

THE ALMOST COMPLETE CONVERGENCE OF THE HIGH-RISK POINT KERNEL FUNCTIONAL CONDITIONAL ESTIMATE FOR ASSOCIATED DATA

HAMZA DAOUDI and IMED BOUAKER

Tahri Mohamed University, Bar, Algeria
Salhi Ahmed University Center, Naama, Algeria.

ABSTRACT

The maximum of the conditional hazard function is a parameter of great importance in statistics studies, because it constitutes the maximum risk of occurrence of an earthquake in a given interval of time. Using the kernel nonparametric estimates of the first derivative of the conditional hazard function, we establish a convergence properties of an estimate of the maximum in the context of quasi-associated data.

1. INTRODUCTION

The estimation of the conditional density function and its derivatives, in statistics functional, was introduced by Ferraty et al. (2006) . These authors obtained almost complete convergence in the case i.i.d. Since this article, an abundant literature has developed on the estimation of the conditional density and its derivatives, in particular in order to use it to estimate the conditional mode. Indeed, considering α -mixing observations, Ferraty et al. (2005) established the almost complete convergence of a kernel estimator of the conditional mode defined by the random variable maximizing the conditional density. Alternatively, Ezzahrioui et al. (2005, 2006) estimated the conditional mode by the point that cancels the derivative of the kernel density estimator.

The latter focused on the asymptotic normality of the estimator proposed in both cases (i.i.d. and mixing). The accuracy of the dominant terms of the quadratic error of the kernel density estimator has been obtained by Laksaci (2007). We refer to Laksaci et al. (2009) for the question of the choice of the smoothing parameter in the estimation of conditional density to functional explanatory variable.

The associated random variables play an important role in a wide variety of areas, including reliability theory, mathematical physics, multivariate statistical analysis, life sciences and in percolation theory. Many works were treated data under positive and negative dependant random variables, one can quote, Newman (1984), Matula (1992) and Roussas (1999).

The concept of quasi-association is a special case of weak dependence introduced by Doukhan et al. (1999) for real-valued stochastic processes. It was applied by Bulinski et al. (2001) to real valued random fields, and it provides a unified approach to studying families of both positively and negatively dependent random variables. To the best of our knowledge, there is few papers dealing with the nonparametric estimation for quasi-associated random variables. We quote, Douge (2010) studied a limit theorem for quasi-associated Hilbertian random variables, Attaoui et al. (2015) studied Asymptotic Results for an M-Estimator of the Regression Function for Quasi-Associated Processes, Tabti et al. (2016) studied Estimation and simulation of conditional hazard function in the quasi-associated framework when the observations are linked via a functional single-index structure, the asymptotic normality of this last estimator was studied by Hamza et al. (2020).

In case of relative regression, Mechab and Laksaci (2016) studied the nonparametric estimate for

associated random variables. Daoudi et al. (2019) studied the asymptotic normality of the non-parametric conditional density function estimate with functional variables for quasi-associated data and the asymptotic normality of the nonparametric conditional distribution function estimate studied by Daoudi and Mechab (2019).

The main contribution of this work is the study of the asymptotic normality of the estimator of the conditional hazard function in case of quasi-associated data. Note that, like all asymptotic statistics nonfunctional parametric, our result is related to the phenomenon of concentration of the probability measure of the explanatory variable and regularity of the functional space of the model.

The paper is organized as follows : in the next section, we present our model. Section 3 is dedicated to fixing notations and hypotheses. We state our main results in Section 4. An application on simulated data is given to validate our theoretical result in Section 5. The auxiliary results and proofs are given in Section 6. We finalize the paper with a conclusion in Section 7.

2. THE MODEL

We consider the random field $Z_i = (X_i, Y_i)$, $i \in \mathbb{N}^N$ with values in $\{\mathcal{F}\} \times \mathcal{Y}$, where $N \in \mathbb{N}^*$ and $(\{\mathcal{F}\}, d)$ is a semi-metric space of possibly dimension infinite. In this context, $(X_i)_{i \in \mathbb{N}^N}$ can be a functional random variable.

Note that since over ten years, the statistical community has concerned with the development of models and methods adapted to these functional data situations. While the first studies in this direction have mainly focused on linear models (see Bosq [?], Ramsay and Silverman [?]), recent developments (see Ferraty and Vieu [?]) report models non-parametric suitable for this type of data.

Subsequently, we fix a point x in $\{\mathcal{F}\}$ (respectively, a compact $S \in \mathcal{F}$), we assume that the spatial observations $(X_i, Y_i)_{i \in \mathbb{N}^N}$ have the same distribution as $Z := (X, Y)$ and that the regular version of the probability conditional of Y knowing $X = x$ exists and admits a density bounded with respect to the Lebesgue measure on \mathcal{Y} , noted f^x . The functional parameter studied in this article, noted h^x , is defined, for any $y \in \mathcal{Y}$ such that $F^x(y) < 1$, by :

$$h^x(y) = \frac{f^x(y)}{1 - F^x(y)},$$

where F^x is the conditional distribution function of Y knowing $X = x$.

And on the other hand, we assume that the field functional randomness is observed on the set $I_{\mathbf{n}} = \{\mathbf{i} = (i_1, \dots, i_N) \in \mathbb{N}^N, 1 \leq i_k \leq n_k, k = 1, \dots, N\}$, $\mathbf{n} = (n_1, \dots, n_N) \in \mathbb{N}^N$ and we estimate the conditional distribution function by :

$$F^x(y) = \frac{\sum_{\mathbf{i} \in I_{\mathbf{n}}} K(h_K^{-1}d(x, X_{\mathbf{i}}))H(h_H^{-1}(y - Y_{\mathbf{i}}))}{\sum_{\mathbf{i} \in I_{\mathbf{n}}} K(h_K^{-1}d(x, X_{\mathbf{i}}))}, \quad \forall y \in \mathcal{Y}$$

where K is a kernel, H a distribution function and $h_K = h_{K, \mathbf{n}}$ (respectively, $h_H = h_{H, \mathbf{n}}$) a sequence of positive real numbers.

We deduce from F^x the density estimator conditional, denoted f^x , defined by :

$$f^x(y) = \frac{h_H^{-1} \sum_{\mathbf{i} \in I_{\mathbf{n}}} K(h_K^{-1}d(x, X_{\mathbf{i}}))H'(h_H^{-1}(y - Y_{\mathbf{i}}))}{\sum_{\mathbf{i} \in I_{\mathbf{n}}} K(h_K^{-1}d(x, X_{\mathbf{i}}))}, \quad \forall y \in \mathcal{Y}$$

where H' is the derivative of H . Note that, if $N = 1$, we obtains the same estimators from Ferraty et al. [?]. The natural estimator of the hazard function conditional, denoted \hat{h}^x , is :

$$\hat{h}^x(y) = \frac{\hat{f}^x(y)}{1 - \hat{F}^x(y)}. \quad \forall y \in \mathcal{Y}$$

Estimation of the high risk point : We will estimate the point at high risk in S , denoted by $\theta(x)$, defined by

$$h^x(\theta(x)) = \max_{y \in S} h^x(y). \quad (1)$$

This model is of great interest in statistics, particularly in seismic risk analysis. In our functional context, it is assumed that it exist a single point $\theta(x)$ in S verifying (??). The estimator natural of $\theta(x)$, denoted by $\hat{\theta}(x)$, is given by :

$$h^x(\hat{\theta}(x)) = \max_{y \in S} h^x(y). \quad (2)$$

Usually, this estimator is not unique. So, throughout the rest of this article $\hat{\theta}(x)$ will denote any random variable checking (??).

In order to study the almost complete convergence of the estimator $\hat{\theta}(x)$ we keep the assumptions of the previous section and we assume that the function h^x is of class C^2 with respect to y and such that :

$$h^{x'}(\theta(x)) = 0 \quad \text{and} \quad h^{x''}(\theta(x)) < 0. \quad (3)$$

Consider $Z_i = (X_i, Y_i)_{1 \leq i \leq n}$ be a n quasi-associated random identically distributed as the random $Z = (X, Y)$, with values in $\mathcal{H} \times \mathbb{R}$, where \mathcal{H} is a separable real Hilbert space with the norm $\| \cdot \|$ generated by an inner product $\langle \cdot, \cdot \rangle$.

We consider the semi-metric d defined by $\forall x, x' \in \mathcal{H} / d(x, x') = \| x - x' \|$. In the following x will be a fixed point in \mathcal{H} and \mathcal{N}_x will denote a fixed neighborhood of x and \mathcal{S} will be fixed compact subset of \mathbb{R} .

We intend to estimate conditional hazard $h^x(y)$ using n dependent observations $(X_i, Y_i)_{i \in \mathbb{N}}$ draw from a random variables with the same distribution with $Z := (X, Y)$.

To estimate the conditional distribution function and the conditional density we consider the following functional kernel estimators :

$$\hat{F}^x(y) = \frac{\sum_{i=1}^n K \left(h_K^{-1} d(x, X_i) \right) H \left(h_H^{-1} (y - Y_i) \right)}{\sum_{i=1}^n K \left(h_K^{-1} d(x, X_i) \right)}, \quad \forall y \in \mathbb{R} \quad (4)$$

where K is the kernel, H is a given distribution function and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is a sequence of positive real numbers which converges to 0 when $n \rightarrow \infty$. We put

$$K_i(x) = K \left(h_K^{-1} d(x, X_i) \right) \quad \text{and} \quad H_i(y) = H \left(h_H^{-1} (y - Y_i) \right)$$

we can write

$$\hat{F}^x(y) = \frac{\hat{F}_N(y, x)}{\hat{F}_D(x)}$$

with

$$\hat{F}_N(y, x) = \frac{1}{n \mathbb{E} [K_1(x)]} \sum_{i=1}^n K_i(x) H_i(y)$$

and for

$$\hat{F}_D(x) = \frac{1}{n \mathbb{E} [K_1(x)]} \sum_{i=1}^n K_i(x)$$

We define the kernel estimator $\hat{f}^x(y)$ of $f^x(y)$ by :

$$\hat{f}^x(y) = \frac{h_H^{-1} \sum_{i=1}^n K \left(h_K^{-1} d(x, X_i) \right) H' \left(h_H^{-1} (y - Y_i) \right)}{\sum_{i=1}^n K \left(h_K^{-1} d(x, X_i) \right)}, \quad \forall y \in \mathbb{R} \quad (5)$$

where H' is the derivative of H . We can write

$$\hat{f}^x(y) = \frac{\hat{f}_N(y, x)}{\hat{F}_D(x)}$$

where

$$\hat{f}_N(y, x) = \frac{1}{nh_H \mathbb{E}[K_1(x)]} \sum_{i=1}^n K_i(x) H'_i(y)$$

Finally, the estimator of the conditional hazard function is $\hat{h}^x(y)$ defined by

$$\hat{h}^x(y) = \frac{\hat{f}^x(y)}{1 - \hat{F}^x(y)}, \quad \forall y \in \mathbb{R} \tag{6}$$

3. NOTATIONS AND HYPOTHESES

All along the paper, when no confusion will be possible, we will denote by C or/and C' some strictly positive generic constants whose values are allowed to change. The variable x is a fixed point in \mathcal{H} , \mathcal{N}_x is a fixed neighborhood of x . We assume that the random pair $Z_i = \{(X_i, Y_i), i \in \mathbb{N}\}$ is stationary quasi-associated processes.

Let λ_k the covariance coefficient defined as :

$$\lambda_k = \sup_{s \geq k} \sum_{|i-j| \geq s} \lambda_{i,j}$$

where

$$\lambda_{i,j} = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} |cov(X_i^k, X_j^l)| + \sum_{k=1}^{\infty} |cov(X_i^k, Y_j)| + \sum_{l=1}^{\infty} |cov(Y_i, X_j^l)| + cov(Y_i, Y_j) |.$$

X_i^k denotes the k^{th} component of X_i defined as $X_i^k := \langle X_i, e^k \rangle$.

For $h_K > 0$, let $B_\theta(x, h_K) := \{x' \in \mathcal{H} / d(x', x) < h_K\}$ be the ball of center x and radius h_K .

Now, we will state the following assumptions that are necessary to show our main result :

(H1) $\mathbb{P}(X \in B(x, h_K)) = \phi(x, h_K) > 0$ and there exists a function $\beta(x, \cdot)$ such that :

$$\forall \epsilon \in [0, 1], \lim_{h \rightarrow 0} \frac{\phi(x, sh_K)}{\phi(x, h_K)} = \beta(x, \epsilon).$$

(H2) For $l \in \{0, 2\}$, the functions $\Phi_l(s) = \mathbb{E}[\frac{\partial^l F^x(y)}{\partial y^l} - \frac{\partial^l F^x(y)}{\partial y^l} \mid d(x, X) = s]$ are differentiable at $s = 0$.

(H3) The conditional distribution $F^x(y)$ satisfies the Hölder condition, that is : $\forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x, \forall (y_1, y_2) \in \mathcal{S}^2$

$$|F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C \left(d^{b_1}(x_1, x_2) + |y_1 - y_2|^{b_2} \right), \quad b_1 > 0, b_2 > 0.$$

where \mathcal{S} is a fixed compact subset of \mathbb{R} .

(H4) H is a cumulative distribution and H' is bounded and lipschitzian function, such that $\int H'(t) dt = 1, \int |t|^{b_2} H'(t) dt < \infty$ and $\int H'^2(t) dt < \infty$.

(H5) K is a bounded, Lipschitzian and differentiable function such that : there exist two constants C and C' with

$$C \mathbb{I}_{[0,1]}(\cdot) < K(\cdot) < C' \mathbb{I}'_{[0,1]}(\cdot)$$

where $\mathbb{I}_{[0,1]}$ is the indicator function on $[0, 1]$, and its derivative K' is such that $-\infty < C < K'(t) < C' < 0$ for $0 \leq t \leq 1$.

(H6) The bandwidths (h_K, h_H) satisfied :

$$(i) \lim_{n \rightarrow \infty} h_K = 0, \lim_{n \rightarrow \infty} h_H = 0 \text{ and } \lim_{n \rightarrow \infty} (h_H^{b_2} + h_K^{b_1}) \sqrt{nh_H \phi(x, h_K)} = 0$$

(H7) The sequence of random pairs $(X_i, X_j), i \in \mathbb{N}$ is quasi-associated with covariance coefficient $\lambda_k, k \in \mathbb{N}$ satisfying :

$$\exists \alpha > 0, \exists C > 0, \text{ such that } \lambda_k \leq Ce^{-\alpha k}.$$

(H8) $0 < \sup_{i \neq j} \mathbb{P}[(X_i, X_j) \in B(x, h_K) \times B(x, h_K)] = O(\phi^2(x, h_K)).$

4. MAIN RESULT

Theorem 1 Under the assumptions (H1)-(H7) and (??), we have,

$$|\widehat{\theta}(x) - \theta(x)| = O\left(h_K^{b_1/2} + h_H^{b_2/2}\right) + O\left(\left(\frac{\log n}{n h_H \phi_x(h_K)}\right)^{\frac{1}{4}}\right), \quad a.co.$$

Lemma 2 Under the assumptions (H1)-(H7), we have,

$$y \in \mathbb{S} |\widehat{h}^x(y) - h^x(y)| = O\left(h_K^{b_1} + h_H^{b_2}\right) + O\left(\left(\frac{\log n}{n h_H \phi_x(h_K)}\right)^{\frac{1}{2}}\right), \quad a.co.$$

The proof of Lemma 4.2 is based on the following results :

Lemma 3 Under the conditins of theorem 4.1, we have,

$$\sup_{y \in \mathbb{S}} |f_N^x(y) - E f_N^x(y)| = O\left(\left(\frac{\log n}{n \phi_x(h_K)}\right)^{\frac{1}{2}}\right). \quad a.co.$$

$$\sup_{y \in \mathbb{S}} |F_N^x(y) - E F_N^x(y)| = O\left(\left(\frac{\log n}{n \phi_x(h_K)}\right)^{\frac{1}{2}}\right). \quad a.co.$$

Lemma 4 Under the assumptions (H1) and (H4)-(H7), we have,

$$\sup_{y \in \mathbb{S}} |F^x(y) - \overset{x}{N}(y)| = O\left(h_K^{b_1} + h_H^{b_2}\right).$$

$$\sup_{y \in \mathbb{S}} |f^x(y) - \overset{x}{N}(y)| = O\left(h_K^{b_3} + h_H^{b_4}\right).$$

Lemma 5 Under the assumptions (H1)-(H3), (H5) and (H7), we have,

$$\overset{x}{D} - \overset{x}{D} = O\left(\left(\frac{\log n}{n \phi_x(h_K)}\right)^{\frac{1}{2}}\right), \quad a.co.$$

Lemma 6 Under the conditions of lemma 4.2, we have,

$$\exists \eta > 0 \quad \text{such that} \quad \sum_{i=1}^n P\left\{\inf_{y \in \mathbb{S}} |1 - \widehat{F}^x(y)| < \eta\right\} < \infty.$$

Corollary 7 Under the assumptions of Lemma 4.2, we have,

$$\sum_{i=1}^{\infty} \mathbb{P}\left(|\widehat{F}_D^x(y)| < 1/2\right) < \infty$$

5. REFERENCES

- Aaidi, A., Ferraty, F., Kassa, R. (2008). Cross-validated estimation in the single functional index model. *Statistics*. 42 : 475-494.
- Attaoui, S., Laksaci, A., Ould-Sa E. (2011). A note on the conditional density estimate in the single functional index model. *Statist. Probab. Lett.* 81 : 45-53.
- Attaoui, S., Laksaci, A., Ould-Sa E. (2015). Asymptotic Results for an M-Estimator of the Regression Function for Quasi-Associated Processes. *Functional Statistics and Applications, Contributions to Statistics*, 10.1007/978-3-319-22476-3-1.
- Attaoui, S., Ling, N. (2016). Asymptotic results of a nonparametric conditional cumulative distribution estimator in the single functional index modeling for time series data with applications. *Metrika*. 79 : 485-511 DOI 10.1007/s00184-015-0564-6.
- Bulinski, A., Suquet, C. (2001). Normal approximation for quasi-associated random fields. *Statist. Probab. Lett.* 54 : 215-226.
- Bulinski, A., Shabanovich, E. (1998). Asymptotical behaviour of some functionals of positively and negatively dependent random fields. (in Russian). *Fundam. Appl. Math.* 4 : 479-492.
- Douge, L. (2010). Ths limites pour des variables quasi-associ hilbertiennes. *Ann. I.S.U.P.* 54 : 51-60.
- Doukhan, P., Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stoch. Proc. Appl.* 84 : 313-342.
- Esary, J., Proschan, F., Walkup, D. (1967). Association of random variables with applications. *Ann. Math. Statist.* 38 : 1466-1476.
- Ferraty, F., Peuch, A., Vieu, P. (2003). Mod dice fonctionnel simple. *C. R. A. S. Mathtiques Paris*. 336 : 1025-1028.
- Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis : theory and practice*, Springer Series in Statistics, New York.
- Laksaci, A., Mechab, B. (2010). Estimation non-paramique de la fonction de hasard avec variable explicative fonctionnelle : cas des donn spatiales. *Rev. Roumaine Math. Pures Appl.* 55 : 35-51.
- Laksaci, A., Mechab, B. (2014). Conditional hazard estimate for functional random fields. *Journal of Statistical Theory and Practice*. 8 : 192-200.
- Laksaci, A., Mechab, W. (2016). Nonparametric relative regression for associated random variables. *Metron*, 10.1007/s40300-016-0084-9.
- Newman, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables : In Inequalities in Statistics and Probability, IMS Lect. Notes-Monographs Series. 5 : 127-140.
- Quintela-del-Río, A. (2008). Hazard function given of functional variable : Nonparametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* 20 : 413-430.
- Tabti, H., Ait Saïdi, A. (2018). Estimation and simulation of conditional hazard function in the quasi-associated framework when the observations are linked via a functional single-index structure. *Commu. Stat. Theory and Methods*. 47 : 816-838.