Chapter 2
Asymptotic Behavior of the Universally Consistent Conditional U-Statistics for Nonstationary and Absolutely Regular Processes

Echarif Elharfaoui, Michel Harel, and Madan L. Puri

Abstract A general class of conditional U-statistics was introduced by W. Stute as a generalization of the Nadaraya–Watson estimates of a regression function. It was shown that such statistics are universally consistent. Also, universal consistencies of the window and k_n-nearest neighbor estimators (as two special cases of the conditional U-statistics) were proved. Later, (Harel and Puri, Ann Inst Stat Math 56(4):819–832, 2004) extended his results from the i.i.d. case to the absolute regular case. In this paper, we extend these results from the stationary case to the nonstationary case.

2.1 Introduction

Let $\{Z_i = (X_i, Y_i);\ i \in \mathbb{N}^*\}$ be a sequence of random vectors with continuous distribution functions $H_i(z)$, $i \in \mathbb{N}^*$, $z \in \mathbb{R}^d \times \mathbb{R}^s$, defined on some probability space $(\Omega, \mathcal{A}, P)$.

Assume that H_i admits a strictly positive density and H_i has the two marginals F_i and G_i.

Let h be a function of k-variates (the U kernel) such that for some $r > 2$, $h \in \mathcal{L}_r^*$, which means that $E\{\sup_\beta |h(Y_\beta)|^r\} < +\infty$ (where sup extends over all permutations $\beta = (\beta_1, \ldots, \beta_k)$ of length k, that is, over all pairwise distinct $\beta_1, \ldots, \beta_k$ taken from $\mathbb{N}^*$) which implies that for all integers $i_1, i_2, \ldots, i_k$ $(i_1 < i_2 < \ldots < i_k)$ $h(Y_{i_1}, \ldots, Y_{i_k}) \in \mathcal{L}_r$, the space of all random variables Z for which $|Z|^r$ is integrable. In order to measure the impact of a few X's, say $(X_1, \ldots, X_k)$, on a

E. Elharfaoui (✉)
Département de Mathématiques et Informatique, Université Chouaîb Doukkali, Faculté des Sciences El Jadida, Rte Ben Maachou, B.P. 20, 24000, Maroc, and IMT (UMR CNRS 5219), Université Paul Sabatier, Toulouse Cedex 31062, France
e-mail: elharfa@cict.fr

M.T. Wells and A. SenGupta (eds.), *Advances in Directional and Linear Statistics*,
DOI 10.1007/978-3-7908-2628-9_2,

function $h(Y_1, \ldots, Y_k)$ of the pertaining Y's, set

$$m(\mathbf{x}) \equiv m(x_1, \ldots, x_k) := E[h(Y_1, \ldots, Y_k)|X_1 = x_1, \ldots, X_k = x_k] \tag{2.1}$$

where m is defined on $\mathbb{R}^{dk}$.

For estimation of $m(\mathbf{x})$, [7] proposed a statistic of the form

$$u_n(\mathbf{x}) = u_n(x_1, \ldots, x_k) = \frac{\sum_\beta h(Y_{\beta_1}, \ldots, Y_{\beta_k}) \prod_{j=1}^k K[(x_j - X_{\beta_j})/h_n]}{\sum_\beta \prod_{j=1}^k K[(x_j - X_{\beta_j})/h_n]} \tag{2.2}$$

where u_n is defined on $\mathbb{R}^{dk}$, K is the so-called smoothing kernel satisfying $\int K(u)\, du = 1$ and $\{h_n,\ n \geq 1\}$ is a sequence of bandwidth tending to zero at appropriate rates. Here summation extends over all permutations $\beta = (\beta_1, \ldots, \beta_k)$ of length k, that is, over all pairwise distinct $\beta_1, \ldots, \beta_k$ taken from $1, \ldots, n$. Stute [7] proved the asymptotic normality and weak and strong consistency of $u_n(\mathbf{x})$ when the random variables $\{(X_i, Y_i),\ i \geq 1\}$ are independent and identically distributed. Harel and Puri [3] extended the results of [7] from independent case to the case when the underlying random variables are absolutely regular. Stute [9] also derived the $\mathcal{L}_r$ convergence of the conditional U-statistics under the i.i.d. set up.

If a number of the X_i's in the random sample are exactly equal to x which can happen if X is a discrete random variable, $P^Y(\cdot|X = x)$ can be estimated by the empirical distribution of the Y_i's corresponding to X_i's equal to x. If few or none of the X_i's are exactly equal to x, it is necessary to use Y_i's corresponding to X_i's near x. This leads to estimators $\widehat{P}_n^Y(\cdot|X = x)$ of the form

$$\widehat{P}_n^Y(\cdot|X = x) = \sum_{i=1}^n W_{ni}(x) 1\!\!1_{[Y_i \in \cdot]}$$

where $W_{ni}(x) = W_{ni}(x, X_1, \ldots, X_n)$ $(1 \leq i \leq n)$ weights those values of i for which X_i is close to x more heavily than these values of i for which X_i is far from x and $1\!\!1_A$ denotes the indicator function of A.

Let g be a Borel function on $\mathbb{R}^s$ such that $g(Y) \in \mathcal{L}_r$. Corresponding to W_n is the estimator $l_n(x)$ of $l(x) = E(g(Y)|X = x)$ defined by

$$l_n(x) = \sum_{i=1}^n W_{ni}(x) g(Y_i).$$

More generally if we now consider the estimates of $m(x)$ defined in (2.4), this leads to weighting those values of β for which $\mathbf{X}_\beta = (X_{\beta_1}, \ldots, X_{\beta_k})$ is close to $\mathbf{x}$ more heavily than the values of β for which $\mathbf{X}_\beta$ is far from $\mathbf{x}$.

This is why, as in [8], we study a fairly general class of conditional U-statistics of the form

$$m_n(\mathbf{x}) = \sum_{\beta} W_{\beta,n}(\mathbf{x}) h(\mathbf{Y}_\beta) \tag{2.3}$$

designed to estimate $m(\mathbf{x})$, where $W_{\beta,n}(\mathbf{x})$ is defined from a function $W_n(\mathbf{x},\mathbf{y})$ by $W_{\beta,n}(\mathbf{x}) = W_n(\mathbf{x},\mathbf{X}_\beta)$, $\mathbf{Y}_\beta = (Y_{\beta_1},\ldots,Y_{\beta_k})$, and the summation in (2.3) takes place over all permutations $\beta = (\beta_1,\ldots,\beta_k)$ of length k such that $1 \le \beta_i \le n,\ i = 1,\ldots,k$.

Remark 2.1. The estimator defined in (2.2) is a special case of the estimator defined in (2.3), see (2.23).

In order to make $m_n(\mathbf{x})$ a local average, $W_{\beta,n}(\mathbf{x})$ has to give larger weights to those $h(\mathbf{Y}_\beta)$ is close to $\mathbf{x}$. For this general class of conditional U-statistics (defined in (2.3)) and for i.i.d. random variables, [8] derived the universal consistency. Harel and Puri [4] extended his results from the i.i.d. case to the absolute regular case. In this paper, we extend it to the nonstationary case and absolutely regular r.v.'s which allow broader applications that include, among others, hidden Markov models (HMM) described in detail in [4].

We shall call $W_{\beta,n}$ *universally consistent* if and only if

$$m_n(\mathbf{X}) \to m(\mathbf{X}) \ \text{ in } \mathcal{L}_r$$

under no conditions on h (up to integrability) or the distribution of $\{(X_i,Y_i),\ i \ge 1\}$. Here $\mathbf{X} = (X_1^0,\ldots,X_k^0)$ is a vector of X's with the same distribution as $(X_1,\ldots,X_k)$ and independent of $\{(X_i,Y_i),\ i \ge 1\}$.

2.2 Preliminaries

Let $(Z_i)_{i\ge1}$ be a stochastic process indexed by the positive integers, taking value in a finite dimensional Euclidean space $\mathbb{H}$. Identifying $\mathbb{H}$ with a product of a finite number copies or the real line, we write H_i for the distribution function of Z_i. We will assume that the process has some form of asymptotic stationarity, implying that the sequence H_i converges in a sense to be made precise to a limiting distribution function H.

For $i \le j$, let $\mathcal{A}_i^j$ denote the σ-algebra of events generated by $Z_i,\ldots,Z_j$. We shall say that the nonstationary stochastic process is absolutely regular if

$$\sup_{n\in\mathbb{N}^*} \max_{1\le j\le n-k} E\left\{ \sup_{A\in\mathcal{A}_{j+k}^\infty} \left| P(A \mid \mathcal{A}_1^j) - P(A) \right| \right\} = \beta(k)^\star \downarrow 0 \ \text{ as } \ n \to \infty$$

where $\mathbb{N}^* = \{1,2,\ldots\}$.

All along the paper, we assume that $(^\star)$ holds with a geometrical rate;

$$\sum_{m\geq 1} m\beta^{\frac{\delta}{1+\delta}}(m) < \infty \quad \text{for some } \delta > 0. \tag{2.4}$$

We consider a parameter ξ in $\mathbb{H}$ whose components can be naturally estimated by U-statistics. To be more formal and precise, we assume that ξ is defined as follows. Let k be an integer, to be the degree of the U-statistics. Let ϕ be a function from $\mathbb{H}^k$ into $\mathbb{H}$, invariant by permutation of its arguments. We are interested in parameters of the form

$$\xi = \int_{\mathbb{H}^k} \phi \, \mathrm{d}H^{\otimes k} = \int_{\mathbb{H}^k} \phi(z_1, \ldots, z_k) \prod_{l=1}^{k} dH(z_l). \tag{2.5}$$

and the function ϕ is called the kernel of the parameter ξ.

Example 2.1. Take $\mathbb{H}$ to be $\mathbb{R}$. The mean vector corresponds to taking $k = 1$ and ϕ is the identity.

Example 2.2. Take $\mathbb{H}$ to be $\mathbb{R}^2$. Consider ξ to be the two-dimensional vector whose components are the marginal variances. We take $k = 2$ and ϕ is going to be a function defined on $(\mathbb{R}^2)^2$. It has two arguments, each being in $\mathbb{R}^2$, and it is defined by

$$\phi\left((u,v),(u',v')\right) = \left(\frac{u^2+u'^2}{2} - uu', \frac{v^2+v'^2}{2} - vv'\right).$$

Such a parameter can be estimated naturally by a U-statistics, essentially replacing $H^{\otimes k}$ in (5) by an empirical counterpart. By using the invariance of ϕ, the estimator of ξ is then of the form

$$\widehat{\xi}_n = \binom{n}{k}^{-1} \sum_{\beta} \phi(Z_{\beta_1}, \ldots, Z_{\beta_k}). \tag{2.6}$$

To specify our assumption on the process, it is convenient to introduce copies of $\mathbb{H}$. Hence we write $\mathbb{H}_i$, $i \geq 1$, an infinite sequence of copies of $\mathbb{H}$. The basic idea is to think of the process at time i as taking value in $\mathbb{H}_i$ and we think of each $\mathbb{H}_i$ as the ith component of $\mathbb{H}^\infty$. We then agree on the following definition.

Definition 2.1. A canonical p-subspace of $\mathbb{H}^\infty$ is any subspace of the form $\mathbb{H}_{i_1} \oplus \ldots \oplus \mathbb{H}_{i_p}$ with $1 \leq i_1 < \cdots < i_p$. We write $\mathbb{S}_p$ for a generic canonical p-subspace.

Remark 2.2. For $(i_1, \ldots, i_p) \neq (j_1, \ldots, j_p)$, if we note $\mathbb{S}_p = \mathbb{H}_{i_1} \oplus \ldots \oplus \mathbb{H}_{i_p}$ and $\mathbb{S}'_p = \mathbb{H}_{j_1} \oplus \ldots \oplus \mathbb{H}_{j_p}$, we have $\mathbb{S}_p \neq \mathbb{S}'_p$, with $\mathbb{S}_p \subset \mathbb{H}^\infty$ and $\mathbb{S}'_p \subset \mathbb{H}^\infty$.

The origin of this terminology is that when $\mathbb{H}$ is the real line, then a canonical p-subspace is a subspace spanned by exactly p distinct vectors of the canonical basis of $\mathbb{H}^\infty$. We write $\sum_{\mathbb{S}_p \subset \mathbb{H}^n}$ for a sum over all canonical p-subspaces included in $\mathbb{H}^n$.

To such a canonical subspace $\mathbb{S}_p = \mathbb{H}_{i_1} \oplus \ldots \oplus \mathbb{H}_{i_p}$ we can associate the distribution function $H_{\mathbb{S}_p}$ of $(Z_{i_1}, \ldots, Z_{i_p})$ as well as the distribution function with the same marginals

$$H^{\otimes \mathbb{S}_p} = \otimes_{1 \le j \le p} H_{i_j} = \otimes_{\mathbb{H}_i \subset \mathbb{S}_p} H_i. \tag{2.7}$$

Clearly the marginal of $H^{\otimes \mathbb{S}_p}$ are independent, while that of $H_{\mathbb{S}_p}$ are not.

Consider two nested canonical subspace $\mathbb{S}_p$ and $\mathbb{S}_{k-p}$ where $\mathbb{S}_{k-p} \subset \mathbb{H}^n \ominus \mathbb{S}_p$. For a function ϕ symmetric in its argument and defined on $\mathbb{S}_p \oplus \mathbb{S}_{k-p}$, we can define its projection onto the functions defined on $\mathbb{S}_p$ by

$$z \in \mathbb{S}_p \to \phi(z, \mathbb{S}_{k-p}) = \int_{\mathbb{S}_{k-p}} \phi(z, y)\, \mathrm{d}H^{\otimes \mathbb{S}_{k-p}}(y). \tag{2.8}$$

Identifying $\mathbb{S}_p \oplus \mathbb{S}_{k-p}$ with $\mathbb{H}^k$ and $\mathbb{H}^p$ with $\mathbb{S}_p$, that allows to project functions defined on $\mathbb{H}^k$ onto functions on $\mathbb{H}^p$. However, with this identification, the projection depends on the particular choice of $\mathbb{S}_{k-p}$ in $\mathbb{H}^n$. To remove the dependence in $\mathbb{S}_{k-p}$, we sum over all choices of $\mathbb{S}_{k-p}$ in $\mathbb{H}^n \ominus \mathbb{S}_p$ by

$$\phi_{\mathbb{S}_p}(z) = \binom{n-p}{k-p}^{-1} \sum_{\mathbb{S}_{k-p} \subset \mathbb{H}^n \ominus \mathbb{S}_p} \phi(z, \mathbb{S}_{k-p}). \tag{2.9}$$

Let k be an integer and for each $n \geq k$, consider a kernel $\phi_n \equiv \phi$ of degree k depending on n.

A U-statistics of degree k is defined by

$$U_n = \binom{n}{k}^{-1} \sum_{\beta} \phi_n(Z_{\beta_1}, \ldots, Z_{\beta_k}), \tag{2.10}$$

we can then define an analogue of Hoeffding decomposition when the random variables come from a nonstationary process. For this purpose, consider, firstly, an expectation of U_n if the process had no dependence, namely,

$$U_{n,0} = \binom{n}{k}^{-1} \sum_{\mathbb{S}_k \subset \mathbb{H}^n} \int_{\mathbb{S}_k} \phi\, \mathrm{d}H^{\otimes \mathbb{S}_k}. \tag{2.11}$$

Then for any $p = 1, \ldots, k$, we define

$$U_{n,p} = \binom{n}{p}^{-1} \sum_{\mathbb{S}_p \subset \mathbb{H}^n} \int_{\mathbb{S}_p} \phi_{\mathbb{S}_p}\, \mathrm{d} \otimes_{\mathbb{H}_i \subset \mathbb{S}_p} (\delta_{Z_i} - H_i) \tag{2.12}$$

where $\delta_{\{.\}}$ is the Dirac function.

Finally, for $p > k$, we set

$$U_{n,p} = 0. \tag{2.13}$$

The analogue of Hoeffding decomposition is the equality

$$U_n = \sum_{0 \le p \le k} \binom{k}{p} U_{n,p}. \tag{2.14}$$

Note that this decomposition makes an explicit use of convention (2.13), and this is why this convention was introduced.

We now need to specify exactly what we mean by asymptotic stationary of a process. For this, recall the following notion of distance between probability measures.

Definition 2.2. The distance in total variation between two probability measures P and Q defined on the same σ-algebra $\mathcal{A}$ is

$$|P - Q|_{\mathcal{A}} = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|.$$

If $\mathbb{S}_p$ is a canonical subspace of $\mathbb{H}^\infty$, we write $\sigma_{\mathbb{S}_p}$ the σ-algebra generated by the Z_i's with $\mathbb{H}_i \subset \mathbb{S}_p$. We write P the probability measure pertaining to the process $(Z_i)_{i \ge 1}$, which is a probability measure on $\mathbb{H}^\infty$.

Definition 2.3. The process $(Z_i)_{i \ge 1}$ with probability measure P on $\mathbb{H}^\infty$ is geometrically asymptotically stationary if there exists a strictly stationary process with distribution Q on $\mathbb{H}^\infty$, and a positive τ less than 1, such that for $i \ge 1$,

$$|P - Q|_{\sigma_{\mathbb{H}_i}} \le \tau^i. \tag{2.15}$$

We suppose that there exists a strictly stationary process $(Z_i^*)_{i \ge 1}$ with probability measure Q on $\mathbb{H}^\infty$, which is absolutely regular with the same rate as the process $(Z_i)_{i \ge 1}$. H is the distribution function of Z_i^*, H admits a strictly positive density and H has the two marginals F and G.

We define the function ϕ^* on $\mathbb{H}_1$ by

$$z \in \mathbb{H}_1 \mapsto \phi^*(z, \mathbb{H}^k \ominus \mathbb{H}_1) = \int_{\mathbb{H}^k \ominus \mathbb{H}_1} \phi(z, y) \, \mathrm{d}H^{\otimes(k-1)}. \tag{2.16}$$

Next, we denote

$$U_{n,1}^* = n^{-1} \sum_{i=1}^{n} \int_{\mathbb{H}_1} \phi^* d(\delta_{Z_i^*} - H).$$

2.3 Assumptions and Main Results

In this section, we identify $\mathbb{H} = \mathbb{H}' \times \mathbb{H}''$ with $\mathbb{R}^d \times \mathbb{R}^s$. For a generic canonical p-subspace S_p of $\mathbb{H}^\infty$, we write $S_{1,k}$ and $S_{2,k}$ its projections respectively in $\mathbb{H}'^\infty$ and $\mathbb{H}''^\infty$.

We consider the nonstationary sequence of random vectors $\{Z_i = (X_i, Y_i);\ i \in \mathbb{N}^*\}$ with values in $\mathbb{R}^d \times \mathbb{R}^s$ and continuous distribution functions H_i and H_i has the two marginals F_i and G_i.

We assume that the sequence $\{Z_i\}_{i\geq 1}$ is absolute regular with rates (2.4) and (2.15) is satisfied with its associated stationary sequence $\{\widetilde{Z}_i = (\widetilde{X}_i, \widetilde{Y}_i);\ i \in \mathbb{N}^*\}$ of stationary random vectors.

For the ease of convenience, we shall write W_β for $W_{\beta,n}$.

Consider the following set of assumptions:

(i) There exists functions $V_n(\mathbf{x}, \mathbf{y})$ on $\mathbb{R}^{2dk}$ such that for each $l \in \mathcal{L}_r^*$, $\mathbf{z}^{(n)} = (z_1, \ldots, z_n) \in \mathbb{R}^{dn}$ and $\mathbf{y}^{(n)} = (y_1, \ldots, y_n) \in \mathbb{R}^{sn}$

$$\sum_\beta W_n(\mathbf{x}, \mathbf{z}_\beta) l(\mathbf{y}_\beta) = \frac{\sum_\beta V_n(\mathbf{x}, \mathbf{z}_\beta) l(\mathbf{y}_\beta)}{\sum_\beta V_n(\mathbf{x}, \mathbf{z}_\beta)}$$

where $\mathbf{z}_\beta = (z_{\beta_1}, \ldots, z_{\beta_k})$ and $\mathbf{y}_\beta = (y_{\beta_1}, \ldots, y_{\beta_k})$.

(ii) There exists a function $V(\mathbf{x})$ on $\mathbb{R}^{dk}$ such that for each scalar function q on $\mathbb{R}^{dk}$ verifying

$$\int |V(\mathbf{x}) q(\mathbf{x})|^r \, dF^{\otimes k}(\mathbf{x}) < \infty$$

we have

$$\lim_{n\to\infty} \binom{n}{k}^{-1} \sum_{S_{1,k} \subset \mathbb{H}'^n} \int_{S_{1,k}} q(\mathbf{z}) V_n(\mathbf{x}, \mathbf{z}) dF^{\otimes S_{1,k}}(\mathbf{z}) = q(\mathbf{x}) \tilde{f}(\mathbf{x}) \int V(\mathbf{z}) d\mathbf{z}$$

where $\tilde{f}(\mathbf{x}) = \prod_{j=1}^k f(x_j)$ and f is the density function of F.

(iii) Define the kernel of degree k by

$$\phi_n(\mathbf{z}, \mathbf{y}) = h(\mathbf{y}) V_n(\mathbf{x}, \mathbf{z}) \Big/ \int V_n(\mathbf{x}, \mathbf{u}) dF^{\otimes k}(\mathbf{u}).$$

Suppose that

$$\sup_{\mathbb{S}_k \subset \mathbb{H}^\infty} \int_{\mathbb{S}_k} |\phi_n|^{2+2\delta} \, dP_{\sigma_{\mathbb{S}_k}} < \infty \tag{2.17}$$

$$\sup_{\mathbb{S}_k \subset \mathbb{H}^\infty} \int_{\mathbb{S}_k} |\phi_n|^{2+2\delta} \, dQ_{\sigma_{\mathbb{S}_k}} < \infty \tag{2.18}$$

where $\delta > 0$.

Remark 2.3. Our conditions (i) and (ii) are completely different from conditions (ii) to (v) in [8]. Our conditions are more general and more easy to verify. More, the condition (i) in [8] is not necessary.

The following theorems generalize Theorems 2.1, 2.2, 2.3 and 2.4 in [4] from the stationary dependent case to the nonstationary dependent case.

Theorem 2.1. *Assume that $h \in \mathcal{L}_r^*$. Then under (i)–(iii), (2.4) and (2.15),*

$$m_n(\mathbf{X}) \to m(\mathbf{X}) \ \ in\ \mathcal{L}_r,\ \ as\ n \to \infty,$$

where $r = 2 + 2\delta$, that is

$$E\left[|m_n(\mathbf{x}) - m(\mathbf{x})|^r \mu(d\mathbf{x})\right] \to 0 \text{ as } n \to \infty \tag{2.19}$$

where μ denotes the distribution of $(X_1, X_2, \ldots, X_k)$.

Corollary 2.1. *Under the conditions of Theorem 2.1 and (2.34) in Sect. 2.4, $m_n(\mathbf{x}) \to m(\mathbf{x})$ with probability one for μ-almost all $\mathbf{x}$.*

Remark 2.4. In [4], we supposed that h is bounded, this condition is not necessary now.

Theorems 2.2 and 2.3 deal with two special cases: window weights and *NN*-weights. Consistency of window estimates for the regression function has been obtained by [2] and [5]. *NN*-weights for the regression function have been studied in [6], Theorem 2.

In what follows, $|\cdot|$ denotes the maximum norm on $\mathbb{R}^d$. We also write

$$\|\mathbf{X}_\beta - \mathbf{x}\| := \max_{1 \le i \le k} |X_{\beta_i} - x_i|.$$

To define window weights, put (see [8])

$$W_\beta(\mathbf{x}) = \begin{cases} \mathbb{1}_{[\|\mathbf{X}_\beta - \mathbf{x}\| \le h_n]} / \sum_\beta \mathbb{1}_{[\|\mathbf{X}_\beta - \mathbf{x}\| \le h_n]} & \text{if well defined} \\ 0 & \text{otherwise.} \end{cases} \tag{2.20}$$

Here $h_n > 0$ is a given window size to be chosen by the statistician. Then we have the following results:

Theorem 2.2. *Assume $h_n \to 0$ and $nh_n^d \to \infty$ as $n \to \infty$. Then, under the conditions (2.4) and (2.15), we have*

$$m_n(\mathbf{X}) \to m(\mathbf{X}) \ \ in\ \mathcal{L}_r,$$

where $W_\beta(\mathbf{x})$ in (2.3) is given by (2.20).

For the *NN*-weights, recall that X_j is among the $k_n NN$ of $x \in \mathbb{R}^d$ if $d_j(x) := \|X_j - x\|$ is among the k_n-smallest ordered values $d_{1:n}(x) \le \ldots \le d_{n:n}(x)$ of the d's. Ties may be broken by randomization.

For a given $1 \le k_n \le n$, set

$$W_\beta(\mathbf{x}) = \begin{cases} k_n^{-d} & \text{if } X_{\beta_i} \text{ is among the } k_n - NN \text{ of } x_i \text{ for } 1 \le i \le k \\ 0 & \text{otherwise.} \end{cases} \tag{2.21}$$

Theorem 2.3. *Assume that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. Then, under the conditions (2.4) and (2.15), we have*

$$m_n(\mathbf{X}) \to m(\mathbf{X}) \ \text{ in } \mathcal{L}_r,$$

where $W_\beta(\mathbf{x})$ in (2.3) is given by (2.21).

We now consider as estimator of $m(x)$, the statistics of the form

$$m_n(\mathbf{x}) = u_n(\mathbf{x}) \tag{2.22}$$

where $u_n(\mathbf{x})$ is defined in (2.2). Then, in view of (2.3) we have

$$W_{\beta,n}(\mathbf{x}) = \frac{\prod_{j=1}^{k} K[(x_j - X_{\beta_j})/h_n]}{\sum_\beta \prod_{j=1}^{k} K[(x_j - X_{\beta_j})/h_n]} \tag{2.23}$$

where $K(\mathbf{x})$ is a so-called smoothing kernel satisfying $\int K(\mathbf{u})d\mathbf{u} = 1$ and $\lim_{\mathbf{u}\to\infty} |\mathbf{u}|K(\mathbf{u}) = 0$ and $\{h_n,\ n \geq 1\}$ is a sequence of bandwidths tending to zero. This special case was studied by [7] for i.i.d. random variables, and from Theorem 2.1, we can generalize his result for nonstationary dependent random variables. The following theorem establishes that the universal consistency still holds for conditional U-statistics involving kernel K and a sequence of bandwidth h_n.

Theorem 2.4. *Assume that $h_n \to 0$ and $nh_n^d \to \infty$ as $n \to \infty$. Then, under the conditions (2.4) and (2.15), we have*

$$m_n(\mathbf{X}) \to m(\mathbf{X}) \ \text{ in } \mathcal{L}_r,$$

where $m(\boldsymbol{x})$ is given (2.1).

2.4 Proof of Theorems and Corollary 2.1

First, we show that m_n is the ratio of two U-statistics. Let $\mathbf{x} = (x_1, \ldots, x_k)$ be fixed throughout. Let

$$U_n(h,\mathbf{x}) = U_n(\mathbf{x}) = U_n = \binom{n}{k}^{-1} \sum_\beta h(\mathbf{Y}_\beta)V_n(\mathbf{x},\mathbf{X}_\beta)\Big/\int V_n(\mathbf{x},\mathbf{u})dF^{\otimes k}(\mathbf{u}).$$

Hence $m_n(\mathbf{x}) = U_n(h,\mathbf{x})\big/U_n(1,\mathbf{x})$ and $U_n(h,\mathbf{x})$, for each $n \geq k$, is a nonstationary U-statistic as defined in (2.10) with a hind depending on n.

Consider the sequence of functionals

$$\theta_n(h, \mathbf{x}) \equiv \theta_n = \binom{n}{k}^{-1} \sum_{S_k \subset \mathbb{H}^n} \int_{S_k} \phi_n dH^{\otimes S_k}$$

where ϕ_n is defined in (iii).

Note that $\theta_n = E(U_n)$.

The decomposition defined in (2.14) can be written as

$$U_n = \theta_n + \sum_{p=1}^{k} \binom{k}{p} U_{n,p}$$

where $U_{n,p}$ is defined as in (2.12).

To prove Theorem 2.1, the following lemmas are needed.

Lemma 2.1. *Under the conditions of Theorem 2.1*

$$E(U_{n,p})^2 = \mathcal{O}(n^{-2}).$$

Proof. We shall consider the case $p = 2$. The proofs in the cases $c = 3, \ldots, k$ are analogous and so they are omitted.

We first note that

$$U_{n,2} = \binom{n}{2}^{-1} \sum_{S_2 \subset \mathbb{H}^n} \int_{S_2} \phi_{S_2} d_{\otimes_{\mathbb{H}_i \subset S_2}} (\delta_{Z_i} - H_i)$$

so we have

$$E(U_{n,2})^2 = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \sum_{1 \le l_1 < l_2 \le n} J((i_1, i_2), (l_1, l_2))$$

where

$$J((i_1, i_2), (l_1, l_2)) = E\left(\int_{S_2} \phi_{S_2} d_{\otimes_{1 \le j \le 2} \mathbb{H}_{i_j}} (\delta_{Z_{i_j}} - H_{i_j}) \right.$$
$$\left. \times \int_{S_2'} \phi_{S_2'} d_{\otimes_{1 \le m \le 2} \mathbb{H}_{l_m}} (\delta_{Z_{l_m}} - H_{l_m}) \right)$$

$$S_2 = \mathbb{H}_{i_1} \oplus \mathbb{H}_{i_2} \text{ and } S_2' = \mathbb{H}_{l_1} \oplus \mathbb{H}_{l_2}.$$

So from condition (2.4) and condition (iii), we have from Lemma 2.1 in [10] the inequalities:

(i) If $1 \leq i_1 < i_2 \leq l_1 < l_2$, then

$$J((i_1, i_2), (l_1, l_2)) \leq M\{\beta(l_2 - l_1)\}^{\frac{\delta}{1+\delta}};\ l_2 - l_1 \geq i_2 - i_1 \tag{2.24}$$

and

$$J((i_1, i_2), (l_1, l_2)) \leq M\{\beta(i_2 - i_1)\}^{\frac{\delta}{1+\delta}};\ i_2 - i_1 \geq l_2 - l_1 \tag{2.25}$$

where M is a finite positive constant.

Thus, using (2.24) and (2.25), we obtain

$$\left| \sum_{1 \leq i_1 < i_2 \leq l_1 < l_2 \leq n} J((i_1, i_2), (l_1, l_2)) \right| = \mathcal{O}(n^2). \tag{2.26}$$

Similarly
(ii) If $1 \leq i_1 < l_1 < i_2 < l_2 \leq n$, then

$$\left| \sum_{1 \leq i_1 < l_1 < i_2 < l_2 \leq n} J((i_1, i_2), (l_1, l_2)) \right| = \mathcal{O}(n^2). \tag{2.27}$$

(iii) If $1 \leq i_1 < l_1 \leq l_2 < i_2 \leq n$, then

$$\left| \sum_{1 \leq i_1 < l_1 \leq l_2 < i_2 \leq n} J((i_1, i_2), (l_1, l_2)) \right| = \mathcal{O}(n^2). \tag{2.28}$$

From (2.26), (2.27) and (2.28), we obtain

$$E(U_{n,2})^2 = \mathcal{O}(n^{-2}).$$

Thus the result for the case $p = 2$ is proved. □

Lemma 2.2. *Under the condition of Theorem 2.1, for μ-almost all $\boldsymbol{x}$*

$$\frac{\theta_n(h, \boldsymbol{x})}{\theta_n(1, \boldsymbol{x})} \to m(\boldsymbol{x}),\ n \to \infty.$$

Proof. By definition, we have

$$\theta_n(h, \mathbf{x}) \equiv \theta = \binom{n}{k}^{-1} \sum_{S_k \subset \mathbb{H}^n} \int_{S_k} \phi_n \, d\mathbb{H}^{\otimes S_k}.$$

Put

$$\theta_n'(h,\mathbf{x}) = \binom{n}{k}^{-1} \sum_{S_k \subset \mathbb{H}'^n} \int_{S_k} h(\mathbf{y}) V_n(\mathbf{x},\mathbf{z}) dF^{\otimes S_k}(\mathbf{y},\mathbf{z}).$$

From condition (2.15), we deduce

$$|\theta_n(h,\mathbf{x}) - \theta_n'(h,\mathbf{x})| \to 0 \;\text{ as } n \to \infty. \tag{2.29}$$

From condition (ii) in Sect. 2.3, we have

$$\lim_{n\to\infty} \binom{n}{k}^{-1} \sum_{S_{1,k} \subset \mathbb{H}'^n} \int_{S_{1,k}} m(\mathbf{z}) V_n(\mathbf{x},\mathbf{z}) dF^{\otimes S_{1,k}}(\mathbf{z}) = m(\mathbf{x}) \tilde{f}(\mathbf{x}) \int V(\mathbf{z}) d\mathbf{z} \tag{2.30}$$

and so

$$\lim_{n\to\infty} \binom{n}{k}^{-1} \sum_{S_{1,k} \subset \mathbb{H}'^n} \int_{S_{1,k}} V_n(\mathbf{x},\mathbf{z}) dF^{\otimes S_{1,k}}(\mathbf{z}) = \tilde{f}(\mathbf{x}) \int V(\mathbf{z}) d\mathbf{z}. \tag{2.31}$$

By definition, we have

$$\begin{aligned}
\theta_n(h,\mathbf{x}) &= \binom{n}{k}^{-1} \sum_{S_k \subset \mathbb{H}'^n} \int_{S_k} h(\mathbf{y}) V_n(\mathbf{x},\mathbf{z}) dF^{\otimes S_k}(\mathbf{y},\mathbf{z}) \\
&= \binom{n}{k}^{-1} \sum_{S_k \subset \mathbb{H}'^n} \int_{S_{1,k}} \left(\int_{S_{2,k}} E(h(\mathbf{y})|\mathbf{X} = \mathbf{z}) V_n(\mathbf{x},\mathbf{z}) dG^{\otimes S_{2,k}}(\mathbf{y}) \right) \\
&\quad \times dF^{\otimes S_{1,k}}(\mathbf{z}) \\
&= \binom{n}{k}^{-1} \sum_{S_{1,k} \subset \mathbb{H}'^n} \int_{S_{1,k}} m(\mathbf{z}) V_n(\mathbf{x},\mathbf{z}) dF^{\otimes S_{1,k}}(\mathbf{z}).
\end{aligned} \tag{2.32}$$

From (2.29)–(2.32), we deduce easily that

$$\frac{\theta_n(h,\mathbf{x})}{\theta_n(1,\mathbf{x})} \to m(\mathbf{x}),\; n \to \infty.$$

To prove Theorem 2.1, from Lemmas 2.1 and 2.2, we now have to show that for μ-almost all $\mathbf{x}$,

$$U_{n,1}(h,\mathbf{x}) \to 0 \text{ in probability.}$$

Since

$$\begin{aligned}U_{n,1}(h,\mathbf{x}) &= n^{-1}\sum_{S_1\subset\mathbb{H}^n}\int_{S_1}\phi_{S_1}d\otimes_{\mathbb{H}_i\subset S_1}(\delta_{Z_i}-H_i)\\ &= n^{-1}\sum_{i=1}^{n}\int_{\mathbb{H}_i}\phi_{\mathbb{H}_i}d(\delta_{Z_i}-H_i)\end{aligned}$$

we have

$$\begin{aligned}E(U_{n,1})^2 &= n^{-2}E\left(\sum_{i=1}^{n}\int_{\mathbb{H}_i}\phi_{\mathbb{H}_i}d(\delta_{Z_i}-H_i)\right)^2\\ &= n^{-2}\sum_{i=1}^{n}E\left(\int_{\mathbb{H}_i}\phi_{\mathbb{H}_i}d(\delta_{Z_i}-H_i)\right)^2\\ &\quad+2n^{-2}\sum_{1\le i<j\le n}E\left\{\left(\int_{\mathbb{H}_i}\phi_{\mathbb{H}_i}d(\delta_{Z_i}-H_i)\right)\right.\\ &\quad\left.\times\left(\int_{\mathbb{H}_j}\phi_{\mathbb{H}_j}d(\delta_{Z_j}-H_j)\right)\right\}.\end{aligned}$$

From Lemma 2.1 of [10] and condition (iii), we have

$$\begin{aligned}E(U_{n,1})^2 &\le 2n^{-2}nM(2,h)+4n^{-2}M^{\frac{1}{1+\delta}}(r,h)\sum_{p=1}^{n}(p+1)\beta^{\frac{\delta}{1+\delta}}(p)\\ &= \mathcal{O}(n^{-1})\end{aligned}$$

where $M(t,h)=E\{\sup_\beta|\phi_n(\mathbf{X}_\beta,\mathbf{Y}_\beta)|^t\}$, which implies

$$E(U_{n,1})^2=\mathcal{O}(n^{-1}). \tag{2.33}$$

From Lemmas 2.1 and 2.2 and from (2.33), we have

$$U_n(h,\mathbf{x})\to m(\mathbf{x})\tilde{f}(\mathbf{x})\int V(\mathbf{z})d\mathbf{z}$$

and

$$U_n(1,\mathbf{x})\to\tilde{f}(\mathbf{x})\int V(\mathbf{z})d\mathbf{z}\text{ in probability}$$

as $n\to\infty$ for μ-almost all $\mathbf{x}$.
It remains to prove the uniform integrability.

It is an easy convergence of the Jensen's inequality

$$\sup_{n\in\mathbb{N}^*} E\left\{\left[\sum_\beta V_n(\mathbf{X},\mathbf{X}_\beta)|h(\mathbf{Y}_\beta)|/\sum_\beta V_n(\mathbf{X},\mathbf{X}_\beta)\right]^r\right\}$$

$$\leq \sup_{n\in\mathbb{N}^*} E\left\{\sum_\beta V_n(\mathbf{X},\mathbf{X}_\beta)|h(\mathbf{Y}_\beta)|^r/\sum_\beta V_n(\mathbf{X},\mathbf{X}_\beta)\right\}$$

$$\leq E\left\{\sup_\beta |h(\mathbf{Y}_\beta)|^r\right\} < +\infty.$$

and Theorem 2.1 is proved.

The proof of Corollary 2.1 is a consequence of Lemma 2.1 and Lemma 2.3 below.

For a d-dimensional vector V, consider the norm $\|V\| = \max_{1\leq j\leq d} |V^{(j)}|$. This norm is equivalent to the Euclidian norm and easy to work with here. We will use this norm in Lemma 2.3 below see also [1].

Lemma 2.3. *Let $(V_n)_{n\geq 1}$ be a sequence of d-dimensional centered absolutely regular and non necessarily stationary random vectors with rate satisfying*

$$\sum_{i\geq 1}(i)^{\frac{r-\delta}{2}}[\beta(i)]^{\frac{\delta}{r}} < \infty \tag{2.34}$$

$$\sup_{i\geq 1} E(\|V_i\|^r) < \infty. \tag{2.35}$$

Then

$$n^{-1}\sum_{i=1}^n V_i \to 0 \textit{ with probability 1, as } n\to\infty.$$

Proof. For $\epsilon > 0$,

$$P\left(\frac{1}{n}\sum_{i=1}^n \|V_i\| \geq \epsilon\right) = P\left(\max_{1\leq j\leq d}\left|\frac{1}{n}\sum_{i=1}^n V_i^{(j)}\right| \geq \epsilon\right)$$

$$\leq \sum_{1\leq j\leq d} P\left(\left|\frac{1}{n}\sum_{i=1}^n V_i^{(j)}\right| \geq \epsilon\right). \tag{2.36}$$

For all $1\leq j\leq d$, one has from Markov's inequality that

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n V_i^{(j)}\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^r n^r} E\left(\left|\sum_{i=1}^n V_i^{(j)}\right|^r\right). \tag{2.37}$$

By Lemma 5.2 of [3], one has that

$$E\left(\left|\sum_{i=1}^{n} V_i^{(j)}\right|^r\right) \leq C n^{r/2}. \tag{2.38}$$

From the above two inequalities, one deduces that

$$E\left(\left|\sum_{i=1}^{n} V_i^{(j)}\right|^r\right) \leq \frac{C}{\epsilon^r} n^{r/2}. \tag{2.39}$$

Since $r/2 > 1$, the last inequality implies that for all $1 \leq j \leq d$,

$$\sum_{n \geq 1} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} V_i^{(j)}\right| \geq \epsilon\right) < \infty$$

which, in turn, implies that

$$\sum_{n \geq 1} P\left(\left\|\frac{1}{n}\sum_{i=1}^{n} V_i\right\| \geq \epsilon\right) < \infty.$$

Lemma 4.3 then follows by Borel–Cantelli theorem.

The proofs Theorems 2.2 to 2.4 are also consequences of Theorem 2.1 by using technics similar as in the proofs of Theorem 2.2 to Theorem 2.4 in [4]: that is to verify that conditions (i)–(iii) are satisfied.

References

1. Bennett G (1962) Probability inequalities for the sum on independent random variables. J Am Stat Assoc 19:33–45
2. Devroye LP, Wagner TJ (1980) Distribution-free consistency results in nonparametric discrimination and regression function estimate. Ann Stat 8:231–239.
3. Harel M, Puri LM (1996) Conditional U-statistics for dependent random varaibles. J Multivar Anal 57:84–100
4. Harel M, Puri LM (2004) Universally consistent conditional U-statistics for absolutely regular processes and its applications for hidden markov models. Ann Inst Stat Math 56(4):819–832
5. Speigelman G, Sacks J (1980) Consistent window estimation inonparametric regression. Ann Stat 8:240–246
6. Stone G (1977) Consistent nonparametric regression, Ann Stat 5:595–620
7. Stute W (1991) Conditional U-statistics. Ann Probab 19:812–825
8. Stute W (1994a) Universally consistent conditional U-statistics. Ann Stat 22:460–473
9. Stute W (1994b) L^p convergence of conditional U-statistics. J Multivar Anal 51:71–82
10. Yoshihara KL (1976) Limiting behavior of U-statistics for stationary absolutely regular processes. Z Wahrsch Verw Gebiete 35:237–252

http://www.springer.com/978-3-7908-2627-2

Advances in Directional and Linear Statistics
A Festschrift for Sreenivasa Rao Jammalamadaka
Wells, M.T.; Sengupta, A. (Eds.)
2011, XIII, 321 p., Hardcover
ISBN: 978-3-7908-2627-2
A product of Physica-Verlag Heidelberg