JOURNAL OF MULTIVARIATE ANALYSIS **57**, 84–100 (1996)
ARTICLE NO. 0023

# Conditional $U$-Statistics for Dependent Random Variables

MICHEL HAREL

*Université de Limoges, U.R.A. C.N.R.S. 745, Toulouse, France*

AND

MADAN L. PURI*

*Indiana University*



independent case to the dependent case. 

## 1. INTRODUCTION

Stute [10] introduced a class of so-called conditional $U$-statistics, which may be viewed as a generalization of the Nadaraya–Watson estimates of a regression function. This extension is similar to Hoeffding's [3] generalization of sample means to what we now call $U$-statistics.

Assume that $(X_i, Y_i)$ are random vectors in the space $\mathbb{R}^p \times \mathbb{R}^m$, where $X_i = (X_{i1}, ..., X_{ip})$ and $Y_i = (Y_{i1}, ..., Y_{im})$, $i = 1, ..., n$. Let $h$ be any function of $k$-variates (the $U$-kernel), $k \leqslant n$ such that $h(Y_1, ..., Y_k)$ is integrable. We are interested in the estimation of

$$m(x_1, ..., x_k) = E[h(Y_1, ..., Y_k) \mid X_1 = x_1, ..., X_k = x_k]. \tag{1.1}$$

When $p = m = k = 1$ and $h = I_d$, then

$$m(x_1) = E[Y_1 \mid X_1 = x_1]$$

is the regression of $Y_1$, given $X_1 = x_1$.

Received January 25, 1994; revised March 1995.

AMS subject classifications: 60F05, 60F15, 62J99.

Key words and phrases: Conditional $U$-statistics, asymptotic normality, strong convergence, absolute regularity.

* Research supported by the Office of Naval Research Contract N00014-91-J-1020.



0047-259X/96 $18.00

For estimation of $m(x_1)$, Nadaraya [4] and Watson [11] independently proposed:

$$m_n(x_1)=\frac{\sum_{i=1}^n Y_i K[(x_1-X_i)/a_n]}{\sum_{i=1}^n K[(x_1-X_i)/a_n)]}.$$

Here $K$ is the so-called smoothing kernel satisfying $\int K(u)\,du=1$ and $(a_n)$ is a sequence of bandwidths tending to zero at appropriate rates.

Schuster [7] under conditions requiring the existence of the density $f$ of $X_i$ and finiteness of $E(|Y_1|^3)$, proved the central limit theorem for $m_n(x)$. See also Rosenblatt [6]. Then, Stute [8] proved the asymptotic normality of $m_n(x)$ only under the condition of the finiteness of $E(Y_1^2)$, while $X_1$ need not have a density at all. Later Yoshihara [13] proved the central limit theorem for $m_n(x)$ when the r.v.'s are $\varphi$-mixing under finiteness of $E(|Y_1|^{2+\delta})$ $(\delta>0)$. But the $\varphi$-mixing condition has applications which are too limited; for example, an ARMA process is never $\varphi$-mixing but generally geometrically absolutely regular.

For an arbitrary $k$, we now consider statistics of the form

$$\begin{aligned} u_n(\mathbf{x}) &= u_n(x_1, ..., x_k)\\ &= \frac{\sum_\beta h(Y_{\beta_1}, ..., Y_{\beta_1}) \prod_{j=1}^k K[(x_j-X_{\beta_j})/a_n]}{\sum_\beta \prod_{j=1}^k K[(x_j-X_{\beta_j})/a_n]}\\ &= \frac{\left[\begin{array}{l}\sum_\beta h((Y_{\beta_1 1}, ..., Y_{\beta_1 m}), ..., (Y_{\beta_k 1}, ..., Y_{\beta_k m}))\\ \qquad\qquad \times \prod_{j=1}^k K[(x_{j1}-X_{\beta_j 1})/a_n, ..., (x_{jp}-X_{\beta_j p})/a_n]\end{array}\right]}{\sum_\beta \prod_{j=1}^k K[(x_{j1}-X_{\beta_j 1})/a_n, ..., (x_{jp}-X_{\beta_j p})/a_n]}. \end{aligned}$$

Here the summation extends over all permutations $\beta=(\beta_1, ..., \beta_k)$ of length $k$. Stute [10] derived the limit distribution of $u_n(x)$ when the r.v.'s are independent under finiteness of $E(|h(Y_1, ..., Y_k)|^{2+\delta})$ $(\delta>0)$. In this paper, we extend the result of Stute [10] for absolutely regular r.v.'s. When $p=m=k=1$ and $h=I_k$ our result extends the result of Yoshihara [13] from the $\varphi$-mixing condition to the absolutely regular condition which permits a broad range of applications.

In what follows, we assume that the function $h$ is symmetric and the sequence of r.v.'s $(X_1, Y_1), ..., (X_n, Y_n)$ is absolutely regular with rates

$$\beta(m)=O(\rho^m) \qquad \text{for some} \quad 0<\rho<1. \tag{1.2}$$

Recall that a sequence of random vectors $\{X_{ni,\,1\leqslant i\leqslant n,\,n\geqslant 1}\}$ is absolutely regular if

$$\sum_{m\leqslant n} \max_{1\leqslant j\leqslant n-m} E\{\sup_{A\in\sigma(X_{ni},\, i\geqslant j+m)} |P(A\mid\sigma(X_{ni}, A\leqslant i\leqslant j))-P(A)|\}=\beta(m)\downarrow 0.$$

Here $\sigma(X_{ni}, 1 \leqslant i \leqslant j)$ and $\sigma(X_{ni}, i \geqslant j+m)$ are the $\sigma$-fields generated by $(X_{n1}, ..., X_{nj})$ and $(X_{n,j+m}, X_{n,j+m+1}, ..., X_{nn})$, respectively. Also recall that $\{X_{ni}\}$ satisfies the strong mixing condition if $\sup_{m \leqslant n} \sup_{1 \leqslant j \leqslant n-m} \{|P(A \cap B) - P(A) P(B)|;\ A \in \sigma(X_{ni}, 1 \leqslant i \leqslant j),\ B \in \sigma(X_{ni}, i \geqslant j+m)\} = \alpha(m) \downarrow 0$. Since $\alpha(m) \leqslant \beta(m)$, it follows that if $\{X_{ni}\}$ is absolutely regular, then it is also strong mixing.

In Section 4, we will show how our results can be applied to some Markov processes and particularly to some ARMA processes.

## 2. Asymptotic Normality

Let $\mathbf{x} = (x_1, ..., x_k)$ be fixed throughout. In this section, $h$ will be assumed to be square-integrable. Set

$$U_n(h, \mathbf{x}) \equiv U(\mathbf{x}) \equiv U_n = \frac{(n-k)!}{n!} \sum_{\beta} h(Y_{\beta_1}, ..., Y_{\beta_k}) \times \prod_{j=1}^{k} K\left(\frac{x_j - X_{\beta_j}}{a_n}\right) \Big/ \prod_{j=1}^{k} EK\left(\frac{x_j - X_1}{a_n}\right). \tag{2.1}$$

Then

$$u_n(\mathbf{x}) = U_n(h, \mathbf{x}) / U_n(1, \mathbf{x}).$$

Note that $U_n(h, \mathbf{x})$ for each $k \geqslant 1$ is a classical $U$-statistic with a kernel depending on $n$.

Next, we denote the distribution function (d.f.) of $(X_i, Y_i)$ by $H$ and the marginals by $F$ and $G$. Consider a sequence of functionals

$$\begin{aligned}\theta_n(h, \mathbf{x}) &\equiv \theta_n \\ &= \int m(z_1, ..., z_k) \prod_{j=1}^{k} K\left(\frac{x_j - z_j}{a_n}\right) \\ &\quad \times F(dz_1) \ldots F(dz_k) \Big/ \prod_{j=1}^{k} \int K\left(\frac{x_j - x}{a_n}\right) F(dx) \\ &= \frac{E[h(\tilde{Y}_1, ..., \tilde{Y}_k) \prod_{j=1}^{k} K((x_j - \tilde{X}_j)/a_n)]}{E[\prod_{j=1}^{k} K((x_j - \tilde{X}_j)/a_n)]},\end{aligned} \tag{2.2}$$

where $(\tilde{X}_i, \tilde{Y}_i)$, $i = 1, ..., k$ are i.i.d. random vectors with d.f. $H$.

We also suppose that the r.v.'s $(Y_1, \mid X_1)$, $(X_2 \mid Y_2)$, ..., $(Y_n \mid X_n)$ are independent and we denote by $\tilde{G}(x; \cdot)$ the conditional d.f. of $(Y_1 \mid X_1 = x)$.

We also assume that

$$E\,|h(Y_1, ..., Y_k)|^{2+\delta} < +\infty \qquad \text{for some} \quad \delta > 0. \tag{2.3}$$

For every $c$ $(0 \leqslant c \leqslant k)$, put

$$\begin{aligned} g_{c,n}&((z_1, y_1), ..., (z_c, y_c)) \\ &\equiv g_c((z_1, y_1), ..., (z_c, y_c)) \\ &= \int h(y_1, ..., y_k) \prod_{j=1}^{k} K\left(\frac{x_j - z_j}{a_n}\right) \Big/ \prod_{j=1}^{k} EK\left(\frac{x_j - X_1}{a_n}\right) \\ &\quad \times \prod_{j=c+1}^{k} \tilde{G}(z_j; dy_j)\, F(dz_j). \end{aligned} \tag{2.4}$$

We have $g_0 = \theta_n$ and

$$\begin{aligned} g_k&((z_1, y_1), ..., (z_k, y_k)) \\ &= h(y_1, ..., y_k) \prod_{j=1}^{k} K\left(\frac{x_j - z_j}{a_n}\right) \Big/ \prod_{j=1}^{k} EK\left(\frac{x_j - X_1}{a_n}\right) \end{aligned} \tag{2.5}$$

Let $n^{-[r]} = \{n(n-1) \cdots (n-r+1)\}^{-1}$. Set

$$\begin{aligned} U_n^{(c)} = n^{-[c]} \sum_{\beta^{(c)}} \int & g_c((z_1, y_1), ..., (z_c, y_c)) \\ & \times \prod_{j=1}^{c} d(I_{[(X_{\beta j},\, Y_{\beta j}) \leqslant (z_j,\, y_j)]} - H(z_j, y_j)), \end{aligned} \tag{2.6}$$

where $\beta^{(c)}$ is the summation over all the permutations $\beta^{(c)} = (\beta_1, ..., \beta_c)$ of length $c$. Then

$$U_n = \theta_n + \sum_{c=1}^{k} \binom{k}{c} U_n^{(c)}. \tag{2.7}$$

Let

$$\sigma^2 = \sigma^2(h, \mathbf{x}) = \lim_{n \to \infty} a_n^p \{E(g_1^2(X_1, Y_1)) - \theta_n^2\} \tag{2.8}$$

if the limit exists. We note also that $\lim_{n \to \infty} a_n^p \theta_n^2 = 0$.

LEMMA 2.1. *Assume that*

(i) $a_n \to 0$ *and* $na_n \to \infty$.

(ii) *$K$ is bounded and has compact support.*

(iii) *F admits a density f which is continuous at each* $x_j$, $1 \leqslant j \leqslant k$, *with* $f(x_j) > 0$.

*Then, we have* $\sigma^2 < \infty$ *and* $(na_n^p)^{1/2}\, U_n^{(1)} \to \mathcal{N}(0, \sigma^2)$ *in distribution, where* $\sigma^2$ *is defined in* (2.8).

*Proof.* First, we show that

$$\lim_{n \to \infty} E((na_n^p)^{1/2}\,(U_n^{(1)}))^2 = \sigma^2. \tag{2.9}$$

From (2.6), we have

$$\begin{aligned}
U_n^{(1)} &= n^{-1} \sum_{i=1}^{n} g_1(X_i, Y_i)\, d(I_{[(X_i, Y_i) \leqslant (x_i, y_i)]} - H(x_i, y_i)) \\
&= n^{-1} \sum_{i=1}^{n} g_1(X_i, Y_i) \\
&\quad - n^{-1} \sum_{i=1}^{n} \int h(y_1, ..., y_k) \prod_{j=1}^{k} K\left(\frac{x_j - z_j}{a_n}\right) \Big/ \prod_{j=1}^{k} EK\left(\frac{x_j - X_1}{a_n}\right) \\
&\quad \times \prod_{j=1}^{k} \tilde{G}(z_j; dy_j)\, F(dz_j) \\
&= n^{-1} \sum_{i=1}^{n} g_1(X_i, Y_i) \\
&\quad - n^{-1} \sum_{i=1}^{n} \int m(z_1, ..., z_k) \Big/ \prod_{j=1}^{k} EK\left(\frac{x_j - X_1}{a_n}\right) \prod_{j=1}^{k} F(dz_j) \\
&= n^{-1} \sum_{i=1}^{n} (g_1(X_i, Y_i) - \theta_n).
\end{aligned}$$

We can write

$$\begin{aligned}
&na_n^p E(U_n^{(1)})^2 \\
&\quad = n^{-1} a_n^p E\left(\sum_{i=1}^{n} (g_1(X_i, Y_i) - \theta_n)\right)^2 \\
&\quad = n^{-1} a_n^p \sum_{i=1}^{n} E(g_1(X_i, Y_i) - \theta_n)^2 \\
&\qquad + n^{-1} a_n^p \sum_{1 \leqslant i \neq j \leqslant n} E\{(g_1(X_j, Y_j) - \theta_n)\}
\end{aligned}$$

$$= n^{-1} a_n^p \sum_{i=1}^{n} E(g_1(X_i, Y_i) - \theta_n)^2$$

$$+ 2n^{-1} a_n^p \sum_{\substack{1 \leqslant i < j \leqslant n \\ j-i \leqslant r}} E\{(g_1(X_i, Y_i) - \theta_n)(g_1(X_j, Y_j) - \theta_n)\}$$

$$+ 2n^{-1} a_n^p \sum_{\substack{1 \leqslant i < j \leqslant n \\ j-i > r}} E\{(g_1(X_i, Y_i) - \theta_n)(g_1(X_j, Y_j) - \theta_n)\}.$$

First, note that

$$\lim n^{-1} a_n^p \sum_{i=1}^{n} E(g_1(X_i, Y_i) - \theta_n)^2 = \lim_{n \to \infty} a_n^p E(g_1(X_1, Y_1) - \theta_n)^2 = \sigma^2. \tag{2.10}$$

For the sake of brevity we set

$$M = \prod_{j=1}^{k} E\left(K\left(\frac{x_j - X_1}{a_n}\right)\right).$$

From condition (ii), we easily deduce that

$$M = O(a_n^{kp}).$$

Then, we have

$$E\{(g_1(X_i, Y_i) - \theta_n)(g_1(X_j, Y_j) - \theta_n)\}$$

$$= M^{-2} \int (m(z_1, ..., z_k) - \theta_n)(m(z_{k+1}, ..., z_{2k}) - \theta_n)$$

$$\times \prod_{m=1}^{k} K\left(\frac{x_m - z_m}{a_n}\right) \prod_{l=1}^{k} K\left(\frac{x_{k+l} - z_{k+l}}{a_n}\right)$$

$$\times \prod_{m=2}^{k} F(dz_m) \prod_{\substack{l=k+2 \\ l \neq j}}^{k} F(dz_{k+l}) F_{i,j}(dz_1, dz_{k+1})$$

$$= M^{-2} a_n^{2kp} \int (m(x_1 - a_n u_1, ..., x_k - a_n u_k) - \theta_n)$$

$$\times (m(x_{k+1} - a_n u_{k+1}, ..., x_{2k} - a_n u_{2k}) - \theta_n)$$

$$\times \prod_{m=1}^{k} K(u_m) \prod_{l=k+1}^{2k} K(u_l)$$

$$\times \frac{f_{i,j}(F^{-1}(x_1 - a_n u_1), F^{-1}(x_{k+1} - a_n u_{k+1}))}{f \circ F^{-1}(x_1 - a_n u_1) f \circ F^{-1}(x_{k+1} - a_n u_{k+1})} du_1 \cdots d_2 k$$

$$\leqslant C a_n^{2kp} M^{-2} \qquad \text{(from conditions (2.4) and (iii))},$$

$$\leqslant C, \tag{2.11}$$

where $F_{i,j}$ and $f_{i,j}$ are respectively the distribution function and the density function of $(X_i, X_j)$ and, by convention, $x_i - a_n u_i = (x_{i1} - a_n u_{i1}, ..., x_{ip} - a_n u_{ip})$.

For any $\delta > 0$

$$\begin{aligned} E\,|g_1(X_i, Y_i) - \theta_n|^{2+\delta} &\leqslant M^{-(2+\delta)} \int |m(z_1, ..., z_k) - \theta_n|^{2+\delta} \left| \prod_{m=1}^{k} K\left(\frac{x_m - z_m}{a_n}\right)\right|^{2+\delta} \prod_{m=1}^{k} F(dz_m) \\ &\leqslant M^{-(2+\delta)} a_n^{kp} \int |m(x_1 - a_n u_1, ..., x_k - a_n u_k) - \theta_n|^{2+\delta} \\ &\quad \times \left| \prod_{m=1}^{k} K(u) \right|^{2+\delta} \prod_{m=1}^{k} f(x_m - a_n u_m)\, du_1 \cdots du_k \\ &\leqslant c a_n^{-k(1+\delta)p}. \end{aligned} \tag{2.12}$$

From (2.11), (2.12), and Lemma 5.1 in the Appendix it follows that

$$\begin{aligned} n^{-1} a_n^p & \sum_{1 \leqslant i \neq j \leqslant n} E\{(g_1(X_i, Y_i) - \theta_n)(g_1(X_j, Y_j) - \theta_n)\} \\ & \leqslant 2Cra_n^p + 2(Ca_n^{-k(1+\delta)p})^{2/(2+\delta)} \sum_{i=r+1}^{n} [\beta(i)]^{\delta/(2+\delta)}. \end{aligned}$$

If we take $r = [a_n^{-(1/2)p}]$, we get (2.9) from (1.2) and (2.10).

From Lemma 5.2 in the Appendix, we obtain the following inequality:

$$\begin{aligned} E & \left| \sum_{i=1}^{m} g_1(X_i, Y_i) - \theta_n \right|^{2+\delta} \\ & \leqslant Cm^{(2+\delta)/2} \sup_{1 \leqslant i \leqslant m} E\,|g_1(X_i, Y_i) - \theta_n|^{2+\delta}, \qquad m \leqslant n. \end{aligned} \tag{2.13}$$

Let now $r = [n^{2/3}]$, $q = [n^{1/3}]$, and $l = [n/(r+1)]$. Put

$$\eta_j = \sum_{i=(j-1)(r+q)+1}^{(j-1)(r+q)+r} (g_1(X_i, Y_i) - \theta_n), \qquad j = 1, ..., l,$$

$$\theta_j = \sum_{i=(j-1)(r+q)+r+1}^{(j-1)(r+q)+r} (g_1(X_i, Y_i) - \theta_n), \qquad j = 1, ..., l,$$

$$\theta_{k+1} = \sum_{i=l(r+q)+1}^{n} (g_1(X_i, Y_i) - \theta).$$

Then, we have

$$n^{-1/2}a_n^{p/2}\sum_{i=1}^{n}(g_1(X_i, Y_i)-\theta_n)=n^{-1/2}a_n^{p/2}\sum_{j=1}^{l}\eta_j+n^{-1/2}a_n^{p/2}\sum_{j=1}^{l+1}\theta_j.$$

From (2.13), we deduce

$$n^{-1/2}a_n^{p/2}\sum_{j=1}^{l+1}\theta_j \xrightarrow{p} 0.$$

To prove Lemma 2.1, it remains to show that $n^{-1/2}a_n^{p/2}\sum_{j=1}^{l}\eta_j$ converges in law to $\mathcal{N}(0, \sigma^2)$ random variable.

From Lemma 5.1 in the Appendix, we obtain

$$\left|E\left\{\exp\left(itn^{-1/2}a_n^{p/2}\sum_{j=1}^{l}\eta_j\right)\right\}-\prod_{j=1}^{k}\left[E\{\exp(itn^{-1/2}a_n^{p/2}\eta_j)\}\right]\right|\leqslant Cl\beta(q).$$

Hence it suffices to show that

$$\prod_{j=1}^{k}\left[E\{\exp(itn^{-1/2}a_n^{p/2}\eta_j)\}\right] \text{ converges to } e^{-t^2\sigma^2/2}. \tag{2.14}$$

Using (2.13), we obtain

$$\begin{aligned}
&E\{\exp(itn^{-1/2}a_n^{p/2}\eta_j)\}\\
&\quad=1-\frac{t^2a_n^p}{2n}E(\eta_j)^2+O\left(\frac{|t|^{2+\delta}a_n^{(2+\delta)p/2}}{n^{(2+\delta)/2}}E(\eta_j)^{2+\delta}\right)\\
&\quad=1-\frac{a_n^p t^2}{2n}E(\eta_n)^2+o(|t|^{2+\delta}n^{-(1/6)(2+\delta)}a_n^{-(k-1)p+((1-2k)/2)\delta p}).
\end{aligned}$$

From condition (ii) and

$$\lim\frac{la_n^p}{n}E(\eta_j)^2=\sigma^2 \qquad \text{(from (2.10))}$$

we get (2.14). The proof follows.

Lemma 2.2. *Under the conditions of Lemma* 2.1

$$(na_n^p)^{1/2}(U_n-\theta_n)\to\mathcal{N}(0, k\sigma^2) \qquad \textit{in distribution.}$$

*Proof.* From the decomposition (2.7) and Lemma 2.1, it is sufficient to prove that

$$E(U_n^{(c)})^2=O(n^{-2}), \qquad 2\leqslant c\leqslant k. \tag{2.15}$$

We shall only consider the case $c=2$. The proofs for the cases $c=3, ..., k$ are analogous and are therefore omitted.

We first note that

$$U_n^{(2)}=n^{-[2]} \sum_{1\leqslant i_1<i_2\leqslant n} \{g_2((X_{i_1}, Y_{i_1}), (X_{i_2}, Y_{i_2})) - g_1(X_{i_1}, Y_{i_1})-g_1(X_{i_2}, Y_{i_2})+\theta_n\}.$$

So we have

$$E(U_n^{(2)})^2=\sum_{1\leqslant i_1<i_2\leqslant n} \sum_{1\leqslant j_1<j_2\leqslant n} J((i_1, i_2), (j_1, j_2)), \tag{2.16}$$

where

$$\begin{aligned} &J((i_1,i_2),(j_1, j_2)) \\ &\quad = E\{g_2((X_{i_1}, Y_{i_1}), (X_{i_2}, Y_{i_2}))-g_1(X_{i_1}, Y_{i_1})-g_1(X_{i_2}, Y_{i_2})+\theta_n\} \\ &\qquad \times \{g_2((X_{j_1}, Y_{j_1}), (X_{j_2}, Y_{j_2}))-g_1(X_{j_1}, Y_{j_1})-g_1(X_{j_2}, Y_{j_2})+\theta_n\}. \end{aligned}$$

Since

$$\int \{g_2((z_1, y_1), (z_2, y_2))-g_1(z_1, y_2)-g_1(z_2, y_2)+\theta_n\}\ H(dz_1, dy_1)=0,$$

therefore, from Lemma 5.1 we have the following inequalities: If $1\leqslant i_1<i_2\leqslant j_1<j_2\leqslant n$ and $j_2-j_1\geqslant i_2-i_1$, then

$$J((i_1, i_2), (j_1, j_2))\leqslant M\beta^{\delta/(2+\delta)}(j_2-j_1) \tag{2.17}$$

and, similarly, if $1\leqslant i_1<i_2\leqslant j_1<j_2\leqslant n$ and $i_2-i_1\geqslant j_2-j_1$, then

$$J((i_1, i_2), (j_1, j_2))\leqslant M\beta^{\delta/(2+\delta)}(i_2-i_1). \tag{2.18}$$

Thus, from (2.17), (2.18), and assumption (1.2)

$$\begin{aligned} &\left|\sum_{1\leqslant i_1<i_2\leqslant j_1<j_2\leqslant n} j((i_1, i_2), (j_1, j_2))\right| \\ &\quad \leqslant \left\{\sum_{\substack{1\leqslant i_1<i_2\leqslant j_1<j_2\leqslant n \\ i_2-i_1\geqslant j_2-j_1}} + \sum_{\substack{1\leqslant i_1<i_2\leqslant j_1<j_2\leqslant n \\ i_1-i_1\leqslant j_2-j_1}}\right\} |J((i_1, i_2), (j_1, j_2))| \\ &\quad \leqslant Cn^2 \sum_{r=1}^{n} (r+1)\, \beta^{\delta/(2+\delta)}(r)=O(n^2). \end{aligned} \tag{2.19}$$

Similarly, we have

$$\left| \sum_{1 \leqslant i_1 < j_1 \leqslant i_2 < j_2 \leqslant n} J((i_1, i_2), (j_1, j_2)) \right| = O(n^2), \tag{2.20}$$

$$\left| \sum_{1 \leqslant i_1 < j_1 < j_2 < i_2 \leqslant n} J((i_1, i_2), (j_1, j_2)) \right| = O(n^2), \tag{2.21}$$

and

$$\left| \sum_{1 \leqslant i_1, j_1 \leqslant n} \sum_{i_2 = 1}^{n} J((i_1, i_2), (j_1, j_2)) \right| \leqslant Cn^2 \left( 1 + \sum_{r=1}^{n} \beta^{\delta/(2+\delta)}(r) \right) = O(n^2). \tag{2.22}$$

Hence from (2.19)–(2.22) and (2.16), we have (2.15) for $c = 2$. ▮

In the following, we shall investigate the asymptotic behavior of the two-dimensional random vector

$$(U_n(h_1, \mathbf{x}) - \theta_n(h_1),\ U_n(h_2, \mathbf{x}) - \theta_n(h_2)),$$

where $h_1$ and $h_2$ are two kernels satisfying the smoothness assumptions of Lemma 2.2. We would like to apply the Cramér–Wald device. So, let $c_1$, $c_2$ denote any two real numbers. Clearly,

$$c_1 U_n(h_1, \mathbf{x}) + c_2 U_n(h_2, \mathbf{x}) = U_n(c_1 h_1 + c_2 h_2, \mathbf{x}) \equiv U_n(h, \mathbf{x}),$$

where $h = c_1 h_1 + c_2 h_2$ and Lemma 2.2 applies. Specification of $\sigma^2(h)$ immediately leads to the following.

LEMMA 2.3. *Under the stated assumptions*

$$(na_n^p)^{1/2} [U_n(h_1, \mathbf{x}) - \theta_n(h_1),\ U_n(h_2, \mathbf{x}) - \theta_n(h_2)] \to \mathcal{N}(0, \Sigma)$$

*in distribution, with*

$$\Sigma = \begin{bmatrix} \sigma^2(h_1, h_1) & \sigma^2(h_1, h_2) \\ \sigma^2(h_1, h_2) & \sigma^2(h_2, h_2) \end{bmatrix}$$

*and where for two functions $h_1$ and $h_2$*

$$\sigma^2(h_1, h_2) = \lim_{n \to \infty} a_n^p E\{(g_1((X_1, Y_1); h_1) - \theta_n(h_1))(g_1((X_1, Y_1); h_2) - \theta_n(h_2))\}.$$

From this lemma, we will deduce the limit distribution of $U_n(\mathbf{x})$.

THEOREM 2.1. *Under the assumptions of Lemma* 2.1, *we have*

$$(na_n^p)^{1/2}\,(u_n(\mathbf{x})-\theta_n)\to\mathcal{N}(0,\rho^2)\qquad \textit{in distribution},$$

*where*

$$\rho^2=k\{\sigma^2(h,h)-2m(\mathbf{x})\,\sigma^2(h,1)+m^2(\mathbf{x})\,\sigma^2(1,1)\}.$$

*Proof.* We have

$$u_n(\mathbf{x})=U_n(h,\mathbf{x})/U(1,\mathbf{x}).$$

Define $g(x_1,x_2)=x_1/x_2$ for $x_2\neq 0$. Then

$$D=\left(\frac{\partial g}{\partial x_1},\frac{\partial g}{\partial x_2}\right)=(x_2^{-1},\ -x_1x_2^{-2}).$$

Since $\theta_n(h,\mathbf{x})\to m(\mathbf{x})$ and $\theta_n(1,\mathbf{x})=1$, we may infer from Lemma 2.3 that

$$(na_n^p)^{1/2}\,(u_n(\mathbf{x})-\theta_n(h,\mathbf{x}))\to\mathcal{N}(0,\rho^2)\qquad \text{in distribution},$$

where

$$\rho^2=(1,\ -m(\mathbf{x}))\,\Sigma\begin{pmatrix}1\\ -m(\mathbf{x})\end{pmatrix}$$

and

$$\Sigma=\begin{pmatrix}\sigma^2(h,h) & \sigma^2(h,1)\\ \sigma^2(h,1) & \sigma^2(1,1)\end{pmatrix}. \quad \blacksquare$$

Under appropriate smoothness assumptions on the marginal density $f$, Theorem 2.1 immediately yields asymptotic normality of $u_n(\mathbf{x})-m(\mathbf{x})$. Now assume that

$$f \text{ is twice differentiable in neighborhoods of } x_j,\qquad 1\leqslant j\leqslant k, \tag{2.23}$$

and

$$K \text{ is symmetric at zero;} \tag{2.24}$$

$m$ admits an expansion

$$m(\mathbf{y}+\boldsymbol{\Delta})=m(\mathbf{y})+\{m'(\mathbf{y})^t\,\boldsymbol{\Delta}\}+\tfrac{1}{2}\,\boldsymbol{\Delta}^t\{m''(\mathbf{y})\,\boldsymbol{\Delta}+o(\boldsymbol{\Delta}^t\boldsymbol{\Delta})\} \tag{2.25}$$

as $\boldsymbol{\Delta}\to 0$, for all $\mathbf{y}$ in the neighborhood of $\mathbf{x}$. Then we have the following.

COROLLARY 2.1. *If, in addition to conditions of Lemma* 2.1, (2.23)–(2.25) *hold, then*

$$(na_n^p)^{1/2}\,(u_n(\mathbf{x})-m(\mathbf{x}))\to\mathcal{N}(0,\rho^2)\qquad\text{in distribution,}$$

*provided that* $na_n^{5p}\to 0$.

*Proof.* See Corollary 2.4 of Stute [10].

## 3. CONSISTENCY

As for the regression estimators, we need to develop consistency results. We provide them here for the dependent case similar to the results established by Stute [10] for the independent case. Our conditions on the $U$-kernel $h$ are not so restrictive as the conditions of Stute [10] on his Theorems 2 and 3.

THEOREM 3.1. *Under the conditions of Lemma* 2.1, *we have for* $\mu_1\otimes\cdots\otimes\mu_k$, *for almost all* $\mathbf{x}$

$$u_n(\mathbf{x})\to m(\mathbf{x})\qquad\text{in probability,}$$

*where* $\mu$ *is the probability measure defined by the d.f.* $F$.

*Proof.* We know that almost surely

$$\theta_n(\mathbf{x})\to m(\mathbf{x}).\tag{3.1}$$

We also know that

$$u_n(\mathbf{x})=U_n(h,\mathbf{x})/U_n(1,\mathbf{x}).$$

So, we have to show that

$$U_n(h,\mathbf{x})\to m(\mathbf{x}),\qquad U_n(1,\mathbf{x})\to 1\quad\text{in probability.}$$

Since $U_n(1,\mathbf{x})$ is a special case of $U_n(h,\mathbf{x})$, we have only to deal with $U_n(h,\mathbf{x})$.

From the decomposition (2.7) and (3.1), we have only to prove that

$$\sum_{c=1}^{k}\binom{k}{c}U_n^{(c)}\to 0\qquad\text{in probability.}$$

But this is a consequence of (2.15). Theorem 3.1 follows.

THEOREM 3.2. *In addition to the conditions of Theorem* 3.1, *assume that*

$$\sum_{n=1}^{\infty} n^{1-\gamma} \exp(-na_n^p) < \infty \qquad \textit{for some} \quad 0 < \gamma < 1 \tag{3.2}$$

*and suppose that h is bounded. Then, for almost all* $\mathbf{x}$

$$u_n(\mathbf{x}) \to u(\mathbf{x}) \qquad \textit{with probability } 1.$$

*Proof.* From (2.15), we have

$$E(U_n - kU_n^{(1)})^2 = O(n^{-2}).$$

Then, from the Borel–Cantelli lemma, it suffices to show that

$$U_n^{(1)} \to 0 \qquad \text{with probability 1.} \tag{3.3}$$

Clearly,

$$U_n^{(1)} = n^{-1} \sum_{i=1}^{n} \{T_{i,n} - E(T_{i,n})\},$$

where

$$T_{i,n} = \int h(Y_i, y_2, ..., y_k) \prod_{j=2}^{k} K\left(\frac{x_j - z_j}{a_n}\right) \times K\left(\frac{x_1 - X_i}{a_n}\right) \Big/ \prod_{j=1}^{k} EK\left(\frac{x_j - X_1}{a_n}\right) \prod_{j=2}^{k} \tilde{G}(z_j; dy_j)\, F(dz_j).$$

We note that there exist two positive constants $b$ and $c$ such that

$$|T_{i,n}| \leqslant b/a_n^p$$
$$E(T_{i,n}^2) \leqslant c/a_n^p.$$

If $U_1, U_2, ..., U_n$ are independent random variables with $|U_i| \leqslant m$, $E(U_i) = 0$, and $E(U_i^2) \leqslant \sigma_i^2$, then an inequality due to Bennett [1, p. 39] states that

$$P\left[\left|n^{-1} \sum_{i=1}^{n} U_i\right| \geqslant \varepsilon\right] \leqslant 2 \exp\{-n\varepsilon^2/2(\sigma^2 + m\varepsilon)\},$$

where $\sigma^2 = n^{-1} \sum_{i=1}^{n} \sigma_i^2$. Put $q = q_n = [n^{1-\gamma}] + 1$ and write

$$U_n^{(1)} = \sum_{j=1}^{q} V_{n,j},$$

where

$$V_{n,j} = \sum_{p=0}^{l_j} \{T_{j+pq,n} - E(T_{j+pq,n})\}$$

and $l_j$ is the largest integer such that $j + l_j q \leqslant n$. Then

$$P[|U_n^{(1)}| \geqslant \varepsilon] \leqslant P\left[n^{-1} \sum_{j=1}^{q} |V_{n,j}| \geqslant \varepsilon\right] \leqslant \sum_{j=1}^{q} P[|V_{n,j}| \geqslant \varepsilon n]. \quad (3.4)$$

For any $j$, $1 \leqslant j \leqslant q$, define

$$B_j = \left\{(y_1, ..., y_{l_j}) \left| \sum_{p=1}^{l_j} y_p \right| \geqslant \varepsilon n q^{-1}\right\}$$

and put

$$g(y_1, ..., y_{l_j}) = \begin{cases} 1 & \text{if} \quad (y_1, ..., y_{l_j}) \in B_j \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 5.1,

$$\begin{aligned} P[|V_{n,j}| \geqslant \varepsilon n] &= Eg(T_{j,n} - E(T_{j,n}), ..., T_{j+l_j q,n} - E(T_{j+l_j q,n})) \\ &\leqslant P\left[\left|\sum_{p=0}^{l_j} \tilde{T}_{j+pq,n} - E(\tilde{T}_{j+pq,n})\right| \geqslant n\varepsilon q^{-1}\right] + 2l_j\beta(q) \\ &\leqslant 2\exp\left\{\frac{-n\varepsilon^2 a_n^p}{2c + 4b\varepsilon}\right\} + 2l_j\beta(q) \\ &\leqslant 2\exp\{-\alpha n a_n^p\} + 2l_j\beta(q) \qquad \left(\alpha = \frac{\varepsilon^2}{(2c + 4b\varepsilon)}\right), \end{aligned} \quad (3.5)$$

where $\tilde{T}_{1,n}, ..., \tilde{T}_{n,n}$ are independent and $\tilde{T}_{i,n}$ is distributed as $T_{i,n}$.
From (3.4) and (3.5), we deduce

$$P[|U_n^{(1)}| \geqslant \varepsilon] \leqslant 2n^{1-\gamma}\{\exp\{-\alpha n a_n^p\} + n^\gamma \beta([n^{1-\gamma}] + 1)\}.$$

From the Borel–Cantelli lemma and conditions (1.2) and (3.2), we deduce (3.3) and Theorem 3.2 is proved.

## 4. APPLICATIONS

Consider the model

$$Y_n = \psi(X_n) + \epsilon_n, \qquad n \geqslant 1, \quad (4.1)$$

where $X_n$ denotes a $\mathbb{R}^p$-vector of observed values, $\psi$ is measurable known function, $\epsilon_n$ is a multivariate white noise corresponding to the measurement errors (that is, $\{\epsilon_n, n \in \mathbb{N}\}$ is a sequence of i.i.d. random $\mathbb{R}^m$-vectors with strictly positive density) and $Y_n$ is an $\mathbb{R}^m$ predictor vector. If the sequence $(X_n)_{n \geqslant 1}$ of random vectors is absolutely regular with a geometrically rate, $E(|\psi(X_n)|^{2+\delta}) < +\infty$ and condition (ii) of Lemma 2.1 is satisfied. Thus we can apply Theorems 2.1, 3.1, and 3.2 for appropriate functions $h$ and $K$ and appropriate sequence $a_n$.

It is well known that any Markov process which is Harris recurrent, aperiodic, and geometrically aperiodic is absolutely regular with a geometrical rate.

For example, consider the model

$$X_n + \sum_{j=1}^{p_1} A_j X_{n-j} = e_n + \sum_{j=1}^{p_2} B_l e_{n-l}, \qquad n \in \mathbb{Z}, \tag{4.2}$$

where $A_1, ..., A_{p_1}$ and $B_1, ..., B_{p_2}$ are $p \times p$ real matrices, $A_{p_1}$ and $B_{p2}$ are invertible and $e_n = (e_{n1}, ..., e_{np})$ is a multivariate white noise, where each $e_{nj}$, $n \geqslant 1$, $1 \leqslant j \leqslant p$, admits the same density $g$ such that $\int |x|^\delta g(x)\, dx < \infty$ and $\int |g(x) - g(x-\theta)|\, dx = O(|\theta|^\gamma)$ for some $\delta > 0$ and $\gamma > 0$.

From Pham and Tran [5], $X_n$ admits a Markovian representation

$$X_n = HZ_n, \qquad Z_n = FZ_{n-1} + Ge_n,$$

where $Z_n$ is a sequence of random vectors and $H$, $F$, $G$ are appropriate matrices. If the eigenvalues of the matrices $H$ have a modulus less than 1, then $X_n$ is absolutely regular with a geometrical rate.

If $p=1$, $m=1$, and $k=2$, the example of Stute [10] can be applied to the particular model

$$Y_n = aX_n + \epsilon_n, \qquad a \in \mathbb{R}, \tag{4.3}$$

where $X_n$ is an ARMA process defined by

$$X_n = bX_{n-1} + e_n, \qquad \text{where} \quad |b| < 1. \tag{4.4}$$

Example 4.1. Put $h(y_1, y_2) = y_1 y_2$. Then

$$m(x_1, x_2) = E(Y_1 \mid X_1 = x_1)\, E(Y_2 \mid X_2 = x_2)$$

$$= a^2 x_1 x_2.$$

When $x_1 = x_2$, the variance $\rho^2$ defined in Theorem 2.1 yields

$$\rho^2 = 4\,\mathrm{Var}(Y_1 \mid X_1 = x_1)\, a^2 x_1^2 \int K^2(u)\, du/f(x_1)$$

$$= 4\sigma^2 a^2 x_1^2 \int K^2(u)\, du/f(x_1)$$

while for $x_1 \neq x_2$, we get

$$\rho^2 = [\mathrm{Var}(Y_1 \mid X_1 = x_1)\, a^2 x_1^2/f(x_1) + \mathrm{Var}(Y_1 \mid X_1 = x_2)\, a^2 x_2^2/f(x_2)] \times \int K^2(u)\, du$$

$$= \sigma^2 \int K^2(u)\, du[a^2 x_1^2/f(x_1) + a^2 x_2^2/f(x_2)].$$

Example 4.2. Suppose $E(\varepsilon_1^4) < +\infty$. For $h(y_1, y_2) = \frac{1}{2}(y_1 - y_2)^2$, we obtain $m(x_1, x_1) = \mathrm{Var}(Y_1 \mid X_1 = x_1) = \sigma^2$. In this case

$$\rho^2 = \{E[(Y - ax_1)^4 \mid X = x_1] - \mathrm{Var}(Y \mid X = x_1)\} \int K^2(u)\, du/f(x_1)$$

$$= (\tau_4 - \sigma^4) \int K^2(u)\, du/f(x_1),$$

where $\tau_4 = E(\varepsilon_1^4)$.

We have seen how the examples of Stute [10] can be applied now to more general models, but it is obvious from (4.1) and (4.2) that we have the possibility of using our results for a much larger set of models and applications when $p > 1$ and $m > 1$.

## Appendix

Lemma 5.1 (Davydov [2]). *Let $\{X_{ni}, 1 \leqslant i \leqslant n, n \geqslant 1\}$ be a nonstationary sequence of r.v.'s which is strong mixing. Let $Z$ be $\sigma(X_{ni}, 1 \leqslant i \leqslant j)$-measurable $(1 \leqslant j \leqslant n)$ and let $V$ be $\sigma(X_{ni}, i \geqslant j + m)$-measurable. If $E(|Z|^p) < \infty$, $E(|V|^q) < \infty$, and $r^{-1} + p^{-1} + q^{-1} = 1$ $(r, p, q > 0)$ then*

$$|E(ZV) - E(Z)\, E(V)| \leqslant (\alpha(m))^{1/r} \{E\,|Z|^p\}^{1/p} \{E\,|V|^q\}^{1/q},$$

*where $C$ is some constant $> 0$, and of course if the sequence is absolutely regular, the inequality (5.1) holds if we replace $(\alpha(m))^{1/r}$ by $(\beta(m))^{1/r}$.*

LEMMA 5.2 (Theorem 1 of Yokoyama [12]). *Let* $\{X_{ni}, 1 \leqslant i \leqslant n, n \geqslant 1\}$ *be a nonstationary sequence of r.v.'s which is strong mixing with* $E(X_{ni}) = 0$, $1 \leqslant i \leqslant n$, $n \geqslant 1$, *and* $\sup_{1 \leqslant i \leqslant n} E\,|X_{ni}|^{r+\delta} < C_n$ *for some* $r > 2$ *and* $\delta > 0$. *If*

$$\sum_{i=0}^{\infty} (i+1)^{r/2} [\alpha(i)]^{\alpha/(r+\delta)} < \infty$$

*then*

$$E\left|\sum_{i=1}^{n} X_{ni}\right|^r \leqslant C_n n^{r/2}, \qquad n \geqslant 1.$$

## REFERENCES


[1] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **19** 33–45.

[2] DAVYDOV, YU. A. (1970). Invariance principle for empirical stationary processes. *Theory Probab. Appl.* **14** 487–498.

[3] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

[4] NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

[5] PHAM, T. D., AND TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process Appl.* **19** 297–303.

[6] ROSENBLATT, M. (1969). Conditional probability density and regression estimates. In *Multivariate Analysis II* (Krishnaiah, Ed.), pp. 25–31. Academic Press, New York.

[7] SCHUSTER, E. F. (1972). Joint asymptotic distribution of the estimate regression function at a finite number of distinct points. *Ann. Math. Statist.* **43** 84–88.

[8] STUTE W. (1984). Asymptotic normality on nearest neighbor regression function estimates. *Ann. Statist.* **12** 917–918.

[9] STUTE W. (1986). On almost sure convergence of conditional empirical distribution functions, *An. Probab.* **14** 891–901.

[10] STUTE W. (1991). Conditional *U*-statistics. *Ann. Probab.* **19**, No. 2, 812–825.

[11] WATSON, G. S. (1964). Smooth regression analysis. *Snakhyā Ser. A* **26** 359–372.

[12] YOKOYAMA, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch. Verw. Geb.* **52** 45–57.

[13] YOSHIHARA, K. (1988). Asymptotic normality of nearest neighbor regression function estimates based on some dependent observations. *Yokohama Math. J.* **36** 55–68.