Journal of Applied Mathematics and Stochastic Analysis, **10**:1 (1997), 3-20.

LIMITING BEHAVIOR OF THE PERTURBED EMPIRICAL DISTRIBUTION FUNCTIONS EVALUATED AT U-STATISTICS FOR STRONGLY MIXING SEQUENCES OF RANDOM VARIABLES

SHAN SUN
Texas Tech University
Department of Mathematics
Lubbock, TX 79409 USA

CHING-YUAN CHIANG
James Madison University
Department of Mathematics
Harrisonburg, VA 22807 USA

(Received November, 1994; Revised April, 1996)

We prove the almost sure representation, a law of the iterated logarithm and an invariance principle for the statistic $\widehat{F}_n(U_n)$ for a class of strongly mixing sequences of random variables $\{X_i, i \geq 1\}$. Stationarity is not assumed. Here $\widehat{F}_n$ is the perturbed empirical distribution function and U_n is a U-statistic based on $X_1, \ldots, X_n$.

Key words: Perturbed Empirical Distribution Functions, Strong Mixing, Almost Sure Representation, U-statistic, Law of the Iterated Logarithm, Invariance Principle.

AMS subject classifications: 62E20, 60F15.

1. Introduction

Let $\{X_i, i \geq 1\}$ be a sequence of nonstationary random variables with c.d.f.'s $\{F_i, i \geq 1\}$ defined on the real line $\mathbb{R}$ and assume $F_i \rightarrow F$ as $i \rightarrow \infty$ for some fixed distribution function F. Also, let $\widetilde{F}_n(x)$ be the corresponding empirical distribution function based on $X_1, \ldots, X_n$, that is, $\widetilde{F}_n(x) = n^{-1}\sum_{i=1}^n u(x - X_i)$, where

$$u(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Consider the sequence of perturbed empirical distribution functions given by

$$\widehat{F}_n(x) = n^{-1}\sum_{i=1}^n K_n(x - X_i), \quad n \geq 1, \quad x \in \mathbb{R},$$

where $\{K_n\}$ is a sequence of continuous c.d.f.'s converging weakly to the c.d.f. with unit mass at zero. Such an $\widehat{F}_n$ can be expressed as the integral (or c.d.f.)

$$\widehat{F}_n(x) = \int_{-\infty}^{x} \widehat{f}_n(t)dt, \quad x \in \mathbb{R},$$

of a kernel density estimator $\widehat{f}_n$ of the type

$$\widehat{f}_n(x) = n^{-1}\sum_{i=1}^{n} a_n^{-1} k\left(\frac{x - X_i}{a_n}\right)$$

suggested by Rosenblatt [16] and [12], where k is a probability density function and $\{a_n\}$ is a sequence of positive real constants tending to the limit zero. The study of the asymptotic properties of $\widehat{F}_n$ was first elucidated by [11]. For related investigations in this direction we refer to [18, 19, 20] and [22].

Let $g(x_1,\ldots,x_m)$, symmetric in its arguments, be a measurable kernel (of degree m), and let U_n be the corresponding U-statistics given by

$$U_n = \binom{n}{m}^{-1} \sum_{C_{n,m}} g(X_{i_1},\ldots,X_{i_m}) \tag{1.1}$$

where $C_{n,m}$ denotes the set of all $\binom{n}{m}$ combinations of m distinct elements $(i_1,\ldots,i_m)$ from $(1,\ldots,n)$.

Now consider the perturbed empirical distribution $\widehat{F}_n$ evaluated at U_n, which is quite useful; e.g., in the estimation of $F(\xi)$ when F is unknown and $\xi = \int_{\mathbb{R}^m} g(x_1,\ldots,x_m)\prod_{i=1}^{m} dF(x_i)$. As with many problems in probability and statistics, the study of the asymptotic behavior of $\widehat{F}_n(U_n)$ has previously been conducted mainly in the i.i.d. case. Thus, under the i.i.d. set-up, [14] proved the asymptotic normality of $\widehat{F}_n(U_n)$, and [10] established the almost sure representation, a law of iterated logarithm and an invariance principle for $\widehat{F}_n(U_n)$. In recent years, however, there has been much interest in the cases of dependence in probability and statistics in general and mixing conditions in particular. The latter represent degrees of weak dependence in the sense of asymptotic independence of past and distant future ([6] and [5]). In this connection, recently [17] has proved the asymptotic normality of $\widehat{F}_n(U_n)$ for the case where X_i's form an absolutely regular stationary sequence.

In this paper we study the asymptotic behavior of $\widehat{F}_n(U_n)$ in the case where the sequence $\{X_i, i \geq 1\}$ is strong mixing, which is more general then the absolutely regular case and is about the weakest mixing condition (see, e.g., [3] and [5]). Moreover, we do not assume stationarity. Specifically, we give the almost sure representation, a law of the iterated logarithm and an invariance principle for $\widehat{F}_n(U_n)$. Thus the results obtained here extend or generalize those of [14], [10] and [17].

We adopt the following notation and general assumptions. Let $\{X_n, n \geq 1\}$ be a sequence of random variables on some probability space $(\Omega, \mathcal{A}, P)$ satisfying the strong mixing condition. We will use the mixing coefficients $\alpha_n(k)$, defined by

$$\alpha_n(k) = \sup\{\,|\,P(A\cap B) - P(A)P(B)\,| : A \in \sigma(X_i: 1 \leq i \leq m), \tag{1.2}$$

$$B \in \sigma(X_i: m+k \leq i \leq n, 1 \leq m \leq n-k)\}, \quad k \leq n-1$$

$\alpha_n(k) = 0$ for $k \geq n$.

The coefficient of strong mixing introduced by [15] then can be written as

$$\alpha(k)=\sup_{n\in\mathbb{N}}\alpha_n(k) \text{ for } k\in\mathbb{N}. \tag{1.3}$$

2. Basic Lemmas

Let $p>1$ and $1\le i_1, i_2<\ldots<i_p$ be arbitrary integers. For any j $(1\le j\le p-1)$, we define

$$F_{i_1,i_2,\ldots,i_p}(x_1,x_2,\ldots,x_p)=P(X_{i_1}\le x_1, X_{i_2}\le x_2,\ldots,X_{i_p}\le x_p). \tag{2.1}$$

Lemma 2.1: *For any integer $p>1$ and integers $(i_1,\ldots,i_p)$ such that $1\le i_1<i_2<\ldots<i_p$, let g be a Borel function such that*

$$\int_{\mathbb{R}} |g(x_1,\ldots,x_p)|^{1+\delta}dF_{i_1,i_2,\ldots,i_j}(x_1,\ldots,x_j)dF_{i_{j+1},\ldots,i_p}(x_{j+1},\ldots,x_p)\le M$$

for some $\delta>0$ and some $M>0$. Then we have

$$\begin{aligned}&\Big|\int_{\mathbb{R}^p} g(x_1,\ldots,x_p)dF_{i_1,i_2,\ldots,i_p}(x_1,\ldots,x_p)\\ &-\int_{\mathbb{R}^p} g(x_1,\ldots,x_p)dF_{i_1,\ldots,i_j}(x_1,\ldots,x_j)dF_{i_{j+1},\ldots,i_p}(x_{j+1},\ldots,x_p)\Big|\\ &\le 4M^{\frac{1}{1+\delta}}[\alpha(i_{j+1}-i_j)]^{\frac{\delta}{1+\delta}}.\end{aligned} \tag{2.2}$$

Moreover, if g is bounded, say $|g(x_1,\ldots,x_p)|\le M^*$, then we can replace the right-hand side of (2.2) *by* $2M^*\alpha(i_{j+1}-i_j)$.

Proof: Let $P_0^{(p)}=F_{i_1,\ldots,i_p}(x_1,\ldots,x_p)$ and

$$P_j^{(p)}=F_{i_1,i_2,\ldots,i_j}(x_1,\ldots,x_j)F_{i_{j+1},\ldots,i_p}(x_{j+1},\ldots,x_p). \tag{2.3}$$

For fixed j $(1\le j<p)$, put

$$A=\{(x_1,\ldots,x_p): |g(x_1,\ldots,x_p)|\le M^{-\beta}[\alpha(d)]^{-\beta}\}$$

where $d=i_{j+1}-i_j$ and $\beta=\frac{1}{1+\delta}$. Then, it follows from the definition of strong mixing that

$$\begin{aligned}&\left|\int\underset{A}{\cdots}\int g(x_1,\ldots,x_p)dP_0^{(p)}-\int\underset{A}{\cdots}\int g(x_1,\ldots,x_p)dP_j^{(p)}\right|\\ &\le 2M^{\beta}[\alpha(d)]^{1-\beta}=2M^{\beta}[\alpha(d)]^{\delta\beta}.\end{aligned} \tag{2.4}$$

Let A^c be the complement of A. Then we have

$$\begin{aligned}&\left|\int\underset{A^c}{\cdots}\int g(x_1,\ldots,x_p)dP_i^{(p)}\right|\le M^{-\beta\delta}[\alpha(d)]^{\beta\delta}\int\underset{A^c}{\cdots}\int |g(x_1,\ldots,x_p)|^{1+\delta}dP_i^{(p)}\\ &\le M^{1-\beta\delta}[\alpha(d)]^{\beta\delta}=M^{\beta}[\alpha(d)]^{\beta\delta},\ (i=0,1).\end{aligned} \tag{2.5}$$

Combining (2.4) and (2.5), we obtain Lemma 2.1.

We shall take $\alpha(n) = \rho^n$, $0 < \rho < 1$, in the remainder of this paper.

The following lemma gives the Bernstein-type of inequality for the rv's satisfying the strong mixing condition.

Lemma 2.2: *Let $\{Y_i, i \geq 1\}$ be a sequence of strong mixing random variables with mean 0. Assume that $\sup_{i \geq 1} |Y_i| \leq M_0$ and $\sup_{i \geq 1} |\operatorname{Var} Y_i|^{\frac{1}{2}} \leq M$. Then for any $\lambda > 0$ and $m \leq n$,*

$$P\left[n^{-1}\left|\sum_{i=1}^{n} Y_i\right| \geq \lambda\right] \leq 2m \exp\left(-\frac{\lambda^2 n^2}{4mnM^2 + \frac{2}{3}M_0\lambda mn}\right) + 2n\alpha(m). \tag{2.6}$$

Proof: Let $\{Y_i^*, i \geq 1\}$ be an independent copy of Y_i's with common $df\ F$. That is, $\{Y_i^*,\ i \geq 1\}$ is a sequence of i.i.d. rv's with common $df\ F$. Then, from the Berstein inequality (see [2]), we have for any $m \in \mathbb{Z}^+$

$$P\left(\left|\sum_{i=1}^{m} \tilde{Y}_i\right| \geq m^{-1}x\right) \leq 2\exp\left(-\frac{x^2}{2m^3M^2 + \frac{2}{3}M_0 xm}\right), \quad \forall x > 0. \tag{2.7}$$

For $j = 1, 2, \ldots, m$, let

$$S^{(j)} = \sum_{i=1}^{k_j m} Y_{j+i}$$

where $k_j = k_{j,n}$ is the largest integer for which $j + k_j m \leq n$. Let

$$h_j(y_1, \ldots, y_{j+1}) = \begin{cases} 1, & \text{if} \left|\sum_{i=1}^{j+1} y_i\right| \geq m^{-1}x \\ 0, & \text{otherwise.} \end{cases}$$

First, we prove by induction that for $\ell = 1, 2, \ldots$

$$E(h_\ell(Y_j, Y_{j+m}, \ldots, Y_{j+\ell m})) \leq \int_{\mathbb{R}^{\ell+1}} h_\ell(y_1, \ldots, y_{\ell+1}) \prod_{i=1}^{\ell+1} dF(y_i) + 2\ell\alpha(m). \tag{2.8}$$

Clearly, (2.8) is true by Lemma 2.1 when $\ell = 1$. Assuming (2.8) is true for ℓ, then applying Lemma 2.1 again, we have

$$\begin{aligned} &Eh_{\ell+1}(Y_j, Y_{j+m}, \ldots, Y_{j+(\ell+1)m}) \\ &\leq E\left[\int_{\mathbb{R}^{\ell+1}} h_{\ell+1}(y_1, \ldots, y_{\ell+1}, Y_{j+(\ell+1)m}) dF_{j,\ldots,j+\ell m}(y_1, \ldots, y_{\ell+1})\right] \\ &\qquad + 2\alpha(m) \\ &\leq 2\alpha(m) + E\left[2\ell\alpha(m) + \int_{\mathbb{R}^{\ell+1}} h_{\ell+1}(y_1, \ldots, y_{\ell+1}, Y_{j+(\ell+1)m}) \prod_{i=1}^{\ell+1} dF(y_i)\right]. \end{aligned} \tag{2.9}$$

This proves (2.8). Therefore (2.7) and (2.8) imply

$$\begin{aligned} P(|S^{(j)}| \geq m^{-1}x) &\leq P\left(\left|\sum_{i=0}^{k_j} Y_{j+im}^*\right| \geq m^{-1}x\right) + 2k_j\alpha(m) \\ &\leq 2\exp\left(-\frac{x^2}{2m^2M^2(k_j+1) + \frac{2}{3}M_0 xm}\right) + 2k_j\alpha(m). \end{aligned} \tag{2.10}$$

Moreover, for $k_j + 1 \le \frac{n+1}{m}$, we obtain

$$P\left(\left|\sum_{i=1}^{n} Y_i\right| \ge x\right) \le \sum_{j=1}^{m} P\left(\mid S^{(j)} \mid \ge m^{-1}x\right). \tag{2.11}$$

Therefore, (2.6) holds by using (2.10) and replacing x by $n\lambda$ in (2.11).

Next, we provide the following lemma as a generalization of a result of [1].

Lemma 2.3: *For some $\delta > 0$, if $\{F'_i, i \ge 1\}$ is uniformly bounded in $(\xi - \delta, \xi + \delta)$, let $A_n^* = \sup_{\mid t - \xi \mid < \delta/2} A_n(t)$, where*

$$A_n(t) = \sup_{\mid x - t \mid \le d_n} \mid \widetilde{F}_n(x) - \widetilde{F}_n(t) - \bar{F}_n(x) + \bar{F}_n(t) \mid, \tag{2.12}$$

$d_n \sim dn^{-\frac{1}{2}}(\log\ \log n)^{\frac{1}{2}}$ as $n \to \infty$, for some constant d, $\bar{F}_n = \frac{1}{n}\sum_{i=1}^{n} F_i$ and F'_i denotes the density of F_i. Then, for any $\eta \in (0, \frac{1}{8})$, we have

$$\lim_{n \to \infty} n^{-\frac{3}{4} + \eta} A_n^* = 0 \ a.s. \tag{2.13}$$

Proof: For $t \in [\xi - \frac{\delta}{2}, \xi + \frac{\delta}{2}]$, define

$$B_n(x) = \widetilde{F}_n(x) - \widetilde{F}_n(t) - \bar{F}_n(x) + \bar{F}_n(t), \quad x \in \mathbb{R}, \tag{2.14}$$

and let $\{c_n, n \ge 1\}$ be a sequence of positive integers such that $c_n \sim n^{\frac{1}{4}}$ as $n \to \infty$. For a fixed n, we partition the interval $[t - d_n, t + d_n]$ by the points $t_{i,n} = t + i\frac{d_n}{c_n}$, $i = 0, \pm 1, \ldots, \pm c_n$ and let

$$q_{i,n} = \bar{F}_n(t_{i+1,n}) - \bar{F}_n(t_{i,n}).$$

Since $\widetilde{F}_n$ and $\bar{F}_n$ are nondecreasing, it follows from (2.14) that we have for $x \in [t_{i,n}, t_{i+1,n}]$

$$\begin{aligned} B_n(x) &\le \widetilde{F}_n(t_{i+1,n}) - \widetilde{F}_n(t) - \bar{F}_n(t_{i,n}) + \bar{F}_n(t) \\ &= B_n(t_{i+1,n}) + q_{i,n} \end{aligned} \tag{2.15}$$

and

$$B_n(x) \ge B_n(t_{i,n}) - q_{i,n}. \tag{2.16}$$

From (2.12), (2.15) and (2.16), we have

$$A_n(t) \le T_n + V_n \tag{2.17}$$

where

$$T_n = \max_{i = 0, \pm i, \ldots, \pm c_n} \mid B_n(t_{i,n}) \mid, \qquad V_n = \max_{i = 0, \pm 1, \ldots, \pm c_n} q_{i,n}. \tag{2.18}$$

Let $\gamma > 0$ be arbitrary; to prove (2.13), we will show that for $\eta \in (0, \frac{1}{8})$

$$P\left(n^{\frac{3}{4} - \eta} A_n^* \ge \gamma \text{ i.o.}\right) = 0. \tag{2.19}$$

Since $\{F'_i, i \ge 1\}$ is uniformly bounded in a neighborhood of t, there exist $C_0 > 0$ and N_0 such that for all $n \ge N_0$,

$$V_n \le C_0 n^{-\frac{3}{4}}(\log\log n)^{\frac{1}{2}}. \tag{2.20}$$

Moreover, we have

$$P\left(| B_n(t_{i,n}) | \geq \epsilon_n^{(\eta)} \right) = P\left(\left| \frac{1}{n}\sum_{j=1}^{n} Y_j \right| \geq \epsilon_n^{(\eta)} \right) \tag{2.21}$$

where

$$Y_j = \mathbb{I}_{\{(t_{i,n} \wedge t, t_{i,n} \vee t]\}}(X_j) - E\left((\mathbb{I}_{\{(t_{i,n} \wedge t, t_{i,n} \vee t]\}}(X_j) \right)$$

and $\epsilon_n^{(\eta)} = \gamma n^{-\frac{3}{4}+\eta}$.

Also, there exists $N_1 > N_0$ such that for $j = 1, 2, \ldots, n$ and $n \geq N_1$, $\operatorname{Var} Y_j \leq | \bar{F}_n(t_{i,n}) - \bar{F}_n(t) | \leq C_2 d_n$, where

$$C_2 = \sup_{x \in (\xi - \delta, \xi + \delta)} | F_i'(x) | .$$

Now, applying Lemma 2.2 with $m = [n^\eta]$, we obtain

$$P(| B_n(t_{i,n}) | \geq \epsilon_n^{(\eta)}) \leq 2n^\eta \exp(- \theta_n^{(\eta)}) + 2n\alpha([n^\eta]), \tag{2.22}$$

where

$$\theta_n^{(\eta)} = \frac{n^2(\epsilon_n^{(\eta)})^2}{4n^{1+\eta}C_2 d_n + \frac{2}{3}\epsilon_n^{(\eta)} n^{1+\eta}}$$

and $[x]$ is the integer part of x. From (2.18), (2.21) and (2.22) it follows that there exists $N_2 > N_1$ such that for $n \geq N_2$

$$\begin{aligned} P(T_n \geq \epsilon_n^{(\eta)}) &\leq (2c_n + 1)(2n^\eta \exp(- \theta_n^{(\eta)}) + 2n\alpha([n^\eta])) \\ &\leq n^{-(s+\frac{1}{2})}, \quad \text{for some } s \geq 1. \end{aligned} \tag{2.23}$$

Hence, there exists an N_3 such that for $n \geq N_3$

$$P(A_n(t) \geq 2\epsilon_n^{(\eta)}) \leq n^{-(\frac{1}{2}+s)}. \tag{2.24}$$

Now, let $\xi_{j,n} = \xi + \frac{j\delta}{2\sqrt{n}}$, $j = 0, \pm 1, \ldots, \pm [n^{\frac{1}{2}}]$, then from (2.24), for $n \geq N_3$ and some C_5, we obtain

$$P\left(\max_{-\sqrt{n} \leq j \leq \sqrt{n}} A_n(\xi_{j,n}) \geq 2\epsilon_n^{(\eta)} \right) \leq C_5 n^{\frac{1}{2}} n^{-(\frac{1}{2}+s)} = C_5 n^{-s}.$$

Using the triangular inequality, we obtain

$$A_n^* \leq 3 \max_{-\sqrt{n} \leq j \leq \sqrt{n}} A_n(t_{j,n}). \tag{2.25}$$

Combining (2.24) and (2.25), we have for some C_6,

$$\sum_{n=N_3}^{\infty} P(A_n^* \geq \gamma n^{-\frac{3}{4}+\eta}) \leq C_6 \sum_{n=N_3}^{\infty} n^{-s} < \infty. \tag{2.26}$$

This proves (2.19).

We introduce the following two sets of conditions:

(A) (i) $\{F_i', i \geq 1\}$ is uniformly bounded in the neighborhood of ξ.

(ii) $\int_{-\infty}^{\infty} | x | k(x)dx < \infty$.

(B) (i) $\{F_i, i \geq 1\}$ is twice differentiable with uniformly bounded $\{F_i'', i \geq 1\}$ in the neighborhood of ξ.

(*ii*) $\int_{-\infty}^{\infty} x^2 k(x)dx < \infty$ and there exists a $\gamma > 0$ such that $k(x) = k(-x)$, $|x| < \gamma$.

The almost sure representation theorem given in section 4 highly relies on the following two lemmas.

Lemma 2.4: *Suppose $\{F_i, i \geq 1\}$ and k satisfy (A) and $\alpha(n) = \rho^n$, for $0 < \rho < 1$. Then there exists an $\epsilon > 0$ such that*

$$\sup_{|x-\xi| < \epsilon} |\widehat{F}_n(x) - \widetilde{F}_n(x)| = o(n^{-\frac{3}{4}+\eta}) + O(n^{\frac{1}{2}} a_n (\log\log n)^{-\frac{1}{2}}) \ a.s. \quad (2.27)$$

If $\{F_i, i \geq 1\}$ and k satisfy (B), then (2.27) holds with $O(n^{\frac{1}{2}} a_n (\log\log n)^{-\frac{1}{2}})$ replaced by $O(na_n^2(\log\log n)^{-1})$.

Proof: Suppose $\{F_i, I \geq 1\}$ and k satisfy (A) and let $\delta > 0$ be such that $\{F'_i, i \geq 1\}$ is uniformly bounded on $(\xi - \delta, \xi + \delta)$, that is, there exists M_1 such that $|\bar{F}_n(x-t) - \bar{F}_n(x)| \leq M_1 |t|$ for $|t| \leq \frac{\delta}{2}$ and $|x - \xi| < \frac{\delta}{2}$. Then, for any x with $|x - \xi| < \delta/2$, we have

$$|\widehat{F}_n(x) - \widetilde{F}_n(x)| = \left| \int_{-\infty}^{\infty} [\widetilde{F}_n(x-t) - \widetilde{F}_n(x)] k_n(t) dt \right|$$

$$\leq \int_{|t| \leq d_n} |\widetilde{F}_n(x-t) - \widetilde{F}_n(x) - \bar{F}_n(x-t) + \bar{F}_n(x)| k_n(t) dt \quad (2.28)$$

$$+ \int_{|t| > d_n} k_n(t) dt + \int_{|t| \leq d_n} |\bar{F}_n(x-t) - \bar{F}_n(x)| k_n(t) dt,$$

where $k_n(t) = \frac{1}{a_n} k(\frac{t}{a_n})$.

For large n, the last term of (2.28) is bounded by

$$M_1 \int_{-\infty}^{\infty} |t| k_n(t) dt = O(a_n) \quad (2.29)$$

and the second term of (2.28) is bounded by

$$n^{\frac{1}{2}} a_n (\log\log n)^{-\frac{1}{2}} \int_{|t| > \frac{d_n}{a_n}} |t| k(t) dt = O\left(n^{\frac{1}{2}} a_n (\log\log n)^{-\frac{1}{2}} \right). \quad (2.30)$$

Therefore, (2.27) holds by applying Lemma 2.3, (2.29) and (2.30). For $\{F'_i, i \geq 1\}$ and k satisfying (B), using the Taylor expansion in the last term of (2.28), we obtain

$$\int_{|t| \leq d_n} |\bar{F}_n(x-t) - \bar{F}_n(x)| k_n(t) dt$$

$$= \int_{|t| \leq d_n} \left| \bar{F}'_n(x) a_n t + \frac{\bar{F}''_n(x_t)}{2} a_n^2 t^2 \right| k_n(t) dt \quad (2.31)$$

$$\leq a_n^2 M_2 \int_{-\infty}^{\infty} t^2 k(t) dt = O(a_n^2)$$

where x_t lies between $x-t$ and x and $\sup_{x\in(\xi-\delta,\xi+\delta)}|F_i''(x)|\leq M_2$.

Similar to (2.30), the second term of (2.28) can be bounded by

$$na_n^2(\log\log n)^{-1}\int_{|t|>\frac{d_n}{a_n}} t^2k(t)dt = O(na_n^2(\log\log n)^{-1}). \tag{2.32}$$

Combining (2.31) and (2.32), and applying Lemma 2.3, we complete the proof.

Lemma 2.5: *Suppose $\{F_i, i\geq 1\}$ and k satisfy (A) and $\alpha(n)=\rho^n$, for $0<\rho<1$. Then we have*

$$\sup_{|x-\xi|<d_n}|\widehat{F}_n(x)-\widehat{F}_n(\xi)-\overline{F}_n(x)+\overline{F}_n(\xi)| = o(n^{-\frac{3}{4}+\eta})+O(a_n)\ a.s. \tag{2.33}$$

If $\{F_i, i\geq 1\}$ and k satisfy (B), then (2.33) holds with $O(a_n)$ replaced by $O(a_n^2)$.

Proof: Let $\epsilon>0$; for $|x-\xi|\leq d_n$, we have

$$|\widehat{F}_n(x)-\widehat{F}_n(\xi)-\overline{F}_n(x)+\overline{F}_n(\xi)|$$

$$\leq \int_{|t|<\epsilon} |\widetilde{F}_n(x-t)-\widetilde{F}_n(\xi-t)-\overline{F}_n(x-t)+\overline{F}_n(\xi-t)|k_n(t)dt \tag{2.34}$$

$$+2\int_{|t|\geq\epsilon}k_n(t)dt+2\sup_{|x-\xi|\leq d_n}\int_{|t|<\epsilon}|\overline{F}_n(x-t)-\overline{F}_n(x)|k_n(t)dt.$$

The proof is completed by applying Lemma 2.3 and similar reasons as in (2.29)-(2.32).

Remark 2.1: It is true that under the following global conditions (i) $[F_i, i\geq 1\}$ is twice differentiable on the real line with bounded second derivative $\{F_i'', i\geq 1\}$ and (ii) $\int_{-\infty}^{\infty}xk(x)dx = 0$ and $\int_{-\infty}^{\infty}x^2k(x)dx<\infty$ (instead of local condition (B)), the second parts of Lemma 2.4 and Lemma 2.5 remain true.

Since we do not assume stationarity, we will need to introduce some new notations for the U-statistics to obtain the following lemma, which generalizes the result of [21].

For every $p(1\leq p\leq m)$ and $n\geq m$, let $1\leq i_1\neq\ldots\neq i_p\leq n$ be arbitrary integers and put

$$\mathcal{U}_{p,n}(x_1,\ldots,x_p)=\sum_{(i_{p+1},\ldots,i_m)\in I}\lambda(x_1,\ldots,x_p;i_{p+1},\ldots,i_m) \tag{2.35}$$

where

$$\lambda(x_1,\ldots,x_p;i_{p+1},\ldots,i_m)=\int_{\mathbb{R}^{m-p}}g(x_1,\ldots,x_m)\prod_{j=p+1}^{m}dF_{i_j}(x_j)$$

and

$$I=\{(i_{p+1},\ldots,i_m):\ 1\leq i_{p+1}\neq\ldots\neq i_m\leq n,\ \ i_j\notin(i_1,\ldots,i_p), p+1\leq j\leq m\}.$$

Also, denote

$$\mathcal{U}_0(x_1,\ldots,x_p)=\sum_{(i_1,\ldots,i_m)\in I_0}\int_{\mathbb{R}^m}g(x_1,\ldots,x_m)\prod_{j=1}^{m}dF_{i_j}(x_j) \tag{2.36}$$

where $I_0=\{(i_1,\ldots,i_m),\ 1\leq i_1\neq\ldots\neq i_m\leq n\}$. For every $p(1\leq p\leq m)$, set

$$U_n^{(p)} = n^{-[m]} \sum_{1 \le i_1 \ne \ldots \ne i_p \le n} \int_{\mathbb{R}^p} \mathfrak{U}_{p,n}(x_1,\ldots,x_p) \prod_{j=1}^{p} d\left(I_{[X_{i_j} \le x_j]} - F_{i_j}(x_j)\right), \tag{2.37}$$

where $n^{-[m]} = (n(n-1)\ldots(n-m+1))^{-1}$.

Lemma 2.6: *If there exists a positive number δ such that*

$$\sup_n \max_{(i_n,\ldots,i_m) \in C_{n,m}} \int_{\mathbb{R}^m} | g(x_1,\ldots,x_m) |^{4+\delta} \prod_{j=1}^{m} dF_{i_j}(x_j) \le M_0 < \infty \tag{2.38}$$

$$\sup_n \max_{(i_1,\ldots,i_m) \in C_{n,m}} E(| g(X_{i_1},\ldots,X_{i_m}) |^{4+\delta}) \le M_0 < \infty, \tag{2.39}$$

and for some δ' $(0 < \delta' < \delta)$, $\alpha(n) = O(n^{-3(4+\delta')/(2+\delta')})$, then we have

$$E(U_n^{(2)})^4 = O(n^{-3-\gamma}) \tag{2.40}$$

where $\gamma = 6(\delta - \delta')/(4+\delta)(2+\delta') > 0$ and

$$E(U_n^{(p)})^2 = O(n^{-3}), \ \ 3 \le p \le m. \tag{2.41}$$

Proof: We proceed as in Lemma 3 of [21]. Since

$$E(U_n^{(2)})^4 = (n^{-[m]})^4 \sum_{j=1}^{4} \sum_{1 \le i_{j1}, i_{j2} \le n} J((I_{11},i_{12}),(i_{21},i_{22}),(i_{31},i_{32}),(i_{41},i_{44})) \tag{2.42}$$

where

$$J((i_{11},i_{12}),(i_{21},i_{22}),(i_{31},i_{32}),(i_{41},i_{44})) \tag{2.43}$$

$$= E\left\{\prod_{j=1}^{4} \int_{\mathbb{R}^2} \mathfrak{U}_{2,n}(x_{j1},x_{j2}) d\left(I_{[X_{i_{j1}} \le x_{j1}]} - F_{i_{j1}}(x_{j1})\right) d\left(I_{[X_{i_{j2}} \le x_{j2}]} - F_{i_{j2}}(x_{j2})\right)\right\}$$

where $\mathfrak{U}_{2,n}(x_{j1},x_{j2})$ is defined as in (2.35).

Now, let $i_{\gamma k}$ $(\le n)$ $(\gamma = 1,2,3,4, k = 1,2)$ be mutually different and reorder $i_{\gamma k}$ as $1 \le k_1 < k_2 < \ldots < k_8 \le n$, then, we have

$$J((i_{11},i_{12}),(i_{21},i_{22}),(i_{31},i_{32}),(i_{41},i_{42})) = E[g(x_{k1},\ldots,X_{k8})]. \tag{2.44}$$

Since

$$\int_{\mathbb{R}^8} g(x_1,\ldots,x_8) dP_j^{(8)} = 0, \ \ j = 1,7$$

where $P_j^{(p)}$ is defined as in (2.3). Applying Lemma 2.1 we obtain

$$E[g(X_{k1},\ldots,X_{k8})] \le \left(\frac{n^{[m]}}{n(n-1)}\right)^4 M_0[\alpha(k_8 - k_7)]^{\frac{2+\delta}{4+\delta}}, \ \text{if } k_8 - k_7 = d^{(1)} \tag{2.4.5}$$

and

$$E[g(X_{k1},\ldots,X_{ki})] \le \left(\frac{n^{[m]}}{n(n-1)}\right)^4 M_0[\alpha(k_2 - k_1)]^{\frac{2+\delta}{4+\delta}}, \text{if } k_2 - k_1 = d^{(1)}, \tag{2.46}$$

where $d^{(k)}$ is the kth largest difference among $(k_{j+1} - k_j)$, $j = 1,\ldots,7$. Therefore,

$$\sum_{\substack{1 \le k_1 < \ldots < k_8 \le n \\ k_8 - k_7 = d^{(1)} \text{ or } k_2 - k_1 = d^{(1)}}} Eg(X_{k_1},\ldots,X_{k_8})$$

$$\leq M_0\left(\frac{n^{[m]}}{n(n-1)}\right)^4 n^4\sum_{j=1}^{n}(j+1)^3[\alpha(j)]^{\frac{2+\delta}{4+\delta}}. \tag{2.47}$$

Also, for some j_i $(2 \leq j_i \leq 6, 1 \leq i \leq 4)$,

$$k_{j_i+1}-k_{j_i}=d^{(i)} \quad (1 \leq i \leq 4).$$

Then applying again Lemma 2.1 we obtain

$$Eg(X_{k_1},\dots,X_{k_8}) \leq \left(\frac{n^{[m]}}{n(n-1)}\right)^4 M_0\sum_{i=1}^{4}[\alpha(k_{j_i+1}-k_{j_i})]^{\frac{2+\delta}{4+\delta}} \tag{2.48}$$

and

$$\sum_{\substack{1 \leq k_1 < \dots < k_8 \leq n \\ k_{j_i+1}-k_{j_i}=d^{(i)}, 1 \leq i \leq 4}} Eg(X_{k_1},\dots,X_{k_8})$$

$$\leq 4\left(\frac{n^{[m]}}{n(n-1)}\right)^4 M_0 n^4\sum_{j=1}^{n}(j+1)^3[\alpha(j)]^{\frac{2+\delta}{4+\delta}}. \tag{2.49}$$

Hence,

$$\sum_{1 \leq k_1 < \dots < k_8 \leq n} Eg(X_{k_1},\dots,X_{k_8}) \leq M\left(\frac{n^{[m]}}{n(n-1)}\right)^4\sum_{j=1}^{n}(j+1)^3(\alpha(j))^{\frac{2+\delta}{4+\delta}}$$

$$= 0\left(n^{4m-3-\gamma}\right) \tag{2.50}$$

where $M > 0$ is some constant.

A similar method can be used to estimate the sums in the other cases and thus obtain (2.40). The proof of (2.41), which is analogous, is omitted.

Lemma 2.7: *Under the conditions of Lemma 2.6, we have*

$$U_n - \xi = mU_n^{(1)} + R_n \tag{2.51}$$

where $R_n = O(n^{-1}(\log\log n)^{\frac{1}{2}})$ *a.s. as* $n \to \infty$.

Proof: The proof of the lemma follows by applying Lemma 2.6 and the approach of Theorem 1 of [21].

Lemma 2.8: *Let* $\{Y_{ni}, 1 \leq i \leq n, n \geq 1\}$ *be a sequence of strong mixing random variables with mean* 0. *Suppose for any* n, m *such that* $m \geq m$ *and any* $J \subset \{1,\dots,n\}$ *with* Card $J = m$

$$E\left(\sum_{j \in J} Y_{nj}\right)^2 = m\sigma^2(1+o(1)) \tag{2.52}$$

for some $\sigma^2 > 0$. *Then, the process* $\{Y_{ni}, 1 \leq i \leq n, n \geq 1\}$ *obeys the law of the iterated logarithm if the following conditions are satisfied for some* δ *and* δ' *such that*

$$\sup_{n \geq 1}\max_{1 \leq i \leq n} E\,|\,Y_{ni}\,|^{2+\delta} = M < \infty \tag{2.53}$$

and

$$\sum_{n=1}^{\infty}\{\alpha(n)\}^{\frac{\delta'}{2+\delta'}} < \infty. \tag{2.54}$$

Proof: The lemma was proved by [7], Lemma 5.6 for a sequence of random variables satisfying the strong mixing condition with mean zero.

Lemma 2.9: *Let* $[Y_{ni}, 1 \leq i \leq n, n \geq 1\}$ *be a sequence of random variables satisfying the strong mixing condition with coefficient* $\alpha(n)$. *Let* X *be* $\sigma(Y_{ni}, 1 \leq i \leq j \leq n)$ *measurable and* Y *be* $\sigma(Y_{ni} < j+m \leq i \leq n)$ *measurable. If* $E\,|\,X\,|^{2+\gamma} < \infty$ *and* $E\,|\,Y\,|^{2+\gamma} < \infty$, *where* γ *is a positive number. Then*

$$|\operatorname{Cov}(X,Y)| \leq 10(E|X|^{2+\gamma})^{\frac{1}{2+\gamma}}(E|Y|^{2+\gamma}[\alpha(m)]^{\frac{\gamma}{2+\gamma}}. \tag{2.55}$$

Proof: This is Proposition 2.8 of [4]. Also, see [9].

Lemma 2.10: *Let $[X_n, n \geq 1\}$ be a sequence of random variables satisfying the strong mixing condition. Also, assume $EX_n = 0$ and $EX_n^2 < \infty$ for all $n \in \mathbb{N}$, and there exists $\beta > 2$ with*

$$\sup_{n \in \mathbb{N}} E|X_n|^\beta < \infty, \sum_{k \in \mathbb{N}} [\alpha(k)]^{1-\frac{2}{\beta}} < \infty.$$

If

$$ES_n^2/n \to \sigma^2 \text{ for some } \sigma > 0, \text{ as } n \to \infty$$

then

$$W_n \xrightarrow{D} W \text{ on } (c[0,1], d) \tag{2.56}$$

where $S_n = \sum_{j=1}^n X_j$, $W_n(t,w) = S_{[nt]}(w)/\sigma n^{\frac{1}{2}}$, $t \in [0,1]$, $w \in \Omega$, d is a uniform metric, and W is a standard Wiener process on $[0,1]$.

Proof: This is Theorem 0 of [8]. A similar theorem is also proved by [5], p. 46, Theorem 1, for stationary, strong mixing random variables.

3. A Law of the Iterated Logarithm for U_n

Let $\{F_n^*, n \geq 1\}$ be any sequence of continuous cumulative distribution functions on $\mathbb{R}^2$ with marginals F, and denote

$$\rho(1) = \int_{\mathbb{R}^m} g^2(x_1, \ldots, x_m) \prod_{i=1}^m dF(x_i) - \xi^2 \tag{3.1}$$

$$\rho(i) = 2\left[\int_{\mathbb{R}^{2m}} g(x_1, \ldots, x_m) g(x_{m+1}, \ldots, x_{2m}) dF_i^*(x_1, x_{m+1}) \prod_{n=2}^m dF(x_n) \prod_{k=m+2}^{2m} dF(x_k) - \xi^2\right], \quad \forall i \geq 2.$$

Let $F_{i,j}$ be the distribution function of (X_i, X_i), $1 \leq i < j$. We have the following law of the iterated logarithm for the nonstationary U-statistics. Such a result generalizes the result of [21], who proved a similar theorem for the stationary U-statistic under a strong mixing set-up.

Theorem 3.1: *Suppose the sequence $\{X_i, i \geq 1\}$ is strongly mixing with $\alpha(n)$ satisfying $\alpha(n) = \rho^n$, $0 < \rho < 1$. Furthermore, assume that for any $n > 1$, there exists a continuous d.f. F_n^* on $\mathbb{R}^2$ with marginals F such that*

$$\| F_{i,j} - F_{j-i}^* \| = O\left(\rho_0^{\max(i,-1)}\right), \quad 1 \leq i < j, \tag{3.3}$$

for some $0 < \rho_0 < 1$, where $\| \quad \|$ denotes the total variation. Suppose also, conditions (2.38) and (2.39) are satisfied. Then, if

$$\sigma^2 = \rho(1) + 2\sum_{i=2}^\infty \rho(i)$$

exists and is finite, we have

$$\limsup_{n \to \infty} \frac{n^{\frac{1}{2}}(U_n - \xi)}{2^{\frac{1}{2}} m\sigma(\log\log n)^{\frac{1}{2}}} = \pm 1 \text{ a.s.} \tag{3.4}$$

Proof: From Lemma 2.7, it suffices to prove that

$$\limsup_{n\to\infty}\frac{n^{-1}\sum_{i=1}^{n}X_{ni}^{*}}{2^{\frac{1}{2}}m\sigma(\log\log n)^{\frac{1}{2}}}=\pm 1 \text{ a.s.} \tag{3.5}$$

where

$$X_{ni}^{*}=n^{-[m-1]}\left(\mathcal{U}_{1,n}(X_i)-\int_{\mathbb{R}}\mathcal{U}_{1,n}(x)dF_i(x)\right)$$

and $\mathcal{U}_{1,n}(x)$ is defined as in (2.35).

We will apply Lemma 2.8 to prove (3.5). That is, we have to verify that the sequence $\{X_{ni}^{*},\ 1\le i\le n,\ n\ge 1\}$ satisfies conditions (2.52)-(2.54).

It is clear that conditions (2.53) follows from (2.38), and condition (2.54) follows from the assumption that $\alpha(n)=\rho^n$, $0<\rho<1$. Therefore, we have only to show that $\{X_{ni}^{*}, 1\le i\le n,\ n\ge 1\}$ satisfies (2.52). That is, for any n,m such that $n\ge m$ and $J\supset\{1,\ldots,n\}$ with $\operatorname{Card}J=m$

$$E\left(\sum_{i\in J}X_{ni}^{*}\right)^2=m\sigma^2(1+o(1)). \tag{3.6}$$

Let us consider first the case when $\operatorname{Card}J=n$; the case when $\operatorname{Card}J<n$ has a similar proof.

Now, it is easy to see that

$$\begin{aligned}\left|\frac{1}{n}E\left(\sum_{i=1}^{n}X_{ni}^{*}\right)^2-\sigma^2\right|&=\left|\frac{1}{n}(n^{-[m-1]})^2\sum_{i=0}^{n-1}\sum_{j=1}^{n-i}\varphi(i,j)-\sum_{i=1}^{\infty}\rho(i)\right| \\ &\le\left|\frac{1}{n}(n^{-[m-1]})^2\sum_{i=0}^{n-1}\sum_{j=1}^{n-i}\varphi(i,j)-\frac{1}{n}\sum_{i=1}^{\infty}(n-i)\rho(i)\right| \\ &\quad+\sum_{i=n+1}^{\infty}|\rho(i)|+\sum_{i=1}^{n}\sum_{k=i}^{\infty}|\rho(k)|=\sum_{i=1}^{3}I_{ni}\end{aligned} \tag{3.7}$$

where

$$\varphi(i,i)=\operatorname{Var}[\mathcal{U}_{1,n}(X_i)],\ i\ge 1, \tag{3.8}$$

and

$$\varphi(i,j)=2\operatorname{Cov}(\mathcal{U}_{1,n}(X_i),\mathcal{U}_{1,n}(X_j)),\ i<j, \tag{3.9}$$

$\mathcal{U}_{1,n}(x)$ being defined as in (2.35).

From condition (3.3), we have $|I_{n1}|=o(1)$. From condition (2.38), we deduce that $|\rho(i)|\le(\alpha(i))^{\frac{\delta'}{2+\delta'}}M_0^{\frac{2}{2+\delta'}}$; by assumption $\alpha(n)=\rho^n$, $0<\rho<1$, we obtain that $I_{ni}\to 0$ as $n\to\infty$ for $i=2$ and 3. Theorem 3.1 is proved.

4. Almost Sure Representation and a Law of Iterated Logarithm for $\widehat{F}_n(U_n)$

Theorem 4.1: *Suppose that $\forall i\ge 1$, $F_i''(\xi)$ exists and is finite and let $\{F_i, i\ge 1\}$ and k satisfy (A). Furthermore, suppose the assumptions of Lemma 2.6 are satisfied and let $\eta>0$ and $\alpha(n)=\rho^n$, $0<\rho<1$. Then*

$$\widehat{F}_n(U_n)=\widehat{F}_n(\xi)+\frac{m\overline{F}'_n(\xi)}{n}\sum_{i=1}^{n}X_{ni}^{*}+R_n \tag{4.1}$$

where X^*_{ni} *is defined as in* (3.5) *and*

$$R_n = o(n^{-\frac{3}{4}+\eta}) + O(a_n) \text{ a.s.} \tag{4.2}$$

as $n \to \infty$ *and* $0 < \eta < \frac{1}{8}$. *Moreover, if* $\{F_i\}$ *and* k *satisfy* (B), *then we may replace* $O(a_n)$ *by* $O(a_n^2)$ *in* (4.2).

Proof: If $\{F_i, i \geq 1\}$ and k satisfy (A), using notations in (2.37) and Hoeffding's projection method (see, e.g. [21]), we obtain

$$U_n = \xi + \sum_{c=1}^{n} \binom{m}{c} U_n^{(c)} = \xi + mU_n^{(1)} + \sum_{c=2}^{m} \binom{m}{c} U_n^{(c)}. \tag{4.3}$$

From Theorem 3.1, we have

$$|U_n - \xi| \leq (1+\epsilon) 2^{\frac{1}{2}} m\sigma n^{-\frac{1}{2}} (\log\log n)^{\frac{1}{2}} \text{ a.s.} \tag{4.4}$$

as $n \to \infty$, for some $\epsilon > 0$. With the aid of Lemma 2.5, we have

$$\widehat{F}_n(U_n) - \widehat{F}_n(\xi) = \bar{F}_n(U_n) - \bar{F}_n(\xi) + o(n^{-\frac{3}{4}+\eta}) + O(a_n) \text{ a.s.}, \tag{4.5}$$

as $n \to \infty$.

Using Young's form of Taylor's theorem and (4.4), we get

$$\bar{F}_n(U_n) - \bar{F}_n(\xi) = (U_n - \xi)\bar{F}'_n(\xi) + O(n^{-1}\log\log n) \text{ a.s.} \tag{4.6}$$

The proof can be completed using (4.5), (4.6) and Lemma 2.6.

If $\{F_i, i \geq 1\}$ and k satisfy (B), the proof is similar.

Theorem 4.2: *Suppose that* $\{F_i''(\xi), i \geq 1\}$ *exist and are finite and let* $\{F_i, i \geq 1\}$, *and* k *satisfy* (A). *Furthermore, suppose the assumptions of Lemma* 2.6 *are satisfied and* $\alpha(n) = \rho^n$, *for* $0 < \rho < 1$. *Then*

$$\widehat{F}_n(U_n) = n^{-1}\sum_{i=1}^{n} Y_{ni} + R_n \tag{4.7}$$

where

$$Y_{ni} = u(\xi - X_i) + m\bar{F}'_n(\xi)X^*_{ni},$$

X^*_{ni} *is defined as in* (3.5) *and*

$$R_n = o(n^{-\frac{3}{4}+\eta}) + O(n^{\frac{1}{2}} a_n (\log\log n)^{-\frac{1}{2}}) \text{ a.s.}$$

as $n \to \infty$ *and* $0 < \eta < \frac{1}{8}$. *Moreover, if* F_i *and* k *satisfy* (B), *then* (4.7) *holds with* $O(n^{\frac{1}{2}} a_n (\log\log n)^{-\frac{1}{2}})$ *replaced by* $O(n a_n^2 (\log\log n)^{-1})$.

Proof: Let $\{F_i, i \geq 1\}$ and k satisfy (A). From Theorem 4.1, we have

$$\widehat{F}_n(U_n) = \widehat{F}_n(\xi) + m\frac{\bar{F}'_n(\xi)}{n}\sum_{i=1}^{n} X^*_{ni} + R_n \tag{4.9}$$

where

$$R_n = o(n^{-\frac{3}{4}+\eta}) + O(a_n) \text{ a.s. as } n \to \infty.$$

Moreover, from Lemma 2.4 it follows that

$$\widehat{F}_n(\xi) = \widetilde{F}_n(\xi) + o(n^{-\frac{3}{4}+\eta}) + O(n^{\frac{1}{2}} a_n (\log\log n)^{-\frac{1}{2}}) \text{ a.s.} \tag{4.10}$$

Combining (4.9) and (4.10) leads to (4.7). The proof under (B) is similar.

As an application of Theorem 4.2, we have the following law of iterated logarithm for $\widehat{F}_n(U_n)$.

Theorem 4.3: *Suppose that $F_i''(\xi)$ exists and is finite for each $i \geq 1$ and let F_i and k satisfy (A). Furthermore, suppose the assumptions of Lemma 2.6 are satisfied, $\alpha(n) = \rho^n$, for $0 < \rho < 1$, and $a_n = o(n^{-1}\log\log n)$. Then*

$$\limsup_{n\to\infty} \frac{n^{\frac{1}{2}}(\widehat{F}_n(U_n) - F(\xi))}{\sqrt{2\tau^2 \log\log n}} = 1 \quad a.s. \tag{4.11}$$

where

$$\tau^2 = \int_{\mathbb{R}} A^2(x)dF(x) + 2\sum_{k=1}^{\infty} \int_{\mathbb{R}^2} A(x)A(y)dF_k^*(x,y)$$

and

$$A(x) = u(\xi - x) - F(\xi) + mF'(\xi)(g_1(x) - \xi), \quad g_1(x) = \int g(x_1,\ldots,x_m)\prod_{i=2}^{m} dF(x_i).$$

Moreover, if $\{F_i, i \geq 1\}$ and k satisfy (B) and $a_n = o(n^{-\frac{3}{4}}(\log\log n)^{\frac{3}{4}})$, then (4.11) *holds.*

Proof: Using Theorem 4.2, the proof follows easily by applying Lemma 2.8.

5. An Invariance Principle for $\widehat{F}_n(U_n)$

Let

$$Y_{ni} = u(\xi - X_i) + m\bar{F}'_n(\xi)X_{ni}^* \tag{5.1}$$

where

$$X_{ni}^* = n^{-[m-1]}(\mathfrak{U}_{1,n}(X_i) - E\mathfrak{U}_{1,n}(X_i)).$$

Denote

$$Z_{ni} = Y_{ni} - F_i(\xi), \tag{5.2}$$

then it is clear that $EZ_{ni} = 0$ and $EZ_{ni}^2 < \infty$. Let $S_n = \sum_{i=1}^{n} Z_{ni}$. Then, we shall prove the following fact

$$\lim_{n\to\infty} \tfrac{1}{n}ES_n^2 \text{ exists and is finite.} \tag{5.3}$$

Hence, we write

$$\tau^{*2} = \lim_{n\to\infty} ES_n^2/n.$$

We will use (3.6) and Lemma 2.9 to prove (5.3). Now, note that

$$ES_n^2 = E\left[m\bar{F}'_n(\xi)\sum_{i=1}^{n} X_{ni}^* + \sum_{i=1}^{n}[u(\xi - X_i) - F_i(\xi)]\right]^2 \tag{5.4}$$

$$= m^2(\bar{F}'_n(\xi))^2 E\left(\sum_{i=1}^{n} X_{ni}^*\right)^2 + E\left[\sum_{i=1}^{n}(u(\xi - X_i) - F_i(\xi))\right]^2$$

$$+ 2m\bar{F}'_n(\xi)E\left[\sum_{i=1}^{n} X_{ni}^* \sum_{i=1}^{n}(u(\xi - X_i) - F_i(\xi))\right]$$

$$= I_1 + I_2 + I_3.$$

So $ES_n^2/n = I_1/n + I_2/n + I_3/n$. We first estimate I_1/n. By (3.6),

$$\frac{I_1}{n} = \frac{m^2[\bar{F}'_n(\xi)]^2}{n} n\sigma^2(1+o(1)) \tag{5.5}$$

$$= m^2[\bar{F}'_n(\xi)]^2\sigma^2(1+o(1)) \to m^2(\bar{F}'(\xi))^2\sigma^2 < \infty \text{ as } n\to\infty.$$

Note that by assumption, $\alpha(n) = \rho^n$, $0 < \rho < 1$, we have

$$\sum_{n=1}^{\infty} [\alpha(n)]^{\frac{\delta'}{2+\delta'}} < \infty, \quad \text{for } 0 < \delta' < 1.$$

Next, we write I_2/n as

$$\frac{I_2}{n} = \frac{1}{n}\sum_{i=1}^{n} E(u(\xi - X_i) - F_i(\xi))^2 + \frac{1}{n}\sum_{i\neq j}\sum \operatorname{Cov}(u(\xi - X_i), u(\xi - X_j)) \tag{5.6}$$

$$= \frac{1}{n}\sum_{i=1}^{n} F_i(\xi)(1 - F_i(\xi)) + \frac{1}{n}\sum_{i\neq j}\sum \operatorname{Cov}(u(\xi - X_i),\ u(\xi - X_j)).$$

Since the first term of (5.6) converges to $F(\xi)(1 - F(\xi))$, we will estimate the second term of (5.6) by applying Lemma 2.9. Note that $E(u(\xi - X_i) - F_i(\xi)) = 0$ and $E \mid u(\xi - X_i))^{2+\delta'} \leq 1$, for some $\delta' > 0$, so we obtain:

$$\frac{2}{n}\left|\sum_{j>i}\sum \operatorname{Cov}(u(\xi - X_i), u(\xi - X_j))\right| \tag{5.7}$$

$$\leq \frac{2}{n}\sum_{j>i}\sum 10[\alpha(j-i)]^{\frac{\delta'}{2+\delta'}} = \frac{20}{n}\sum_{i=1}^{n-1}\sum_{k=1}^{n-i}(\alpha(k))^{\frac{\delta'}{\delta'+2}}$$

$$\to 20\sum_{j=1}^{\infty}[\alpha(j)]^{\frac{\delta'}{2+\delta'}} < \infty, \text{ as } n\to\infty.$$

Finally, the last term of (5.4) can be bounded by using Lemma 2.9 again. That is, for $n \geq N_0$ with N_0 sufficiently large and $E \mid X^*_{ni} \mid \leq M_0$, we obtain:

$$\frac{1}{n}\left| E\left[2m\bar{F}_n(\xi)\sum_{i=1}^{n} X^*_{ni}\sum_{i=1}^{n}(u(\xi - X_i) - F_i(\xi))\right]\right| \tag{5.8}$$

$$\leq \frac{2m}{n}\bar{F}'_n(\xi)E\sum_{i=1}^{n}\left| X^*_{ni}(u(\xi - X_i) - F_i(\xi))\right|$$

$$+ \frac{2m}{n}\bar{F}'_n(\xi)\sum_{i\neq j}\sum \mid E[X^*_{ni}(u(\xi - X_j) - F_j(\xi))] \mid$$

$$\leq \frac{2m}{n}\bar{F}'_n(\xi)nM_0 + \frac{4m}{n}\bar{F}'_n(\xi)\sum_{j>i}\sum[\alpha(j-i)]^{\frac{\delta'}{2+\delta'}}$$

$$\leq 2m\bar{F}'_n(\xi)M_0 + 4M_0 m\bar{F}'_n(\xi)\sum_{k=1}[\alpha(k)]^{\frac{\delta'}{2+\delta'}}$$

$$\to 2m\bar{F}'(\xi)M_0 + 4mM_0\bar{F}'(\xi)\sum_{n=1}^{\infty}[\alpha(n)]^{\frac{\delta'}{2+\delta'}} < \infty, \text{ as } n\to\infty.$$

From (5.5)-(5.8), we have (5.3). This ensures the finiteness of τ^{*2}.

Next, let W_n be a random function of $[0,1]$ defined as

$$W_n\left(\frac{i}{n}\right) = \begin{cases} 0, & \text{if } 0 \le i \le m-1 \\ \dfrac{i(\widehat{F}_i(U_i) - F_i(\xi))}{\tau^* n^{\frac{1}{2}}}, & \text{if } m \le i \le n \end{cases}$$

and $W_n(t)$, $0 \le t \le 1$ defined elsewhere, by linear interpolation.

Thus, W_n has continuous sample paths and belongs to $C[0,1]$, the space of all continuous functions on $[0,1]$.

Theorem 5.1: *Assume that (i) $F_i''(\xi)$ exists (finite), $i \ge 1$, (ii) F_i and k satisfy (A), and (iii) $a_n = o(n^{-1}(\log\log n)^{\frac{1}{2}})$ as $n \to \infty$. Then*

$$W_n \xrightarrow{D} W \text{ on } (C[0,1], d) \tag{5.10}$$

where d is uniform metric, and W is a standard Wiener process on $[0,1]$. Moreover, if F_i and k satisfy (B), then (5.9) holds if $a_n = o(n^{-\frac{3}{4}}(\log\log n)^{\frac{1}{2}})$ as $n \to \infty$.

Proof: Define the random function W_n^*, $n \ge 1$, on $[0,1]$ by $W_n^*(0) = 0$.

$$W_n^*\left(\frac{i}{n}\right) = \frac{\sum_{j=1}^{i}[Y_{nj} - F_i(\xi)]}{\tau^* n^{\frac{1}{2}}} \quad \text{if } 1 \le i \le n \tag{5.11}$$

where Y_{nj} and τ^* are defined as in (5.1) and (5.3), respectively. Note that W_n^* also have continuous sample paths and they belong to $C[0,1]$.

If conditions (A) are satisfied, then by Lemma 2.10 and (5.3), we have $W^* \xrightarrow{D} W$ on $(C[0,1], d)$. Therefore, it suffices to show that

$$d(W_n, W_n^*) \xrightarrow{P} 0 \text{ as } n \to \infty. \tag{5.12}$$

Note that

$$d(W_n, W_n^*) \le \max_{m \le i \le n} \frac{i^{\frac{1}{2}} | V_i |}{\tau^* n^{\frac{1}{2}}} + \frac{1}{\tau^* n^{\frac{1}{2}}} \sum_{i=1}^{m} | Y_{ni} | \tag{5.13}$$

where

$$V_i = n^{\frac{1}{2}}(\widehat{F}_n(U_n) - n^{-1}\sum_{i=1}^{n} Y_{ni}). \tag{5.14}$$

So (5.12) holds if we prove the following two facts:

$$\text{For any } \epsilon > 0, \quad P\left(\frac{1}{\tau^* n^{\frac{1}{2}}} \sum_{i=1}^{m} | Y_{ni} | > \epsilon\right) \to 0 \text{ as } n \to \infty \tag{5.15}$$

and

$$\max_{m \le i \le n} \frac{i^{\frac{1}{2}} | V_i |}{n^{\frac{1}{2}} \tau^*} \xrightarrow{P} 0 \text{ as } n \to \infty. \tag{5.16}$$

(5.15) holds by applying the Chebyshev inequality and noting that $E | Y_{ni} | \le M$,

$$P\left(\frac{1}{\tau^* n^{\frac{1}{2}}} \sum_{i=1}^{m} | Y_{ni} | > \epsilon\right) \le \frac{\sum_{i=1}^{m} E | Y_{ni} |}{\tau^* \epsilon n^{\frac{1}{2}}} \to 0 \text{ as } n \to \infty.$$

Next, we prove (5.16). To this end, we first use Theorem 4.2 and $a_n = o(n^{-1}(\log\log n)^{\frac{1}{2}})$ to obtain

$$\lim_{n \to \infty} V_n = 0 \quad \text{a.s.} \tag{5.17}$$

Therefore, by Egoroff's Theorem, for any $\delta > 0$, $\epsilon > 0$, there exists a measurable set A such that $P(A^c) < \frac{1}{2}\delta$, where A^c is the complement of A, and V_n converges to zero uniformly on A; so there exists a positive integer n_0 such that for $n \geq n_0$, $|V_n(\omega)| < \epsilon$ for all $\omega \in A$. For $n \geq n_0$, we have

$$P\left\{\max_{m \leq i < n} \frac{i^{\frac{1}{2}}|V_n|}{n^{\frac{1}{2}}\tau^*} > \epsilon\right\}$$

$$\leq \sum_{i=m}^{n_0} P\left\{i^{\frac{1}{2}}|V_i| > n^{\frac{1}{2}}\tau^*\epsilon\right\} + P\left\{\max_{n_0+1 \leq i \leq n} \frac{i^{\frac{1}{2}}|V_i|}{n^{\frac{1}{2}}} > \tau^*\epsilon\right\}. \quad (5.18)$$

Now, clearly, the first sum on the right side of (5.18) tends to 0 as $n \to \infty$.

On the other hand, since

$$P\left(\left\{\max_{n_0+1 \leq i \leq n} |V_i| > \epsilon\right\} \cap A\right) = 0,$$

it follows that

$$P\left(\max_{n_0+1 \leq i \leq n} \frac{i^{\frac{1}{2}}|V_i|}{n^{\frac{1}{2}}} > \tau^*\epsilon\right)$$

$$\leq P\left(\left\{\max_{n_0+1 \leq i \leq n} |V_i| > \epsilon\right\} \cap A\right) + P(A^c) < \tfrac{1}{2}\delta. \quad (5.19)$$

The proof of the theorem follows from (5.18) and (5.19). If conditions (B) are satisfied, the proof is similar.

Remark 5.1: Theorem 5.1 clearly contains the Central Limit Theorem for $\widehat{F}_n(U_n)$, which extends the result of [17] for absolutely regular stationary random variables. In addition, Theorem 5.1 provides information such as for $x > 0$,

$$P\left(\max_{1 \leq i \leq n} i(\widehat{F}(u) - F_i(\xi)) \geq x\tau^* n^{\frac{1}{2}}\right) \to 2[1 - \Phi(x)], \text{ as } n \to \infty.$$

Acknowledgements

The authors thank the associate editor and referee for their helpful comments that led to an improved presentation.

References

[1] Bahadur, R.R., A note on quantiles in large samples, *Ann. Math. Statist.* **37** (1966), 577-580.

[2] Bennett, G., Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.* **57** (1962), 33-45.

[3] Bradley, R., Basic properties of strong mixing conditions, In: *Dependence in*

Probability and Statistics: A Survey of Recent Results (ed. by E. Eberlein and M.S. Taqqu), Birkhäuser, Boston (1986), 165-192.

[4] Doukhan, P. and Portal, F., Principe d'invariance faible pour la fonction de répartition empirique dans un cadre multidmensionnel et mélangeant, *Prob. and Math. Stat.* **8**:2 (1987), 117-132.

[5] Doukhan, P., *Mixing: Properties and Examples*, Lecture Notes in Statistics, Springer-Verlag 1994.

[6] Eberlein, E. and Taqqu, M.S., *Dependence in Probability and Statistics: A survey of Recent Results*, Birkhäuser, Boston 1986.

[7] Harel, M. and Puri, M.L., Law of the iterated logarithm for perturbed empirical distribution functions evaluated at a random point for nonstationary random variables, *J. of Theoretical Probability* (1994), to appear.

[8] Herrndorf, N., A functional central limit theorem for strongly mixing sequences of random variables, *Z. Wahrsch. Verw. Gebiete* **69** (1985), 541-550.

[9] Ibragimov, J.A. and Linnik, Yu.V., *Independent and Stationary Sequences of Random Variables*, Wolters-Noerdhoff, Groningen 1971.

[10] Lea, C. and Puri, M.L., Asymptotic properties of perturbed empirical distribution functions evaluated at a random point, *J. Stat. Planning and Inference* **19** (1988), 201-215.

[11] Nadaraya, E.A., Some new estimates for distribution function, *Theory Probab. Appl.* **9** (1964), 497-500.

[12] Parzen, E., On estimation of a probability density function and mode, *Ann. Math. Statist.* **33** (1962), 1065-1076.

[13] Puri, M.L. and Ralescu, S., On Berry-Esséen rates, a law of the iterated logarithm and an invariance principle for the proportion of the sample below the sample mean, *J. Multivariate Anal.* **14** (1983), 231-247.

[14] Puri, M.L. and Ralescu, S., Central limit theorem for perturbed empirical distribution functions evaluated at a random point, *J. Multivariate Anal.* **16** (1986), 273-279.

[15] Rosenblatt, M., A central limit theorem and a strong mixing condition, *Proc. Natl. Acad. Sci. USA* **42** (1956), 43-47.

[16] Rosenblatt, M., Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.* **27** (1956), 832-837.

[17] Sun, S., Asymptotic behavior of the perturbed empirical distribution function evaluated at a random point for absolutely regular sequences, *J. Multivariate Anal.* **47**:2 (1993), 230-249.

[18] Winter, B.B., Strong uniform consistency of integrals of density estimators, *Canad. J. Statist.* **1** (1973), 247-253.

[19] Winter, B.B., Convergence rate of perturbed empirical distribution functions, *J. Appl. Probab.* **16** (1979), 163-173.

[20] Yamato, H., Uniform convergence of an estimator of a distribution function, *Bull. Math. Statist.* **15** (1973), 69-78.

[21] Yoshihara, K., Limiting behavior of U-statistics for stationary, absolutely regular processes, *Z. Wahrsch. Verw. Gebiete* **35** (1976), 237-252.

[22] Yukitsch, J.E., A note on limit theorems for perturbed empirical processes, *Stoch. Process. Appl.* **33** (1989), 163-173.