MATHEMATICAL MODELING AND COMPUTING, Vol. 10, No. 3, pp. 625–637 (2023)

Generalized regression function for surrogate scalar response

Boumahdi M.[1,*], Ouassou I.[1], Rachdi M.[2]

[1] *Université Cadi Ayyad, Ecole Nationale des Sciences Appliquées, Marrakech, Morocco*
[2] *Université Grenoble Alpes, Grenoble Cedex 09, France*
[*] *Corresponding author: mounirboumahdi1210112999@gmail.com*

(Received 15 February 2023; Accepted 5 July 2023)

In this paper we develop and generalize the estimator of regression function for surrogate scalar response variable given a functional random one. Then, we build up some asymptotic properties in terms of the almost complete convergences, depending in the result we show the superiority of our estimator in term of prediction.

Keywords: *surrogate response; functional variable; almost complete convergence; kernel estimators; scalar response; entropy; semi-metric space.*

2010 MSC: 62G08 **DOI:** 10.23939/mmc2023.03.625

1. Introduction

Studying the link between a scalar response variable Y given a new value for the explanatory variable X is an important subject in nonparametric statistics, and there are several ways to explain this link. For examples, the conditional expectation, the conditional distribution, the conditional density and the conditional hazard function.

We are interested in estimating the nonparametric regression for surrogated scalar response. We are based on the following model:

$$Y = m(X) + \varepsilon,$$

where m is the regression operator, X is a functional covariate which belongs to a semi-metric space $(\mathcal{F}, d)$, Y is scalar response variable and ε is a random error satisfies $\mathbb{E}(\varepsilon|X) = 0$.

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. 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) [1], Carroll and Wand (1991) [2] and Pepe (1992) [3].

To estimate the generalized regression function for surrogate data $\hat{m}_R(x)$ we adopt an approach based on validation data ideas. In fact, the idea is to introduce the information contain in the validation data and surrogate variable $\tilde{Y}$ of Y. Inside the simulation study of Section 4, 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 but we can clearly notice that the quality of our $\hat{m}_R$ depends on the size n of the validation data and ρ. Specifically, our estimator greatly better as the value of n and ρ increases.

The main objective of this paper is to purpose the uniform almost complete convergence (with rate) of our estimator $\hat{m}_R$ and we study its performance against $\hat{m}_V$ in term of prediction section 4.

Section 3 is dedicated to some probability tools for functional variable and the uniform rates of convergence are stated therein. The remark 1 show that the rates of convergence of our estimator generalizes the results given by Ferraty and Vieu (2006) [4] and F. Ferraty, A. Laksaci, A. Tadj, P. Vieu (2010) [5]. This paper begins with the construction of the estimator of $m_R(x)$ in section 2, in section 3 we present the almost-complete, then we study the performance of the estimator by computing the relative mean squared error (RMSE) by using a testing data. Finally, we display the superiority of our estimator in term of prediction when we are lacking complete data by using simulated data.

The choice of $\tilde{Y}$ (the surrogate variable of Y) in practice is difficult but it is important for the quality of our estimator in effect 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 lot of statistical models when we are lacking complete data.

2. Estimation procedure

Let $(X, Y) \in \mathcal{F} \times \mathbb{R}$ denotes a random vector, where $(\mathcal{F}, d)$ is a semi-metric space equipped with the semi-metric d, we are concerned with the estimation of a generalized regression function defined as following:

$$m(x) = \mathbb{E}\left[\varphi(Y)|X = x\right] \quad \forall x \in \mathcal{F}. \tag{1}$$

Where φ is a known real-value Borel function. The model 1 has been studied by [6] when $\varphi(Y) = Y$. Therefore, let $(X_1, Y_1), \ldots, (X_N, Y_N)$ be a random sample consisting of independent and identically distributed (i.i.d) variable from the distribution of (X, Y).

Let $\hat{m}_C(x)$ be the classical kernel estimator which is obtained with the complete data (Ferraty and Vieu (2006) [4])

$$\hat{m}_C(x) = \sum_{i=1}^{N} Y_i W_{1,n,i}(x),$$

where

$$W_{1,n,i}(x) = \frac{K\left(\frac{d(X_i,x)}{h}\right)}{\sum_{l=1}^{N} K\left(\frac{d(X_l,x)}{h}\right)}, \tag{2}$$

with $(X_1, Y_1), \ldots, (X_N, Y_N)$ is a random sample consisting of independent and identically distributed variable from the distribution of (X, Y). But the problem here is the unavailability of some data in the response variable:

$$\begin{gathered} (X_1(t); Y_1) \\ \vdots \\ (X_i(t); ???) \\ \vdots \\ (X_j(t); ???) \\ \vdots \\ (X_k(t); ???) \\ \vdots \\ (X_N(t); Y_N) \end{gathered}$$

Consequently, we are concerned with the estimation of a regression function for surrogate functional response, we can write:

$$m(x) = \mathbb{E}\left[\varphi(Y)|X = x\right] = \mathbb{E}\big[\mathbb{E}\big(\varphi(Y)|X, \tilde{Y}\big)|X = x\big] \quad \forall x \in \mathcal{F}.$$

Where $\tilde{Y}$ is a surrogate variable of Y. So, we propose the regression function for surrogate functional response as following:

$$\hat{m}_R(x) = \sum_{i \in V} \varphi(Y_i) W_{1,n,i}(x) + \sum_{j \in \bar{V}} U(X_j, \tilde{Y}_j) W_{1,n,j}(x), \tag{3}$$

where

$$U(X_j,\tilde{Y}_j)=\sum_{i\in V}\varphi(Y_i)W_{2,n,i}(X_j,\tilde{Y}_j) \quad \forall j\in\bar{V}, \tag{4}$$

with

$$W_{2,n,i}(X_j,\tilde{Y}_j)=\frac{W\left(\frac{d(X_j,X_i)}{h},\frac{\tilde{Y}_i-\tilde{Y}_j}{b}\right)}{\sum_{l\in V}W\left(\frac{d(X_j,X_l)}{h},\frac{\tilde{Y}_l-\tilde{Y}_j}{b}\right)}. \tag{5}$$

Where K is a kernel function and both $h=h_N$, $b=b_N$ are a sequence of positive reals that tends to zero when N goes to infinity. 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 $\{1,2,\dots,N\}$. Where W is a kernel function which is defined on $\mathbb{R}^2$ and b is sequence of real numbers which tends to zero. For sake of simplicity, we will use only one kernel. In sense that $W(\cdot,\cdot)=K(\cdot)K(\cdot)$. This consideration is because the choice of the kernel has less influence in the performance of the estimator. Remarkably our goal is then to improve the modeling by filling/recovering some of the information missed in the response variable with this surrogate variable as following:

$$\begin{array}{ccc}
(X_1(t);Y_1) & (X_1(t);Y_1) & (X_1(t);Y_1)\\
\vdots & \vdots & \vdots\\
(X_i(t);???) & (X_i(t);Y_i) & \big(X_i(t);U(X_i,\tilde{Y}_i)\big)\\
\vdots & \vdots & \vdots\\
(X_j(t);???) & (X_j(t);Y_j) & \big(X_j(t);U(X_j,\tilde{Y}_j)\big)\\
\vdots & \vdots & \vdots\\
(X_k(t);???) & (X_k(t);Y_k) & \big(X_k(t);U(X_k,\tilde{Y}_k)\big)\\
\vdots & \vdots & \vdots\\
(X_N(t);Y_N) & (X_N(t);Y_N) & (X_N(t);Y_N)\\
\Downarrow & \Downarrow & \Downarrow\\
\hat{m}_V(x) & \hat{m}_C(x) & \hat{m}_R(x)
\end{array}$$

3. Some asymptotic properties

In the sequel, when no confusion is possible, we will denote by C and C' some strictly positive generic constants, $\mathcal{F}$ is a semi metric space and:

$$m^{\tilde{Y}}(x)=\mathbb{E}\big[\varphi(Y)|X=x,\tilde{Y}\big].$$

Recall that a semi-metric (sometimes called pseudo-metric) is just a metric violating the property

$$[d(x,y)=0]\Rightarrow[x=y].$$

Now $S_{\mathcal{F}}$ is a fixed subset of $\mathcal{F}$ and for $\eta>0$ we consider the following η-neighborhood of $S_{\mathcal{F}}$:

$$S_{\mathcal{F}}^{\eta}=\left\{x\in\mathcal{F},\exists x',d(x',x)\leqslant\eta\right\}.$$

Definition 1. *One says that the rate of almost complete convergence of $(X_n)_{n\in\mathbb{N}}$ to X is of order u_n if and only if*

$$\exists\varepsilon_0>0,\quad \sum_{n\in\mathbb{N}}P\big(|X_n-X|>\varepsilon_0u_n\big)<\infty,$$

and we write

$$X_n-X=O_{a.co}(u_n).$$

We define the Kolmogorov's entropy as follows.

Definition 2. *Let $S_{\mathcal{F}}$ be a subset of a semi-metric space $\mathcal{F}$, and let $\varepsilon>0$ be given. A finite set of point $x_1,x_2,\dots,x_{n_0}$ in $\mathcal{F}$ is called an ε-net for $S_{\mathcal{F}}$ if $S_{\mathcal{F}}\subset\cup_{k=1}^{N_0}B(x_k,\varepsilon)$.*

The quantity $\psi_{S_{\mathcal{F}}}=\log(N_\varepsilon)$, where N_ε is the minimal number of open balls in $\mathcal{F}$ of radius ε which is necessary to cover S, is called the Kolmogorov's ε-entropy of the set $S_{\mathcal{F}}$.

This concept was introduced by Kolmogorov in the mid-1950's see [7], 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 at uniform (over S) asymptotic results. For more examples see [5].

We consider the following assumptions:

(H1) For all x in the subset $S_{\mathcal{F}}$,
$$0 < C\phi(h) \leqslant P(X \in B(x,h)) \leqslant C'\phi(h) < \infty,$$
$$0 < C_1'\phi(b) \leqslant P\big(\tilde{Y} \in [\tilde{y}-b;\tilde{y}+b]\big) \leqslant C_2'\phi(b) < \infty,$$
$$C\phi(h)\phi(b) < \mathbb{E}\big[K\big(h^{-1}d(x,X_i)\big)K\big(b^{-1}(\tilde{y}-\tilde{Y}_1)\big)\big] < C'\phi(h)\phi(b);$$

(H2) $\forall x_1, x_2 \in S_{\mathcal{F}}$ and $\forall i \in N$
$$\left|m^{\tilde{y}}(x_1)K_i(x_1) - m^{\tilde{y}}(x_2)K_i(x_2)\right| \leqslant C\left|K_i(x_1) - K_i(x_2)\right|,$$
and there exists $\beta_1 > 0$ such that $\forall x_1, x_2 \in S_{\mathcal{F}}$ and
$$|m(x_1) - m(x_2)| \leqslant C d^{\beta_1}(x_1,x_2);$$

(H3) $\forall m \geqslant 2$, $\mathbb{E}\big(|\varphi(Y)|^m|X\big) < C < \infty$;

(H4) K is a bounded and Lipschitz kernel on its support $[0,1]$, such that $-\infty < C < K'(t) < C' < 0$,

(H5) The functions ϕ and $\psi_{S_{\mathcal{F}}}$ are such that:

(H5a) $\exists C>0$, $\exists \eta_0 > 0$, $\forall \eta < \eta_0$, $\phi'(\eta) < C$, and
$$\exists C > 0,\ \exists \eta_0 > 0,\ \forall 0 < \eta < \eta_0,\ \int_0^\eta \phi(u)\,du > C\eta\,\phi(\eta),$$

(H5b) for n large enough:
$$\frac{(\log n)^2}{n\,\phi(h)} < \frac{(\log n)^2}{n\,\phi(b)\,\phi(h)} < \psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) < \frac{n\phi(b)\phi(h)}{\log n} < \frac{n\phi(h)}{\log n}.$$

(H6) The Kolmogorov's ε-entropy of $S_{\mathcal{F}}$ satisfies
$$\sum_{n=1}^{\infty} \exp\left\{(1-\beta)\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)\right\} < \infty,\ \text{for some } \beta > 1.$$

Note that (H5a) implies that for n large enough
$$0 \leqslant \phi(h) \leqslant Ch. \tag{6}$$
The condition (H5b) implies that:
$$\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)} \to 0, \quad \text{and} \quad \frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(b)\,\phi(h)} \to 0. \tag{7}$$
The condition (H6) implies that:
$$\sum_{n=1}^{\infty} N_\varepsilon(S_{\mathcal{F}})^{1-\beta} < \infty. \tag{8}$$
The following Theorem states the rate of convergence of $\hat{m}_R(x)$ and $\hat{m}_S(x)$ for the surrogated scalar response, uniformly over the set $S_{\mathcal{F}}$. 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) [4])

Theorem 1. *Under the hypotheses (H1)–(H6), we have*
$$\sup_{x\in S_{\mathcal{F}}}\left|\hat{m}_R(x) - m(x)\right| = O\big(h^{\beta_1}\big) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\,\phi(h)}}\right) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log N}{N}\right)}{N\phi(h)}}\right) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\,\phi(h)\,\phi(b)}}\right),$$
where h is the concentration of the probability measure of the functional variable X in the ball with center x and radius h and $\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)$ is the entropy function.

Remark 1. This paper has stated uniform consistency results in functional setting. The fact to be able to state results on the quantity $\sup_{x\in S_{\mathcal{F}}} \left|\hat{m}_R(x)-m(x)\right|$ allows directly to obtain result on quantity $\left|\hat{m}_R(x)-m(x)\right|$. 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. For $N = n$ (without surrogate data) we get the estimator presented by F. Ferraty, A. Laksaci, A. Tadj, P. Vieu (2010) [5]:

$$\sup_{x\in S_{\mathcal{F}}} |\hat{m}_R(x)-m(x)| = O(h^{\beta_1}) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\,\phi(h)}}\right).$$

By building a suitable projection-based semi-metric, the entropy function becomes $\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) = O(\log n)$ and for $N = n$ (without surrogate data) we get the estimator of Ferraty and Vieu (2006) [4]

$$|\hat{m}_R(x)-m(x)| = O(h^{\beta_1}) + O_{a.co.}\left(\sqrt{\frac{\log n}{n\,\phi(h)}}\right).$$

4. Numerical examples

Let $\hat{m}_V(x)$ be the classical kernel estimator which is obtained with the true observations in the validation data set V

$$\hat{m}_V(x) = \frac{\sum_{i\in V} K\big(h^{-1}d(x,X_i)\big)\varphi(Y_i)}{\sum_{i\in V} K\big(h^{-1}d(x,X_i)\big)}. \tag{9}$$

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

$$\hat{m}_C(x) = \frac{\sum_{i=1}^{N} K\big(h^{-1}d(x,X_i)\big)\varphi(Y_i)}{\sum_{i=1}^{N} K\big(h^{-1}d(x,X_i)\big)}.$$

Within this section we will evaluate the interest of using $\hat{m}_R(x)$ over $\hat{m}_V(x)$. We choose K the Gaussian kernel as follow:

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

We generate 400 observations $(X_i, Y_i)_i$ using 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

$$X_i(t) = \sin(2\pi W_i t)\cos(2\pi W_i t) + W_i t + b_i,$$

where $b_i \sim N(0,4)$ and $W \sim U(0,4)$.

The response variable Y is generated by taking as a regression operator:

$$m(x) = \pi \int_0^1 x^2(t)\,dt.$$

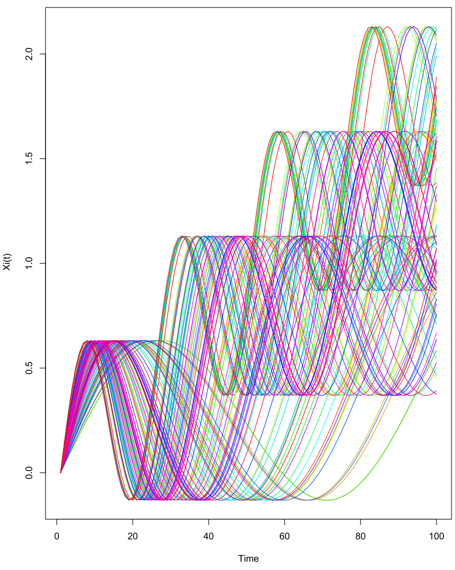

Fig. 1. The panel presents 100 smooth curves of X_i.

Let $I_0 = \{1,\ldots,300\}$ and $I_1 = \{301,\ldots,400\}$ be two subsets of indices. Then, we choose $\Delta = (X_i, Y_i)_{i\in I_0}$ as the learning sample and $\Gamma = \{(X_i,Y_i)\}_{i\in I_1}$ 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 $\varepsilon_i \sim N\big(0, \sqrt{1-\rho^2}\big)$, 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. In the sequel of this simulation study, we take $\rho = 0.75$. 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 $\hat{m}_V(x)$ of $m(x)$. The estimator $\hat{m}_R(x)$ is then constructed by using the surrogate data $\{(X_i, Y_i)\}_{i \in \bar{V}}$ with the help of the validation data, where $\bar{V} \cup V = \{1, \ldots, N\}$. It should be pointed out that for $N = n$ (complete observations), we have $\hat{m}_V(x) = \hat{m}_R(x) = \hat{m}_C(x)$. We evaluate the performance of the estimator $\hat{m}_R(x)$ in terms of prediction, by computing the relative mean squared error (RMSE) on the test sample:

$$\mathrm{RMSE}(\hat{m}_R^x) = \sqrt{\frac{\sum_{i \in \Gamma} \left(\hat{m}_C(X_i) - \hat{m}_R(X_i)\right)^2}{n}}.$$

We have run 100 replicates of the simulation process for various values of n. We computed, for two estimators $\hat{m}_R(x)$ and $\hat{m}_V(x)$ 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 sample:

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

The results are summarized in Table 1 obviously the quality of the prediction of two estimators depends on the size n of the validation data.

Table 1. $\hat{m}_R(x)$ and $\hat{m}_V(x)$ whereas $\hat{m}_C(x)$ ($\rho = 0.75$).

estimator	$p(V)$	Mean	RMSE
$\hat{m}_V(x)$	60%	23.44264	0.082
$\hat{m}_R(x)$	60%	24.31731	0.079
$\hat{m}_C(x)$	—	24.34636	—
$\hat{m}_V(x)$	80%	24.4267	0.076
$\hat{m}_R(x)$	80%	24.3352	0.066
$\hat{m}_C(x)$	—	24.34636	—

Specifically, RMSE decrease as the value of n increases. On the other hand, for $n = 180$ that means the percentage of validation data in sample is 60% our estimator $\hat{m}_R(x)$ is better than $\hat{m}_V(x)$ in term of RMSE inferior.

In addition for $n = 240$ that means that we know 80% of data, our $\hat{m}_R(x)$ still greatly better as result of RMSE $= 0.066$. Nearly with the same mean of $\hat{m}_C(x)$.

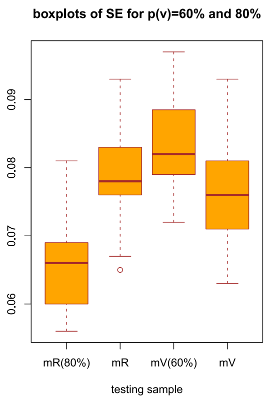

Fig. 2. A boxplots of the SE of $\hat{m}_V$ and $\hat{m}_R$.

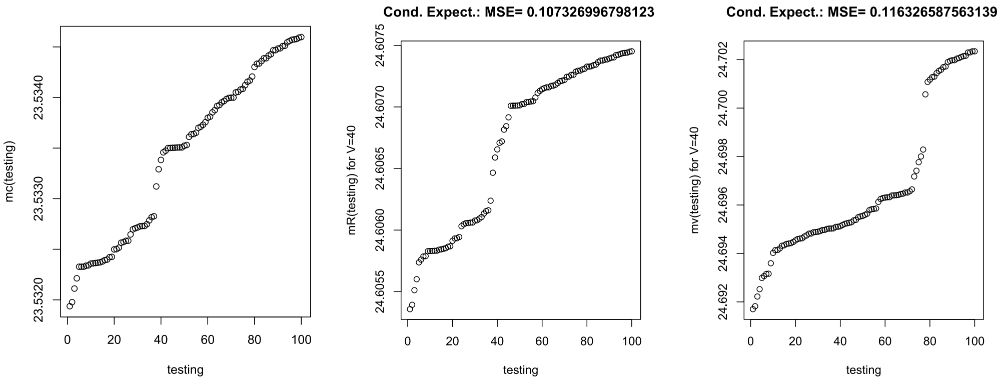

Fig. 3. Prediction by $\hat{m}_V$ (right panel), $\hat{m}_R$ (middle panel) and $\hat{m}_C$ (left panel).

Figure 3 shows the Prediction of a scalar response from testing curves.

It can be noticed from Figure 3 that our $\hat{m}_R(x)$ is closer than $\hat{m}_V(x)$ to the curve $\hat{m}_C(x)$ which represents the estimator with the complete sample and consequently, even the percentage of validation data in sample is 60% our estimator $\hat{m}_R(x)$ performing better than $\hat{m}_V(x)$.

We present in this paper the almost-complete convergence of Generalized regression function for surrogate scalar response given a functional random by using validation and simulated sample set. In addition we show the performance of our estimator to reduce RMSE by employing testing data. That verifies the effectiveness of the theoretical results, this latter gives us an exact rate of convergence of estimator.

However, lot of issues are possible, such comparison of our estimator to other estimator for missing data, additionally we can extend these results to the linear model.

5. Proof of Theorem 1

Firstly, note that:

$$i \in V \Rightarrow i \in \{1,\dots,n\},$$
$$j \in \bar{V} \Rightarrow j \in \{n+1,\dots,N\},$$
$$m^{\tilde{Y}}(x) = \mathbb{E}\big[\varphi(Y)|X=x,\tilde{Y}\big].$$

From now on, we will denote by C is a generic nonnegative real constant, and we will take

$$\varepsilon = \frac{\log n}{n}.$$

Observe that, according to (H1) and (H4) we have

$$\forall\, x \in S_{\mathcal{F}} \quad C\phi(h) < \mathbb{E}[K_1(x)] < C'\phi(h). \tag{10}$$

Note that (H5a) implies that for n large enough

$$0 \leqslant \phi(h) \leqslant Ch. \tag{11}$$

The condition (H5b) implies that:

$$\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)} \to 0 \quad \text{and} \quad \frac{\log n}{n\,h} \to 0. \tag{12}$$

The condition (H6) implies that:

$$\sum_{n=1}^{\infty} N_{\varepsilon}(S_{\mathcal{F}})^{1-\beta} < \infty. \tag{13}$$

Firstly, we write

$$\begin{aligned}\hat{m}_R(x) - m(x) = {} & \sum_{i\in V}\varphi(Y_i)W_{1,n,i}(x) - \sum_{i\in V} m^{\tilde{Y}_i}(x)W_{1,n,i}(x) - \sum_{j\in\bar{V}} m^{\tilde{Y}_j}(x)W_{1,n,j}(x) \\ & + \sum_{j\in\bar{V}} U(X_j,\tilde{Y}_j)W_{1,n,j}(x) + \sum_{i=1}^{N} m^{\tilde{Y}_i}(x)W_{1,n,i}(x) - m(x)\end{aligned}$$

with

$$\begin{cases} E_1 = \sum\limits_{i\in V}\big(\varphi(Y_i) - m^{\tilde{Y}_i}(X_i)\big)W_{1,n,i}(x), \\ E_2 = \sum\limits_{j\in\bar{V}}\big(U(X_j,\tilde{Y}_j) - m^{\tilde{Y}_j}(X_j)\big)W_{1,n,j}(x), \\ E_3 = \sum\limits_{i=1}^{N}\big(m^{\tilde{Y}_i}(X_i) - m(x)\big)W_{1,n,i}(x). \end{cases}$$

Furthermore, we put

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

and we define

$$\begin{cases} \hat{r}_1(x) = \dfrac{1}{n}\sum\limits_{i\in V}\Delta_i(x), \\ \tilde{r}_1(x) = \dfrac{1}{N}\sum\limits_{i=1}^{N}\Delta_i(x), \\ \hat{r}_2(x,y) = \dfrac{1}{n}\sum\limits_{i\in V}\big(\varphi(Y_i) - m^{\tilde{Y}_i}(X_i)\big)\Delta_i(x), \\ \hat{r}_3(x) = \dfrac{1}{N}\sum\limits_{i=1}^{N}\big(m^{\tilde{Y}_i}(X_i) - m(x)\big)\Delta_i(x). \end{cases}$$

By the definition of $\hat{r}_1$ and $\hat{r}_2$:

$$E_1 = \frac{1}{\hat{r}_1(x)}\big(\hat{r}_2(x,y) - \mathbb{E}(\hat{r}_2(x,y))\big) + \frac{\mathbb{E}(\hat{r}_2(x,y))}{\hat{r}_1(x)},$$
$$E_3 = \frac{1}{\tilde{r}_1(x)}\big(\hat{r}_3(x,y) - \mathbb{E}(\hat{r}_3(x,y))\big) + \frac{\mathbb{E}(\hat{r}_3(x,y))}{\tilde{r}_1(x)}.$$

Lemma 1. *Under the hypotheses (H1) and (H4)–(H6), we have*

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

and

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

Proof of this lemma is detailed in [5].

Lemma 2. *Under the hypotheses (H1), (H2) and (H4)–(H6), we have*

$$\sup_{x\in S_{\mathcal{F}}} |\mathbb{E}[\hat{r}_2(x)]| = 0, \quad \sup_{x\in S_{\mathcal{F}}} |\mathbb{E}[\hat{r}_3(x)]| = O(h^{\beta_1}).$$

Proof. By stationarity,

$$\begin{aligned}
\left|\mathbb{E}\left[\hat{r}_2(x)\right]\right| &= \left|\mathbb{E}\left[\Delta_1(x)\mathbb{E}\left[\left(\varphi(Y_i)-m^{\tilde{Y}_1}(X_1)\right)|X_1\right]\right]\right| \\
&= \left|\mathbb{E}\left[\Delta_1(x)\mathbb{E}[\varphi(Y_i)|X_1]-m(X_1)\right]\right| \\
&= \left|\mathbb{E}\left[\Delta_1(x)m(X_1)-m(X_1)\right]\right| \\
&= 0,
\end{aligned}$$

and

$$\begin{aligned}
\left|\mathbb{E}\left[\hat{r}_3(x)\right]\right| &= \left|\mathbb{E}\left[\Delta_1(x)\mathbb{E}\left[\left(m^{\tilde{Y}_i}(X_1)-m(x)\right)|X_1\right]\right]\right| \\
&= \left|\mathbb{E}\left[\Delta_1(x)\mathbb{E}\left[\varphi(Y_i)|X_1\right]-m(x)\right]\right]\right| \\
&= \left|\mathbb{E}\left[\Delta_1(x)m(X_1)-m(x)\right]\right]\right|.
\end{aligned}$$

Under (H2) we obtain

$$\forall x\in S_{\mathcal{F}},\ \left|\mathbb{E}\left[\hat{r}_3(x)\right]\right| \leqslant Ch^{\beta_1}.$$ ■

Lemma 3. *Under the assumptions (H1)–(H6), we have*

$$\sup_{x\in S_{\mathcal{F}}} \left|\hat{r}_2(x)-\mathbb{E}[\hat{r}_2(x)]\right| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\,\phi(h)}}\right), \quad \sup_{x\in S_{\mathcal{F}}} \left|\hat{r}_3(x)-\mathbb{E}[\hat{r}_3(x)]\right| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log N}{N}\right)}{N\phi(h)}}\right).$$

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

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

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

$$\forall\, x\in S_{\mathcal{F}} \quad C\phi(h) < \mathbb{E}[K_1(x)] < C'\phi(h). \tag{14}$$

Next, we denote by $x_1,\ldots,x_{N_\varepsilon(S_{\mathcal{F}})}$ an ε-net (see Kolomogorov and Tikhomirov (1959) [7]) for $S_{\mathcal{F}}$. Furthermore, for all x in $S_{\mathcal{F}}$ we put

$$k(x) = \arg\min_{k\in\{1,2,\ldots,N_\varepsilon(S_{\mathcal{F}})\}} d(x,x_k).$$

Now we use the following decomposition

$$\begin{aligned}
|\hat{r}_2(x)-\mathbb{E}[\hat{r}_2(x)]| \leqslant{}& \underbrace{\sup_{x\in S_{\mathcal{F}}}\left|\hat{r}_2(x)-\hat{r}_2(x_{k(x)})\right|}_{T_1} + \underbrace{\sup_{x\in S_{\mathcal{F}}}\left|\hat{r}_2(x_{k(x)})-\mathbb{E}[\hat{r}_2(x_{k(x)})]\right|}_{T_2} \\
&+ \underbrace{\sup_{x\in S_{\mathcal{F}}}\left|\mathbb{E}[\hat{r}_2(x_{k(x)})]-\mathbb{E}[\hat{r}_2(x)]\right|}_{T_3}.
\end{aligned}$$

For the term T_1:

$$T_1 = \sup_{x\in S_{\mathcal{F}}} \left| \sum_{i\in V} \left(\frac{1}{n\,\mathbb{E}[K_1(x)]} \left(\varphi(Y_i)\big[K_i(x) - K_i(x_k)\big] + \big[m_R(x_k)K_i(x_k) - m_R(x)K_i(x)\big] \right) \right) \right|$$
$$\leqslant \sup_{x\in S_{\mathcal{F}}} \frac{C}{n\phi(h)} \sum_{i\in V} |\varphi(Y_i)|\,|K_i(x) - K_i(x_{k(x)}|\,\mathbb{1}_{B(x,h)\cup B(x_{k(x)},h)}(X_i).$$

Note that we have used the fact:

$$\big|m_R(x_1)K_i(x_1) - m_R(x_2)K_i(x_2)\big| \leqslant C\big|K_i(x_1) - K_i(x_2)\big|.$$

The Lipschitzianity of the kernel K on $[0,1]$ gives

$$T_1 \leqslant \frac{C}{n}\sum_{i=1}^{n} Z_i \quad \text{with} \quad Z_i = \frac{\varepsilon\,\varphi(Y_i)}{h\,\phi(h)}\,\mathbb{1}_{B(x,h)\cup B(x_{k(x)},h)}(X_i).$$

By (H3) we have

$$\mathbb{E}\left(|\varphi(Y)|^m\right) = \mathbb{E}\left(\mathbb{E}\left(|\varphi(Y)|^m|X\right)\right) < C < \infty.$$

So, we get

$$\mathbb{E}(|Z_1|^m) \leqslant \frac{C\varepsilon^m}{h^m\phi(h)^{m-1}}. \tag{15}$$

By using the result (11) together with the definition of ε we have for n large enough: $\frac{\varepsilon}{h} \leqslant C$.

So, we get:

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

Now, By applying Corollary A.8 in Ferraty and Vieu (2006) [4] with $a^2 = \frac{\varepsilon}{h\phi(h)}$, one can get:

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

Finally, applying (15) for $m=1$

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

Using (H5b) together with (12) and the fact that:

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

we get:

$$T_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)}}\right). \tag{16}$$

Similar steps allow to get:

$$T_3 = O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)}}\right). \tag{17}$$

It remains to evaluate T_2. Indeed, we write

$$P\left(T_2 > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)}}\right) = P\left(\max_{k\in\{1,\dots,N_\varepsilon(S_{\mathcal{F}})\}} \big|\hat{r}_2(x_k) - \mathbb{E}\hat{r}_2(x_k)\big| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)}}\right)$$
$$\leqslant N_\varepsilon(S_{\mathcal{F}}) \max_{k\in\{1,\dots,N_\varepsilon(S_{\mathcal{F}})\}} P\left(\big|\hat{r}_2(x_k) - \mathbb{E}\hat{r}_2(x_k)\big| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)}}\right)$$
$$\leqslant N_\varepsilon(S_{\mathcal{F}}) \max_{k\in\{1,\dots,N_\varepsilon(S_{\mathcal{F}})\}} P\left(\left|\frac{1}{n}\sum_{i\in V}\Gamma_i\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)}}\right),$$

where
$$\Gamma_i = \frac{1}{\mathbb{E}[K_1(x)]}\big[K_i(x_k)\varphi(Y_i) - \mathbb{E}\big(K_i(x_k)\varphi(Y_i)\big) + E\big(m_R(x_k)K_i(x_k)\big) - m_R(x_k)K_i(x_k)\big].$$
The same argument as those invoked for proving Lemma 6.3 in Ferraty and Vieu (2006, p. 65) [4] can be used to show that $\mathbb{E}|\Gamma_i|^m = O\left(\phi(h)^{-m+1}\right)$. By applying the exponential inequality given by Corollary A.8.ii in Ferraty and Vieu (2006) [4].

For all $\eta > 0$:
$$P\left(\left|\hat r_2(x_k) - \mathbb{E}\,\hat r_2(x_k)\right| > \eta\sqrt{\tfrac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\,\phi(h)}}\right) \leqslant 2\exp\left\{-C\eta^2\psi_{S_\mathcal{F}}(\varepsilon)\right\}.$$
Therefore, by choosing $C\eta^2 = \beta$
$$P\left(\left|\hat r_2(x_k) - \mathbb{E}\,\hat r_2(x_k)\right| > \eta\sqrt{\tfrac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\,\phi(h)}}\right) \leqslant N_\varepsilon(S_\mathcal{F}) \max_{k\in\{1,\dots,N_\varepsilon(S_\mathcal{F})\}} P\left(\left|\hat r_2(x_k) - \mathbb{E}\,\hat r_2(x_k)\right| > \eta\sqrt{\tfrac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\,\phi(h)}}\right)$$
$$\leqslant C'\big(N_\varepsilon(S_\mathcal{F})\big)^{1-C\eta^2}.$$
By (13), $\sum_{n=1}^{\infty} N_\varepsilon(S_\mathcal{F})^{1-C\eta^2} < \infty$, we obtain:
$$T_2 = O_{a.co.}\left(\sqrt{\tfrac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\,\phi(h)}}\right). \tag{18}$$
For the term $\hat r_3(x,y) - \mathbb{E}[\hat r_3(x,y)]$ we use the same decomposition and we fix $\varepsilon = \frac{\log N}{N}$ to get:
$$\sup_{x\in S_\mathcal{F}}\sup_{y\in S_\mathcal{R}}\left|\hat r_3(x,y) - \mathbb{E}[\hat r_3(x,y)]\right| = O_{a.co.}\left(\sqrt{\tfrac{\psi_{S_\mathcal{F}}\left(\frac{\log N}{N}\right)}{Ng\,\phi(h)}}\right).$$
So, this Lemma can be easily deduced from (16)–(18) . ∎

Lemma 4. *Under the assumptions of Theorem (H1)–(H6), we have* $\forall j \in \bar V$
$$\sup_{x\in S_\mathcal{F}}\left|m^{\tilde Y_j}(X_j) - U(X_j,\tilde Y_j)\right| = O(h^{\beta_1}) + O_{a.co.}\left(\sqrt{\tfrac{\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right)}{n\,\phi(h)\,\phi(b)}}\right).$$

Proof. To simplify we put $X_j = x$, $\tilde Y_j = \tilde y$. The proof is based on the following decomposition
$$U(x,\tilde y) - m_R(x) = \frac{\big[U_2(x,\tilde y) - \mathbb{E}[U_2(x,\tilde y)]\big]}{U_1(x,\tilde y)} + \frac{\big[\mathbb{E}[U_2(x,\tilde y)] - m_R(x)\big]}{U_1(x,\tilde y)} + [1 - U_1(x,\tilde y)]\,\frac{m_R(x)}{U_1(x,\tilde y)},$$
where
$$U_1(x,\tilde y) = \frac{1}{n\,\mathbb{E}\big[K\big(h^{-1}d(x,X_1)\big)K\big(b^{-1}(\tilde y - \tilde Y_1)\big)\big]}\sum_{i\in V} K\big(h^{-1}d(x,X_i)\big)K\big(b^{-1}(\tilde y - \tilde Y_i)\big),$$
and
$$U_2(x,\tilde y) = \frac{1}{n\,\mathbb{E}\big[K\big(h^{-1}d(x,X_1)\big)K\big(b^{-1}(\tilde y - \tilde Y_1)\big)\big]}\sum_{i\in V} K\big(h^{-1}d(x,X_i)\big)K\big(b^{-1}(\tilde y - \tilde Y_i)\big)\varphi(Y_i).$$
Once again the proof is based on separate treatment of the different terms. In particular, we use the same ideas of Lemma 3
$$\left|U_1(x,\tilde y) - \mathbb{E}[U_1(x,\tilde y)]\right| \leqslant \underbrace{\sup_{x\in S_\mathcal{F}}\left|U_1(x,\tilde y) - U_1(x_k,\tilde y)\right|}_{R_1} + \underbrace{\sup_{x\in S_\varepsilon}\left|U_1(x_k,\tilde y) - \mathbb{E}[U_1(x_k,\tilde y)]\right|}_{R_2}$$
$$+ \underbrace{\sup_{x\in S_\mathcal{F}}\left|\mathbb{E}[U_1(x_k,\tilde y)] - \mathbb{E}[U_1(x,\tilde y)]\right|}_{R_3}.$$
For the term R_1 we employ the Lipschitzianity of the kernel K on $[0,1]$ with (H1) and (H2) lead directly
$$R_1 \leqslant \frac{C}{n}\sum_{i=1}^{n} Z_i \quad \text{with} \quad Z_i = \frac{\varepsilon}{h\,\phi(h)\,\phi(b)}\,\mathbb{1}_{B(x,h)\cup B(x_{k(x)},h)}(X_i)\,\mathbb{1}_{[\tilde y - b\leqslant \tilde Y_i\leqslant \tilde y + b]}.$$

It is clear that the assumption (H1) permits to write that

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

So, we get

$$\mathbb{E}(|Z_1|^m) \leqslant \frac{C\varepsilon^m}{h^m\left[\phi(h)\,\phi(b)\right]^{m-1}}.$$

By using the result (11) together with the definition of ε we have for n large enough: $\frac{\varepsilon}{h} \leqslant C$. So, we get:

$$\mathbb{E}(|Z_1|^m) \leqslant \frac{C\varepsilon^{m-1}}{h^{m-1}[\phi(h)\,\phi(b)]^{m-1}}. \tag{19}$$

Now, by applying Corollary A.8 in Ferraty and Vieu (2006) [4] with $a^2 = \frac{\varepsilon}{h\phi(h)\phi(b)}$, we get:

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

Finally, applying (19) for $m = 1$ one gets

$$R_1 = O\left(\frac{\varepsilon}{h}\right) + O_{a.co.}\left(\sqrt{\frac{\varepsilon \log n}{n\,h\,\phi(h)\,\phi(b)}}\right).$$

Using (H5b) together with (12) and the fact that:

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

we get

$$R_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)\,\phi(b)}}\right).$$

Thus, we deduce that

$$R_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)\,\phi(b)}}\right) \quad \text{and} \quad R_3 = O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)\,\phi(b)}}\right). \tag{20}$$

It remains to evaluate R_3. Indeed, we write

$$\begin{aligned}
P\left(R_2 > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(b)\,\phi(h)}}\right) &= P\left(\max_{k\in\{1,\dots,N_\varepsilon(S_{\mathcal{F}})\}} \left|U_1(x_k,\tilde{y}) - \mathbb{E}U_1(x_k,\tilde{y})\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(b)\phi(h)}}\right) \\
&\leqslant N_\varepsilon(S_{\mathcal{F}}) \max_{k\in\{1,\dots,N_\varepsilon(S_{\mathcal{F}})\}} P\left(\left|U_1(x_k,\tilde{y}) - \mathbb{E}\,U_1(x_k,\tilde{y})\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(b)\,\phi(h)}}\right) \\
&\leqslant N_\varepsilon(S_{\mathcal{F}}) \max_{k\in\{1,\dots,N_\varepsilon(S_{\mathcal{F}})\}} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(b)\,\phi(h)}}\right),
\end{aligned}$$

where

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

It follows from the fact that the kernel K is bounded, that $E|\Gamma_i|^2 \leqslant C\,(\phi(b)\phi(h))^{-1}$. Thus, we apply the Bernstein exponential inequality we obtain:

$$P\left(\left|U_1(x_k,\tilde{y}) - \mathbb{E}\,U_1(x_k,\tilde{y})\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(b)\,\phi(h)}}\right) \leqslant 2\exp\left\{-C\eta^2\psi_{S_{\mathcal{F}}}(\varepsilon)\right\}.$$

Therefore, by choosing $C\eta^2 = \beta$, we have:

$$N_\varepsilon(S_{\mathcal{F}}) \max_{k\in\{1,\dots,N_\varepsilon(S_{\mathcal{F}})\}} P\left(\left|U_1(x_k,\tilde{y}) - \mathbb{E}\,U_1(x_k,\tilde{y})\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)\,\phi(b)}}\right) \leqslant C'\big(N_\varepsilon(S_{\mathcal{F}})\big)^{1-C\eta^2}.$$

Finally, we obtain by (H5) and (13)

$$R_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(b)\,\phi(h)}}\right). \tag{21}$$

We get

$$\sup_{x\in S_{\mathcal{F}}}\left|U_1(x,\tilde{y})-1\right| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\,\phi(b)\,\phi(h)}}\right).$$

The first lemma 1 allows to conclude

$$\sum_{n=1}^{\infty} P\left(\inf_{x\in S_{\mathcal{F}}} U_1(x,\tilde{y}) < \frac{1}{2}\right) < \infty.$$

By using the same decomposition:

$$\left|U_2(x,y,\tilde{y}) - \mathbb{E}[U_2(x,y,\tilde{y})]\right| \leqslant \underbrace{\sup_{x\in S_{\mathcal{F}}}\left|U_2(x,y,\tilde{y}) - U_2(x_k,y,\tilde{y})\right|}_{S_1} + \underbrace{\sup_{x\in S_{\mathcal{F}}}\left|U_2(x_k,y,\tilde{y}) - \mathbb{E}[U_2(x_k,y,\tilde{y}]\right|}_{S_2}$$
$$+ \underbrace{\sup_{x\in S_{\mathcal{F}}}\left|\mathbb{E}[U_2(x_k,y,\tilde{y})] - \mathbb{E}[U_2(x,y,\tilde{y})]\right|}_{S_3}.$$

For the term S_1:

$$S_1 = \sup_{x\in S_{\mathcal{F}}}\left|\sum_{i\in V}\left(\frac{1}{n\,\mathbb{E}\left[K_1(x)K_1(\tilde{y})\right]}\left(\varphi(Y_i)\left[K_i(x)-K_i(x_k)\right] + \left[m^{\tilde{y}}(x_k)K_i(x_k) - m^{\tilde{y}}(x)K_i(x)\right]\right)\right)\right|$$
$$\leqslant \sup_{x\in S_{\mathcal{F}}}\frac{C}{n\,\phi(h)\,\phi(b)}\sum_{i\in V}|\varphi(Y_i)|\left|K_i(x)-K_i(x_{k(x)})\right| 1\!\!1_{B(x,h)\cup B(x_{k(x)},h)}(X_i)\, 1\!\!1_{\left[\tilde{y}-b\leqslant \tilde{Y}_i\leqslant \tilde{y}+b\right]}.$$

The Lipschitzianity of the kernel K on $[0,1]$ and (H2) implies that:

$$S_1 \leqslant \frac{C}{n}\sum_{i=1}^{n} Z_i \quad \text{with} \quad Z_i = \frac{\varepsilon|\varphi(Y_i)|}{h\,\phi(h)\,\phi(b)}\, 1\!\!1_{B(x,h)\cup B(x_{k(x)},h)}(X_i)\, 1\!\!1_{\left[\tilde{y}-b\leqslant \tilde{Y}_i\leqslant \tilde{y}+b\right]},$$

So, we can follow the same steps as T_1 and T_3 to get:

$$S_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)\,\phi(b)}}\right) \quad \text{and} \quad S_3 = O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(h)\,\phi(b)}}\right). \tag{22}$$

Following same idea of T_2 to get:

$$S_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\varepsilon)}{n\,\phi(b)\,\phi(h)}}\right), \tag{23}$$

$$\left|\mathbb{E}[U_2(x,y,\tilde{y})] - m_R(x)\right| \leqslant C\,\mathbb{E}\left[K\left(\tfrac{d(x,X_i)}{h}\right)K\left(\tfrac{\tilde{y}-\tilde{Y}_1}{b}\right)\mathbb{E}\left[\left(|\varphi(Y_i) - m^{\tilde{y}}(x)|\right)|(X_1,\tilde{y})|\right]\right]$$
$$\leqslant C\,\mathbb{E}\left[K\left(\tfrac{d(x,X_i)}{h}\right)K\left(\tfrac{\tilde{y}-\tilde{Y}_1}{b}\right)\mathbb{E}\left[|\varphi(Y_i)|\,|(X_1,\tilde{y}) - m^{\tilde{y}}(x)|\right)\right]$$
$$\leqslant C\mathbb{E}\left[K\left(\tfrac{d(x,X_i)}{h}\right)K\left(\tfrac{\tilde{y}-\tilde{Y}_1}{b}\right)\mathbb{E}\left[|m^{\tilde{y}}(X_1) - m^{\tilde{y}}(x)|\right]\right].$$

By (H2) we have

$$\left|\mathbb{E}[U_2(x,y,\tilde{y})] - m^{\tilde{y}}(x)\right| = O\big(h^{\beta_1}\big). \tag{24}$$

So, the Lemma 4 can be easily deduced from (20)–(24). ■

Acknowledgment

The authors greatly thank the Editor in chief and the reviewers and the MMC journal.

[1] Duncan G. J., Hill D. H. An investigation of the extent and consequences of measurement error in labor-economic survey data. Journal of Labor Economics. **3** (4), 508–532 (1985).

[2] Carroll R. J., Wand M. P. Semiparametric estimation in logistic measurement error modelss. Journal of the Royal Statistical Society. **53** (3), 573–585 (1991).

[3] Pepe M. S. Inference using surrogate outcome data and validation sample. Biometrika. **79** (2), 355–365 (1992).

[4] Ferraty F., Vieu P. Nonparametric Functional Data Analysis. Theory and Practice. New York, Springer Series in Statistics (2006).

[5] Ferraty F., Laksaci A., Tadj A., Vieu P. Rate of uniform consistency for nonparametric estimates with functional variables. Journal of Statistical Planning and Inference. **140** (2), 335–352 (2010).

[6] Firas I., Ali Hajj H., Rachdi M. Regression model for surrogate data in high dimensional statistics. Journal of Communications in Statistics – Theory and Methods. **49** (13), 3206–3227 (2019).

[7] Kolmogorov A. N., Tikhomirov V. M. ε-entropy and ε-capacity of sets in function spaces. Uspekhi Matematicheskikh Nauk. **14** (2 (86)), 3–86 (1959).

Узагальнена функція регресії для сурогатної скалярної реакції

Бумахді М.[1], Уассу І.[1], Рахді М.[2]

[1] *Університет Каді Айяда, Національна школа прикладних наук, Марракеш, Марокко*
[2] *Університет Гренобль-Альпи, Гренобль Цедекс 09, Франція*

У цій статті розробляється та узагальнюється оцінка функції регресії для сурогатної скалярної змінної відповіді, яка задана функціонально випадковою. Після цього конструюються деякі асимптотичні властивості в термінах майже повної збіжності, залежно від результату показується перевага запропонованої оцінки в термінах передбачення.

Ключові слова: *сурогатна відповідь; функціональна змінна; майже повна збіжність; оцінки ядра; скалярний відгук; ентропія; напівметричний простір.*