On the real-time kernel estimation of some conditional models for functional covariates

Imane Bouazza[1]

[1]Department of Mathematics, Faculty of Sciences, University of Dr. Moulay Tahar, Saïda, Algeria

imen.bouazza@univ-saida.dz

Routinely, in non-parametric statistical applications, the conditional models frequently arise both in a theoretical framework and application given their crucial significance in prediction issues. In this context, an easily computable, smooth nonparametric estimate of these models is generated using the kernel approach. In situations when data items arrive sequentially and when a large amount of data may be generated quickly, real-time updating algorithms have recently gained significant favor. It demonstrated its efficacy in numerous applications and fields, including economics, finance, and others. We then shall study in this summary some real-time conditional model estimators when the response Y is a real-valued random variable and the covariate $X = x$ belongs to an infinite-dimensional space $\mathcal{H}$ (whose distance is defined by $d_{\mathcal{H}}(x, X_k) = \|x - X_k\|$). Afterwards, under stationary and ergodic conditions, we establish the almost sure convergence (with rates) on a compact set of these estimators, by using exponential inequality adapted to this context. Finally, a simple numerical experiment is realized in the presentation to evaluate the performance of the suggested estimators in comparison to its natural ones, in terms of reducing the computational time without significantly impacting the accuracy.

The main conditional model we are interested in is the conditional distribution function of Y given $X = x$ defined by

$$F_{Y/X}(y/x) = \mathbb{E}\left[\mathbb{I}_{(Y\leq y)}/X = x\right], \ \forall y \in \mathbb{R}.$$

By a simple modification of the traditional Nadaraya-Watson estimation based on the observed samples, we can define a real-time version given by

$$\widehat{F}_n^x(y) = \frac{\sum_{k=1}^{n} L_1\left(a_k^{-1} d_{\mathcal{H}}(x, X_k)\right) L_2\left(b_k^{-1}(y - Y_k)\right)}{\sum_{k=1}^{n} L_1\left(a_k^{-1} d_{\mathcal{H}}(x, X_k)\right)} = \frac{\widehat{\Psi}_n(x,y)}{\widehat{\Upsilon}_n(x)}.$$

where

$$\widehat{\Psi}_n(x,y) = \frac{1}{n\psi_n(x, a_n)} \sum_{k=1}^{n} L_1\left(a_k^{-1} d_{\mathcal{H}}(x, X_k)\right) L_2\left(b_k^{-1}(y - Y_k)\right),$$

and

$$\widehat{\Upsilon}_n(x) = \frac{1}{n\psi_n(x, a_n)} \sum_{k=1}^{n} L_1\left(a_k^{-1} d_{\mathcal{H}}(x, X_k)\right).$$

with $\Psi(\cdot,\cdot)$ is the joint probability function assumed to be continuous and bounded, $\Upsilon(\cdot)$ is the marginal one, the functions L_1 and L_2 are kernels and a_k, b_k are two positive real numbers which decrease to zero as n tends to infinity.

In addition $\psi_n(x, a_n) = n^{-1} \sum_{k=1}^{n} \phi_k(x, a_k)$ where for a ball B of radius a_k centered at x;

$$\phi_k(x, a_k) \leq \mathbb{P}[X_k \in B(x, a_k)/\wp_{k-1}]$$

and for all $k = 1, \dots, n$, $\wp_{k-1}$ is the $\sigma-$field generated by $((X_1, Y_1), \dots, (X_{k-1}, Y_{k-1}))$.

Also, in order to control the estimation based on functional ergodic data, we should use the following version of exponential inequality for partial sums of unbounded martingale differences.

Lemma 1. *Let $(Z_n)_{n\geq 1}$ be a sequence of real martingale differences with respect to the sequence of $\sigma-$ fields $\wp_n = \sigma(Z_1, Z_2, \dots, Z_n)_{n\geq 1}$ generated by the random variables $Z_1, Z_2, \dots, Z_n$. Set $S_n = \sum_{k=1}^{n} Z_k$. For any $p \geq 2$ and for any $n \geq 1$, assume that there exist some nonnegative constants C and d_n such that*

$$\mathbb{E}(Z_n^p/\wp_{n-1}) \leq C^{p-2} p! d_n^2 \text{ almost surely.}$$

Then, for any $\epsilon > 0$, we have

$$\mathbb{P}(|S_n| > \epsilon) \leq 2 \exp\left(-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right), \text{ where } D_n = \sum_{k=1}^{n} d_k^2.$$