STATISTICS IN TRANSITION new series, June 2023
Vol. 24, No. 3, pp. 117–137, https://doi.org/10.59170/stattrans-2023-039
Received – 13.01.2022; accepted – 12.01.2023

Conditional density function for surrogate scalar response

Mounir Boumahdi [1], Idir Ouassou [2], Mustapha Rachdi [3]

Abstract

This paper presents the estimator of the conditional density function of surrogated scalar response variable given a functional random one. We construct a conditional density function by using the available (true) response data and the surrogate data. Then, we build up some asymptotic properties of the constructed estimator in terms of the almost complete convergences. As a result, we compare our estimator with the classical estimator through the Relatif Mean Square Errors (RMSE). Finally, we end this analysis by displaying the superiority of our estimator in terms of prediction when we are lacking complete data.

Key words: Density function, surrogate response, functional variable, almost complete convergence, kernel estimators, scalar response, entropy, semi-metric space.

1. Introduction

There are many situations that may study the link between two variables, with the main goal to be able to predict new values. This predicted problem has been widely studied in the literature when both variables are of finite dimensions. Of course, the same problem can occur when some of the variables are functional. Our aim is to investigate this problem when the explanatory variable is functional and the response one is still real.
We are based in the following model:

$$Y = m(X) + \varepsilon. \tag{1}$$

Where m is the regression operator, X is a functional covariate which belongs to a semi-metric space (E,d) and Y is the response variable, ε is a random error.
Our goal is to build the conditional density function for surrogate data by using the true response data and the surrogate data. By following the work of Wang (2006), Firas et al. (2019) and based on the work of Ramsay and Silverman (2002), Ferraty and Vieu (2006), Horvath and Kokoszka (2012), Cuevas (2014), Zhang (2014), Bongiorno et al. (2014), Hsing and Eubank (2015), Goia, Vieu (2016) and Wang, Chiou, and Müller (2016) and the references therein, we construct our estimator $\hat{f}_R^X(y)$.

The problem we are addressing in this work i.e., the unavailability of some data in the response variable, can be motivated both from a practical and a theoretical point of view.

[1]University Cadi Ayyad, The Eminent University of Science and Knowledge, Marrakech, Morocco. E-mail:mounirboumahdi1210112999@gmail.com. ORCID:https://orcid.org/0000-0002-1017-3760.
[2]University Cadi Ayyad, The Eminent University of Science and Knowledge, Marrakech, Morocco. E-mail:i.ouassou@uca.ac.ma. ORCID:https://orcid.org/0000-0001-5573-5053.
[3]Université Grenoble Alpes, Grenoble Cedex 09, France.
E-mail: MUSTAPHA.RACHDI@univ-grenoble-alpes.fr. ORCID:https://orcid.org/0000-0003-2887-4212.

In fact, it may be difficult or expensive to exactly measure some response observations Y. Our goal is then to improve the modeling by filling/recovering some of the information missed in the response variable with this surrogate variable. In this case, one solution is to use the help of validation data to capture the underlying relation between the true variables and surrogate ones. Some examples where validation data are available can be found in Duncan and Hill (1985), Carroll and Wand (1991) and Pepe (1992).

This paper aims to study the conditional density for missing response by the kernel method, we explore in this work the aspect of missing data in the response variable to estimate the conditional density function for surrogate data. We adopt an approach based on validation data ideas. In fact, the idea is to introduce the information contained in both the validation data and the surrogate data.

The unavailable observation of Y will be replaced by the estimator of $\mathbb{E}\left(Y \mid X, \tilde{Y}\right)$, denoted by $U(X_j, \tilde{Y}_j)$ for all $j \in \bar{V}$ that corresponds to the size of the missing data, where $\tilde{Y}$ is surrogate variable of Y. To estimate $\mathbb{E}\left(Y \mid X, \tilde{Y}\right)$ we adopt an approach based on validation data and the brut data (the primary data), which includes surrogate data and the corresponding observations of the covariate X.

Inside the simulation study of section 4, the surrogate variable $\tilde{Y}_i$ of Y_i was generated from $\tilde{Y}_i = \rho Z_i + \varepsilon_i$, where Z_i is the standard score of Y_i and $\varepsilon_i \sim N\left(0, \sqrt{1-\rho^2}\right)$, in such a way that the correlation coefficient between Y_i and $\tilde{Y}_i$ is approximately equal to p which would not be controllable in practice, but we can clearly notice that the quality of our $\hat{f}_R^x$ depends on the size n of the validation data and ρ. Specifically, our estimator greatly better as the value of n and ρ increases.

We already know the convergence almost complete of the classical kernel estimator $\hat{f}_C^x(y)$(Ferraty and Vieu (2006)) towards the real $f_Y^x(y)$, In fact, within the section 4, we calculated and represented graphically the conditional density function estimator for surrogate data and we conduct a computational study on a simulated data in order to show advantages of using $\hat{f}_R^x(y)$ over $\hat{f}_V^x(y)$.

Effectively, we are in a position to give the alternative estimator of $\hat{f}_C^x(y)$ (estimator of Ferraty and Vieu) when we are lacking complete data with the help of $\tilde{Y}$(the surrogate variable of Y), so in reality the choice of $\tilde{Y}$ is important to improve the quality of our estimator. In practice we can cite as an example two diseases (Y and $\tilde{Y}$) presenting similar symptoms, more that there is a strong correlation between these two diseases, more our estimator is better. So, there exists a wide scope of applied scientific fields for which our approach could be of interest for examples Biometrics, Genetics or Environmetrics and this approach can be helpful for a lot of statistical models when we are lacking complete data.

The main objective of this paper is to estimate the conditional density function for surrogate data. Then, we present the almost complete convergence of our estimator $\hat{f}_R^X(y)$ and we study its performance against $\hat{f}_V^X(y)$ by computing the relative mean squared error (RMSE) using simulated data.

2. Estimation procedure

Let $(X,Y) \in \mathscr{F} \times \mathbb{R}$ denote a random vector, where $(\mathscr{F},d)$ is a semi-metric space equipped with the semi-metric d. We are concerned with the estimation of the conditional density function for surrogate data. Therefore, let $(X_1,Y_1),...,(X_N,Y_N)$ be a random sample consisting of independent and identically distributed (i.i.d) variable from the distribution of (X,Y).

The regression function for surrogate data defined in [**?**] as follows

$$\hat{m}_R(x) = \sum_{i\in V} Y_i W_{1,n,i}(x) + \sum_{j\in\bar{V}} U(X_j,\tilde{Y}_j) W_{1,n,j}(x), \tag{2}$$

with

$$U(X_j,\tilde{Y}_j) = \sum_{i\in V} Y_i W_{2,n,i}(X_j,\tilde{Y}_j), \qquad \forall j \in \bar{V}. \tag{3}$$

We can estimate the conditional c.d.f $F_Y^x(y)$ for surrogate data as follows

$$\hat{F}_R^x(y) = \sum_{i\in V} H\left(\frac{y-Y_i}{g}\right) W_{1,n,i}(x) + \sum_{j\in\bar{V}} R(X_j,y,\tilde{Y}_j) W_{1,n,j}(x), \tag{4}$$

where

$$W_{1,n,i}(x) = \frac{K\left(\dfrac{d(X_i,x)}{h}\right)}{\sum_{l=1}^{N} K\left(\dfrac{d(X_l,x)}{h}\right)}, \tag{5}$$

and

$$R(X_j,y,\tilde{Y}_j) = \sum_{i\in V} H\left(\frac{y-Y_i}{g}\right) W_{2,n,i}(X_j,\tilde{Y}_j), \qquad \forall j \in \bar{V}. \tag{6}$$

With

$$W_{2,n,i}(X_j,\tilde{Y}_j) = \frac{W\left(\dfrac{d(X_j,X_i)}{h}, \dfrac{\tilde{Y}_i-\tilde{Y}_j}{b}\right)}{\sum_{l\in V} W\left(\dfrac{d(X_j,X_l)}{h}, \dfrac{\tilde{Y}_l-\tilde{Y}_j}{b}\right)}. \tag{7}$$

The conditional density function can be obtained by derivating the conditional c.d.f. Since we have now at hand some estimator $\hat{F}_R^x(y)$ of $F_Y^x(y)$, it is natural to propose the following estimate:

$$\hat{f}_R^x(y) = \frac{\partial \hat{F}_R^x(y)}{\partial y}.$$

Assuming the differentiability of H, we build our new estimator of conditional density function for surrogate data as following:

$$\hat{f}_R^x(y) = \sum_{i\in V} \Omega_i(y) W_{1,n,i}(x) + \sum_{j\in\bar{V}} L(X_j,y,\tilde{Y}_j) W_{1,n,j}(x).$$

Where

$$\Omega_i(y) = g^{-1} K_0(g^{-1}(y-Y_i)),$$

and

$$L(X_j, y, \tilde{Y}_j) = \sum_{i \in V} g^{-1} K_0(g^{-1}(y - Y_i)) W_{2,n,i}(X_j, \tilde{Y}_j), \qquad \forall j \in \bar{V}. \tag{8}$$

Where K is a kernel function and both $h = h_N$ and $g = g_N$ are a sequence of positive reals that tends to zero when N goes to infinity.

$$\forall u \in \mathbb{R}, \quad H(u) = \int_{-\infty}^{u} K_0(v) dv. \tag{9}$$

K_0 is a function from $\mathbb{R}$ into $\mathbb{R}^+$ such that $\int K_0 = 1$. To give an estimator of F_Y^x when there are surrogate data in the response variable, let us introduce the integer n $(n < N)$ that corresponds to the size of the validation set V. Let $\bar{V}$ be the complementary set of V in the set $\{1, 2, ..., N\}$.

W is a kernel function which is defined on $\mathbb{R}^2$ and b is sequence of real numbers which tend to zero. To simplify, we will use only one kernel. In sense that $K = K_0$ and $W(\cdot, \cdot) = K(\cdot)K(\cdot)$. This consideration is because the choice of the kernel has less influence on the performance of the estimator.

3. Some asymptotic properties

In the sequel, when no confusion is possible, we will denote by C and C' some strictly positive generic constants, we denote by $f^{x_1, \tilde{y_1}}(y)$ the conditional distribution function of Y given $(X, \tilde{Y})$:

$$f^{x_1, \tilde{y_1}}(y) = \frac{\partial F^{x_1, \tilde{y_1}}(y)}{\partial y},$$

with

$$F^{x_1, \tilde{y_1}}(y) = P(Y \leq y \mid x_1, \tilde{y_1}).$$

Recall that a semi-metric (sometimes called pseudo-metric) is just a metric violating the property $[d(x,y) = 0] \Rightarrow [x = y]$. We define the Kolmogorov's entropy as follows:

Definition 3.1 *Let $S_{\mathscr{F}}$ be a subset of a semi-metric space $\mathscr{F}$, and let $\varepsilon > 0$ be given. A finite set of point $x_1, x_2, ..., x_{n_0}$ in $\mathscr{F}$ is called an ε-net for $S_{\mathscr{F}}$ if $S_{\mathscr{F}} \subset \bigcup_{k=1}^{N_0} B(x_k, \varepsilon)$. The quantity $\psi_{S_{\mathscr{F}}} = \log(N_\varepsilon)$, where N_ε s the minimal number of open balls in $\mathscr{F}$ of radius ε which is necessary to cover S, is called the Kolmogorov's ε-entropy of the set $S_{\mathscr{F}}$.*

This concept was introduced by Kolmogorov (see, Kolmogorov and Tikhomirov, 1959) and it represents a measure of the complexity of a set, in sense that, high entropy means that much information is needed to describe an element with an accuracy ε. Therefore, the choice of the topological structure (with other words, the choice of the semi-metric) will play a crucial role when one is looking the uniform (over S) asymptotic results. For more precision about this concept, see Ferraty et al. (2010).

We consider the following assumptions:

(H1) For all x in the subset $S_{\mathscr{F}}$ we have,

$$0 < C\phi(h) \leq P(X \in B(x,h)) \leq C'\phi(h) < \infty.$$

For all $\tilde{y}$ in the subset $S_{\mathscr{R}}$

$$0 < C\phi(b) \leq P(Y \leq \tilde{y} \leq Y + b) \leq C'\phi(b) < \infty.$$

For all $x, \tilde{y}$ in the subset $S_{\mathscr{F}} \times S_{\mathscr{R}}$

$$C\phi(h)\phi(b) < \mathbb{E}[K(h^{-1}d(x,X_i))K(b^{-1}(\tilde{y} - Y_1)] < C'\phi(h)\phi(b).$$

(H2) There exists $b_1, b_2, b_3 > 0$ such that $\forall x_1, x_2 \in S_{\mathscr{F}}$, $\forall y_1, y_2 \in S_{\mathscr{R}}$ and $\forall \tilde{y_1}, \tilde{y_2} \in S_{\mathscr{R}}$

$$|f^{x_1}(y_1) - f^{x_2}(y_1)| \leq C\left(d^{\beta_1}(x_1,x_2) + |y_1 - y_2|^{\beta_2}\right),$$

$$\text{and} \quad \left|f^{x_1,\tilde{y_1}}(y_1) - f^{x_2,\tilde{y_2}}(y_1)\right| \leq C\left(d^{\beta_1}(x_1,x_2) + |y_1 - y_2|^{\beta_2} + |\tilde{y_1} - \tilde{y_2}|^{\beta_3}\right).$$

(H3) K and K_0 are bounded and Lipschitz kernel on its support $[0,1]$, such that $-\infty < C < K'(t) < C' < 0$.

(H4) The functions ϕ and $\psi_{S_{\mathscr{F}}}$ are such that:

(H4a) $\exists C > 0, \exists \eta_0 > 0, \forall \eta < \eta_0, \ \phi'(\eta) < C$, and

$$\exists C > 0, \exists \eta_0 > 0, \ \forall 0 < \eta < \eta_0, \ \int_0^{\eta} \phi(u)\,du > C\,\eta\,\phi(\eta),$$

(H4b) For some $\gamma \in (0,1)$, $\gamma' \in (0,1)$ and $\gamma^{"} \in (0,1)$ $\lim_{n\to+\infty} n^{\gamma}h = \infty$, ,$\lim_{n\to+\infty} n^{\gamma'}g = \infty$ and $\lim_{n\to+\infty} n^{\gamma^{"}}b = \infty$, and for n large enough:

$$\frac{(\log n)^2}{ng\phi(h)} < \frac{(\log n)^2}{ng\phi(b)\phi(h)} < \psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right) < \frac{ng\phi(b)\phi(h)}{\log n} < \frac{ng\phi(h)}{\log n}.$$

(H5) The Kolmogorov's ε-entropy of $S_{\mathscr{F}}$ satisfies

$$\sum_{n=1}^{\infty} n^{2\gamma+1} \exp\left\{(1-\beta)\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)\right\} < \infty, \text{ for some } \beta > 1,$$

and

$$\sum_{n=1}^{\infty} n^{2\gamma^{"}+1} \exp\left\{(1-\beta)\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)\right\} < \infty, \text{ for some } \beta > 1.$$

Note that (H4a) implies that for n large enough

$$0 \leq \phi(h) \leq Ch. \tag{10}$$

The condition (H4b) implies that:

$$\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)} \to 0, \text{ and } \frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(b)\phi(h)} \to 0. \tag{11}$$

The condition (H4b) implies that:

$$\sum_{n=1}^{\infty} n^{2\gamma''+1} N_{\varepsilon}(S_{\mathscr{F}})^{1-\beta} < \infty, \text{ and } \sum_{n=1}^{\infty} n^{2\gamma+1} N_{\varepsilon}(S_{\mathscr{F}})^{1-\beta} < \infty. \tag{12}$$

Conditions (H2)-(H3) are very standard in the nonparametric setting. Concerning (H4a), the boundness of the derivative of ϕ around zero allows to consider ϕ as a Lipschitzian function. Hypothesis (H4b) deals with topological considerations by controlling the entropy of $S_{\mathscr{F}}$. For a radius not too large, one requires that $\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)$ is not too small and not too large. Moreover (H4b) implies that $\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)} \to 0$ *and* $\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)\phi(b)} \to 0$ tends to 0 when n tends to $+\infty$, in some "usual" cases, one has $\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right) \sim C \log n$. The assumption (H5) acts on the Kolmogorov ε-entropy of $S_{\mathscr{F}}$.

The following Theorem states the rate of convergence of $\hat{f}_R^x$ for the surrogated scalar response, uniformly over the set $S_{\mathscr{F}}$ and $S_{\mathscr{R}}$. The asymptotics are stated in terms of almost complete convergence (denoted by a.co.), which imply both weak and strong convergences (see Section A-1 in Ferraty and Vieu, 2006).

Theorem 3.1 *Under the hypotheses (H1)-(H5), we have*

$$\sup_{x\in S_{\mathscr{F}}} \sup_{y\in S_{\mathscr{R}R}} |\hat{f}_R^x(y) - f_Y^x(y)| = O(h^{\beta_1}) + O(g^{\beta_2})$$

$$+ O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)}{ng\phi(h)}}\right) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log N}{N}\right)}{Ng\phi(h)}}\right)$$

$$+ O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)\phi(b)}}\right) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)}{ng\phi(h)\phi(b)}}\right).$$

4. Numerical results

In this section, we evaluate the performance of the proposed estimator by conducting a number of simulation studies. Let $\hat{f}_V^x(y)$ be the classical conditional density function estimator which is obtained with the true observations in the validation data set V:

$$\hat{f}_V^x(y) = \frac{\sum_{i\in V} K\big(h^{-1}d(x,X_i)\big) g^{-1} K_0\big(g^{-1}(y-Y_i)\big)}{\sum_{i\in V} K\big(h^{-1}d(x,X_i)\big)},$$

and $\hat{f}_C^x(y)$ the classical kernel estimator which is obtained with the complete data for (such as an example with $N = 300$ in the simulation below)

$$\hat{f}_C^x(y) = \frac{\sum_{i=1}^{N} K\left(h^{-1}d(x,X_i)\right) g^{-1} K_0\left(g^{-1}(y-Y_i)\right)}{\sum_{i=1}^{N} K\left(h^{-1}d(x,X_i)\right)}.$$

Within this section we will calculate and represent graphically the conditional density function estimator for surrogate data and we conduct a computational study on a simulated data in order to show advantages of using $\hat{f}_R^x(y)$ over $\hat{f}_V^x(y)$.

We choose K and K_0 the Gaussian kernel as follows

$$K_0(u) = K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-u^2/2}.$$

We generate 400 observations $(X_i, Y_i)_i$ using the following model:

$$Y_i = m(X_i) + \varepsilon.$$

Where the errors ε_i are i.i.d. according to the normal distribution $N(0;5)$. More precisely, the functional regressors $X_i(t)$ are defined, for any $t \in [0,1]$, by

$$X_i(t) = \sin(2\pi\, t) + W_i * t.$$

Where $W_i \sim U(0.5;2)$. The scalar response variable Y is generated by taking as a regression operator:

$$m(x) = 2\pi * \sin(b_i) \times \int_0^1 x^2(t)dt + \varepsilon.$$

Where: $\varepsilon_i \sim N(0,2)$ and $b_i \sim N(0,0.1)$. Let $I_0 = \{1,....,300\}$ and $I_1 = \{301,....,400\}$ be two subsets of indices. Then, we choose $\Delta = (X_i,Y_i)_{i\in I_0}$ as the learning sample and $\Gamma = \{(Xi,Yi)\}_{i\in I_1}$ as the testing sample. The surrogate variable $\tilde{Y}_i$ of Y_i, for all $i \in I_0$ was generated from $\tilde{Y}_i = \rho Z_i + \varepsilon_i$, where Z_i is the standard score of Y_i and $\varepsilon_i \sim N(0, \sqrt{1-\rho^2})$, in such a way that the correlation coefficient between Y_i and $\tilde{Y}_i$ is approximately equal to ρ which would not be controllable in practice.

In the sequel of this simulation study, we take $p = 0.75$. From the learning sample containing $N = 300$ functional data, we randomly choose a set V of n validation data $\{(X_i,Y_i)\}_{i\in V}$ which allows to build the estimator $\hat{f}_V^x(y)$ of $f_Y^x(y)$.

The estimator $\hat{f}_R^x(y)$ is then constructed by using the surrogate data $\{(X_i,Y_i)\}_{i\in\bar{V}}$ with the help of the validation data, where $\bar{V} \cup V = \{1,....N\}$. It should be pointed out that for $N = n$ (complete observations), we have $\hat{f}_V^x(y) = \hat{f}_R^x(y) = \hat{f}_C^x(y)$.

We evaluate the performance of the estimator $\hat{f}_R^x(y)$ in terms of prediction, by computing the relative mean squared error (RMSE) on the test sample:

$$RMSE(\hat{f}_R^x) = \sqrt{\frac{\sum_{i\in\Gamma}\left(\hat{f}_C^x(Y_i) - \hat{f}_R^x(Y_i)\right)^2}{100}}.$$

We have run 100 replicates of the simulation process for various values of n. We computed, for the two estimators $\hat{f}_R^x(y)$ and $\hat{f}_V^x(y)$ the mean and relative mean squared error (RMSE) over this 100 replications.

The comparison study results, for different values of percentage of validation data in samples:

$$p(V) = \frac{card(V)}{N} \times 100\% = \frac{n}{N} \times 100\%.$$

The results are summarized in the following Table 1.

Table 1. $\hat{f}_R^x(Y_i)$ and $\hat{f}_V^x(Y_i)$ whereas $\hat{f}_C^x(Y_i)$ for $n = 100$ and $n = 210$.

estimator	$p(V)$	Mean	RMSE
$\hat{f}_V^x(y)$	33%	0.048	0.34
$\hat{f}_R^x(y)$	33%	0.03767	0.02356
$\hat{f}_C^x(y)$	-	0.03960	-
$\hat{f}_V^x(y)$	70%	0.044	0.0368
$\hat{f}_R^x(y)$	70%	0.03882	0.001
$\hat{f}_C^x(y)$	-	0.03960	-

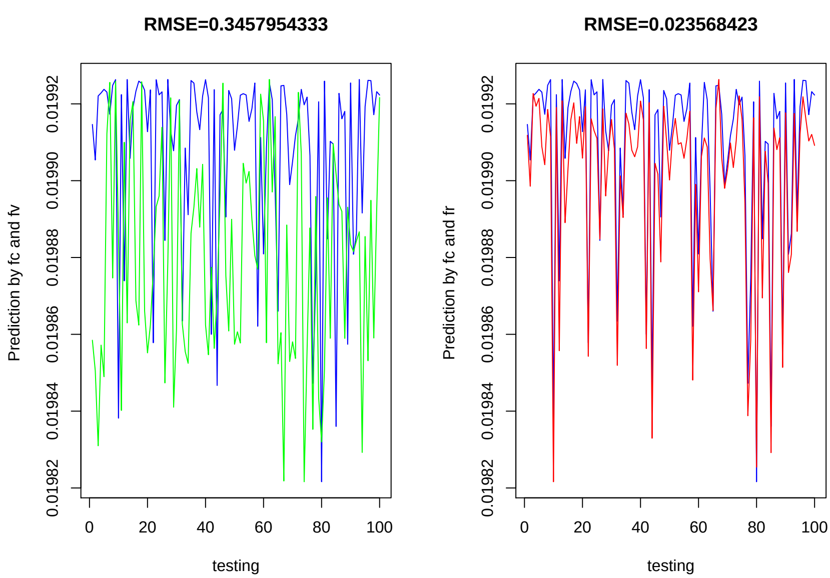

Figure 1. $\hat{f}_R^x$ (the red line) and $\hat{f}_V^x$ (the green line) with $\hat{f}_C^x$ (blue line) for Card(V)=n=100.

Obviously the quality of the prediction of the two estimators depend on the size n of the validation data. Specifically, RMSE decrease when the value of n increases. On the other hand, for $n = 100$ that means the percentage of validation data in a sample is 33% our estimator $\hat{f}_R^x(y)$ is better than $\hat{f}_V^x(y)$ in terms of RMSE inferior. In addition for $n = 210$ that means that we know 70% of data, our $\hat{f}_R^x(y)$ still greatly better as result of $RMSE = 0.001$. Nearly with the same mean of $\hat{f}_C^x(y)$.

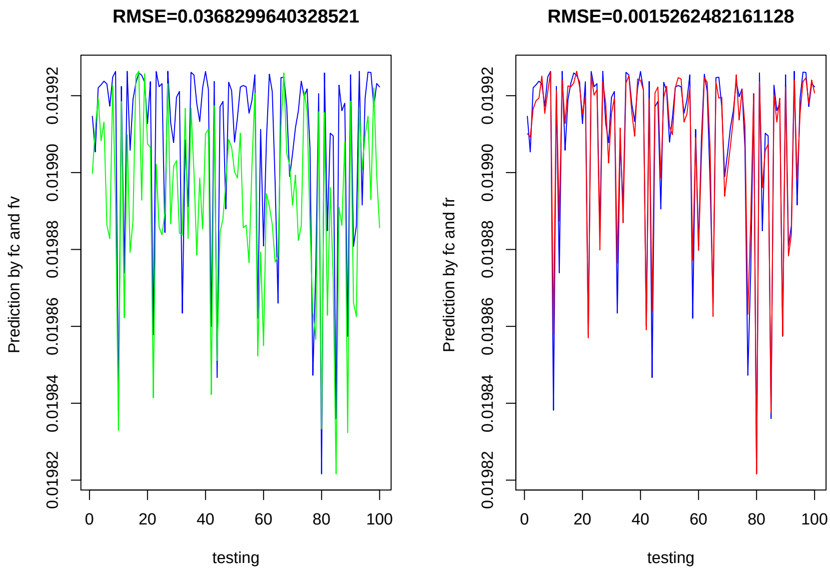

Figure 2. $\hat{f}_R^x$ (the red line) and $\hat{f}_V^x$ (the green line) with $\hat{f}_C^x$ (blue line) for $Card(V) = n = 210$.

It can be noticed from Figure 1 and Figure 2 that our $\hat{f}_R^x(y)$ is closer than $\hat{f}_V^x(y)$ to the curve $\hat{f}_C^x(y)$ which represents the estimator with the complete samples. Consequently, even if the percentage of validation data in sample increases from 33% to 70%, the estimator $\hat{f}_R^x(y)$ keeps performing better than $\hat{f}_V^x$.

5. Remarks and Conclusion

This paper has stated uniform consistency results when X is functional and Y is scalar. The fact to be able to state results on the quantity

$$\sup_{x\in S_{\mathscr{F}}} \sup_{y\in S_{\mathscr{R}}} |\hat{f}_R^x(y) - f_Y^x(y)|.$$

Allows directly to obtain results on quantity

$$|\hat{f}_R^x(y) - f_Y^x(y)|.$$

The entropy function represents a measure of the complexity of a set, in sense that, high entropy means that much information is needed to describe an element with an accuracy $\varepsilon = \frac{\log n}{n}$, in fact, the quality of the prediction of this estimator depends on the size n of the validation data. By building a suitable projection-based semi-metric, the entropy function becomes $\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right) = O(\log n)$ and for $N = n$ (without surrogate data) we get the estimator of Ferraty and Vieu (2006)

$$\sup_{x\in S_{\mathscr{F}}} \sup_{y\in S_{\mathscr{R}}} |\hat{f}_R^x(y) - f_Y^x(y)| = O(h^{\beta_1}) + O(g^{\beta_2}) + O_{a.co.}\left(\sqrt{\frac{\log n}{ng\phi(h)}}\right).$$

We present in this paper the almost complete convergence of conditional density function for surrogated scalar response variable given a functional random by using validation sample set. In addition, we show the performance of our estimator $\hat{f}_R^x(y)$ than $\hat{f}_V^x(y)$ to reduce RMSE by using the simulated data. This confirms that our estimator is a good alternative to the $\hat{f}_C^x(y)$ estimator (see Ferraty and Vieu, 2006) when we lack complete data.

Proof of Theorem

We note that :

$$i \in V \Rightarrow i \in \{1, ...,n\}, \text{ and } j \in \bar{V} \Rightarrow j \in \{n+1, ...,N\}.$$

We can write

$$\begin{aligned}
\hat{f}_R^x(y) - f_Y^x(y) &= \sum_{i\in V} \Omega_i(y) W_{1,n,i}(x) - \sum_{i\in V} f_Y^{X_i,\tilde{Y}_i}(y) W_{1,n,i}(x) \\
&\quad - \sum_{j\in\bar{V}} f_Y^{X_j,\tilde{Y}_j}(y) W_{1,n,j}(x) + \sum_{j\in\bar{V}} L(X_j,\tilde{Y}_j) W_{1,n,j}(x) \\
&\quad + \sum_{i=1}^{N} f_Y^{X_i,\tilde{Y}_i}(y) W_{1,n,i}(x) - f_Y^x(y). \\
&= E_1 + E_2 + E_3,
\end{aligned}$$

with

$$\begin{cases}
E_1 = \sum_{i\in V} \left(\Omega_i(y) - f_Y^{X_i,\tilde{Y}_i}(y)\right) W_{1,n,i}(x), \\
E_2 = \sum_{j\in\bar{V}} \left(L(X_j,\tilde{Y}_j) - f_Y^{X_i,\tilde{Y}_i}(y)\right) W_{1,n,j}(x), \\
E_3 = \sum_{i=1}^{N} \left(f_Y^{X_i,\tilde{Y}_i}(y) - f_Y^x(y)\right) W_{1,n,i}(x).
\end{cases}$$

And

$$\Omega_i(y) = g^{-1} K_0(g^{-1}(y - Y_i)).$$

Furthermore, we put

$$\Delta_i(x) = \frac{K\left(\dfrac{d(X_i,x)}{h}\right)}{\mathbb{E}\left[K\left(\dfrac{d(X_i,x)}{h}\right)\right]},$$

and we define

$$\begin{cases}
\hat{r}_1(x) = \dfrac{1}{n} \sum_{i\in V} \Delta_i(x), \\
\tilde{r}_1(x) = \dfrac{1}{N} \sum_{i=1}^{N} \Delta_i(x), \\
\hat{r}_2(x,y) = \dfrac{1}{n} \sum_{i\in V} \left(\Omega_i(y) - f_Y^{X_i,\tilde{Y}_i}(y)\right) \Delta_i(x), \\
\hat{r}_3(x) = \dfrac{1}{N} \sum_{i=1}^{N} \left(f_Y^{X_i,\tilde{Y}_i}(y) - f_Y^x(y)\right) \Delta_i(x).
\end{cases}$$

By the definition of $\hat{r}_1$, $\tilde{r}_1$, $\hat{r}_2$, and $\hat{r}_3$ we have:

$$E_1 = \frac{1}{\hat{r}_1(x)}\left(\hat{r}_2(x,y) - \mathbb{E}\left(\hat{r}_2(x,y)\right)\right) + \frac{\mathbb{E}\left(\hat{r}_2(x,y)\right)}{\hat{r}_1(x)},$$

and

$$E_3 = \frac{1}{\tilde{r}_1(x)}\left(\hat{r}_3(x,y) - \mathbb{E}\left(\hat{r}_3(x,y)\right)\right) + \frac{\mathbb{E}\left(\hat{r}_3(x,y)\right)}{\tilde{r}_1(x)}.$$

The numerators in this decomposition will be treated directly by using Lemma 6.2 and Lemma 6.3 below, while the denominators are treated directly by using Lemma 6.1 together with part i) of Proposition A.6 defined in p232 of Ferraty and Vieu, 2006. For the term E_2 will be treated by using Lemma 6.4.
Finally, the Theorem 3.2 is consequence of the following intermediate results

Lemma 5.1 *Under the hypotheses (H1) and (H3)-(H5), we have*

$$\sup_{x\in S_{\mathscr{F}}} |\hat{r}_1(x) - 1| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right),$$

and

$$\sum_{n=1}^{\infty} P\left(\inf_{x\in S_{\mathscr{F}}} \hat{r}_1(x) < \frac{1}{2}\right) < \infty.$$

The Proof of this Lemma is detailed in [**?**]

Lemma 5.2 *Under the hypotheses (H1),(H2) and (H4)-(H5), we have*

$$\sup_{x\in S_{\mathscr{F}}} \sup_{y\in S_{\mathscr{R}}} |\mathbb{E}[\hat{r}_2(x,y)]| = O\left(g^{\beta_2}\right),$$

and

$$\sup_{x\in S_{\mathscr{F}}} \sup_{y\in S_{\mathscr{R}}} |\mathbb{E}[\hat{r}_3(x,y)]| = O\left(h^{\beta_1}\right).$$

Proof of Lemma 5.2

By stationarity, we have

$$\begin{aligned} |\mathbb{E}[\hat{r}_2(x,y)]| &= \left|\mathbb{E}\left[\Delta_1(x)\mathbb{E}\left[\left(\Omega_1(y) - f_Y^{X_1,\tilde{Y}_1}(y)\right)|X_1\right]\right]\right| \\ &= \left|\mathbb{E}\left[\Delta_1(x)\mathbb{E}[\Omega_1(y)|X_1] - \mathbb{E}[f_Y^{X_1,\tilde{Y}_1}(y)|X_1]\right]\right| \\ &= \left|\mathbb{E}\left[\mathbb{I}_{B(x,h)}(X_1)\Delta_1(x)\mathbb{E}\left[\Omega_1(y)|X_1\right] - f_Y^{X_1}(y)\right]\right|. \end{aligned}$$

The fact that $\int_{\mathbb{R}} K_0(u)du = 1$ allows us to write:

$$\begin{aligned} \mathbb{E}\left[\Omega_1(y)|X_1\right] - f_Y^{X_1}(y) &= \int_{\mathbb{R}} g^{-1}K_0\left(g^{-1}(y-u)\right)\left(f_Y^{X_1}(y) - f_Y^{X_1}(u)\right)du \\ &= \int_{\mathbb{R}} K_0(v)\left(f_Y^{X_1}(y) - f_Y^{X_1}(y-vg)\right)dv \end{aligned}$$

Thus, under (H3) we obtain uniformly

$$\left|\mathbb{E}\left[\Omega_1(y)|X_1\right] - f_Y^{X_1}(y)\right| \leq Cg^{\beta_2}.$$

Hence, we get

$$\forall x \in S_{\mathscr{F}}, \quad |\mathbb{E}[\hat{r}_2(x,y)]| \leq Cg^{\beta_2}.$$

$$\begin{aligned} |\mathbb{E}[\hat{r}_3(x,y)]| &= \left|\mathbb{E}\left[\Delta_1(x)\mathbb{E}\left[\left(f_Y^{X_i,\tilde{Y}_i}(y) - f_Y^x(y)\right)|X_1\right]\right]\right| \\ &= \mathbb{E}\left[\mathbb{1}_{B(x,h)}(X_1)\Delta_1(x)\left|f_Y^{X_1}(y) - f_Y^x(y)\right|\right] \leq Ch^{\beta_1}. \end{aligned}$$

Lemma 5.3 *Under the assumptions of the Theorem, we have*

$$\sup_{x\in S_{\mathscr{F}}} \sup_{y\in S_{\mathscr{R}}} |\hat{r}_2(x,y) - \mathbb{E}[\hat{r}_2(x,y)]| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)}{ng\phi(h)}}\right),$$

and

$$\sup_{x\in S_{\mathscr{F}}} \sup_{y\in S_{\mathscr{R}}} |\hat{r}_3(x,y) - \mathbb{E}[\hat{r}_3(x,y)]| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log N}{N}\right)}{Ng\phi(h)}}\right).$$

Proof of Lemma 5.3

We treat only the first case, the second result can be treated by the same arguments. Firstly, we simplify the notation by denoting for all $i = 1,\ldots,n$, by

$$K_i(x) = K(h^{-1}d(x,X_i)).$$

Observe that, according to (H1) and (H3) we have

$$\forall x \in S_{\mathscr{F}} \qquad C\phi(h) < \mathbb{E}[K_1(x)] < C'\phi(h). \tag{13}$$

Next, we denote by $x_1,\ldots,x_{N_\varepsilon(S_{\mathscr{F}})}$ an ε-net (see Kolomogorov and Tikhomirov (1959)) for $S_{\mathscr{F}}$ and by $t_1,\ldots,t_{d_n}$ some l_n-net for the compact $S_{\mathscr{R}}$. Furthermore, for all x in $S_{\mathscr{F}}$ and y in $S_{\mathscr{R}}$ we put

$$k(x) = \arg\min_{k\in\{1,2,\ldots,N_\varepsilon(S_{\mathscr{F}})\}} d(x,x_k) \text{ and } j(y) = \arg\min_{j\in\{1,2,\ldots,d_n\}} |y - t_j|.$$

Now, we fix $\varepsilon = \frac{\log n}{n}$ and $l_n = n^{-2\gamma-1}$ and we use the following decomposition

$$\begin{aligned}
|\hat{r}_2(x,y) - \mathbb{E}[\hat{r}_2(x,y)]| \leq & \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} \left|\hat{r}_2(x,y) - \hat{r}_2(x_{k(x)},y)\right|}_{T_1} \\
& + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} \left|\hat{r}_2(x_{k(x)},y) - \hat{r}_2(x_{k(x)},t_{j(y)})\right|}_{T_2} \\
& + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} \left|\hat{r}_2(x_{k(x)},t_{j(y)}) - \mathbb{E}[\hat{r}_2(x_{k(x)},t_{j(y)})]\right|}_{T_3} \\
& + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} \left|\mathbb{E}[\hat{r}_2(x_{k(x)},t_{j(y)})] - \mathbb{E}[\hat{r}_2(x_{k(x)},y)]\right|}_{T_4} \\
& + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} \left|\mathbb{E}[\hat{r}_2(x_{k(x)},y)] - \mathbb{E}[\hat{r}_2(x,y)]\right|}_{T_5}.
\end{aligned}$$

For the term T_1 we employ the Lipschitzianity of the kernel K on $[0,1]$ which gives

$$T_1 \leq \frac{C}{n}\sum_{i=1}^{n} Z_i \text{ with } Z_i = \frac{\varepsilon}{hg\,\phi(h)}\, \mathbb{I}_{B(x,h)\cup B(x_{k(x)},h)}(X_i),$$

Therefore, it is clear that the assumption (H3) permits to write that

$$Z_1 = O\left(\frac{\varepsilon}{h\phi(h)}\right),\ \mathbb{E}[Z_1] = O\left(\frac{\varepsilon}{hg}\right) \text{ and } var(Z_1) = O\left(\frac{\varepsilon^2}{h^2g^2\phi(h)}\right).$$

So, we get

$$\mathbb{E}(|\, Z_1\,|^m) \leq \frac{C\varepsilon^m}{h^m g^m \phi(h)^{m-1}}. \tag{14}$$

By using the result (10) together with the definition of ε we have for n large enough:

$$\frac{\varepsilon}{hg} \leq C.$$

So, we get:

$$\mathbb{E}(|\, Z_1\,|^m) \leq \frac{C\varepsilon^{m-1}}{h^{m-1}g^{m-1}\phi(h)^{m-1}}.$$

Now, by applying Corollary A.8 in Ferraty and Vieu (2006) with $a^2 = \dfrac{\varepsilon}{hg\phi(h)}$, we get:

$$\frac{1}{n}\sum_{i=1}^{n} Z_i = EZ_1 + O_{a.co.}\left(\sqrt{\frac{\varepsilon\log n}{ngh\phi(h)}}\right).$$

Finally, applying (14) for $m=1$ one gets

$$T_1 = O\left(\frac{\varepsilon}{hg}\right) + O_{a.co.}\left(\sqrt{\frac{\varepsilon\log n}{nhg\phi(h)}}\right).$$

By (10) and the definition of ε for n large enough :

$$C\frac{(\log n)^2}{(ng\phi(h))^2} \geq \frac{\varepsilon \log n}{nhg\phi(h)}$$

Using (H4b) together with (11) and the fact that:

$$\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Z_i\right| > \sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)}}\right\} \subset \left\{\left|\frac{1}{n}\sum_{i=1}^{n} Z_i\right| > \sqrt{\frac{C(\log n)^2}{(ng\phi(h))^2}}\right\},$$

we get

$$T_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)}}\right). \tag{15}$$

Thus, by Assumption (H4b) we deduce that

$$T_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)}}\right) \text{ and } T_5 = O\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)}}\right). \tag{16}$$

We use the same ideas to treat R_2. In fact, we use the Lipschitz condition on the kernel K and the assumption (H2) to write that

$$\begin{aligned}\left|\hat{r}_2(x_{k(x)},y) - \hat{r}_2(x_{k(x)},t_{j(y)})\right| &\leq \frac{C}{n\,\phi(h)}\sum_{i=1}^{n} K_i(x_{k(x)})\left(\left|\Omega_i(y) - \Omega_i(t_{j(y)})\right|\right. \\ &\qquad \left. + |f_Y^{X_i,\tilde{Y}_i}(y) - f_Y^{X_i,\tilde{Y}_i}(y)|\right) \\ &\leq \frac{C}{n}\sum_{i=1}^{n} Z_i,\end{aligned}$$

where $Z_i = \dfrac{l_n K_i(x_{k(x)})\,\mathbb{1}_{B(x_{k(x)},h)}(X_i)}{g^2\,\phi(h)}$.

It is clear that the assumption (H3) permits to write that

$$Z_1 = O\left(\frac{l_n}{g^2\phi(h)}\right),\ \mathbb{E}[Z_1] = O\left(\frac{l_n}{g^2}\right) \text{ and } \ var(Z_1) = O\left(\frac{l_n^2}{g^4\phi(h)}\right).$$

Invoking the same idea in (15), allows to get:

$$T_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{gn\phi(h)}}\right) \text{ and } T_4 = O\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{gn\phi(h)}}\right). \tag{17}$$

It remains to evaluate R_3. Indeed, we write

$$
\begin{aligned}
P&\left(T_3>\eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(h)}}\right)\\
&= P\left(\max_{j\in\{1,2,\ldots d_n\}}\max_{k\in\{1,\ldots,N_\varepsilon(S_{\mathscr{F}})\}}\left|\hat{r}_2(x_k,t_j)-\mathbb{E}\hat{r}_2(x_k,t_j)\right|>\eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(h)}}\right)\\
&\leq d_nN_\varepsilon(S_{\mathscr{F}})\max_{j\in\{1,2,\ldots,d_n\}}\max_{k\in\{1,\ldots,N_\varepsilon(S_{\mathscr{F}})\}}P\left(\left|\hat{r}_2(x_k,t_j)-\mathbb{E}\hat{r}_2(x_k,t_j)\right|>\eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(h)}}\right)\\
&\leq d_nN_\varepsilon(S_{\mathscr{F}})\max_{j\in\{1,2,\ldots,d_n\}}\max_{k\in\{1,\ldots,N_\varepsilon(S_{\mathscr{F}})\}}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\right|>\eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(h)}}\right).
\end{aligned}
$$

Where

$$\Gamma_i=\frac{1}{\mathbb{E}[K_1(x)]}\left[K_i(x_k)(\Omega_i(t_j)-f^{X_i,\tilde{Y}_i}(t_j))-E\left(K_i(x_k)(\Omega_i(t_j)-f^{X_i,\tilde{Y}_i}(t_j))\right)\right].$$

It follows, from the fact that the kernel K and K_0 and $f^{X_i,\tilde{Y}_i}$ are bounded, that

$$E|\Gamma_i|^2\leq C\,(\phi(h))^{-1}.$$

Thus, we apply the Bernstein exponential inequality, we obtain for all $j\leq d_n$, that

$$P\left(\left|\hat{r}_2(x_k,t_j)-\mathbb{E}\hat{r}_2(x_k,t_j)\right|>\eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(h)}}\right)\leq 2\exp\left\{-C\eta^2\psi_{S_{\mathscr{F}}}(\varepsilon)\right\}.$$

Therefore, by choosing $C\eta^2=\beta$, and using the fact that $d_n=O(l_n^{-1})$, we conclude that

$$
\begin{aligned}
d_nN_\varepsilon(S_{\mathscr{F}})\max_{j\in\{1,2,\ldots,d_n\}}\max_{k\in\{1,\ldots,N_\varepsilon(S_{\mathscr{F}})\}}&P\left(\left|\hat{r}_2(x_k,t_j)-\mathbb{E}\hat{r}_2(x_k,t_j)\right|>\eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(h)}}\right)\\
&\leq C'd_n\left(N_\varepsilon(S_{\mathscr{F}})\right)^{1-C\eta^2}.
\end{aligned}
$$

Finally, we obtain

$$T_3=O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n^{1-\gamma}\phi(h)}}\right) \tag{18}$$

For the term $\hat{r}_3(x,y)-\mathbb{E}[\hat{r}_3(x,y)]$. First we fix $\varepsilon=\frac{\log N}{N}$ and $l_N=N^{-2\gamma-1}$.

Using the decomposition and invoking the same arguments as for the proof of $\hat{r}_2(x,y)-\mathbb{E}[\hat{r}_2(x,y)]$, we get:

$$\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}}\left|\hat{r}_3(x,y)-\mathbb{E}[\hat{r}_3(x,y)]\right|=O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log N}{N}\right)}{Ng\phi(h)}}\right).$$

Lemma 5.4 *Under the assumptions of Theorem (H1)-(H6), we have* $\forall j \in \bar{V}$

$$\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{F}}} |f^{\tilde{Y}_j,X_j}(y) - L(X_j,\tilde{Y}_j)| = O(h^{\beta_1}) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)\phi(b)}}\right)$$

$$+O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}\left(\frac{\log n}{n}\right)}{ng\phi(b)\phi(h)}}\right).$$

To simplify we put $X_j = x$ and $\tilde{Y}_j = \tilde{y}$.

Proof of Lemma 5.4

The proof is based on the following decomposition

$$L(x,y,\tilde{y}) - f_Y^{X_i,\tilde{y}_i}(y) = \frac{1}{L_1(x,\tilde{y})}\Big[L_2(x,y,\tilde{y}) - \mathbb{E}[L_2(x,y,\tilde{y})]\Big]$$

$$+\frac{1}{L_1(x,\tilde{y})}\Big[\mathbb{E}[L_2(x,y,\tilde{y})] - f_Y^{X_i,\tilde{Y}_i}(y)\Big] + [1 - L_1(x,\tilde{y})]\,\frac{f_Y^{X_i,\tilde{Y}_i}(y)}{L_1(x,\tilde{y})}.$$

Where

$$L_1(x,\tilde{y}) = \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))K(b^{-1}(\tilde{y}-\tilde{Y}_1))]}\sum_{i\in V} K(h^{-1}d(x,X_i))K(b^{-1}(\tilde{y}-\tilde{Y}_i)),$$

and

$$L_2(x,y,\tilde{y}) = \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))K(b^{-1}(\tilde{y}-\tilde{Y}_1))]}\sum_{i\in V} K(h^{-1}d(x,X_i))K(b^{-1}(\tilde{y}-\tilde{Y}_i))\Omega_i(y).$$

$$\begin{aligned}
|L_1(x,\tilde{y}) - \mathbb{E}[L_1(x,\tilde{y})]| \leq\ & \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{\tilde{y}\in S_{\mathscr{R}}} |L_1(x,\tilde{y}) - L_1(x_k,\tilde{y})|}_{R_1} \\
& + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{\tilde{y}\in S_{\mathscr{R}}} \left|L_1(x_k,\tilde{y}) - L_1(x_k,t_{j(\tilde{y})})\right|}_{R_2} \\
& + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{\tilde{y}\in S_{\mathscr{R}}} \left|L_1(x_k,t_{j(\tilde{y})}) - \mathbb{E}[L_1(x_k,t_{j(\tilde{y})})]\right|}_{R_3} \\
& + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{\tilde{y}\in S_{\mathscr{R}}} \left|\mathbb{E}[L_1(x_k,t_{j(\tilde{y})})] - \mathbb{E}[L_1(x_k,\tilde{y})]\right|}_{R_4} \\
& + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{\in S_{\mathscr{R}}} |\mathbb{E}[L_1(x_k,\tilde{y})] - \mathbb{E}[L_1(x,\tilde{y})]|}_{R_5}.
\end{aligned}$$

For the term R_1 we employ the Lipschitzianity of the kernel K on $[0,1]$ with (H1) and (H2) lead directly

$$R_1 \leq \frac{C}{n}\sum_{i=1}^{n} Z_i \text{ with } Z_i = \frac{\varepsilon}{h\phi(h)\phi(b)}\,\mathbb{I}_{B(x,h)\cup B(x_{k(x)},h)}(X_i)\,\mathbb{I}_{\{Y\leq\tilde{y}\leq Y+b\}},$$

It is clear that the assumption (H3) permits to write that

$$Z_1 = O\left(\frac{\varepsilon}{h\phi(h)}\right),\ \mathbb{E}[Z_1] = O\left(\frac{\varepsilon}{h}\right) \text{ and } var(Z_1) = O\left(\frac{\varepsilon^2}{h^2\phi(b)\phi(h)}\right).$$

By using the same steps as (15) we get

$$R_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\phi(h)\phi(b)}}\right) \text{ and } R_5 = O\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\phi(h)\phi(b)}}\right). \tag{19}$$

We use the same ideas to treat R_2. In fact we use the Lipschitz condition on the kernel K and the assumption (H2) to write that

$$\begin{aligned}\left|L_1(x_{k(x)},\tilde{y}) - L_1(x_{k(x)},t_{j(\tilde{y})})\right| &\leq \frac{C}{n\phi(h)\phi(b)}\sum_{i=1}^{n} K_i(x_{k(x)})\left(\left|K_i(\tilde{y}) - k_i(t_{j(\tilde{y})})\right|\right. \\ &\leq \frac{C}{n}\sum_{i=1}^{n} Z_i,\end{aligned}$$

where $Z_i = \dfrac{w_n K_i(x_{k(x)})\,\mathbb{I}_{B(x_{k(x)},h)}(X_i)\,\mathbb{I}_{\{Y\leq\tilde{y}\leq Y+b\}}}{b\phi(h)\phi(b)}$.

It is clear that the assumption (H3) permits to write that

$$Z_1 = O\left(\frac{w_n}{b\phi(b)\phi(h)}\right),\ \mathbb{E}[Z_1] = O\left(\frac{w_n}{b}\right) \text{ and } var(Z_1) = O\left(\frac{w_n^2}{b^2\phi(b)\phi(h)}\right).$$

Similarly, as previously we get

$$R_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\phi(b)\phi(h)}}\right) \text{ and } R_4 = O\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\phi(b)\phi(h)}}\right). \tag{20}$$

It remains to evaluate R_3. Indeed, we write

$$\begin{aligned}&P\left(R_3 > \eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(b)\phi(h)}}\right) \\ &\quad = P\left(\max_{j\in\{1,2,\ldots d_n\}}\max_{k\in\{1,\ldots,N_\varepsilon(S_{\mathscr{F}})\}}\left|L_1(x_k,t_{j(\tilde{y})}) - \mathbb{E}L_1(x_k,t_{j(\tilde{y})})\right| > \eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(b)\phi(h)}}\right) \\ &\leq d_n N_\varepsilon(S_{\mathscr{F}})\max_{j\in\{1,2,\ldots,d_n\}}\max_{k\in\{1,\ldots,N_\varepsilon(S_{\mathscr{F}})\}} P\left(\left|L_1(x_k,t_{j(\tilde{y})}) - \mathbb{E}L_1(x_k,t_{j(\tilde{y})})\right| > \eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(b)\phi(h)}}\right) \\ &\quad \leq d_n N_\varepsilon(S_{\mathscr{F}})\max_{j\in\{1,2,\ldots,d_n\}}\max_{k\in\{1,\ldots,N_\varepsilon(S_{\mathscr{F}})\}} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\right| > \eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(b)\phi(h)}}\right).\end{aligned}$$

Where

$$\Gamma_i = \frac{1}{\mathbb{E}[K_1(x)K_1(\tilde{y})]}\left[K_i(x_k)(K_i(t_{j(\tilde{y})}) - E\left(K_i(x_k)K_i(t_j)\right)\right].$$

It follows from the fact that the kernel K is bounded , that $E|\Gamma_i|^2 \leq C\left(\phi(b)\phi(h)\right)^{-1}$. Thus, we apply the Bernstein exponential inequality we obtain for all $j \leq N_\varepsilon(S_{\mathscr{F}})$, that

$$P\left(\left|L_1(x_k,t_{j(\tilde{y})}) - \mathbb{E}L_1(x_k,t_{j(\tilde{y})})\right| > \eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(b)\phi(h)}}\right) \leq 2\exp\left\{-C\eta^2\psi_{S_{\mathscr{F}}}(\varepsilon)\right\}.$$

Therefore, by choosing $C\eta^2 = \beta$, and using the fact that $d_n = O(w_n^{-1})$, we conclude that

$$\begin{aligned} d_nN_\varepsilon(S_{\mathscr{F}}) \max_{j\in\{1,2,\ldots,d_n\}} \max_{k\in\{1,\ldots,N_\varepsilon(S_{\mathscr{F}})\}} P\left(\left|L_1(x_k,t_j) - \mathbb{E}L_1(x_k,t_j)\right| > \eta\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\,\phi(b)\phi(h)}}\right) \\ \leq C' d_n\left(N_\varepsilon(S_{\mathscr{F}})\right)^{1-C\eta^2}. \end{aligned}$$

Finally, using (H5) and (12) we obtain

$$R_3 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{n\phi(b)\phi(h)}}\right). \tag{21}$$

By using the same decomposition:

$$\begin{aligned} |L_2(x,y,\tilde{y}) - \mathbb{E}[L_2(x,y,\tilde{y})]| \leq\ & \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} |L_2(x,y,\tilde{y}) - L_2(x_k,y,\tilde{y})|}_{S_1} \\ & + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} \left|L_2(x_k,y,\tilde{y}) - L_2(x_k,t_{j(y)},\tilde{y})\right|}_{S_2} \\ & + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} \left|L_2(x_k,t_{j(y)},\tilde{y}) - \mathbb{E}[L_2(x_k,t_{j(y)},\tilde{y}]\right|}_{S_3} \\ & + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} \left|\mathbb{E}[L_2(x_k,t_{j(y)},\tilde{y}] - \mathbb{E}[L_2(x_k,y,\tilde{y})]\right|}_{S_4} \\ & + \underbrace{\sup_{x\in S_{\mathscr{F}}}\sup_{y\in S_{\mathscr{R}}} |\mathbb{E}[L_2(x_k,y,\tilde{y})] - \mathbb{E}[L_2(x,y,\tilde{y})]|}_{S_5}. \end{aligned}$$

So as before we get:

$$S_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)\phi(b)}}\right) \text{ and } S_5 = O\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(h)\phi(b)}}\right). \tag{22}$$

And

$$S_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(b)\phi(h)}}\right) \text{ and } S_4 = O\left(\sqrt{\frac{\psi_{S_{\mathscr{F}}}(\varepsilon)}{ng\phi(b)\phi(h)}}\right). \tag{23}$$

And

$$S_3 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{ng\phi(b)\phi(h)}}\right). \tag{24}$$

$$\begin{aligned}
&|\mathbb{E}[L_2(x,y,\tilde{y})] - f^x(y)| \\
&\qquad\leq C\,\mathbb{E}\left[\left|K(\tfrac{d(x,X_1)}{h})K(\tfrac{\tilde{y}-\tilde{Y}_1}{b})\mathbb{E}\left[\tfrac{1}{g}K_0\left(\frac{\tilde{y}-\tilde{Y}_1}{g}\right) - f_Y^{X_1,\tilde{Y}_1}(y)|\left(X_1,\tilde{Y}_1\right)\right|\right]\right] \\
&\qquad\leq C\,\mathbb{E}\left[\left|K(\tfrac{d(x,X_1)}{h})K(\tfrac{\tilde{y}-\tilde{Y}_1}{b})\mathbb{E}\left[\tfrac{1}{g}K_0\left(\frac{\tilde{y}-\tilde{Y}_1}{g}\right) | \left(X_1,\tilde{Y}_1\right)\right] - f_Y^{X_1,\tilde{Y}_1}(y)\right|\right]
\end{aligned}$$

Moreover, by change of variable:

$$\mathbb{E}\left[g^{-1}K_0\left(\frac{y-Y_1}{g}\right) | \left(X_1,\tilde{Y}_1\right)\right] = \int_{\mathbb{R}} K_0(u) f_Y^{X_1,\tilde{Y}_1}(y-ug)du.$$

Finally, by (H2) we get:

$$\left|\mathbb{E}[L_2(x,y,\tilde{y})] - f_Y^{X_i,\tilde{Y}_i}(y)\right| = O(g^{\beta_1}). \tag{25}$$

So, The Lemma 5.4 can be easily deduced from (19), (20), (22), (23), (24) and (25). □

Acknowledgements

The authors greatly thank the Editor in chief and the reviewers for the careful reading, constructive comments and relevant remarks, which permit us to improve the paper.

References

Carroll, R. J., Knickerbocker, R. K., and Wang, C. Y., (1995). Dimension reduction in a semiparametric regression model with errors in covariates. *The Annals of Statistics*.

Carroll, R. J., Wand, M. P.. (1991). Semiparametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society*.

Duncan, G. J., Hill, D. H.. (1985). An investigation of the extent and consequences of measurement error in labor-economic survey data. *Journal of Labor Economics*.

Firas, I., Ali Hajj H., and Rachdi, M., (2019). Regression model for surrogate data in high dimensional statistics. *Journal of Communications in Statistics – Theory and Methods*.

Ferraty, F., Laksaci, A., Tadj, A., Vieu, P., (2010). Rate of uniform consistency for non-parametric estimates with functional variables. *Journal of Statistical Planning and Inference*.

Ferraty, F., Vieu, P., (2006). Nonparametric functional data analysis. Theory and practice, *NY: Springer Series in Statistics*.

Ferraty, F., Vieu, P., (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.*

Ferraty, F. Vieu, P., (2011). Kernel regression estimation for functional data. *In The Oxford Handbook of Functional Data Analysis (Ed. F. Ferraty and Y. Romain)*. Oxford University Press.

Hsing, T., Eubank R., (2015). Theoretical foundations of functional data analysis, with an intro- duction to linear operators. *Wiley series in probability and statistics*. Chichester, UK: John Wiley and Sons.

Horvath, L., Kokoszka P., (2012). Inference for functional data with applications. New York, *NY: Springer Series in Statistics*.

Goia, A., Vieu P., (2016). An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis*, 146, pp. 1–6.

Kolmogorov A. N., Tikhomirov V. M., (1959). ε-entropy and ε-capacity. *Uspekhi Mat. Nauk.*, 14, pp. 3–86., 2, pp. 277–364.

Lecoutre, J. P., (1990). Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statist. Probab. Lett.*

Loève, M., (1963). *Probability Theory*, 3rd ed. Van Nostranr Princeton.

Pepe, M. S., (1992). Inference using surrogate outcome data and validation sample. *Biometrika 79*.

Rachdi, M., Vieu, P., (2007). Nonparametric regression for functional data: Automatic smoothing parameter selection. *Journal of Statistical Planning and Inference*.

Wang, Q. H., (2000). Estimation of linear error-in-covariables models with validation data under random censorship. *Journal of Multivariate Analysis*.

Wang, Q. H., Rao, J. N. K., (2002). Empirical likelihood-based in linear errors in covariables models with validation data. *Biometrika*.

Wang, Q. H., (2003). Dimension reduction in partly linear error-in-response models with validation data. *Journal of Multivariate Analysis*.

Wang, Q. H., (2006). Nonparametric regression function estimation with surrogate data and validation sampling. *Journal of Multivariate Analysis*.

Wang, J. L., Chiou, J. M., and Muller, H. G., (2016). Review of functional data analysis. *Annual Review of Statistics and Its Application*.

Zhang, J., (2014). Analysis of variance for functional data. *Monographs on Statistics and Applied Probability*.