Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process Based on Dependent Functional Data

Salim Bouzebda, Fethi Madani and Youssouf Souddi

1Alliance Sorbonne Université, Université de Technologie de Compiègne, L.M.A.C., Compiègne, France
2Université Dr Tahar Moulay de Saida, Algérie
Laboratoire des Modèles Stochastiques, Statistique et Applications

Abstract

The purpose of this paper is to establish the invariance principle for the conditional set-indexed empirical process formed by strong mixing random variables when the co-variates are functional. We establish our results under some assumptions on the richness of the index class $C$ of sets in terms of metric entropy with bracketing. We apply our main result for testing the conditional independence, that is, testing whether two random vectors $Y_1$ and $Y_2$ are independent, given $X$. The theoretical results of the present paper are (or will be) key tools for many further developments in functional data analysis.

Keywords. Conditional distribution; Nadaraya-Watson regression estimator; Empirical process; Strong mixing; Functional data; Semi-metric space; Covering number; Small ball probability.


1 Introduction

The theory of empirical processes is one of the major continuing themes in the historical development of mathematical statistics and it has many applications ranging from parameter estimation to hypothesis testing, its history theory dates back to the 1930’s and 1940’s there has been a great deal research works. The asymptotic properties of empirical processes indexed by functions have been intensively studied during the past decades (see, e.g., van der Vaart and Wellner (1996) or Dudley (1999) for self-contained, comprehensive books on the topic with various statistical applications). Vapnik and Červonenkis (1971) characterize, modulo measurability, the classes $C$ of sets for which the Glivenko-Cantelli theorem holds, in the independent framework. In this setting many papers were published, we cite among many others Dudley (1978), Giné and Zinn (1984), Le Cam (1983), Pollard (1982) and Bass

$1$ E-mail: souddiyoucef@yahoo.fr
and Pyke (1984). Dudley (1978) studied the empirical process indexed by a class of measurable sets, that is, he considered $\mathcal{F} = \{ \mathbf{1}_A(\cdot) : A \in \mathcal{A} \}$, where $\mathcal{A}$ is a suitable subset of the Borel $\sigma$-algebra. He obtained several very useful results that go far beyond Donsker’s theorem, more precisely, he stated different assumptions under which weak convergence to a Gaussian process holds, including a so-called metric entropy with inclusion. Generalizing this idea, Ossiander (1987) introduced $L_2$-brackets to approximate the elements of $\mathcal{F}$. These brackets allow to study larger classes of functions as long as a metric entropy integrability condition is satisfied, see Ossiander (1987), Theorem 3.1. To deal with random variables such as time series that are dependent, one naturally asks whether results obtained under the independence assumption remain valid. However, a bracketing condition under strong mixing was stated by Andrews and Pollard (1994). Doukhan et al. (1995) studied the function-indexed empirical process for $\beta$-mixing sequences. The case of Gaussian long-range dependent random vectors was already handled by Arcones (1994), Theorem 9. The assumption on the bracketing number therein is very restrictive and was considerably improved later. In this lines of research in different type of mixing, we may cite Eberlein (1984), Nobel and Dembo (1993) and Yu (1994). The extension of the above exploration to conditional empirical processes is practically useful and technically more challenging, we may refer to Stute (1986a), Stute (1986b), Horváth and Yandell (1988) for the case of independent observations, other authors were interested to the dependent case, for example Yoshihara (1990) established the asymptotic normality when the sequences are $\phi$-mixing. Polonik and Yao (2002) have established uniform convergence and asymptotic normality of set-indexed conditional empirical process in a strictly stationary and strong mixing framework. The results of Polonik and Yao (2002) were extended by Poryvaı˘ı (2005). In the present paper, we are interested in the limiting behavior of conditional set-indexed empirical process when the covariates are functional. Functional data analysis is a field that has been really popularized with the book by Ramsay and Silverman (2005a) and that received a lot of attention in the last 20 years with a general aim of adapting existing multivariate ideas to the functional framework. For good sources of references to research literature in this area along with statistical applications consult Ramsay and Silverman (2005a), Bosq (2000), Ramsay and Silverman (2005b), Ferraty and Vieu (2006), Bosq and Blanke (2007), Shi and Choi (2011), Horváth and Kokoszka (2012), Zhang (2014), Bongiorno et al. (2014), Hsing and Eubank (2015) and Aneiros et al. (2017). Dimensionality effects have tended to slow down the development of nonparametric modelling ideas in infinite-dimensional setting. However, this field has been investigated many years ago by Ferraty and Vieu (2006) and caused up considerable interest since several hundreds of papers have been published in the last decade. More precisely, dimensionality problem links with probability theory in infinite-dimensional space by means of the small ball probability function of the underlying process and with the topological structure on the infinite-dimensional space. More precisely the interest of using a semi-metric-type topology are discussed in details in the book of Ferraty and Vieu (2006), we may refer for recent references to Bouzebda and Nemouchi (2020, 2021); Bouzebda and Nezzal (2021).
This paper extends asymptotic results for multivariate statistics of set-indexed conditional empirical process to the context of functional statistical samples. We establish the uniform convergence and asymptotic normality when the observations assumed are strong mixing taking its values in semi-metric space. It should be noted that even for i.i.d. functional data, no weak convergence has so far been established. To the best of our knowledge, the results presented here, respond to a problem that has not been studied systematically up to the present, which was the basic motivation of the paper.

The remainder of this paper is organized as follows. Section 2, we present the notation and definitions together with the conditional empirical process. Section 2.1, we give our main results. An application of our main result to the test of the conditional independence is given in Section 4. Some concluding remarks and possible future developments are relegated to Section 5. To prevent from interrupting the flow of the presentation, all proofs are gathered in Section ??.

2 The set indexed conditional empirical process

We consider a sample of random elements \((X_1, Y_1), \ldots, (X_n, Y_n)\) copies of \((X, Y)\) that takes its value in a space \(\mathcal{E} \times \mathbb{R}^d\). The functional space \(\mathcal{E}\) is equipped with a semi-metric \(d_{\mathcal{E}}(\cdot, \cdot)\).

We aim to study the links between \(X\) and \(Y\), by estimating functional operators associated to the conditional distribution of \(Y\) given \(X\) such as the regression operator, for some measurable set \(C\) in a class of sets \(\mathcal{C}\),

\[
G(C \mid x) = \mathbb{E} \left( \mathbf{1}_{\{Y \in C\}} \mid X = x \right).
\]

This regression relationship suggests to consider the following Nadaraya Watson-type (Nadaraja (1964) and Watson (1964)) conditional empirical distribution:

\[
G_n(C, x) = \frac{\sum_{i=1}^{n} \mathbf{1}_{\{Y_i \in C\}} K \left( \frac{d_{\mathcal{E}}(x, X_i)}{h_n} \right)}{\sum_{i=1}^{n} K \left( \frac{d_{\mathcal{E}}(x, X_i)}{h_n} \right)},
\]

(2.1)

where \(K(\cdot)\) is a real-valued kernel function from \([0, \infty)\) into \([0, \infty)\) and \(h_n\) is a smoothing parameter satisfying \(h_n \to 0\) as \(n \to \infty\), \(C\) is a measurable set, and \(x \in \mathcal{E}\). By choosing \(C = (-\infty, z], z \in \mathbb{R}^d\), it reduces to the conditional empirical distribution function \(F_n(z \mid x) = G_n((-\infty, z], x)\), refer to Stute (1986a), Stute (1986b), Horváth and Yandell (1988). However, the corresponding class \(\mathcal{C} = \{-\infty, z], z \in \mathbb{R}^d\}\). Concerning the semi-metric topology defined on \(\mathcal{E}\), we will use the notation

\[
B(x, t) = \{x_1 \in \mathcal{E} : d_{\mathcal{E}}(x_1, x) \leq t\},
\]

\footnote{A semi-metric (sometimes called pseudo-metric) \(d(\cdot, \cdot)\) is a metric which allows \(d(x_1, x_2) = 0\) for some \(x_1 \neq x_2\).}
for the ball in $E$ with center $x$ and radius $t$. We denote
\[ F(t; x) = \mathbb{P}(d_E(x, X_i) \leq t) = \mathbb{P}(X_i \in B(x, t)) = \mathbb{P}(D_i \leq t), \]
which is usually called in the literature the small ball probability function when $t$ is decreasing to zero. One is interested in the behavior of $F(u; x)$ as $u \to 0$. Gasser et al. (1998) assume that $F(h; x) = \phi(h_n) f_1(x)$ as $h \to 0$ and refer to $f_1(x)$ as the probability density (functional). When $\mathcal{H} = \mathbb{R}^m$, then $F(h; x) = P[\|x - X_i\| \leq h]$ and it can be seen that in this case $\phi(h_n) = C(m) h^m$ ($C(m)$ is the volume of a unit ball in $\mathbb{R}^m$) and $f_1(x)$ is the probability density of the random variable $X_1$. Indeed, it can be shown directly that \[ \lim_{h \to 0} \left( \frac{1}{h^m} \right) F(h; x) = C(m) f_1(x). \] Motivated by the work of Gasser et al. (1998) and the above argument we make the assumption (H4)(i)-(ii), refer to this discussion and details to Masry (2005).

Often statistical observations are not independent but are not far from being independent. If not taken into account, dependence can have disastrous effects on statistical inference. The notion of mixing quantifies how close to independence a sequence of random variables is, and it can help us to extend classical results for independent sequences to weakly dependent or mixing sequences, refer to Bradley (2007) for more details. Let us specify the dependence that we will consider in the present paper.

**Definition 1.** A sequence $\{\zeta_k, k \geq 1\}$ is said to be $\alpha$-mixing if the mixing coefficient
\[ \alpha(n) \equiv \sup_{k \geq 1} \sup \{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_{n+k}, B \in \mathcal{F}_1 \} \]
converges to zero as $n \to \infty$, where $\mathcal{F}_m = \sigma\{\zeta_l, \zeta_{l+1}, \ldots, \zeta_m\}$ denotes the $\sigma$-algebra generated by $\zeta_l, \zeta_{l+1}, \ldots, \zeta_m$ with $l \leq m$. We use the term geometrically strong mixing if, for some $a > 0$ and $\beta > 1$,
\[ \alpha(j) \leq a^{-j^\beta}, \]
and exponentially strong mixing if, for some $b > 0$ and $0 < \gamma < 1$,
\[ \alpha(k) \leq b \gamma^k. \]

Throughout the sequel, we assume tacitly that sequence of random elements $\{(X_i, Y_i), i = 1, \ldots, n\}$ is strongly mixing.

**2.1 Assumptions and notation**

Throughout this paper $x$ is a fixed element of the functional space $E$. We define metric entropy with inclusion which provides a measure of richness(or complexity) of class of sets
\( \mathcal{C} \). For each \( \varepsilon > 0 \), the covering number is defined as:

\[
\mathcal{N}(\varepsilon, \mathcal{C}, \mathcal{G} \mid x) = \inf \{ n \in \mathbb{N} : \exists C_1, \ldots, C_n \in \mathcal{C} \text{ such that } \forall C \in \mathcal{C} \exists 1 \leq i, j \leq n \text{ with } C_i \subset C \subset C_j \text{ and } \mathcal{G}(C_j \setminus C_i \mid x) < \varepsilon \},
\]

the quantity \( \log(\mathcal{N}(\varepsilon, \mathcal{C}, \mathcal{G} \mid x))) \) is called metric entropy with inclusion of \( \mathcal{C} \) with respect to \( \mathcal{G}(\cdot \mid x) \). Estimates for such covering numbers are known for many classes; see, e.g., Dudley (1984). We will often assume below that either \( \log(\mathcal{N}(\varepsilon, \mathcal{C}, \mathcal{G} \mid x)) \) or \( \mathcal{N}(\varepsilon, \mathcal{C}, \mathcal{G} \mid x)) \) behave like powers of \( \varepsilon^{-1} \). We say that the condition \((R_\gamma)\) holds if

\[
\log(\mathcal{N}(\varepsilon, \mathcal{C}, \mathcal{G} \mid x)) \leq H_\gamma(\varepsilon), \text{ for all } \varepsilon > 0,
\]

(2.2)

where

\[
H_\gamma(\varepsilon) = \begin{cases} 
\log(A\varepsilon) & \text{if } \gamma = 0, \\
A\varepsilon^{-\gamma} & \text{if } \gamma > 0,
\end{cases}
\]

for some constants \( A, r > 0 \). As in Polonik and Yao (2002), it is worth noticing that the condition \((2.2)\), \( \gamma = 0 \), holds for intervals, rectangles, balls, ellipsoids, and for classes which are constructed from the above by performing set operations union, intersection and complement finitely many times. The classes of convex sets in \( \mathbb{R}^d \) \( (d \geq 2) \) fulfill the condition \((2.2)\), \( \gamma = (d - 1)/2 \). This and other classes of sets satisfying \((2.2)\) with \( \gamma > 0 \), can be found in Dudley (1984). In this section, we establish the weak convergence of the process \( \{\tilde{v}_n(C, x) : C \in \mathcal{C}\} \) defined by

\[
\tilde{v}_n(C, x) := \sqrt{n\phi(h_n)} \left( G_n(C, x) - \mathbb{E}G_n(C, x) \right).
\]

(2.3)

In our analysis, we will make use of the following assumptions.

\textbf{(H1)} For all \( t > 0 \), we have \( \phi(t) > 0 \). For all \( t \in (0, 1) \), \( \tau_0(t) \) exists, where

\[
\tau_0(t) = \lim_{r \to 0} \frac{\phi(rt)}{\phi(r)} = \lim_{r \to 0} \frac{\mathbb{P}(d_{\mathcal{E}}(x, X) \leq rt)}{\mathbb{P}(d_{\mathcal{E}}(x, X) \leq t)} < \infty;
\]

\textbf{(H2)} There exist \( \beta > 0 \) and \( \eta_1 > 0 \), such that for all \( x_1, x_2 \in N_x \), a neighborhood of \( x \), we have

\[
|G(C \mid x_1) - G(C \mid x_2)| \leq \eta_1 d_{\mathcal{E}}^\beta(x_1, x_2);
\]

(i) Let \( g_2(u) = \text{Var} \left( I_{\{Y_i \in \mathcal{C}\}} \mid X_j = u \right) \) for \( u \in \mathcal{E} \). Assume that \( g_2(u) \) is independent of \( j \) and is continuous in some neighborhood of \( x \), as \( h \to 0 \),

\[
\sup_{\{u : d(x, u) \leq h\}} \left| g_2(u) - g_2(x) \right| = o(1),
\]

Assume

\[
g_{\psi}(u) = \mathbb{E}(|I_{\{Y_i \in \mathcal{C}\}} - G(C \mid x)|^\psi \mid X_i = u), u \in \mathcal{E},
\]

\[
\psi > 0.
\]
is continuous in some neighborhood of $x$,

(ii) Define, for $i \neq j, u, v \in \mathcal{E}$,

$$
g(u, v; x) = \mathbb{E}((\mathbb{1}_{Y_i \in C} - G(C \mid x))(\mathbb{1}_{Y_j \in C} - G(C \mid x)) \mid X_i = u, X_j = v).$$

Assume that $g(u, v; x)$ does not depend on $i, j$ and is continuous in some neighborhood of $(x, x)$;

\textbf{(H3)} There exist $m \geq 2$ and $\eta_2 > 0$, such that, we have, almost surely

$$
\mathbb{E}(|Y|^m \mid X) \leq \eta_2 < \infty;
$$

\textbf{(H4)}

(i) For all $i \geq 1$,

$$
0 < c_5 \phi(h_n) f_1(x) \leq \mathbb{P}(X_i \in B(x, h)) = F(h; x) \leq c_6 \phi(h_n) f_1(x),
$$

where $\phi(h_n) \to 0$ as $h \to 0$ and $f_1(x)$ is a nonnegative functional in $x \in \mathcal{E}$,

(ii) We have

$$
\sup_{i \neq j} \mathbb{P}((X_i, X_j) \in B(x, h) \times B(x, h)) = \sup_{i \neq j} \mathbb{P}(D_i \leq h, D_j \leq h) \leq \psi(h) f_2(x),
$$

where $\psi(h) \to 0$ as $h \to 0$ and $f_2(x)$ is a nonnegative functional in $x \in \mathcal{E}$. We assume that the ratio $\psi(h)/\phi^2(h)$ is bounded;

\textbf{(H5)} For all $(y_1, y_2) \in \mathbb{R}^{2d}$ and constants $b_3 > 0, \eta_4 > 0$, we have for the conditional density $f(\cdot)$ of $Y$ given $X = x$ the following

$$
|f(y_1) - f(y_2)| \leq \eta_4 \|y_1 - y_2\|^{b_3};
$$

(i) $F(u; x) = \phi(u) f_1(x)$ as $u \to 0$, where $\phi(0) = 0$ and $\phi(u)$ is absolutely continuous in a neighborhood of the origin,

(ii) We have

$$
\sup_{i \neq j} \mathbb{P} \left( D_i \leq u, D_j \leq u \right) \leq \psi(u) f_2(x),
$$

as $u \to 0$, where $\psi(u) \to 0$ as $u \to 0$. We assume that the ratio $\psi(h)/\phi^2(h)$ is bounded;

\textbf{(H6)} The kernel function $K(\cdot)$ is supported within $(0, 1/2)$ and has a continuous first derivative on $(0, 1/2)$. Moreover, there exist constants $0 < \eta_5 \leq \eta_6 < \infty$ such that:

$$
0 < \eta_5 \mathbb{1}_{(0, 1/2)}(\cdot) \leq K(\cdot) \leq \eta_6 \mathbb{1}_{(0, 1/2)}(\cdot),
$$
and
\[ K(1/2) - \int_0^{1/2} K'(s)\tau_0(s)ds > 0, \quad K^2(1/2) - \int_0^{1/2} (K^2)'(s)\tau_0(s)ds > 0; \]

**H7** Assume the class of sets \( \mathcal{C} \) satisfies the condition (2.2);

**H8** (Mixing): for some \( v > 2 \) and \( \delta > 1 - \frac{2}{v} \), we have
\[ \sum_{\ell=1}^\infty \ell^\delta[\alpha(\ell)]^{1 - \frac{2}{\delta}} < \infty; \]

**H9** The smoothing parameter \((h_n)\) satisfies:
\[ \frac{\log n}{n \min(a_n, \phi(h_n))} \to 0, \]

(i) Let \( h_n \to 0 \) and \( n\phi(h_n) \to \infty \) as \( n \to \infty \). Let \( v_n \) be a sequence of positive integers satisfying \( v_n \to \infty \) such that \( v_n = o\left((n\phi(h_n))^{1/2}\right) \) and
\[ (n/\phi(h_n))^{1/2}\alpha(v_n) \to 0 \quad \text{as} \quad n \to \infty. \]

### 2.2 Comments on the assumptions

The Condition (H1) is related to the small ball probabilities, which plays a major role both from theoretical and practical points of view, because the notion of ball is strongly linked with the semi-metric \( d(\cdot, \cdot) \), the choice of this semi-metric will become an important stage when the data are tacking its values in some infinite dimensional space. The second part of (H1) will be used to control the bias of nonparametric estimators, one needs to have some information on the variability of the small-ball probability. Indeed, in many examples, the small ball probability function can be written approximately as the product of two independent functions in terms of \( x \) and \( h \), as in the following examples, which can be found in Proposition 1 of Ferraty et al. (2007):

1. \( \phi(h_n) = Ch_n^v \) for some \( v > 0 \) with \( \tau_0(s) = s^v; \)
2. \( \phi(h_n) = Ch_n^v \exp(-Ch_n^{-p}) \) for some \( v > 0 \) and \( p > 0 \) with \( \tau_0(s) \) is the Dirac’s function;
3. \( \phi(h_n) = C |\ln(h_n)|^{-1} \) with \( \tau_0(s) = |\cdot|_0,1 \) \( (s) \) the indicator function in \( [0,1] \).

The conditions (H2)-(H3) are classical in the nonparametric regression estimation. (H4) is similar to those in Masry (2005). (H5): About the conditions on the density \( f(\cdot) \) is classical Lipschitz-type nonparametric functional model. The conditions on the kernel are not very restrictive. The first part of condition (H6) appears in many kernel functional studies and is easily satisfied for wide classes of kernel functions, the interested reader can refer to \( H_4 \) in Ferraty et al. (2007). The second part of this condition, which is added in
this paper as a necessary tool to get uniform results, is linked to the function \( \tau_0(\cdot) \) and is also rather general. For example, when \( \tau_0(\cdot) \) is identified to be the Dirac mass at 1/2, the second part of \( \tau_0(\cdot) \) is true as long as \( K'(s) \leq 0 \) and \( K(1/2) > 0 \). Other examples can be derived from Proposition 2 in Ferraty et al. (2007). Condition (H8) rules out too large or too small bandwidths for which consistency could not be obtained. It is satisfied with \( h_n = O(\log n)^{-\nu_1} \) (for some suitable \( \nu_1 > 0 \)) as long as the process \( X \) is of the exponential type (that is when the small-ball probability function is exponentially decaying). It is also satisfied with \( h_n = O(n/ \log n)^{-\nu_2} \) (for some suitable \( \nu_2 > 0 \)) for fractal processes (that is, when the small-ball probability is of polynomial decaying). More details can be found in Ferraty and Vieu (2006).

3 Main results

Below, we write \( Z \overset{D}{=} \mathcal{N}(\mu, \sigma^2) \) whenever the random variable \( Z \) follows a normal law with expectation \( \mu \) and variance \( \sigma^2 \), \( \overset{D}{\longrightarrow} \) denotes the convergence in distribution and \( \overset{P}{\longrightarrow} \) the convergence in probability.

**Theorem 1. [Uniform Consistency]** Suppose that the hypotheses (H1)-(H8) hold and that \( (X_t, Y_t) \) is geometrically strong mixing with \( \beta > 2 \). Let \( \mathcal{C} \) be a class of measurable sets for which

\[
\mathcal{N}(\epsilon, \mathcal{C}, G(\cdot | x)) < \infty
\]

for any \( \epsilon > 0 \). Suppose further that \( \forall C \in \mathcal{C} \)

\[
|G(C, y)f(y) - G(C, x)f(x)| \longrightarrow 0, \text{ as } y \rightarrow x.
\]

If \( n\phi(h_n) \rightarrow \infty \) and \( h_n \rightarrow 0 \) as \( n \rightarrow \infty \), then

\[
\sup_{C \in \mathcal{C}} |G_n(C, x) - E(G_n(C, x))| \overset{P}{\longrightarrow} 0.
\]

The proof of this theorem is based on the following relations. Remark that, the proof of Theorem 1 is a direct consequence of the decomposition:

\[
G_n(C, x) - E(G_n(C, x)) = \frac{1}{E(f_n(x))} \left[ \hat{F}_n(C, x) - E(\hat{F}_n(C, x)) \right] - \frac{G_n(C, x)}{E(f_n(x))} \left[ \hat{f}_n(x) - E(\hat{f}_n(x)) \right],
\]

where

\[
\hat{F}_n(C, x) = \frac{1}{n\phi(h_n)} \sum_{i=1}^{n} 1\{Y_i \in C\} K \left( \frac{d_\varepsilon(x, X_i)}{h_n} \right),
\]

\[
\hat{f}_n(x) = \frac{1}{n\phi(h_n)} \sum_{i=1}^{n} K \left( \frac{d_\varepsilon(x, X_i)}{h_n} \right),
\]
and of the Lemmas 1 and 2 below, for which the proofs are given in the Appendix.

**Lemma 1.** Suppose that the hypotheses (H1)-(H8) hold and for every fixed $C \in \mathcal{C}$ as $n \to \infty$ we have:

$$\sup_{C \in \mathcal{C}} \left| \hat{F}_n(C, x) - \mathbb{E} \left( \hat{F}_n(C, x) \right) \right| = o_p(1)$$

**Lemma 2.** Suppose that the hypotheses (H1)-(H8) hold and for every fixed $N_E$ neighborhood of $x$ in the functional space $E$ as $n \to \infty$, we have

$$\sup_{x \in N_E} \left| \hat{f}_n(x) - \mathbb{E} \left( \hat{f}_n(x) \right) \right| = o_p(1).$$

Before to establishing the asymptotic normality define the “bias” term by

$$B_n(x) = \frac{\mathbb{E} \left( \hat{f}_n(x) \right) - G_n(C, x) \mathbb{E} \left( \hat{F}_n(C, x) \right)}{\mathbb{E} \left( \hat{F}_n(C, x) \right)}.$$

By stationarity of order one of the $(X_i)$’s, we have

$$\mathbb{E}(\hat{f}_n(x)) = 1.$$

The following result give the weak convergence of our estimators. Keep in mind that $f_1(x)$ is given in (H5).

**Theorem 2** (Asymptotic normality). Let (H2)-(H5)(i)(ii)-(H6)-(H8)-(H9)(i) hold and $(X_i, Y_i)$ is geometrically strong mixing with $\beta > 2$, then $n\phi(h_n) \to \infty$ as $n \to \infty$. For $m \geq 1$ and $C_1, \ldots, C_m \in \mathcal{C}$,

$$\{\tilde{\nu}_n(C_i, x) \}_{i=1,\ldots,m} \overset{D}{\rightarrow} \mathcal{N}(0, \Sigma),$$

where $\Sigma = \sigma_{ij}(x), i, j = 1, \ldots, m$ and

$$\sigma_{ij}(x) = \frac{C_2}{C_1 f_1(x)} \left( \mathbb{E}(1_{\{Y \in C_i \cap C_j\}} | X = x) - \mathbb{E}(1_{\{Y \in C_i\}} | X = x) \mathbb{E}(1_{\{Y \in C_j\}} | X = x) \right),$$

whenever $f_1(x) > 0$ and

$$C_1 = K(1/2) - \int_0^{1/2} K'(s) \tau_0(s) ds, \quad C_2 = K^2(1/2) - \int_0^{1/2} (K^2)'(s) \tau_0(s) ds.$$

To establish the density of the process, we need to introduce the following function which provides the information on the asymptotic behaviour of the modulus of continuity

$$\Lambda_\gamma(\sigma^2, n) = \left\{ \begin{array}{ll} \sqrt{\sigma^2 \log \frac{1}{\sigma^2}}, & \text{if } \gamma = 0; \\ \max \left( (\sigma^2)^{(1-\gamma)/2}, n\phi(h_n)^{(3\gamma-1)/(2(3\gamma+1))} \right), & \text{if } \gamma > 0. \end{array} \right.$$ 

**Theorem 3.** Suppose that (H1)-(H9) hold and the process $(X_i, Y_i)$ is exponentially strong mixing for each $\sigma^2 > 0$, let $\mathcal{C}_{\sigma} \subset \mathcal{C}$ be a class of measurable sets with
\[ \sup_{C \in \mathcal{C}} G(C, x) \leq \sigma^2 \leq 1, \]

and suppose that \( C \) fulfills \( (R_\gamma) \) with \( \gamma \geq 0 \). Further, we assume that \( \phi(h_n) \to 0 \) and \( n\phi(h_n) \to +\infty \) as \( n \to +\infty \), such that

\[ n\phi(h_n) \leq \left( \Lambda_\gamma(\sigma^2, n) \right)^2, \]

and as \( n \to +\infty \), we have

\[ n\phi \left( \frac{\sigma^2 \log \left( \frac{1}{\sigma^2} \right) \log(n)}{\log(n)} \right)^{1+\gamma} \to \infty. \]

Further we assume that \( \sigma^2 \geq h^2 \). For \( \gamma > 0 \) and \( d = 1, 2 \), the later has to be replaced by \( \sigma^2 \geq \phi(h_n) \log \left( \frac{1}{\phi(h_n)} \right) \) then for every \( \epsilon > 0 \) there exist a constant \( M > 0 \) such that

\[ \mathbb{P} \left( \sup_{C \in \mathcal{C}} |\tilde{v}_n(C, x)| \geq M\Lambda_\gamma(\sigma^2, n) \right) \leq \epsilon, \]

for all sufficiently large \( n \).

By combining Theorem 2 and Theorem 3 we have the following result.

**Theorem 4.** Under conditions of Theorem 2 and Theorem 3, then the process:

\[ \{\tilde{v}_n(C, x) : C \in \mathcal{C}\}, \]

converges in law to a Gaussian process \( \{\tilde{v}(C, x) : C \in \mathcal{C}\}, \) that admits a version with uniformly bounded and uniformly continuous paths with respect to \( \|\cdot\|_2 \)–norm with covariance \( \sigma_{ij}(x) \) given in Theorem 2.

**Remark 1.** Central limit theorems are usually used to establish confidence intervals for the target to be estimated. In the context of non-parametric estimation the asymptotic variance \( \Sigma := \sigma_{ij}(x) \) in the central limit depends on certain functions, including the ones that are estimated. This situation is classic regardless of whether the data is i.i.d. or dependent. As a result, only approximate confidence intervals can be obtained in practice, even when \( \Sigma \) functionally specified. To be more precise let us consider the following particular case of Theorem 2, where \( m = 1 \). In the this situation, \( \Sigma \) is reduced to, for \( A \in \mathcal{C} \),

\[ \sigma^2(x) = \frac{C_2}{C_1^2f_1(x)} \left( \mathbb{E}(\mathbb{1}_{\{Y \in A\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y \in A\}} \mid X = x)^2 \right) = \frac{C_2}{C_1^2f_1(x)} W_2(x). \]

Observe that the limiting variance contains the unknown function \( f_1(\cdot) \) and that the normalization depends on the function \( \phi(h_n) \) which is not identifiable explicitly. Let us introduce the following estimate

\[ F_{x,n}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{d(x, X_i) \leq t\}}, \]
One may estimate $\tau_0(\cdot)$ by
$$
\tau_n(t) = \frac{F_{x,n}(th)}{F_{x,n}(h)}.
$$

This can use to give the following estimates
$$
C_{1,n} = K(1/2) - \int_0^{1/2} K'(s) \tau_n(s)ds, \quad C_{2,n} = K^2(1/2) - \int_0^{1/2} (K^2)'(s) \tau_n(s)ds.
$$

One can estimate $W_{2,n}(x)$ by
$$
W_{2,n}(x) = (G_n(C, x) - G_n^2(C, x)).
$$

The use of Theorem 2, in connection with Slutsky’s theorem, gives
$$
\frac{C_{1,n}}{\sqrt{C_{2,n}}} \sqrt{n} \sum_{j=1}^{n} \{ Y_j \in C \} (G_n(C, x) - G(C, x)) \overset{D}{\rightarrow} N(0, 1).
$$

This result can be used in the construction of the confidence interval in the usual way, we omit the details.

### 3.1 The bandwidth selection criterion

Many methods have been established and developed to construct, in asymptotically optimal ways, bandwidth selection rules for nonparametric kernel estimators especially for Nadaraya-Watson regression estimator we quote among them Hall (1984), Härdle and Marron (1985), Rachdi and Vieu (2007), Bouzebda and El-hadjali (2020) and Bouzebda and Nemouchi (2020). This parameter has to be selected suitably, either in the standard finite dimensional case, or in the infinite dimensional framework for ensuring good practical performances. Let us define the leave-out-$(X_i, Y_i)$ estimator for regression function
$$
G_{n,j}(C, x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \in C\}} K \left( \frac{d_C(x, X_i)}{h_n} \right)}{\sum_{i=1}^{n} K \left( \frac{d_C(x, X_i)}{h_n} \right)}.
$$

In order to minimize the quadratic loss function, we introduce the following criterion, we have for some (known) non-negative weight function $\mathcal{W}(\cdot)$:
$$
CV(C, h) := \frac{1}{n} \sum_{j=1}^{n} \left( \mathbb{1}_{\{Y_j \in C\}} - G_{n,j}(C, X_j) \right)^2 \mathcal{W}(X_j).
$$

Following the ideas developed by Rachdi and Vieu (2007), a natural way for choosing the bandwidth is to minimize the precedent criterion, so let’s choose $\hat{h}_n \in [a_n, b_n]$ minimizing among $h \in [a_n, b_n]$:
$$
\sup_{C \in \Psi} CV(C, h).
$$
The main interest of our results is the possibility to derive the asymptotic properties of our estimate even if the bandwidth parameter is a random variable, like in the last equation. One can replace (3.2) by

\[ CV(C, h_n) := \frac{1}{n} \sum_{j=1}^{n} \left( \mathbb{I}_{\{Y_j \in C\}} - G_{n,j}(C, X_j) \right)^2 \hat{W}(X_j, x). \]  

(3.3)

In practice, one takes, for \( j = 1, \ldots, n \), the uniform global weights \( W(X_j) = 1 \), and the local weights

\[ \hat{W}(X_j, x) = \begin{cases} 
1 & \text{if } d(X_j, x) \leq h_n, \\
0 & \text{otherwise}. 
\end{cases} \]

For sake of brevity, we have just considered the most popular method, that is, the cross-validated selected bandwidth. This may be extended to any other bandwidth selector such the bandwidth based on Bayesian ideas Shang (2014).

4 Testing the independence

Concepts of conditional independence play an important role in unifying many seemingly unrelated ideas of statistical inference, see Dawid (1980). Measuring and testing conditional dependence are fundamental problems in statistics, which form the basis of limit theorems, Markov chain, sufficiency and causality Dawid (1979), among others. Conditional independence also plays a central role in graphical modeling Koller and Friedman (2009), causal inference Pearl (2009) and artificial intelligence Zhang et al. (2011), refer also to Zhou et al. (2020) for recent references. The idea of treating conditional independence as an abstract concept with its own calculus was introduced by Dawid (1979), who showed that many results and theorems concerning statistical concepts such as ancillarity, sufficiency, causality, etc., are just applications of general properties of conditional independence-extended to encompass stochastic and non-stochastic variables together. Let \( \mathcal{C}_1, \mathcal{C}_2 \) be some classes of sets. In this section, we consider a sample of random elements \((X_1, Y_{1,1}, Y_{1,2}), \ldots, (X_n, Y_{n,1}, Y_{n,2})\) copies of \((X, Y_1, Y_2)\) that takes its value in a space \( \mathcal{E} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \) and define, for \((C_1, C_2) \in \mathcal{C}_1 \times \mathcal{C}_2,\)

\[ G_n(C_1 \times C_2, x) = \frac{\sum_{i=1}^{n} \mathbb{I}_{\{Y_{1,i} \in C_1\}} \mathbb{I}_{\{Y_{2,i} \in C_2\}} K\left( \frac{d_{\mathcal{E}}(x, X_i)}{h_n} \right)}{\sum_{i=1}^{n} K\left( \frac{d_{\mathcal{E}}(x, X_i)}{h_n} \right)}, \]  

(4.1)

\[ G_{n,1}(C_1, x) = \frac{\sum_{i=1}^{n} \mathbb{I}_{\{Y_{1,i} \in C_1\}} K\left( \frac{d_{\mathcal{E}}(x, X_i)}{h_n} \right)}{\sum_{i=1}^{n} K\left( \frac{d_{\mathcal{E}}(x, X_i)}{h_n} \right)}, \]  

(4.2)
We would test the following null hypothesis

\[ H_0 : Y_1 \text{ and } Y_2 \text{ are conditionally independent given } X = x. \]
Against the alternative

\[ \mathcal{H}_1 : Y_1 \text{ and } Y_2 \text{ are conditionally dependent.} \]

Statistics of independence those can be used are

\[
S_{1,n} = \sup_{(C_1, C_2) \in \mathcal{C}_1 \times \mathcal{C}_2} |\hat{\nu}_n(C_1, C_2, x)|, \tag{4.8}
\]

\[
S_{2,n} = \sup_{(C_1, C_2) \in \mathcal{C}_1 \times \mathcal{C}_2} |\bar{\nu}_n(C_1, C_2, x)|. \tag{4.9}
\]

A combination of Theorem 4 with continuous mapping theorem we obtain the following result.

**Theorem 5.** We have under condition of Theorem 4, as \( n \to \infty \),

\[
S_{1,n} \to \sup_{(C_1, C_2) \in \mathcal{C}_1 \times \mathcal{C}_2} |\hat{\nu}(C_1, C_2, x)|, \tag{4.10}
\]

\[
S_{2,n} \to \sup_{(C_1, C_2) \in \mathcal{C}_1 \times \mathcal{C}_2} |\bar{\nu}(C_1, C_2, x)|. \tag{4.11}
\]

**Remark 2.** It is well known that Theorem 5 can be used easily through routine bootstrap sampling as in Bouzebda (2012), Bouzebda and Cherfi (2012) and Bouzebda et al. (2018), which we describe briefly as follows. Let \( N \) be a large integer. Let \( S_{1,n}^{(1)}, \ldots, S_{1,n}^{(N)} \) be the bootstrapped versions of \( S_{1,n} \), for \( j = 1, 2 \). With the convention that large values of \( S_{j,n} \), \( j = 1, 2 \), lead to the rejection of the null hypothesis \( \mathcal{H}_0 \), under some regularity conditions, a valid approximation to the P-value for the test based on \( S_{j,n} \), \( j = 1, 2 \), for \( N \) large enough, is given by

\[
\frac{1}{N} \sum_{k=1}^{N} \mathbb{I}\{S_{j,n}^{(k)} \geq S_{j,n}\}.
\]

The investigation of the bootstrap should require a different methodology than that used in the present paper, and we leave this problem open for future research.

### 5 Concluding remarks

In the present work, we have established the invariance principle for the conditional set-indexed empirical process formed by strong mixing random variables when the covariates are functional. Our results are obtained under assumptions on the richness of the index class \( \mathcal{C} \) of sets in terms of metric entropy with bracketing in the framework of mixing data. An application of testing the conditional independence is proposed. Notice that mixing is some kind of asymptotic independence assumption which is commonly used for seek of simplicity but which can be unrealistic in situations where there is strong dependence between the data. Extending non-parametric functional ideas to general dependence structure is a rather underdeveloped field. Note that the ergodic framework avoid the widely used strong mixing condition and its variants to measure the dependency and the very involved...
probabilistic calculations that it implies. It would be interesting to extend our work to the
case of the functional ergodic data, which requires non trivial mathematics, this would go
well beyond the scope of the present paper.

Acknowledgement

The authors are indebted to the Editor-in-Chief, Associate Editor and the referee for their
very valuable comments, suggestions careful reading of the article which led to a consider-
able improvement of the manuscript.

CRediT author statement

Youssouf Souddi: Conceptualization, Methodology, Investigation, Writing - Original Draft,
Writing - Review & Editing.
Salim Bouzebda and Fethi Madani: Conceptualization, Methodology, Investigation, Writing - Original Draft, Writing - Review & Editing.
All authors contributed equally to this work.

References


orems for dependent stochastic processes. *International Statistical Review / Revue Interna-

and related fields*. Selected papers based on the presentations at the 4th international workshop
on functional and operational statistics, IWFOS, Corunna, Spain, June 15–17, 2017. Cham:
Springer.

Arcones, M. A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian


In *Contributions in infinite-dimensional statistics and related topics*, pages 1–5. Esculapio,
Bologna.


