbb{P}\left(\sup_{w\in\mathcal{I}} \sup_{n\in\mathbb{N}} |X_n(w)| > B\right) + \mathbb{P}\left(\sup_{\mathcal{I}} |Y_n - Y| > \frac{\epsilon}{2B}\right)$$
$$\leqslant \mathbb{P}\left(\sup_{w\in\mathcal{I}} |X_n| > B\right) + \mathbb{P}\left(\sup_{\mathcal{I}} |Y_n - Y| > \frac{\epsilon}{2B}\right) \xrightarrow[n\to+\infty]{} 0$$

for enough high B. $\square$

Let's studying now the second part (28) in the expression of $\widehat{F}_n^{(1)}(t|x) - \widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)$. Note that the numerator is equal to:

$$(r_{n,0}(x,t) f_{n,2}(x) - f_{n,1}(x) r_{n,1}(x,t)) \left(f_{n,2}(x) f_{n,0}(x) - f_{n,1}^2(x) - \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) + \widehat{f}_{n,1}^2(x)\right)\cdot$$

The first term $\mathrm{Num}_1(28) = r_{n,0}(x,t) f_{n,2}(x) - f_{n,1}(x) r_{n,1}(x,t)$ of this last expression converges uniformly, thanks to (25):

$$\sup_{t\in\mathbb{R}} \sup_{x\in I} \left|\mathrm{Num}_1(28) - f_X(x)^2 \mu_2(K) F(t|x)\right| \xrightarrow[n\to+\infty]{} 0. \tag{32}$$

With the same arguments as in the study of the numerator of (27), we study the second term

$$\mathrm{Num}_2(28) = f_{n,2}(x) f_{n,0}(x) - f_{n,1}^2(x) - \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) + \widehat{f}_{n,1}^2(x)$$

and show that

$$\alpha_n \sup_{x\in I} \left|\mathrm{Num}_2(28) - f_X(x)\mu_2(K)\left(f_{n,0}(x) - \widehat{f}_{n,0}(x)\right) - f_X(x)\left(f_{n,2}(x) - \widehat{f}_{n,2}(x)\right)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0\cdot \tag{33}$$

Thanks to the boundedness property of f_X on I, and the corollary 5.2, we have, by the Lemma 5.4:

$$\alpha_n \sup_{t\in\mathbb{R}} \sup_{x\in I} \left|\mathrm{Num}(28) - \widehat{j}_n(x,t)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0$$
$$\text{where } \widehat{j}_n(x,t) = f_X(x)^3 \mu_2(K) F(t|x) \left(\mu_2(K)\left(f_{n,0}(x) - \widehat{f}_{n,0}(x)\right) + \left(f_{n,2}(x) - \widehat{f}_{n,2}(x)\right)\right) \tag{34}$$

The denominator in (28) can be expressed as

$$\mathrm{Den}(28) = \mathrm{Den}(27) \frac{1}{f_{n,2}(x) f_{n,0}(x) - f_{n,1}(x)^2}.$$

It is clear that, thanks to (25)

$$\sup_{x\in I}\left|\frac{1}{f_{n,2}(x)f_{n,0}(x)-f_{n,1}(x)^2}-\frac{1}{f_X(x)^2\mu_2(K)}\right| \xrightarrow[n\to+\infty]{} 0\cdot$$

Then, tanks to the boundedness property of f_X on I, the Lemma 5.4 says that

$$\sup_{x\in I}\left|\mathrm{Den}(28)-\frac{1}{f_X(x)^4\mu_2(K)^2}\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0\cdot \tag{35}$$

Finally, we have, thanks to the Lemma 5.4

$$\alpha_n\sup_{t\in\mathbb{R}}\sup_{x\in I}\left|(28)-\frac{F(t|x)}{f_X(x)}\left(f_{n,0}(x)-\widehat{f}_{n,0}(x)\right)-\frac{F(t|x)}{f_X(x)\mu_2(K)}\left(f_{n,2}(x)-\widehat{f}_{n,2}(x)\right)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0\cdot \tag{36}$$

Remember that $\widehat{F}_n^{(1)}(t|x)-\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)=(27)+(28)$. Then, combining (31) and (36) it is easy to see finally that

$$\alpha_n\sup_{t\in\mathbb{R}}\sup_{x\in I}\left|\widehat{F}_n^{(1)}(t|x)-\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)+\frac{F(t|x)}{f_X(x)}\left(\widehat{f}_{n,0}(x)-f_{n,0}(x)\right)-\frac{\widehat{r}_{n,0}(t,x)-r_{n,0}(t,x)}{f_X(x)}\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0\cdot \tag{37}$$

Step 4:

Now choosing $c(x)=\dfrac{1}{f_X(x)}$ and $d(x)=-\dfrac{F(t|x)}{f_X(x)}$ in the definition of $W_{n,0}(x,t)$, the local empirical process, it is easy to show that

$$\begin{aligned} W_{n,0}(x,t) &= \frac{nh}{f_X(x)}\widehat{r}_{n,0}(x,t)-\frac{F(t|x)}{f_X(x)}\times nh\widehat{f}_{n,0}(x)-\frac{nh}{f_X(x)}r_{n,0}(x,t)+\frac{F(t|x)}{f_X(x)}\times nhf_{n,0}(x)\\ &= \frac{nh}{f_X(x)}\left[\widehat{r}_{n,0}(x,t)-r_{n,0}(x,t)-F(t|x)(\widehat{f}_{n,0}(x)-f_{n,0}(x))\right]\cdot \end{aligned}$$

Let $A_n=\sqrt{2nh_n\log(h_n^{-1})}^{-1}$, so

$$A_nW_{n,0}(x,t)=\alpha_n\left(\frac{\widehat{r}_{n,0}(t,x)-r_{n,0}(t,x)}{f_X(x)}-\frac{F(t|x)}{f_X(x)}\left(\widehat{f}_{n,0}(x)-f_{n,0}(x)\right)\right). \tag{38}$$

Then

$$\begin{aligned} &\alpha_n\left(\widehat{F}_n^{(1)}(t|x)-\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)\right)=A_nW_{n,0}(x,t)\\ &+\alpha_n\left(\widehat{F}_n^{(1)}(t|x)-\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)+\frac{F(t|x)}{f_X(x)}\left(\widehat{f}_{n,0}(x)-f_{n,0}(x)\right)-\frac{\widehat{r}_{n,0}(t,x)-r_{n,0}(t,x)}{f_X(x)}\right), \end{aligned}$$

and applying Theorem 5.1

$$\sup_{t\in\mathbb{R}}\sup_{x\in I}\left|\alpha_n\left(\widehat{F}_n^{(1)}(t|x)-\widetilde{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t|x)\right)\right)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma_F(I) \tag{39}$$

with

$$\begin{aligned}
\sigma_F^2(I) &= \sup_{t\in\mathbb{R}}\sup_{x\in I}\mathbb{E}\left(\left[\frac{1\!\!1_{\{Y\leqslant t\}}}{f_X(x)}-\frac{F(t|x)}{f_X(x)}\right]^2|X=x\right)f_X(x)||K||_2^2\\
&= \sup_{t\in\mathbb{R}}\sup_{x\in I}\frac{\mathbb{E}\left(\left[1\!\!1_{\{Y\leqslant t\}}-\mathbb{E}(Y\leqslant t|X=x)\right]^2|X=x\right)}{f_X(x)}||K||_2^2\\
&= \sup_{t\in\mathbb{R}}\sup_{x\in I}\left(\frac{F(t|x)(1-F(t|x))}{f_X(x)}\right)||K||_2^2.
\end{aligned}$$

This finishes the proof of Theorem 2.1.

References

[1] A. Berlinet, L. Devroye, *A comparison of kernel density estimates*, Publ. Inst. Statist. Univ. Paris, 38 (1994), pp. 3-79.

[2] D. Blondin, *Lois limites uniformes et estimation non paramétrique de la régression*, PhD thesis, Université de Paris 6, France, 2004.

[3] G. Collomb, *Estimation non-paramétrique de la régression: revue bibliographique*, Int. Statist. Rev., 49 (1981), pp. 75-93.

[4] P. Deheuvels, *Estimation non-paramétrique de la densité par histogramme généralisé*, La Revue de Statistique Appliquée, 35 (1977), pp. 5-42.

[5] P. Deheuvels, G. Derzko, *Asymptotic certainty bands for kernel density estimators based upon a bootstrap resampling scheme*, Statistical Models and Methods for Biomedical and Technical Systems, 3 (2008), pp. 171-186.

[6] P. Deheuvels, D.M. Mason, *General asymptotic confidence bands based on kernel-type function estimators*, Statist. Infer. Stochastic Process, 7(3) (2004), pp. 225-277.

[7] L. Devroye,*The uniform convergence of the Nadaraya-Watson regression function estimate*, Can. J. Statist., 6 (1978), pp. 179-191.

[8] U. Einmahl, D.M. Mason, *An empirical process approach to the uniform consistency of kernel-type function estimators*, J. Theoret. Probab., 13(1) (2000), pp. 1-37.

[9] U. Einmahl, D.M. Mason, *Uniform in bandwidth consistency of kernel-type function estimators*, Ann. Stat. 33(3) (2005), pp. 1380-1403.

[10] J. Fan, I. Gijbels, *Local polynomial modeling and its applications*. Monographs on Statistics and Applied Probability, Chapman and Hall, Vol. 66, 1996.

[11] A. Gannoun, S. Girard, C. Guinot,J. Saracco, *References curves based on nonparametric quantile regression*, Statistics in Medecine, 21 (2002), pp. 3119-3155.

[12] W. Härdle, *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990.

[13] M. Mint El Mouvid, *Sur l'estimateur linéaire local de la fonction de répartition conditionnelle*, PhD thesis, Université de Montpellier 2, France, 2000.

[14] E.A. Nadaraya, *On estimating regression*, Theor. Probab. Appl. 9 (1964), pp. 141-142.

[15] W. Stute, *On Almost Sure Convergence of Conditional Empirical Distribution Functions*, Ann. Probab. 14(3) (1986), pp. 891-901.

[16] M. Talagrand, *Sharper bounds for Gaussian and empirical processes*, Ann. Probab. 22 (1994), pp. 28-76.

[17] G.S. Watson, *Smooth regression analysis*. Sankhyā Ser.A 26 (1964), pp. 359-372.