NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

UNIVERSITY OF ALBERTA

ASYMPTOTIC BEHAVIOR OF U-STATISTICS FOR STATIONARY STRONG MIXING PROCESSES

By

XIAOLING GU ©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIRMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA

FALL 1990

ISBN 0-315-64976-3

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled

"Asymptotic Behaviour of U-Statistics for

Stationary Strong Mixing Processes"

submitted by Xiao-Ling Gu in partial fulfillment of the requirements for the degree of Master of Science.

K. L. Mehra

Supervisor K.L. Mehra

R. Karunamuni

R. Karunamuni

M. Sudhakara Rao

M. Sudhakara Rao

H. I. Freedman

H.I. Freedman

July 26, 1990

Date

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: XIAOLING GU

TITLE OF THESIS:

ASYMPTOTIC BEHAVIOR OF U-STATISTICS

FOR STATIONARY STRONG MIXING PROCESSES

DEGREE FOR WHICH THESIS WAS PRESENTED: MASTER OF SCIENCE

YEAR THIS DEGREE GRANTED: 1990

(Signed)

Permanent Address:

Math. Dept.
Southwest Jiaotong Univ.
Changdu, Sichuan
P.R.China

Date: Aug 2, 1990

ABSTRACT

Asymptotic behavior of U-statistics based on realizations of various types of weakly dependent processes has been investigated for many years but only up to absolute regularity condition, with very few results in the strong mixing case. The object of this thesis is to study the limiting behavior of U- statistics for strong mixing processes. First we study the asymptotic behavior of a special class of U- statistics possessing a "decomposable structure" , which includes some ordinary estimators like moment estimators, and derive for such statistics central limit theorems and function central limit theorems for their corresponding partial sum processes. Then we extend this class to a larger one satisfying certain strong and weak consistency properties: Under some broad conditions on strong mixing processes we prove that $U_n \to \theta$(a.s.) if h is bounded and a.e. continuous, but for unbounded h under these conditions, we are able to prove only that $U_n \to \theta$(in prob.) if h is a.e. continuous.

ACKNOWLEDGMENTS

I am greatly indebted to my supervisor, Professor K.L. Mehra for introducing me to the problem and for the considerable time and effort that he has spent in assisting me in my work and in preparing this thesis.

I am also grateful to Professor M.S. Rao for many valuable suggestions.

Finally, I would like to thank Dr. R. Karunamuni and Professor H.I. Freedman for their concerns about my thesis.

TABLE OF CONTENTS

CHAPTER I

ASYMPTOTIC BEHAVIOR OF U-STATISTICS IN THE I.I.D. CASE

1.1: PRELIMINARIES

Large-sample theory considers a sample $X = (X_1,\ldots,X_n)$ for fixed $n = 1,2,\ldots$, and assesses the performance of statistical procedures, sequences of estimators etc. based on X as $n \to \infty$. Mathematically, the results are illuminating and useful limit theorems which are needed for statistical applications. Consider the simplest situation: Estimate the mean value EX_1 when $X_1,\ldots,X_n$ are i.i.d. (independent and identically distributed) with $\sigma^2 = \mathrm{Var}\,(X_1) < \infty$. It follows from the classical central limit theorem and strong law of large numbers that the estimator $\overline{X}$ converges to EX_1 with probability 1 and $\sqrt{n}(\overline{X} - EX_1)$ converges weakly to $N(0, \sigma^2)$. So the problem arises: How to extend the above results to more general statistics and non i.i.d. sequences $\{X_n: n=1,2,\ldots\}$?

In recent years, a great deal of work has been devoted to the above two aspects of the problem: Firstly, the extension of the limit theorems from $\overline{X}$ to statistics such as U-statistics, V-statistics, or L-statistics; and the other to the extension of the limit theorems from the i.i.d. case to the case of weakly dependent sequences such as m-dependent processes, ϕ-mixing processes, uniformly mixing processes, absolutely regular processes , or strong mixing processes.

The first important contribution to the first aspect is Hoffeding's work[13](1948). He established the limit theory of U-statistics in the i.i.d. case, which are sums of identically distributed random variables. In 1958, Chernoff and Savage [4] obtained the famous result that the central limit theorem remains true for a rather large class of linear rank-order statistics in the i.i.d. case . Later, using the projection method, which is a very strong and useful tool to study limiting behavior, Hajek extended their work [10](1968) to the non i.i.d. case.

On the second aspect, Rosenblatt is the first to put forward the concept of strong mixing processes and prove a central limit theorem for $\overline{X}$ [17](1957) .After that, many authors such as Ibragimov [13], Philipp [16], Serfling [22] , and many others have contributed to the limiting behavior of $\overline{X}$ based on mixing processes.

Recently , combining the two aspects, many authors have studied the limiting behavior of U-statistics,V-statistics and L-statistics based on various weakly dependent processes. In particular, for U-statistics under weak dependence Sen [21] (1972) considered ϕ–mixing processes and derived a central limit theorem and a law of the iterated logarithm. Fears and Mehra [9] (1974) proved the Chernoff-Savage Theorem concerning the asymptotic normality of two-sample linear rank statistics for sequences of observations which satisfy the ϕ-mixing dependence. Later, Mehra [16](1975) established similar results for the one-sample linear rank statistics under ϕ-mixing. Yoshihara [29] (1976) established the weak convergence of U- and V-statistics for stationary absolutely regular processes. Later Yoshihara [30] (1978) also proved the weak convergence of one sample rank order statistics under similar conditions. Denker and Keller [7] (1983) extended these results and proved central limit theorems and their rates of convergence, functional central limit theorems and a.s. approximation by a Brownian motion. Harel and Puri [12] (1989) extended some of Yoshihara's results to nonstationary case.

Since in general one has the following implications: ϕ–mixing ⇒ uniform mixing in both directions of time ⇒ uniform mixing ⇒ absolutely regular ⇒ strong mixing, so the extension of above mentioned results under "strong mixing " are of great importance. But even after nearly fifteen years, extension of above mentioned results on U-statistics to the strong mixing case has not been available. This fact somehow implies that the last extension is very difficult.

The main object of this thesis is to study the limiting behavior of U- and V-statistics based on strong mixing processes. In Chapter 1 we introduce U- and V-statistics in the i.i.d. case and explain the method of projection. In order to exhibit the beautiful structure and limiting behavior of U-statistics, we summarize the main results about U-statistics. In Chapter 2 we introduce various weakly dependent processes and the central limit theorems in the strong mixing case , and review some basic results about the limiting behavior of U- and V-statistics based on absolutely regular processes, especially Yoshihara's work . By examining the conditions for absolutely regular processes, we prove that the linear processes described in [10] and [28] are absolutely regular. Furthermore,we point out that the main difficulty in the strong mixing case lies in the fact that the "fundamental" approximation lemma of Yoshihara [29], which plays a key role in proving the weak convergence of U- and V-statistics, does not seem to extend to the strong mixing case.

In order to get a similar approximation lemma, we have confined our attention to a special class of U- and V-statistics possessing a "decomposable structure" , which includes some ordinary estimators like moment estimators. So in Chapter 3 we study the limiting behavior of such a class of U- and V-statistics and prove central limit theorems and their rates of convergence, functional central limit theorems and a.s. approximation by a Brownian motion. Because the class of U- and V-statistics stated in Chapter 3 is not large enough to include many interesting statistics ,we have attempted to extend this class to larger one satisfying certain consistency results. However, we are not able to prove for this larger class properties such as central limit theorems. Chapter 4 is devoted to the strong and weak consistency results: Under some conditions we prove that $U_n \to \theta$(a.s.) if h is bounded and a.e. continuous, and $U_n \to \theta$(in prob.) if h is a.e. continuous.

In the remaining portion of this chapter we confine our attention to the i.i.d. case. We mention the concepts of U- and V-statistics, introduce the method of projection, exhibit the martingale structure of U- and V-statistics, and present the main results on the limiting behavior of U- and V-statistics. The ideas and the write-up are mostly as in [26](Serfling).

1.2: DEFINITION OF U-STATISTICS

Let $X_1, X_2, \ldots$ be independent observations from a distribution F. Consider a "parametric function" $\theta = \theta(F)$ for which there is an unbiased estimator. That is, $\theta(F)$ may be represented as

$$\theta(F) = E_F\{h(X_1,\ldots,X_m)\} = \int\ldots\int h(x_1,\ldots,x_m)dF(x_1)\ldots dF(x_m), \qquad (1)$$

for some function $h = h(x_1,\ldots,x_m)$ called a "kernel". Without loss of generality, we may assume that h is symmetric in the arguments. For, if not, it may be replaced by the symmetric kernel

$$\frac{1}{m!}\sum_{p} h(x_{i_1},\ldots,x_{i_m})$$

where $\sum_p$ denotes summation over the m! permutations $(i_1,\ldots,i_m)$ of $(1,\ldots,m)$.

Definition 1: For any kernel h, the corresponding U-statistic for estimation of θ on the basis of a sample $X_1,\ldots,X_n$ of size $n \geq m$ is obtained by averaging the kernel h symmetrically over the observations:

$$U_n = U(X_1,\dots,X_n) = \frac{1}{\binom{n}{m}} \sum_c h(X_{i_1},\dots,X_{i_m}), \tag{2}$$

where $\sum_c$ denotes summation over the $\binom{n}{m}$ combinations of m distinct elements $\{i_1,\dots,i_m\}$ from $\{1,\dots,n\}$. Clearly, U_n is unbiased estimate of θ .

Examples. (i) Let $\theta(F)$ = mean of $F = \mu(F) = \int x \, dF(x)$ with the kernel $h(x) = x$, the corresponding U-statistic is

$$U(X_1,\dots,X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X},$$

the sample mean.

(ii) Let $\theta(F) = \mu^2(F) = [\int x \, dF(x)]^2$. For the kernel $h(x_1,x_2) = x_1 x_2$, the corresponding U-statistic is

$$U(X_1,\dots,X_n) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} X_i X_j .$$

(iii) Let $\theta(F)$ = variance of $F = \int (x - \mu)^2 \, dF(x)$ which can be written as $\theta = \frac{1}{2} \int\int (x_1 - x_2)^2 dF(x_1) dF(x_2)$. For the kernel

$$h(x_1,x_2) = \frac{1}{2}(x_1 - x_2)^2,$$

the corresponding U-statistic is

$$\begin{aligned} U(X_1,\dots,X_n) &= \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(X_i, X_j) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right) \\ &= S^2, \end{aligned}$$

the sample variance.

(iv) Let $\theta(F) = F(t_0) = \int_{-\infty}^{t_0} dF(x) = P_F(X_1 \le t_0)$. For the kernel $h(x)=I(x_1 \le t_0)$,

the corresponding U-statistic is

$$U(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^{n} I(x_1 \le t_0) = F_n(t_0),$$

where F_n denotes the sample distribution function.

(v) Let $\theta(F) = E_F|X_1 - X_2| = \int\int |x_1 - x_2|\, dF(x_1)dF(x_2)$, a measure of spread. For the kernel $h(x_1,x_2) = |x_1 - x_2|$, the corresponding U-statistic is

$$U(X_1,\ldots,X_n) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} |X_i - X_j| ,$$

the statistic known as"Gini's mean difference."

(vi) The Wilcoxon one-sample statistic. For estimation of $\theta(F) = P_F(X_1+X_2 \le 0)$,the kernel is given by $h(x_1,x_2) = I(x_1+x_2 \le 0)$ and the corresponding U-statistic is

$$U(X_1,\ldots,X_n) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} I(X_i + X_j \le 0).$$

Definition 2 : Corresponding to a U-statistic

$$U_n = U(X_1,\ldots,X_n) = \frac{1}{\binom{n}{m}} \sum_{c} h(X_{i_1},\ldots,X_{i_m})$$

for estimation of $\theta(F) = E_F(h)$, the associated von Mises statistic

$$V_n = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(X_{i_1},\ldots,X_{i_m}) \tag{3}$$

$= \theta(F_n)$

is called V-statistic, where F_n denotes the standard sample distribution function.

The extension to the case of several samples is straightforward. Consider k independent collections of independent observations $\{X_1^{(1)}, X_2^{(1)}, \ldots\}, \ldots, \{X_1^{(k)}, X_2^{(k)}, \ldots\}$ taken from distributions $F^{(1)}, \ldots, F^{(k)}$, respectively. Let $\theta = \theta(F^{(1)}, \ldots, F^{(k)})$ denote a parametric function for which there is an unbiased estimator. That is ,

$$\theta = E\{h(X_1^{(1)}, \ldots, X_{m_1}^{(1)}; \ldots; X_1^{(k)}, \ldots, X_{m_k}^{(k)}\}, \tag{4}$$

where h is assumed, without loss of generality, to be symmetric within each of its k blocks of arguments.

Definition 3 : Corresponding to the "kernel" h and assuming $n_1 \geq m_1, \ldots, n_k \geq m_k$, the U-statistic for estimation of θ is defined as

$$U_n = \frac{1}{\prod_{j=1}^{k} \binom{n_j}{m_j}} \sum_c h(X_{i_{11}}^{(1)}, \ldots, X_{i_{1m1}}^{(1)}; \ldots; X_{i_{k1}}^{(k)}, \ldots, X_{i_{kmk}}^{(k)}) ; \tag{5}$$

here $\{i_{j_1}, \ldots, i_{j_{m_j}}\}$ denotes a set of m_j distinct elements of the set $(1,2,\ldots,n_j)$, $1 \leq j \leq k$, and $\sum_c$ denotes summation over all such combinations.

Many statistics of interest are of k-sample U-statistic type ; for example, the Wilcoxon 2-sample statistic. Let $\{X_1, \ldots, X_{n_1}\}$ and $\{Y_1, \ldots, Y_{n_2}\}$ be independent observations from continuous distributions F and G, respectively. Then, for $\theta(F,G) = \int G dF = P(X \leq Y)$, an unbiased estimator is

$$U = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_i \leq Y_j).$$

Definition 4 : We call a kernel h degenerate if for any $x_i \in (1 \le i \le m)$ and $j \in (1,\dots,m)$,

$$E_F h(x_1,\dots,x_{j-1},X_j,x_{j+1},\dots,x_m) = 0 . \qquad (6)$$

Definition 5 : A U-statistic (or a V-statistic) is called degenerate if the corresponding kernel h is degenerate.

1.3: PROJECTION OF U-STATISTICS

Consider a symmetric kernel h satisfying $E_F \mid h \mid < \infty$. We define the associated functions

$$h_c(x_1,\dots,x_c) = E_F h(x_1,\dots,x_{c-1},X_c,X_{c+1},\dots,X_m) \qquad (7)$$

for each c=1,...,m-1 and put $h_m \equiv h$.

We now let

$$\tilde{h}_c = h_c - \theta(F) \qquad (8)$$

and

$$g_1(x_1) = \tilde{h}_1(x_1),$$

$$g_2(x_1,x_2) = \tilde{h}_2(x_1,x_2) - g_1(x_1) - g_1(x_2),$$

$$g_3(x_1,x_2,x_3) = \tilde{h}_3(x_1,x_2,x_3) - \sum_{i=1}^{3} g_1(x_i) - \sum_{1 \le i < j \le 3} g_2(x_i,x_j) ,$$

...,

$$g_m(x_1,\dots,x_m) = \tilde{h}_m(x_1,\dots,x_m) - \sum_{i=1}^{m} g_1(x_i) - \sum_{1 \le i_1 < i_2 \le m} g_2(x_{i_1},x_{i_2})$$

$$- \dots - \sum_{(1 \le i_1 < ,\dots, i_{m-1} \le m)} g_m(x_{i_1},\dots,x_{i_{m-1}}) . \tag{9}$$

Clearly, the g_c's are symmetric in their arguments. Also, it is readily seen that g_c's are degenerate. According to Hoeffding's projection method,every U-statistic can be written as a finite weighted sum of degenerate ones, namely,

$$U_n - \theta = \sum_{c=1}^{m} \binom{m}{c} U_n^c , \tag{10}$$

where U_n^c denotes the U-statistic obtained from the degenerate kernel g_c ,that is,

$$U_n^c = \frac{1}{\binom{n}{c}} \sum_{(1 \le i_1 < ,\dots, i_c \le n)} g_c(X_{i_1},\dots,X_{i_c}) . \tag{11}$$

Further, for each $c = 1,\dots,m$,

$$E_F(S_{cn} \mid X_1,\dots,X_k) = S_{ck} \qquad ,c \le k \le n ,$$

where

$$S_{ck} = \binom{k}{c} U_k^c . \tag{12}$$

Thus ,with $\mathcal{F}_k = \sigma(X_1,\dots,X_k)$, the sequence $(S_{cn}, \mathcal{F}_n)_{n \ge c}$ is a forward martingale.

Now the projection of the U-statistic U_n is defined as

$$\hat{U}_n = \binom{m}{1} U_n^1$$

$$= \frac{m}{n} \sum_{i=1}^{n} g_1(X_i)$$

$$= \frac{m}{n} \sum_{i=1}^{n} \tilde{h}_1(X_i) \tag{13}$$

and the remainder of U_n is defined as

$$R_n = \sum_{c=2}^{m} \binom{m}{c} U_n^c . \tag{14}$$

Hence

$$U_n - \theta = \widehat{U}_n + R_n . \tag{15}$$

In fact R_n is also a U-statistic with a degenerate kernel H(see Serfling [24]) , that is,

$$R_n(H) = \frac{1}{\binom{n}{m}} \sum_{i_1<\dots<i_m}^{n} H(X_{i_1}, \dots, X_{i_m}), \tag{16}$$

where

$$H(X_1, \dots, X_m) = h(X_1, \dots, X_m) - \tilde{h}_1(X_1) \dots - \tilde{h}_1(X_m) - \theta \ . \tag{17}$$

1.4: Main Results

First we treat the remainder R_n .We have the following theorems.

Theorem A :[26] Let v be an even positive integer. If $E_F h^v < \infty$ then

$$E_F(R_n)^v = O(n^{-v}), \text{ as } n \to \infty . \tag{18}$$

Theorem B : [26] Let v be even positive integer. If $E_F h^v < \infty$ then,for any $\delta >$ $1/v$,with probability 1

$$R_n = o(n^{-1}(\log n)^{\delta}) \text{ , as } n \to \infty.$$

Thus, the classical limit theorems for $\overline{X}$ can be generalized to U-statistics using Theorem A , the martingale property of U_n and the decomposition (15) above.

Theorem C: [26] If $E_F|h| < \infty$, then $U_n \to \theta$ (wp1), as $n \to \infty$.

Theorem D: [26] If $E_F h^2 < \infty$ and $\zeta_1 = Var(h_1) > 0$,then

$$\overline{\lim}_{n\to\infty} \frac{n^{1/2}(U_n - \theta)}{(2m^2\zeta_1 \log\log n)^{1/2}} = 1 \text{ .wp1} \quad (19)$$

Theorem E: [26] If $E_F h^2 < \infty$ and $\zeta_1 = Var(h_1) > 0$, then

$$n^{1/2}(U_n - \theta) \to N(0, m^2 \zeta_1).$$

Theorem F: [26] If $v = E_F |h|^3 < \infty$ and $\zeta_1 = Var(h_1) > 0$, then

$$\sup_{-\infty<t<\infty} \left| P\left(\frac{n^{1/2}(U_n - \theta)}{m\zeta_1^{1/2}} \le t\right) - \phi(t) \right| \le Cv(m^2\zeta_1)^{-3/2} n^{-1/2}, \quad (20)$$

where C is an absolute constant.

For the corresponding sequence of U- statistics, $\{U_n\}_{n\ge m}$, we consider two associated sequences of stochastic processes on the unit interval [0,1]. The process pertaining to the future , $\{Z_n(t), 0\le t\le 1\}$, is defined by

$$Z_n(0) = 0;$$

$$Z_n(t_{n,k}) = \frac{U_k - \theta}{(Var\{U_n\})^{1/2}},\ k\ge n, \text{ where } t_{n,k} = \frac{Var(U_k)}{Var(U_n)};$$

$$Z_n(t) = Z_n(t_{n,k}),\ t_{n,k+1} < t < t_{n,k}.$$

Theorem G : [26] $Z_n(t)$ converges weakly in D[0,1] to the standard Wiener measure W.

This result generalizes Theorem E and provides additional information such as the following

Corollary : For $x>0$,

$$\lim_{n\to\infty} P(\sup_{k\ge n}(U_k - \theta) \ge x(var\{U_n\})^{1/2})$$
$$= P(\sup_{0\le t\le 1} W(t) \ge x) = 2(1-\Phi(x)) \quad (21)$$

The process pertaining to the past, $\{Y_n(t), 0 \le t \le 1\}$, is defined by

$$Y_n(0) = 0,\ 0 \le t \le \frac{m-1}{n};$$

$$Y_n\left(\frac{k}{n}\right) = \frac{k(U_k - \theta)}{(m^2\zeta_1)^{1/2} n^{1/2}},\ k = m, m+1, \ldots, n;$$

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

Theorem H : [26] $Y_n(t)$ converges weakly in $D[0,1]$ to the standard Wiener measure W.

The similar results hold for V-statistics from the following Lemma.

Lemma : [26] Let r be a positive integer. Suppose that

$$E_F \left| h(X_{i_1}, \ldots, X_{i_m}) \right|^r < \infty, \text{ all } 1 \le i_1, \ldots, i_m \le n.$$

Then

$$E \left| U_n - V_n \right|^r = O(n^{-r}). \tag{22}$$

CHAPTER II

ASYMPTOTIC BEHAVIOR OF U-STATISTICS IN THE ABSOLUTELY REGULAR CASE

In this chapter ,we first define various weakly dependent processes and state certain central limit theorems based on realizations of such processes.. Then we focus, in particular, on the absolutely regular processes. We mention the "fundamental lemma" and the main theorems of Yoshihara[29], Denker and Keller[7], and discuss some conditions for absolutely regular processes. Using these conditions we improve upon the results of [10] by proving that the strong mixing process described in [10] is also an absolutely regular process. Furthermore, we point out why it seems impossible to extend the "fundamental lemma" of Yoshihara [29] to strong mixing processes.

2.1 DEFINITIONS OF WEAKLY DEPENDENT PROCESSES

Let $X_1, X_2, \ldots$ be a strictly stationary stochastic process of random variables on a measurable space $(\mathcal{X},\mathcal{F})$. The past history of the process X_t is described by the σ−algebra $\mathcal{M}_1^{t-s}$, the future by the σ−algebra $\mathcal{M}_{t+s}^{\infty}$,where $\mathcal{M}_a^b$ is the σ−algebra generated by $\{X_t \mid a \le t \le b\}$. It may be that these σ−algebras are independent , in the sense that, for all $A \in \mathcal{M}_1^{t-s}$,$B \in \mathcal{M}_{t+s}^{\infty}$,

$$P(AB) - P(A)P(B) = 0 .$$

In the general case , the magnitude of the left-hand side measures the dependence between past and future, and it is practical and useful sometimes to assume this to be small, in some sense. This idea is captured in the following definitions:

Definition 1 : [27](Stout) A stationary process $\{X_t\}$ is said to be *-mixing if

$$\phi^*(m) = \sup\{ \frac{1}{P(A)P(B)} | P(AB) - P(A)P(B) | \} \to 0 \quad \text{as } m \to \infty , \qquad (1)$$

where the supremum extends over $n \in N$, $A \in \mathcal{M}_1^n$ and $B \in \mathcal{M}_{m+n}^{\infty}$.

Definition 2 : [27](Stout) A stationary process $\{X_t\}$ is said to be uniformly mixing in both directions of time if the process itself and the time reversed process are uniformly mixing, that is,

$$\phi^+(m) = \sup \max\{ | P(A | B) - P(A) | , | P(B | A) - P(B) | \} \to 0 \quad \text{as } m \to \infty, \qquad (2)$$

where the supremum extends over $n \in N$, $A \in \mathcal{M}_1^n$ and $B \in \mathcal{M}_{m+n}^{\infty}$

Definition 3 : [27](Stout) A stationary process $\{X_t\}$ is said to be ϕ–mixing (uniformly mixing) if

$$\phi(m) = \sup | P(B | A) - P(B) | \to 0 \quad \text{as } m \to \infty , \qquad (3)$$

where the supremum extends over $n \in N$, $A \in \mathcal{M}_1^n$ and $B \in \mathcal{M}_{m+n}^{\infty}$.

Definition 4 : [22](Rozanov)(1961) A stationary process $\{X_t\}$ is said to be absolutely regular if

$$\beta(m) = \sup E[\sup\{ | P(B | \mathcal{M}_1^n) - P(B) | \mid B \in \mathcal{M}_{m+n}^{\infty} \}] \to 0 \quad \text{as } m \to \infty . \qquad (4)$$

Definition 5: [19](Rosenblatt) A stationary process $\{X_t\}$ is said to be strong mixing (completely regular) if

$$\alpha(m) = \sup |P(AB) - P(A)P(B)| \to 0 \quad \text{as } m \to \infty, \tag{5}$$

where the supremum extends over $n \in N$, $A \in \mathcal{M}_1^n$ and $B \in \mathcal{M}_{m+n}^{\infty}$.

Since $\phi^*(m) \le \phi^+(m) \le \phi(m) \le \beta(m) \le \alpha(m)$, it follows that $*$-mixing $\Rightarrow$ uniform mixing in both directions of time $\Rightarrow$ uniform mixing $\Rightarrow$ absolutely regular $\Rightarrow$ strong mixing . But the converse implications are not true.

There are a number of results in literature concerning the strong mixing processes. Rosenblatt [20] gave necessary and sufficient conditions for a process to be strong mixing. He established that a stationary Markov process is strong mixing if and only if it is uniformly pure non-deterministic. He also gave equivalent conditions for strong mixing in terms of transition operator and the invariant probability measure. Kolmogorov and Rozanov [15] showed that for stationary Gaussian sequences, $\alpha(m) \le \rho(m) \le 2\pi\,\alpha(m)$. With the aid of this result Ibragimov and Linnik [14] established that a stationary Gaussian sequences is strong mixing if it has a continuous spectral density that is bounded away from 0. Billingsley [2] showed that if Y_n is a Markov chain satisfying Doeblin's condition(see Doob[8] p.192) then $X_n = f(Y_n)$ is strong mixing. Chanda [3] and Whithers [28] considered strong mixing properties of the process $Y_n = \sum_{j=0}^{\infty} w_j e_{n-j}$, where $\{e_n\}$ are i.i.d. random variables. Under some conditions they proved that Y_n is strong mixing. Athreya and Pantula [1] showed that Y_n is strong mixing where Y_n is Harris-recurrent Markov chain, provided there exists a stationary probability distribution $\pi(.)$ for Y_n.

Examples. (i) [27](Stout) Continued fraction expansions of real numbers when viewed properly provide an interesting example of a *-mixing sequence. Let the mapping $T : (0,1) \to (0,1)$ be defined by

$$Tx = x^{-1} - [\,x^{-1}]$$

where $[y]$ denotes the largest integer $\le y$. Let $a_1(x) = [x^{-1}]$ and $a_{n+1}(x) = a_1(T^n x)$ for $n \ge 1$ and each $x \in (0,1)$. Here T^n is the nth iterate of T. Then $(a_1(x), a_2(x), \ldots)$ is referred to as the continued fraction expansion of x. This is sometimes written as

$$x = \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{a_3(x) + \ldots}}} \,.$$

Let $\Omega = (0,1)$, A be any Borel set, and P be defined by

$$P(A) = \log^{-1}2 \int_A (1+x)^{-1}\, dx \,.$$

Then,analysis shows that $\{a_1, a_2, \ldots\}$ is *-mixing sequence with $\phi(m) = Kq^{\sqrt{m}}$ for some $K<\infty$ and $0<q<1$.

(ii) [1] Let Y_n be an autoregressive series given by $Y_n = \rho\, Y_{n-1} + e_n$,$n = 1,2,\ldots$, where $|\rho| < 1$ and $\{\, e_n \,\}$ are i.i.d. random variables independent of Y_0 .Assume that

(a) $E[\{\log |e_1|\}^+] < \infty$

and

(b) for some $n_0 \ge 1$, $U_{n0} = \sum_{j=0}^{n_0} \rho^j e_j$ has a non-trivial absolutely continuous component. Then for any initial distribution of Y_0 concentrated on a bounded set, Y_n is uniformly mixing iff there exists a finite constant c such that $|e_1| < c$.

2.2: The Central Limit Theorems for Strong Mixing Processes

In this section we mention some central limit theorems for strong mixing processes.

Theorem A:[14](Ibragimov)Let the process $\{X_j\}$ be centered and strong mixing with $E|X_j|^{2+\delta} < \infty$ for some $\delta > 0$. If

$$\sum_{n=1}^{\infty} \alpha(n)^{\delta/(2+\delta)} < \infty , \tag{6}$$

then

$$\sigma^2 = E(X_0^2) + 2 \sum_{n=1}^{\infty} E(X_0 X_j) < \infty , \tag{7}$$

and if $\sigma \neq 0$,then

$$\lim P\{\sigma^{-1} n^{-1/2} \sum_{j=1}^{n} X_j < z\} = \Phi(z).$$

Theorem B : [14](Ibragimov)Let the process $\{X_j\}$ be bounded, centered and strong mixing with

$$\sum_{n=1}^{\infty} \alpha(n) < \infty ; \tag{8}$$

then

$$\sigma^2 = E(X_0^2) + 2 \sum_{n=1}^{\infty} E(X_0 X_j) < \infty ,$$

and if $\sigma \neq 0$,then

$$\lim P\{\sigma^{-1} n^{-1/2} \sum_{j=1}^{n} X_j < z\} = \Phi(z).$$

2.3 : U-STATISTICS OF ABSOLUTELY REGULAR PROCESSES

Suppose that $\{X_i\}$ is a strictly stationary,absolutely regular process . Let $i_1 < i_2 < \ldots < i_k$ be arbitrary integers. For any j ($0 \le j \le k-1$), put

$$P_j^{(k)}(E^{(j)} \times E^{(k-j)}) = P((X_{i_1}, \ldots, X_{i_j}) \in E^{(j)}) P((X_{i_{j+1}}, \ldots, X_{i_k} \in E^{(k-j)})$$

and

$$P_0^{(k)}(E^{(k)}) = P(X_{i_1}, \ldots, X_{i_k}) \in E^{(k)}) ,$$

where $E^{(i)}$ is a Borel set in R^i .

Fundamental lemma : [29](Yoshihara) For any j ($0 \le j \le k-1$), let $h(x_1, \ldots, x_k)$ be a Borel function such that

$$\int \ldots \int |h(x_1, \ldots, x_k)|^{1+\delta} dP_j^{(k)} \le M$$

for some $\delta > 0$. Then

$$\left| \int \ldots \int |h(x_1, \ldots, x_k)|^{1+\delta} dP_0^{(k)} - \int \ldots \int |h(x_1, \ldots, x_k)|^{1+\delta} dP_j^{(k)} \right|$$

$$\le 4M^{1/(1+\delta)} \beta^{\delta/(1+\delta)}(i_{j+1} - i_j) . \tag{10}$$

The proof is simply based on a Rozanov and Volkonskiis' result in [22] and further,we will use their result to explain why it seems that it is impossible to extend this lemma to the strong mixing case. In 1961, Rozanov and Volkonskii found that

$$\beta(n) = \frac{1}{2} V(\varphi_n), \tag{11}$$

where $\varphi_n = P_{0n} - P_{1n}$ is a signed measure, P_{0n} is the measure induced by the process $\{X_n\}$ on the σ-algebra $\mathcal{M}_{-\infty}^{0} \cup \mathcal{M}_n^{\infty}$, P_{1n} is the measure defined for $A \in \mathcal{M}_n^{\infty}$,B $\in \mathcal{M}_{-\infty}^{0}$ by the equality

$$P_{1n}(AB) = P_{0n}(A)P_{0n}(B) ,$$

and $V(\varphi_n)$ is the total variation of φ_n on the whole space and is defined as in [31] (pp. 35 -37)

$$V(\varphi_n) = \overline{V}(\varphi_n) + \underline{V}(\varphi_n),$$

where

$$\overline{V}(\varphi_n) = \sup \{\varphi_n(B) \mid B \in \mathcal{M}_{-\infty}^{0} \cup \mathcal{M}_n^{\infty}\}$$

$$\underline{V}(\varphi_n) = \inf \{\varphi_n(B) \mid B \in \mathcal{M}_{-\infty}^{0} \cup \mathcal{M}_n^{\infty}\} .$$

A well known fact [31](pp.35-37) about $V(\varphi_n)$ is that

$$V(\varphi_n) = \sup_{|f| \le 1} \left| \int f(x)\varphi_n(dx) \right| \qquad (12)$$

where supremum runs over all such f(x) which are $\mathcal{M}_{-\infty}^{0} \cup \mathcal{M}_n^{\infty}$-measurable ,with bound 1. According to this fact and $\beta(n) = \frac{1}{2} V(\varphi_n)$, we obtain that

$$\beta(n) = \frac{1}{2} \sup_{|f| \le 1} \left| \int f(x)\varphi_n(dx) \right| .$$

In fact , there exists such $f_{n0}(x)$ that (see [31] pp.35-37)

$$\beta(n) = \frac{1}{2} \left| \int f_{n0}(x)\varphi_n(dx) \right| .$$

Replacing φ_n with $P_0^{(k)} - P_j^{(k)}$ and letting $f_{j0}(x) = |h_j(x_1,\dots,x_k)|$, we have

$$\beta(i_{j+1}-i_j) = \frac{1}{2}\left|\int\dots\int |h_j(x_1,\dots,x_k)|\,dP_0^{(k)} - \int\dots\int |h_j(x_1,\dots,x_k)|\,dP_j^{(k)}\right|. \quad (13)$$

So , when $\{X_n\}$ is not a absolutely regular process, that is, $\beta(i_{j+1}-i_j) \nrightarrow 0$ as $i_{j+1}-i_j \to \infty$, the right side of (13) $\nrightarrow 0$ for the above h_j . Thus, in order to guarantee that the right side of (13) tends to zero as $i_{j+1}-i_j \to \infty$ in the strong mixing case, we should restrict the class of h and pay attention to the structure of h. In next chapter we will deal with this situation.

According to the fundamental lemma Yoshihara [29] obtained the convergence rates of $E(R_n)^2$ where R_n is defined as before (see (14) or (16) in [1.2]).

Proposition 1 : Let $\{X_n\}$ is absolutely regular. If for some $\delta>0$ and $\varepsilon>0$ holds $\beta(n)^{\delta/(2+\delta)} = O(n^{-2+\varepsilon})$, then

$$E(R_n)^2 = O(n^{-2+\varepsilon}).$$

If we assume that for some $\delta>0$ $\sum \beta(n)^{\delta/(2+\delta)} < \infty$, then

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

Set

$$\sigma_n^2(h) = E\left(\sum_{i=1}^{n} \tilde{h}_1(X_i)\right)^2 \quad (14)$$

and

$$\sigma^2(h) = E(\tilde{h}_1(X_1))^2 + 2\sum_{i>1} E\tilde{h}_1(X_1)\tilde{h}_1(X_i). \quad (15)$$

Since $U_n - \theta = \hat{U}_n + R_n$ where θ and $\hat{U}_n$ are defined as before (see (1) and (13) in [1.2]), using this proposition and Theorem A [2.2] and noting $\beta(n) \le \alpha(n)$, one has

Theorem A : [29](Yoshihara) Let h be a non-degenerate kernel. Then

$$\frac{n}{m\sigma_n(h)}(U_n(h)-\theta) \to N(0,1),$$

provided the following condition (*) is satisfied :

(*) $\{X_n\}$ is absolutely regular with coefficients $\beta(n)^{\delta/(2+\delta)} = O(n^{-2+\varepsilon})$ for some $\delta>0$, $0\le\varepsilon<1$, $\sigma^2(h)>0$ and

$$\sup E\,|\,h(X_{i1},\ldots,X_{im})\,|^{2+\delta}<\infty,$$

where the supremum is over all choices of $1\le i_1<i_2<\ldots<i_m$.

The same statement holds for V-statistics when the above supremum is replaced by the supremum over all choices of $1\le i_j$ $(1\le j\le m)$.

Using Proposition 1 and Theorem 1 (b) in [6] (Dehling) one has the following theorems.

Theorem B : [7](Denker and Keller)Let h be a non-degenerate kernel, then

$$\Delta_n = \sup\left|\,P\left(\frac{n}{m\sigma_n(h)}(U_n-\theta)\le x\right)-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{t^2}{2}\right)dt\,\right| = O(n^{-\lambda}),$$

where $\lambda=\frac{(1-\varepsilon)\delta}{144}$, under the following condition (*):

(*) $\{X_n\}$ is absolutely regular with coefficients $\beta(n)$ satisfying $\beta(n)^{\delta/(2+\delta)} = O(n^{-2+\varepsilon})$ for for some $0<\delta\le 1$, $0\le\varepsilon<1$, $\sigma^2(h)>0$ and

$$\sup E\,|\,h(X_{i1},\ldots,X_{im})\,|^{2+\delta}<\infty,$$

where the supremum is over all choices of $1\le i_1<i_2<\ldots<i_m$.

The same statement holds for V-statistics when the supremum is replaced by the supremum over all choices of $1 \le i_j$ $(1 \le i \le m)$.

Denote by

$$\pi(P\,,\; Q) = \inf\{\varepsilon > 0 \mid P(A) \le Q(A^{\varepsilon}) + \varepsilon \text{ for all closed } A \subset \mathbf{R}\} \tag{16}$$

the Prohorov distance of two distributions P and Q on **R**, where $A^{\varepsilon} = \{x \in \mathbf{R} \mid \operatorname{dist}(x,A) < \varepsilon\}$.

Theorem C : [7](Denker and Keller)If the conditions of Theorem B hold, then

$$\gamma_n = \pi\left(L\left(\frac{n}{m\sigma_n(h)}(U_n - \theta)\right) \quad N(0,1)\right) = O(n^{-\lambda}),$$

where $\lambda = \frac{(1-\varepsilon)\delta}{144}$.

In order to formulate the functional central limit theorem one defines the D[0.1]-valued random functions ξ_n by

$$\xi_n(t) = \sqrt{n}\,\frac{t}{m\sigma(h)}(U_{[nt]} - \theta) \qquad (0 \le t \le 1), \tag{16}$$

where [nt] denotes the greatest integer not exceeding nt.

Using the basic lemma and Proposition 1, Denker and Keller (1983) obtained the following proposition.

Proposition 2 :[7](Denker and Keller) Let $\{X_n\}$ be absolutely regular with coefficients $\beta(n)$ satisfying $\beta(n)^{\delta/(2+\delta)} = O(n^{-2+\varepsilon})$ for some $\delta, \varepsilon > 0$, then

$R_n = O(n^{-3/4+\varepsilon/2}(\log n)^3)$ a.s.

and

$P(\max_{1 \le i \le n} i \mid R_n \mid \ge c_n) = O(n^{1/2+\varepsilon} c_n^{-2} (\log n)^3)$

for any sequence $c_j \in R^+$.

So the following theorems can be deduced from Theorem 1 in [5](Dehling) and Theorem 3 in [6](Dehling) together with Proposition 2.

Theorem D : [7] (Denker)Let h be a non-degenerate kernel, then ξ_n converges weakly in D[0,1] to the standard Wiener measure, if the following condition (*) is satisfied:

(*) $\{X_n\}$ is absolutely regular with coefficients $\beta(n)$ satisfying $\beta(n)^{\delta/(2+\delta)} = O(n^{-2+\varepsilon})$ for for some $0 < \delta \le 1$, $0 \le \varepsilon < 1$, $\sigma^2(h) > 0$ and

$\sup E\,|\,h(X_{i1},\ldots,X_{im})\,|^{2+\delta} < \infty$,

where the supremum is over all choices of $1 \le i_1 < i_2 < \ldots < i_m$.

The same statement holds for V-statistics when the supremum is replaced by the supremum over all choices of $1 \le i_j$ $(1 \le j \le m)$.

Theorem E : [7](Denker) Let h be a non-degenerate kernel. Under the following condition (*), then one can redefine $\{X_n\}$ without changing its distribution on a richer probability space together with a standard Brownian motion B(t) such that

$$\frac{n}{m}(U_n - \theta) - B(n) = O(n^{1/2-\lambda}) \text{ a.s. for some } \lambda > 0 :$$

(*) $\{X_n\}$ is absolutely regular with coefficients $\beta(n)$ satisfying $\beta(n)^{\delta/(2+\delta)} = O(n^{-2+\varepsilon})$ for for some $0 < \delta \le 1$, $0 < \varepsilon < 1/2$, $\sigma^2(h) > 0$ and

$\sup E\,|\,h(X_{i1},\ldots,X_{im})\,|^{2+\delta} < \infty$,

where the supremum is over all choices of $1 \le i_1 < i_2 < \ldots < i_m$.

The same statement holds for V-statistics when the supremum is replaced by the supremum over all choices of $1 \le i_j$ $(1 \le j \le m)$.

2.4 The Absolute Regularity Property of Linear Processes

Since the results in [2.3] exhibit the good limiting behavior of U- and V-statistics on absolutely regular processes, it is valuable to examine conditions for absolutely regular processes. Chanda [3] ,Gorodetskii [10] and Withers [28] discussed the conditions for linear processes to be strong-mixing. By examining their proofs, we find that these conditions also guarantee that the linear processes are absolutely regular.

Let Z_i ; $i = 0, \pm 1, \pm 2, \ldots$, be a sequence of independent random variables with characteristic functions φ_i and probability densities $p_i(x)$, and let $\{g_k\}$, $k = 0,1,2,\ldots$,be a certain sequence of numbers: $g_0 \neq 0$. Define

$$X_j = \sum_{i=0}^{\infty} g_i Z_{j-i} ,$$

where the identity is to be understood in the sense of convergence in distribution. Use the notation

$$S_i(\delta) = \sum_{j=i}^{\infty} |g_j|^{\delta} ,$$

$$\beta^*(k) = \sum_{i=k}^{\infty} (S_i(\delta))^{1/(1+\delta)} , \text{ if } \delta < 2 ,$$

$$= \{ \sum_{i=k}^{\infty} \max (S_i(\delta))^{1/(1+\delta)} , \sqrt{S_i(2) \left| \log S_i(2) \right|} \} , \text{ if } \delta \ge 2 .$$

Let

(i) $$\int_{-\infty}^{\infty} | p_i(x) - p_i(x+\alpha) | dx \le C_1 |\alpha| ;$$

(ii) $E|Z_i|^{\delta} \le C_2 < \infty$ for some $\delta>0$; if $\delta \ge 1$, then we assume that $EZ_i = 0$, and if $\delta \ge 2$,that $\mathrm{Var}Z_i = 1$;

(iii) $$g(z) = \sum_{k=0}^{\infty} \gamma_k z^k \neq 0 \text{ for } |z| \le 1 ;$$

(iv) $\beta^*(0) < \infty$.

Under the above conditions, [illegible]detskii [10] proved that $\{X_i\}$ is strong mixing with $\alpha(k) \le M\beta^*(k)$, where M is a certain constant. We will prove that under the same conditions, $\{X_i\}$ is absolutely regular with $\beta(k) \le M\beta^*(k)$, where M is a certain constant.

Theorem F: If (i),(ii),(iii),and (iv) hold , then $\{X_i\}$ is absolutely regular with $\beta(k) \le M\beta^*(k)$, where M is a certain constant.

Lemma : For i=1,2,... , let $A_i \in \mathcal{M}_{-\infty}^{m}$, $B_i \in \mathcal{M}_{n+m}^{\infty}$ where $\mathcal{M}_{-\infty}^{m}$ is the σ-algebra induced by

$$\{X_j | -\infty \le j \le m\}$$

and $\mathcal{M}_{n+m}^{\infty}$ is the σ-algebra induced by

$$\{X_j | n+m \le j \le \infty\}.$$

If

$$\sum_{i=1}^{r} | P(A_i B_i) - P(A_i)P(B_i) | \le \beta^+(n)$$

for any integer $r>0$ and a sequence $\{\beta^+(n), n=1,2,\dots\}$, where

$$A_i \cap A_j = \emptyset , \ i \neq j ,$$

and

$$\beta^+(n) \to 0, \text{ as } n \to \infty .$$

Then $\{X_i\}$ is absolutely regular with coefficient $\beta^+(n)$.

PROOF. Since

$$\beta(n) = \frac{1}{2} V(\varphi_n) \text{ ,(see (11) [2.3])}$$

where

$$\phi_n(A_iB_i) = P(A_iB_i) - P(A_i)P(B_i),$$

we need to prove

$$V(\varphi_n) \le \beta^+(n).$$

Since

$$V(\varphi_n) = \sup_{|f| \le 1} \left| \int f(x)\varphi_n(dx) \right| \text{ , (see (12) [2.3])}$$

taking simple function

$$f(x) = \sum_{i=1}^{r} c_i I_{[A_iB_i]}(x),$$

where

$$|c_i| \le 1 \text{ , } 1 \le i \le r,$$

we have

$$\left| \int f(x)\varphi_n(dx) \right| \le \beta^+(n) .$$

Since any f(x) with bound 1, which is $\mathcal{M}_{-\infty}^{m} \cup \mathcal{M}_{n+m}^{\infty}$ - measurable, can be approximated by such simple functions, we get

$$\sup_{|f|\le 1} \left| \int f(x)\varphi_n(dx) \right| \le \beta^+(n).$$

PROOF OF THEOREM F.

Let A_i $(1\le i \le r)$ be any (p+1)-dimensional Borel set in the space of $X_{-p},\ldots,X_0$ and B_i $(1\le i\le r)$ be any m-dimensional Borel set in the space of $X_k,\ldots,X_{k+m-1}$ and

$$A_i \cap A_j = \emptyset\ , \ i \ne j$$

Write

$$X = (X_{-p},\ldots,X_0),$$

$$Y = (X_k,\ldots,X_{k+m-1}),$$

$$W_t = \sum_{j=0}^{t-1} g_j Z_{t-j}\ , \ k\le t\le k+m-1,$$

$$V_t = X_t - W_t = \sum_{j=t}^{\infty} g_j Z_{t-j}.$$

Let

$$V = (V_k,\ldots,V_{k+m-1}),$$

$$W = (W_k,\ldots,W_{k+m-1}),$$

so that $Y = W + V$, x,y,v,w are the values taken by the random vectors X,Y,V,W and

$$C = \{v : |v_t| \le \eta_t,\ k\le t\le k+m-1\}.$$

Then

$$P\{(X ,Y)\in \sum_{i=1}^{r} A_iB_i\}$$

$$= \sum_{i=1}^{r} P(X\in A_i ,Y\in B_i)$$

$$= \sum_{i=1}^{r} \{P(X\in A_i ,W + V\in B_i,V\in C)+P(X\in A_i ,W + V\in B_i,V\in C^c)\}.$$

Since W is independent of X and V,

$$P(X\in A_i, W+V\in B_i, V\in C)$$

$$= \int_{A_i\times C} P(W\in B_i - v)dF(x,v),$$

where F(x,v) is the joint distribution function of (X,V).

Let

$$\chi_1 = \inf_{v\in C} P(W\in B_i - v),$$

$$\chi_2 = \sup_{v\in C} P(W\in B_i - v),$$

thus

$$P(X\in A_i, Y\in B_i)$$

$$\geq \chi_1 P(X\in A_i, V\in C^c)+P(X\in A_i, W+V\in B_i, V\in C^c)$$

$$=\chi_1 P(X\in A_i)-\chi_1 P(X\in A_i C^c)+P(X\in A_i, W+V\in B_i, V\in C^c)$$

$$\geq\chi_1 P(X\in A_i)-P(X\in A_i, V\in C^c)$$

and

$$P(X\in A_i, Y\in B_i)$$

$$\leq \chi_2 P(X\in A_i)+P(X\in A_i, W+V\in B_i, V\in C^c)$$

$$\leq\chi_2 P(X\in A_i)+P(X\in A_i, V\in C^c),$$

that is,

$$\chi_1 P(X\in A_i)-P(X\in A_i, V\in C^c)\leq P(X\in A_i, Y\in B_i)\leq\chi_2 P(X\in A_i)+P(X\in A_i, V\in C^c). \quad (18)$$

Take A_i to be the entire space of $\{X_{-p},\ldots,X_0\}$ so that (18) leads to

$$\chi_1 - P(V\in C^c)\leq P(Y\in B_i)\leq\chi_2 + P(V\in C^c).$$

So

$$\chi_1 P(X\in A_i)-P(X\in A_i)P(V\in C^c)$$

$$\leq P(Y\in B_i)P(X\in A_i)$$

$$\leq \chi_2 P(X\in A_i)+P(X\in A_i)P(V\in C^c).$$

It follows that

$$P(X\in A_i, Y\in B_i)-P(Y\in B_i)P(X\in A_i)$$

$$\le\chi_2 P(X\in A_i)-\chi_1 P(X\in A_i)+P(X\in A_i)P(V\in C^c)+P(X\in A_i, V\in C^c),$$

$$P(X\in A_i, Y\in B_i)-P(Y\in B_i)P(X\in A_i)$$

$$\ge \chi_1 P(X\in A_i)-\chi_2 P(X\in A_i)-(P(X\in A_i, V\in C^c)+P(X\in A_i)P(V\in C^c))$$

$$\ge-\{(\chi_2-\chi_1)P(X\in A_i)+P(X\in A_i, V\in C^c)+P(X\in A_i)P(V\in C^c)\},$$

that is,

$$|P(X\in A_i, Y\in B_i)-P(Y\in B_i)P(X\in A_i)|$$

$$\le(\chi_2-\chi_1)P(X\in A_i)+(P(X\in A_i, V\in C^c)+P(X\in A_i)P(V\in C^c))$$

$$\le\sup_{v\in C}|P(W\in B_i-v)-P(W\in B_i)|P(X\in A_i)$$

$$+\{P(X\in A_i, V\in C^c)+P(X\in A_i)P(V\in C^c)\}.$$

So

$$\sum_{i=1}^{r}|P(X\in A_i, Y\in B_i)-P(Y\in B_i)P(X\in A_i)|$$

$$\le\sup_{v\in C}|P(W\in B_i-v)-P(W\in B_i)|\sum_{i=1}^{r}P(X\in A_i)$$

$$+\sum_{i=1}^{r}\{P(X\in A_i, V\in C^c)+P(X\in A_i)P(V\in C^c)\}$$

$$\le\sup_{v\in C}|P(W\in B_i-v)-P(W\in B_i)|+P(V\in C^c).$$

Since Gorodetskii has showed [10] that

$$\sup_{v\in C}|P(W\in B_i-v)-P(W\in B_i)|+P(V\in C^c)\le M\beta^*(k),$$

where M is independent of B_i. So the result $\beta(k)\le M\beta^*(k)$ follows from the above lemma.

CHAPTER III

Central Limit Theorems for a Special Class of U-statistics for Strong Mixing

This chapter is devoted to the weak convergence of U-statistics on strong mixing processes. A special class of U-statistics with "decomposable" kernels is introduced. For this class, central limit theorems and their rate of convergence, functional central limit theorems and a.s. approximation by a Brownian motion are established.

3.1 : Main Results

A kernel h is called "decomposable" if h possesses the following form:

$$h(x_1,\dots,x_m) = \sum_{i_1=1}^{\infty}\cdots\sum_{i_m=1}^{\infty} a_{i_1\dots i_m} f^1_{i_1}(x_1)\dots f^m_{i_m}(x_m) \qquad (1)$$

$$+ \sum_{\tau(m-1)}\sum_{i_1=1}^{\infty}\cdots\sum_{i_{m-1}=1}^{\infty} a_{i_1\dots i_{(m-1)}} g^1_{i_1}(x_{\tau_1^{(m-1)}})\dots g^{m-1}_{i_{m-1}}(x_{\tau_{(m-1)}^{(m-1)}})$$

$$+ \dots$$

$$+ \sum_{\tau}\sum_{i=1}^{\infty} a_i \phi_i(x_\tau),$$

where $\sum_{\tau(m-1)}$ denotes summation over the $\binom{m}{m-1}$ combinations of m-1 distinct elements $\{\tau_1^{(m-1)},\dots,\tau_{m-1}^{(m-1)}\}$ from $\{1,\dots,m\}$, ... $\sum_{\tau}$ denotes summation over the $\binom{m}{1}$ combinations of 1 distinct elements $\{\tau.\}$ from $\{1,\dots,m\}$ and

$$\sum_{i_1=1}^{\infty}\cdots\sum_{i_m=1}^{\infty}|a_{i_1\ldots i_m}|<\infty, \tag{2}$$

$$\sum_{i_1=1}^{\infty}\cdots\sum_{i_{m-1}=1}^{\infty}|a_{i_1\ldots i_{(m-1)}}|<\infty,$$

...

$$\sum_{i=1}^{\infty}|a_i|<\infty.$$

Theorem A : Let the process $\{X_n\}$ be strong mixing and h be a bounded non-degenerate kernel . Suppose that h is "decomposable" and

$$\sup_{i,j}|f^j_{i_j}|<\infty \quad \text{(a.e)},$$

$$\sup_{i,j}|g^j_{i_j}|<\infty \quad \text{(a.e)},$$

...,

$$\sup_i|\phi_i|<\infty \quad \text{(a.e)}. \tag{3}$$

If $\alpha(n)=O(n^{-2+\gamma})$ for $0\le\gamma<1$, and $\sigma(h)>0$ (see (15) [2.3])then

$$\frac{n^{1/2}}{m\sigma(h)}(U_n(h)-\theta)\rightarrow N(0,1). \tag{4}$$

When h is unbounded, a central limit theorem can still be proved for general m without any real difficulty in spirit, but we have to add some moment conditions about $f^{(j)}_{i_j}(1\le j\le m)$ and impose further conditions on the convergence rate of $\alpha(n)$. For simplicity, we only present a theorem in the case m=2. In this case, if h is "decomposable" then h can be expressed with a slight change in notation in the following form:

$$h(x_1,x_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j} f_i(x_1) g_j(x_2) + \sum_{i=1}^{\infty} a_i(\phi_i(x_1) + \phi_i(x_2)) \tag{5}$$

where

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{i,j}| < \infty, \sum_{i=1}^{\infty} |a_i| < \infty. \tag{6}$$

Theorem B : Let the process $\{X_n\}$ be strong mixing and h be a non-degenerate kernel with $E|h^{2+\mu}|<\infty$ for some $\mu \geq \delta/2$ $(0<\delta)$. Suppose that h is "decomposable" and

$$\sup_i E|f_i|^{4+\delta} < \infty, \sup_j E|g_j|^{4+\delta} < \infty, \tag{7}$$

$$\sup_i E|\phi_i| < \infty.$$

If $\alpha(n)^{\delta/(4+\delta)} = O(n^{-2+\gamma})$ for $0 \leq \gamma < 1$, and $\sigma(h)>0$ then

$$\frac{n^{1/2}}{2\sigma(h)}(U_n(h) - \theta) \to N(0,1). \tag{8}$$

Example 3.1:

Obviously, when $h(x_1,\dots,x_m)$ is a symmetric polynomial about $x_1,\dots,x_m$, it is "decomposable" . For example, the kernel $h(x_1,x_2)= \frac{1}{2}(x_1-x_2)^2$ is "decomposable" and the corresponding U-statistic is the sample variance S^2(see 1.1). So, if $E|X|^r<\infty (r>4)$ then

$$\frac{n^{1/2}}{(\mu_4 - \theta^2)}(S^2 - \theta) \to N(0,1),$$

where $\mu_4 = EX^4$ and $\theta = Var(X)$.

Theorem C : If the conditions of theorem B are satisfied and $\alpha(n)^{\delta/(4+\delta)} = O(n^{-2+\gamma})$ with $0 \leq \gamma < 1$ and $0<\delta\leq 2$,then

$$\Delta_n = \sup_x \left| P\left(\frac{n}{m\sigma_n(h)}(U_n - \theta) \leq x\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt \right| = O(n^{-\lambda})$$

where $\lambda=\frac{\delta(1-\gamma)}{400(3-\gamma)}$ and $\sigma_n(h)$ is defined as (14) [2.3].

Theorem D : If the conditions of theorem B are satisfied and $\alpha(n)^{\delta/(4+\delta)} = O(n^{-2+\gamma})$ with $0 \le \gamma<1$ and $0<\delta\le 2$,then

$$\gamma_n= \pi\left(L(\frac{n}{m\sigma_n(h)}(U_n-\theta)) \quad N(0,1)\right)= O(n^{-\lambda})$$

where $\lambda=\frac{\delta(1-\gamma)}{400(3-\gamma)}$ and π is the Prohorov Distance (see (16)[2.3]) .

In order to formulate the functional central limit theorem let us define the D[0,1]-valued random functions ξ_n by

$$\xi_n (t)= \frac{n^{1/2}t}{2\sigma(h)} (U_{[nt]} - \theta) \qquad (0\le t\le 1)$$

where [nt] denotes the greatest integer not exceeding nt.

Theorem E : If the conditions of theorem B are satisfied and $\alpha(n)^{\delta/(4+\delta)} = O(n^{-2+\gamma})$ with $0<\gamma<\frac{1}{2}$, then $\xi_n(t)$ converges weakly in D[0,1] to the standard Wiener measure.

Theorem F : If the conditions of theorem E are satisfied, then we can redefine $\{X_n\}$ without changing its distribution on a richer probability space together with a standard Brownian B(t) such that

$$\frac{n}{2}(U_n - \theta)- B(n) = o((n\log\log n)^{1/2}) \text{ a.s. }.$$

In order to prove the above theorems we need the following propositions.

Proposition 1 : Let h be a bounded non-degenerate kernel . Suppose that h is "decomposable" and

$$\sup_{i,j} | \xi_{ij}^{(j)} | < \infty \quad \text{(a.e)},$$

$\sup_{i,j} |g_{i_j}^{(j)}| < \infty$ (a.e),

....,

$\sup_i |\phi_i| < \infty$ (a.e).

If $\alpha(n)=O(n^{-2+\gamma})$ with $0<\gamma<1$, then

$$E(R_n)^2 \le n^{-2+\gamma}C. \tag{10}$$

If $\alpha(n) \to 0$ as $n\to\infty$, then

$$E(R_n)^2 \to 0.$$

Proposition 2 : Let h be a non-degenerate kernel. Suppose that h is "decomposable" and

$\sup_i E|f_i|^{4+\delta} < \infty$, $\sup_j E|g_j|^{4+\delta} < \infty$,

$\sup_i E|\phi_i| < \infty$,

for some $\delta>0$. If $\alpha(n)^{\delta/(4+\delta)} = O(n^{-2+\gamma})$ with $0<\gamma<1$, then

$$E(R_n)^2 \le n^{-2+\gamma}C.$$

3.2 : Proofs

In order to prove the above propositions we need the following lemma. For simplicity, we use C as a generic constant.

Lemma 1: [13] If ξ is $\mathcal{M}_1^n$ - measurable and η is $\mathcal{M}_{m+n}^{\infty}$ -measurable, then

(i) $|E\xi\eta - E\xi E\eta| \le 12\,\alpha(m)^{1/c}E^{1/a}|\xi|^a E^{1/b}|\eta|^b$,

where $\frac{1}{c}+\frac{1}{a}+\frac{1}{b} = 1$.(ii) Especially, when $|\xi| \le C_1$ and $|\eta| \le C_2$ we have

$$|E\xi\eta - E\xi E\eta| \le 16C_1C_2\alpha(m).$$

PROOF OF PROPOSITION 1 : Suppose $\alpha(n)=O(n^{-2+\gamma})$. For simplicity, we only consider the case m=2. The proof for general m is analogous and so is omitted. Since

$$h(x_1,x_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j} f_i(x_1) g_j(x_2) + \sum_{i=1}^{\infty} a_i(\phi_i(x_1) + \phi_i(x_2))$$

and

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{i,j}| < \infty, \sum_{i=1}^{\infty} |a_i| < \infty,$$

so we have (see (7)-(8)[1.3])

$$h_1(x_1) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j} f_i(x_1) Eg_j + \sum_{i=1}^{\infty} a_i(\phi_i(x_1) + E\phi_i), \quad (11)$$

$$\tilde{h}_1(x_1) = h_1(x_1) - \theta = h_1(x_1) - Eh_1$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j} f_i(x_1)(Eg_j - Ef_i) + \sum_{i=1}^{\infty} a_i(\phi_i(x_1) - E\phi_i), \quad (12)$$

and (see (17) [1.3])

$$H = h(x_1,x_2) - \theta - \tilde{h}_1(x_1) - \tilde{h}_1(x_2)$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j}(f_i(x_1) - Ef_i)(g_j(x_2) - Eg_j). \quad (13)$$

Thus (see (16) [1.3])

$$R_n = \frac{1}{\binom{n}{2}} \sum_{i<j} \sum_{s=1}^{\infty} \sum_{t=1}^{\infty} a_{s,t}(f_s(X_i) - Ef_s)(g_t(X_j) - Eg_t). \quad (14)$$

Put

$$J_{s,t,u,v}((i_1,i_2),(j_1,j_2))$$

$$= a_{s,t}a_{u,v}E[(f_s(X_{i1}) - Ef_s)(g_t(X_{i2}) - Eg_t)(f_u(X_{j1}) - Ef_u)(g_v(X_{j2}) - Eg_v)].$$

So, we have

$$E(R_n)^2 = \frac{1}{\binom{n}{2}^2} \sum_{u=1}^{\infty} \sum_{v=1}^{\infty} \sum_{s=1}^{\infty} \sum_{t=1}^{\infty} \sum_{i_1<i_2} \sum_{j_1<j_2} J_{s,t,u,v}((i_1,i_2),(j_1,j_2)). \tag{15}$$

Thus, from the preceding Lemma part (ii) we have the following inequalities:

(i) If $1 \le i_1 < i_2 \le j_1 < j_2 \le n$ and $j_2 - j_1 \ge i_2 - i_1$, then

$$|J_{s,t,u,v}((i_1,i_2),(j_1,j_2))|$$

$$= |Ea_{s,t}a_{u,v}(f_s(X_{i_1}) - Ef_s)(g_t(X_{i_2}) - Eg_t)(f_u(X_{j_1}) - Ef_u)(g_v(X_{j_2}) - Eg_v)|$$

$$\le |Ea_{s,t}a_{u,v}\xi\eta|$$

$$\le |a_{s,t}a_{u,v}| C\alpha(j_2 - j_1) \tag{17}$$

where

$$\xi = (f_s(X_{i_1}) - Ef_s)(g_t(X_{i_2}) - Eg_t)(f_u(X_{j_1}) - Ef_u) \tag{18}$$

and

$$\eta = (g_v(X_{j_2}) - Eg_v). \tag{19}$$

Similarly, if $1 \le i_1 < i_2 \le j_1 < j_2 \le n$ and $i_2 - i_1 \ge j_2 - j_1$, then

$$|J_{s,t,u,v}((i_1,i_2),(j_1,j_2))| \le |a_{s,t}a_{u,v}| C\alpha(i_2 - i_1). \tag{20}$$

So,

$$\Big| \sum_{1 \le i_1 < i_2 \le j_1 < j_2 \le n} J_{s,t,u,v}((i_1,i_2),(j_1,j_2)) \Big|$$

$$\le \sum_{1 \le i_1 < i_2 \le j_1 < j_2 \le n \text{ and } i_2 - i_1 \ge j_2 - j_1} |J_{s,t,u,v}((i_1,i_2),(j_1,j_2))|$$

$$+ \sum_{1 \le i_1 < i_2 \le j_1 < j_2 \le n \text{ and } j_2 - j_1 \ge i_2 - i_1} |J_{s,t,u,v}((i_1,i_2),(j_1,j_2))|$$

$$\le |a_{s,t}a_{u,v}| Cn^2 \sum_{k=1}^{n} (k+1)\alpha(k) = |a_{s,t}a_{u,v}| O(n^{2+\gamma}). \tag{21}$$

(ii) Similarly, we have

$$\left| \sum_{1 \le i_1 < j_1 \le i_2 < j_2 \le n} J_{s,t,u,v}((i_1,i_2),(j_1,j_2)) \right|$$

$$\le \sum_{1 \le i_1 < j_1 \le i_2 < j_2 \le n \text{ and } j_1 - i_1 \ge j_2 - i_2} |J_{s,t,u,v}((i_1,i_2),(j_1,j_2))|$$

$$+ \sum_{1 \le i_1 < j_1 \le i_2 < j_2 \le n \text{ and } j_2 - i_2 \ge j_1 - i_1} |J_{s,t,u,v}((i_1,i_2),(j_1,j_2))|$$

$$\le |a_{s,t}a_{u,v}| O(n^{2+\gamma}), \tag{22}$$

$$\sum_{1 \le i_1 < j_1 < j_2 < i_2 \le n} J_{s,t,u,v}((i_1,i_2),(j_1,j_2)) |$$

$$\le \sum_{1 \le i_1 < j_1 < j_2 < i_2 \le n \text{ and } j_1 - i_1 \ge i_2 - j_2} |J_{s,t,u,v}((i_1,i_2),(j_1,j_2))|$$

$$+ \sum_{1 \le i_1 < j_1 < j_2 < i_2 \le n \text{ and } i_2 - j_2 \ge j_1 - i_1} |J_{s,t,u,v}((i_1,i_2),(j_1,j_2))|$$

$$\le |a_{s,t}a_{u,v}| O(n^{2+\gamma}), \tag{23}$$

$$\left| \sum_{1 \le i_1 \; j_1 \le n} \sum_{i_2=1}^{n} J_{s,t,u,v}((i_1,i_2),(j_1,i_2)) \right|$$

$$\le \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} |J_{s,t,u,v}((i_1,i_2),(i_1,i_2))| + 2 \sum_{1 \le i_1 < j_1 \le n} \sum_{i_2=1}^{n} |J_{s,t,u,v}((i_1,i_2),(j_1,i_2))|$$

$$\le |a_{s,t}a_{u,v}| Cn^2 (1 + \sum_{k=1}^{n} \alpha(k)) = O(n^2), \tag{24}$$

and

$$\Big|\sum_{1\le i_2\ j_2\le n}\sum_{i_1=1}^{n} J_{s,t,u,v}((i_1,i_2),(i_1,j_2))\Big| \le O(n^2). \tag{25}$$

Since

$$\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}\sum_{s=1}^{\infty}\sum_{t=1}^{\infty}|a_{s,t}a_{u,v}| < \infty, \tag{26}$$

hence from (21)-(26) and (15), we have (10). If $\alpha(n)\to 0$ as $n\to\infty$, similarly we can obtain

$$E(R_n)^2 \to 0\,.$$

PROOF OF PROPOSITION 2 : The proof is the same as the above proof , if we note that when $1\le i_1<i_2\le j_1<j_2\le n$ and $j_2-j_1\ge i_2-i_1$, we have

$$|J_{s,t,u,v}((i_1,i_2),(j_1,j_2))|$$

$$= |Ea_{s,t}a_{u,v}[(f_s(X_{i1})-Ef_s)(g_t(X_{i2})-Eg_t)(f_u(X_{j1})-Ef_u)][(g_v(X_{j2})-Eg_v)]|$$

$$= |Ea_{s,t}a_{u,v}\xi\eta| \text{ (see (18)–(19))}$$

$$\le |a_{s,t}a_{u,v}|\,C\alpha(j_2-j_1)^{\delta/(4+\delta)}, \tag{27}$$

using Lemma 1 part (i) in [4.2].

PROOF OF THEOREM A : From Theorem B in [2.2] we have

$$\frac{n^{1/2}\,\hat{U}_n}{m\sigma(h)} = \frac{1}{n^{1/2}\sigma(h)}\sum_{i=1}^{n}\tilde{h}_1(X_i) \to N(0,1).$$

Since $U_n-\theta = \hat{U}_n + R_n$,thus the result follows from Proposition 1.

PROOF OF THEOREM B : Recall the definition of $\tilde{h}_1$ (see (7)-(8) [1.2]) and $E|h^{2+\mu}|<\infty$, so $E|\tilde{h}_1|^{2+\mu} \leq E|h-\theta|^{2+\mu}<\infty$. Since $\sum_1^\infty \alpha(n)^{\delta/(4+\delta)}<\infty$, $\sum_1^\infty \alpha(n)^{\mu/(2+\mu)}<\infty$ for $\mu\geq\delta/2$. Thus we have

$$\frac{n^{1/2}\hat{U}_n}{m\sigma(h)}=\frac{1}{n^{1/2}\sigma(h)}\sum_{i=1}^{n}\tilde{h}_1(X_i)\rightarrow N(0,1)$$

from Theorem A [2.2]. Since $U_n-\theta=\hat{U}_n+R_n$, thus the result follows from Proposition 2.

PROOF OF THEOREM D: The result can be deduced from Theorem 1 (a) in [6] together with Proposition 2.

PROOF OF THEOREM C: The result follows from Theorem D.

In order to prove Theorem E and Theorem F, we need the following proposition.

Proposition 3: Let h be a non-degenerate kernel. Suppose that h is "decomposable" and

$\sup_i E|f_i|^{4+\delta}<\infty$, $\sup_j E|g_j|^{4+\delta}<\infty$,

$\sup_i E|\phi_i|<\infty$,

for some $0<\delta$. If $\alpha(n)^{\delta/(4+\delta)}=O(n^{-2+\gamma})$ with $0<\gamma<1$, then

$$R_n=O(n^{-(3/4)+\gamma/2}(\log n)^3)\quad \text{a.s.}$$

and

$$P(\max_{1\leq i\leq n} i|R_n|\geq c_n)=O(n^{(1/2)+\gamma}c_n^{-2}(\log n)^3)$$

for any sequence $c_j\in R^+$.

In order to prove Proposition 3, we need the following lemma.

Lemma 2 : Let

$$Z(p,q)=\binom{p+q}{2}R_{p+q}-\binom{p}{2}R_p.$$

Then under the conditions of Proposition 3, we have

$$E(Z(p,q))^2= O(q(p+q)^{3/2+\gamma}).$$

PROOF: The proof is similar to the proof in [7]. If $q>p$, it follows from Proposition 2 that

$$E(Z(p,q))^2\le \{E[(Z(0,p))^2]^{1/2} + E[(Z(0,p+q))^2]^{1/2}\}^2$$

$$=O((p+q)^{2+\gamma})=O(q(p+q)^{1+\gamma}).$$

If $p\ge q\ge p^{1/2}$ we obtain similarly

$$E(Z(p,q))^2=O((p+q)^{2+\gamma})=O(q(p+q)^{3/2+\gamma}.$$

Now, if $q<p^{1/2}$ we have

$$E(Z(p,q))^2= E(\sum_{k=0}^{q-1} Z(p+k,1))^2\le \{\sum_{k=0}^{q-1} [E(Z(p+k,1))^2]^{1/2}\}^2.$$

Using the same technique as the proof of Proposition 2, we can prove

$$E(Z(p+k,1))^2= O((p+q)^{1+\gamma}),$$

thus

$$\{\sum_{k=0}^{q-1} [E(Z(p+k,1))^2]^{1/2}\}^2=\sum_{k=0}^{q-1} O((p+q)^{1+\gamma})= O(q^2(p+q)^{1+\gamma})=O(q(p+q)^{3/2+\gamma}).$$

This finishes the proof.

PROOF OF PROPOSITION 3 : Using Lemma 2 we can prove the result just like the proof of proposition 3 of [7].

PROOF OF THEOREM E : Write

$$\xi_n(t)= \frac{n^{1/2}t}{2\sigma(h)} (\hat{U}_{[nt]} +R_{[nt]})$$

$$= \frac{n^{1/2}t}{\sigma(h)} \sum_{s=1}^{[nt]} \tilde{h}_1(x_s) + \frac{n^{1/2}t}{2\sigma(h)} R_{[nt]}.$$

By Theorem 1 in [5], $\frac{n^{1/2}t}{\sigma(h)} \sum_{s=1}^{[nt]} \tilde{h}_1(x_s)$ converges weakly in D[0,1] to the standard Wiener measure. The maximal inequality of Proposition 3 shows that

$$P(\max_{0\le t\le 1} \frac{n^{1/2}t}{2\sigma(h)} |R_{[nt]}| \ge \varepsilon) \le P(\max_{0\le t\le 1} [nt] |R_{[nt]}| \ge \sigma(h)\varepsilon n^{1/2}) = O(n^{-(1/2)+\varepsilon}).$$

Hence the result follows.

PROOF OF THEOREM F : The result can be deduced from Theorem 4 in [6] together with Proposition 3 just as before.

CHAPTER IV

STRONG AND WEAK CONSISTENCY OF U-STATISTICS FOR STRONG MIXING PROCESSES

This chapter is devoted to the weak and strong consistency of U-statistics for strong mixing processes. As we have pointed out in Chapter 2 there is much difficulty in extending the central limit theorem to strong mixing processes because the basic approximation lemma of Yoshihara [29] for absolutely regular processes is no longer valid. This is why in Chapter 3 we were able to obtain the central limit theorem for only a special class of U-statistics . But when considering the consistency properties of U-statistics, we find that these properties hold for rather big classes of U-statistics. To the best of our knowledge, our results derived below are new. Our main result is that $U_n \rightarrow \theta$ (a.s.) if (i) the kernel $h(x_1,...x_m)$ is a.e. bounded and a.e continuous w.r.to Lebesgue measure on R^m , and (ii) the distribution F of X_1 is absolutely continuous w.r.to Lebesgue measure m on R, and the probability of the closed hull of non-continuity set of h is zero. For unbounded kernels h, however, we are able to prove only weak consistency under similar conditions.

4.1: STRONG CONSISTENCY OF U-STATISTICS

In this chapter we use C as a generic constant and P_0 as the "independent" measure corresponding to P, i.e.,generated by $\{X_i\}$ with i.i.d. r.v.'s X_i $i=1,2,\cdots$.

Lemma 1 : If h is a bounded continuous kernel on $\mathbf{R}^m$, then for any given $\{\varepsilon_t\}\downarrow 0$, as $t\to\infty$, and a compact set $\{K_t^m\}$ in $\mathbf{R}^m$, there exists a symmetric bounded function g_t, which is continuous on K_t^m and "decomposable" (see (1) in (3.1)), such that

$$\sup_{x\in K_t^m} |g_t(x)-h(x)| \le \varepsilon_t/4.$$

PROOF : From the Weierstrass's approximation theorem [17](pp.33) there exists a polynomial p_t such that

$$\sup_{x\in K_t^m} |p_t(x)-h(x)| \le \varepsilon_t/4.$$

Define

$$g_t(x_1,\dots,x_m) = \frac{1}{m!}\sum_{c} p_t(x_{i_1},\dots,x_{i_m}) \quad \text{for} \quad x=(x_1,\dots,x_m)\in K_t^m \tag{1}$$

$$g_t(x_1,\dots,x_m) = 0 \qquad \text{for} \qquad x\in (K_t^m)^c,$$

where $\sum_c$ denotes summation over the m! permutations $(i_1,\dots,i_m)$ of $(1,\dots,m)$. Since h is symmetric, $\sup_{x\in K_t^m} |p_t(x_{i_1},\dots,x_{i_m})-h(x_1,\dots,x_m)| \le \varepsilon_t/4$. Thus we get the g_t just what we need .

Lemma 2 : If h is an a.e bounded variable with $Eh(X_1) = 0$ and $\sum_{i=1}^{n}\alpha(i) = O(n^{1-\lambda})$ $(0<\lambda<1)$,then $E(S_n)^4 \le Cn^{3-\lambda}$ where $S_n = \sum_{i=1}^{n} h(X_i)$.

PROOF : The result follows from the fact that $E(S_n)^4 \le Cn^2 \sum_{i=1}^{n}\alpha(i)$ [see Ibragimov and Linnik[14] or Billingsley[2]].

Recall from [1.2]

$$U_n(h)-\theta = \hat{U}_n(h)+R_n(H)$$

where

$$R_n(H)=\frac{1}{\binom{n}{m}}\sum_{i_1<\dots<i_m}^{n} H(X_{i1},\dots,X_{im})$$

and

$$H(X_1,\dots,X_m)=h(X_1,\dots,X_m)-\tilde{h}_1(X_1)\dots-\tilde{h}_1(X_m)-\theta,$$

we have the following results.

Proposition 1 : If h is a bounded and continuous kernel on $\mathbf{R}^m$ with $E|X_1|<\infty$

and

$$\alpha(n)=O(n^{-(1+\gamma)})\ (0<\gamma<1)\ ,$$

then

$$R_n(H)\rightarrow 0\ (a.s.).$$

The proof of Proposition 1 is given below.

Theorem A : If the conditions of proposition 1 hold, then $U_n(h)-\theta\rightarrow 0$ (a.s.).

PROOF : Since $U_n(h)-\theta=\hat{U}_n(h)+R_n(H)$, the result follows from Proposition 1 and the well-known fact that $\hat{U}_n(h)\rightarrow 0$ (a.s.) in view of the strong law of large numbers for strong mixing sequences(see Ibragimov, I. A. and Linnik, Yu. V. [14]).

Proposition 2 : Let h be a.e bounded and a.e continuous w.r. to Lebesgue measure on $\mathbf{R}^m$ with its non-continuity set denoted by D. Let F be absolutely continuous w.r.to Lebesgue measure on $\mathbf{R}$, $E|X_1|<\infty$ and

$$\alpha(n)=O(n^{-(1+\gamma)})\ (0<\gamma<1)\ .$$

If for any $\{\varepsilon_t\}\downarrow 0$, as $t\rightarrow\infty$, there exist an integer $r_t>0$ and Lebesgue measurable sets B^j_{st} in $\mathbf{R}$ ($1\le s\le r_t$, $1\le j\le m$) $t=1,2,\dots$such that

$$\bigcup_{s=1}^{r_t} B_{s\,t}^1 \times \ldots \times B_{s\,t}^m \supseteq D \cap A_t^m$$

with $A_t^m = \prod_{i=1}^m \{x_i \mid |x_i| \le t\}$, $A_t^m - \bigcup_{s=1}^{r_t} B_{s\,t}^1 \times \ldots B_{s\,t}^m$ a closed set, and

$$P_0\Big(\bigcup_{s=1}^{r_t} B_{s\,t}^1 \times \ldots \times B_{s\,t}^m\Big) \le \varepsilon_t\,,$$

then

$$R_n(H) \to 0 \text{ (a.s.)}.$$

The proof of Proposition 2 is given below.

Theorem B : If the conditions of proposition 2 hold, then $U_n(h)-\theta \to 0$ (a.s.).

PROOF : Since $U_n(h)-\theta = \hat{U}_n(h)+R_n(H)$, the result follows from Proposition 2 and $\hat{U}_n(h) \to 0$ (a.s.) as for Theorem A.

Corollary : Let h be a.e bounded and a.e continuous w.r.to Lebesgue measure on $\mathbf{R}^m$ with its non-continuity set D. Let F be absolutely continuous w.r.to Lebesgue measure on **R** , $E|X_1| < \infty$ and

$$\alpha(n) = O(n^{-(1+\gamma)}) \quad (0<\gamma<1) \ .$$

If the probability of the closed hull of D is zero, then $U_n(h)-\theta \to 0$ (a.s.). In particular, if D is a closed set ,then $U_n(h)-\theta \to 0$ (a.s.).

For example, the kernel $h(x_1,x_2)$ of Wilcoxon one-sample statistic.[2.1] is $I(x_1+x_2 \le 0)$, so $D = \{(x_1,x_2) \mid x_1 + x_2 = 0\}$, thus D is a closed set.

PROOF : In view of Theorem B, we only need to verify the conditions of Proposition 2 . Without loss of generality we may assume D is a closed set, thus

$D \cap A_t^m$ is a closed set too. It is obvious that for any given δ_t there exists an open set $G_t \supset D \cap A_t^m$ such that

$$m(G_t) - m(D \cap A_t^m) = m(G_t) \leq \delta_t .$$

Since $P_0 << m$ (Lebesgue measure), for any given ε_t we can choose such δ_t to guarantee $P_0(G_t) \leq \varepsilon_t$. Thus for any $x = (x_1,\dots,x_m) \in G_t$ there exists an open neighborhood with form $B^1_t(x_1) \times \dots \times B^m_t(x_m)$ such that $B^1_t(x_1) \times \dots \times B^m_t(x_m) \subseteq G_t$. In this way we get an open covering $\{B^1_t(x_1) \times \dots \times B^m_t(x_m) \mid x = (x_1,\dots,x_m) \in G_t\}$ of $D \cap A_t^m$. Since $D \cap A_t^m$ is a bounded closed set, according to the finite covering theorem, there are $r_t < \infty$ such that

$$\bigcup_{s=1}^{r_t} B^1_{s\,t} \times \dots B^m_{s\,t} \supseteq D \cap A_t^m .$$

Noting $B^1_{s\,t} \times \dots B^m_{s\,t} \subseteq G_t$ and $P_0(G_t) \leq \varepsilon_t$, we have

$$P_0\Big(\bigcup_{s=1}^{r_t} B^1_{s\,t} \times \dots B^m_{s\,t} \Big) \leq \varepsilon_t .$$

Thus the conditions of Proposition 2 are satisfied from which the proof follows.

4.2 : Proofs of Propositions

PROOF OF PROPOSITION 1 : For simplicity we consider m=2. Since $E|X_1| < \infty$,for any given $\{\varepsilon_t\} \downarrow 0$ there exists t so that

$$P_0[(A_t \times A_t)^c] \leq \frac{\varepsilon_t}{8} . \tag{2}$$

According to lemma 1, there exists a symmetric, bounded, and continuous function g on R^m with the form (1)[4.1] such that

$$\sup_{x\in A_t^2} | g_t(x)-H(x) | \le \varepsilon_t /4. \tag{3}$$

Now we consider $R_n(g_t)$. recall $E(R_n(g_t))^2 \le n^{-2+\gamma C}$ from Proposition 1 in [3.1], we have $R_n(g_t)\to 0$ (a.s.) for fixed t, as $n\to\infty$ by using the Borel Cantelli lemma. Since

$$R_n(H)= R_n(g_t) + R_n(H- g_t)$$

(see Lemma 2 [4.1] for H), If we can prove that for any given $\{\varepsilon_t\}\downarrow 0$, as $t\to\infty$,

$$P(\overline{\lim}_{n\to\infty} | R_n(H- g_t) | \le \varepsilon_t /2)= 1 ,$$

then we would have

$$P(\overline{\lim}_{n\to\infty} | R_n(H) | \le\varepsilon_t)=1,$$

that is ,

$$R_n(H)\to 0 \text{ (a.s.)}.$$

Denote $H- g_t$ by H_t^* ,then

$$R_n(H- g_t) = R_n(H_t^*) = \frac{\sum_{i=1}^{n} \sum_{i<j}^{n} H_t^*(X_i,X_j)}{\binom{n}{2}}.$$

Without loss of generality, we assume that $|H_t^*| \le 1$. (since if $|H_t^*| \le M$, we can consider $R_n(H_t^*/M)$). Define

$$Y_{ij}^t= 1 \qquad (X_i, X_j)\in (A_t\times A_t)^c$$

$$Y_{ij}^t= 0 \qquad (X_i, X_j)\in (A_t\times A_t),$$

$$\theta_t = P_0(Y_{ij}^t= 1) ,$$

and

$$Y_n^t = \sum_{i=1}^{n} \sum_{i<j}^{n} Y_{ij}^t .$$

Noting $H_t^*(X_i,X_j) \le \varepsilon_t/4$ when $(X_i,X_j) \in (A_t \times A_t)$ (see (3) in [4.2]) and $\theta_t \le \frac{\varepsilon_t}{8}$ (see (2) in [4.2]) , thus

$$\{ |R_n(H_t^*)| > \varepsilon_t/2 \} \subseteq \{ Y_n^t > \binom{n}{2} \frac{\varepsilon_t}{4} \} \subseteq \{ |Y_n^t - \theta_t \binom{n}{2}| \ge \frac{\varepsilon_t}{8} \binom{n}{2} \}.$$

If we can show

$$\sum_{n=1}^{\infty} P\{ |Y_n^t - \theta_t \binom{n}{2}| \ge \frac{\varepsilon_t}{8} \binom{n}{2} \} < \infty,$$

then

$$\sum_{n=1}^{\infty} P\{ |R_n(H_t^*)| > \varepsilon_t/2 \} < \infty.$$

Therefore from the Borel Cantelli lemma, we would have

$$P(\overline{\lim}_{n \to \infty} |R_n(H_t^*)| \le \varepsilon_t /2) = 1 .$$

Define

$$U_n^t = \frac{Y_n^t}{\binom{n}{2}}.$$

Since

$$(A_t \times A_t)^c = A_t \times A_t^c \cup A_t^c \times A_t \cup A_t^c \times A_t^c ,$$

U_n^t is a U-statistic with kernel

$$h^t(x,y) = I(x \in A_t) I(y \in A_t^c) + I(y \in A_t) I(x \in A_t^c) + I(x \in A_t^c) I(y \in A_t^c). \qquad (4)$$

and

$$\theta_t = E_0 h^t (X_1, X_2). \tag{5}$$

Thus

$$U_n^t - \theta_t = \hat{U}_n^t + R_n^t.$$

Now we have (using Markov inequality)

$$P\{\,|\,Y_n^t - \theta_t \tbinom{n}{2}\,| \geq \frac{\varepsilon_t}{8} \tbinom{n}{2}\,\}$$

$$= P\{\,|\,U_n^t - \theta_t\,| \geq \frac{\varepsilon_t}{8}\,\}$$

$$= P\{\,|\,\hat{U}_n^t + R_n^t\,| \geq \frac{\varepsilon_t}{8}\}$$

$$\leq P\{\,|\,\hat{U}_n^t\,| \geq \frac{\varepsilon_t}{16}\} + P\{\,|\,R_n^t\,| \geq \frac{\varepsilon_t}{16}\}$$

$$\leq \frac{E(\hat{U}_n^t)^4}{(\frac{\varepsilon_t}{16})^4} + \frac{E(R_n^t)^2}{(\frac{\varepsilon_t}{16})^2}.$$

Recall from [1.2]

$$\hat{U}_n^t = \frac{2}{n} \sum_{i=1}^{n} \tilde{h}_1(X_i),$$

so it follows from Lemma 2 [4.1]that

$$E(\hat{U}_n^t)^4 \leq \frac{C}{n^{1+\lambda}}.$$

Since the kernel h^t possess the "decomposable " structure , in light of Proposition 1 of [3.1], we obtain

$$E(R_n^t)^2 \leq \frac{C}{n^{1+\gamma}}.$$

Thus

$$P\{\,|Y_n^t - \theta_t \tbinom{n}{2}| \ge \frac{\varepsilon_t}{8}\tbinom{n}{2}\,\} \le \frac{C_t}{n^{1+\min(\lambda,\gamma)}}$$

and it follows that

$$\sum_{n=1}^{\infty} P\{\,|Y_n^t - \theta_t \tbinom{n}{2}| \ge \frac{\varepsilon_t}{8}\tbinom{n}{2}\,\} \le \sum_{n=1}^{\infty} \frac{C_t}{n^{1+\min(\lambda,\gamma)}} < \infty.$$

The proof is complete.

PROOF OF PROPOSITION 2 : The proof is analogous to the proof of proposition 1. The only difference is that here Y_{ij}^t is defined

$$Y_{ij}^t = 1 \qquad (X_i, X_j) \in [\bigcup_{s=1}^{r_t} B_{s\,t}^1 \times B_{s\,t}^2 \cap (A_t \times A_t)] \cup (A_t \times A_t)^c$$

$$Y_{ij}^t = 0 \qquad (X_i, X_j) \in (A_t \times A_t) - \bigcup_{s=1}^{r_t} B_{s\,t}^1 \times B_{s\,t}^2 \ .$$

Thus U_n^t is a U-statistic with kernel

$$h^t(x,y) = I[[\bigcup_{s=1}^{r_t} B_{s\,t}^1 \times B_{s\,t}^2 \cap (A_t \times A_t)] \cup (A_t \times A_t)^c]. \qquad (6)$$

Without loss of generality, we may assume that $h^t(x,y)$ is symmetric. For, if not, it may be replaced by the symmetric kernel

$$\frac{1}{2!} \sum_{p} h^t(x_{i_1}, x_{i_2})$$

where $\sum_p$ denotes summation over the 2! permutations (i_1, i_2) of (1,2). It is easy to check that $h^t(x,y)$ also possess the "decomposable" structure. Noting

$$(A_t \times A_t) - \bigcup_{s=1}^{r_t} B_{s\,t}^1 \times B_{s\,t}^2$$

is a compact set, so by letting

$$K_t = (A_t \times A_t) - \bigcup_{s=1}^{r_t} B_{s\,t}^1 \times B_{s\,t}^2 \,,$$

we have

$$\sup_{x \in K_t} |g_t(x) - h(x)| \leq \varepsilon_t/4.$$

Hence we can get the proof without any difficulty since it is now verbatim the same as for Proposition 1.

4.3 : The Weak Consistency of U-statistics for Strong Mixing Processes

For the case that h is unbounded, we are not able to extend the above a.s. convergence results because our method used in the proof of Proposition 1 is no longer valid. However, by confining our attention to weak consistency of U-statistics, we are able to make some progress.

In this section we shall consider a martingale approximation instead of uniform approximation as a tool to achieve the following results;

Theorem C : Let h be a kernel with

$$\sup_{(i_1,\ldots,i_m)} E|h(X_{i_1},\ldots,X_{i_m})|^2 < \infty .$$

If h is continuous ,then

$$U_n \to \theta \text{ (in prob.)}.$$

Theorem D : Let h be a kernel with

$$\sup_{(i_1,\ldots,i_m)} E\,|h(X_{i_1},\ldots,X_{i_m})|^2 < \infty .$$

If h is a.e. continuous about Lebesgue measure on $\mathbf{R}^m$ and F is absolutely continuous w.r.to Lebesgue measure on $\mathbf{R}$, then

$U_n \to \theta$ (in prob.).

Put $A_{k,j} = \left[\frac{j}{2^k}, \frac{j+1}{2^k}\right)$ when $-k2^k \le j \le (k-1)2^k$,

$$A_{k,-k2^k-1} = (-\infty, -k\ ),$$

and

$$A_{k,(k-1)2^k+1} = [k,\infty) .$$

Let $\mathcal{F}_k$ be the σ–algebra generated by

$$\{\prod_{i=1}^{m} A_{k j_i}) \mid -k2^k-1 \le j_i \le (k-1)2^k+1\} .$$

Denote $\mathcal{F} = \bigvee_{k=1}^{\infty} \mathcal{F}_k$ and $H_k = E_0(H \mid \mathcal{F}_k)$. Then

$$H_k = \sum_{-k2^k-1}^{(k-1)2^k+1} \cdot \sum H_{i,j}\, I(A_{k,i} \times A_{k,j}), \tag{7}$$

where $I(A_{k,i} \times A_{k,j})$ is the indicator function of $A_{k,i} \times A_{k,j}$ and

$$H_{i,j} = \frac{E_0\{HI(A_{k,i} \times A_{k,j})\}}{P_0(A_{k,i} \times A_{k,j})}. \tag{8}$$

It is obvious that $\mathcal{F} \supseteq \mathcal{B} = \{B \mid B$ is any Borel set in $\mathbf{R}^m\}$, so H is $\mathcal{F}$-measurable .Since $E_0|H| < \infty$, by martingale convergence theorem we have

$$H_k \to E_0(H \mid \mathcal{F}) = H, \text{ as } k \to \infty. \quad (\text{a.s. } P_0)$$

Denote

$$H_k^* = H - H_k , \tag{9}$$

thus

$$H_k^* \to 0 \;\; (\text{a.s. } P_0). \tag{10}$$

4.4 : Proofs

For simplicity, we consider m=2 .

Lemma 3 : If h is a.e. continuous on $\mathbf{R}^m$ and F is absolutely continuous w.r.to Lebesgue measure on **R**, then there exists an $\{\varepsilon_s\} \downarrow 0$ as $s \to \infty$ and not dependent on k such that

$$E_k(\varepsilon_s) = \{(x_1,x_2) \mid |H_k^*(x_1, x_2)| \le \varepsilon_s\} \tag{11}$$

is a P_0 –continuity set.

PROOF: First we consider h is continuous. Since

$$H_k^*(x_1, x_2) = H(x_1, x_2) - H_k(x_1, x_2) \quad \text{(see (5)-(7) in [4.3])}$$

and $H(x_1, x_2)$ is continuous, $H(x_1, x_2) - H_{i,j}$ is continuous on $A_{k,i} \times A_{k,j}$. Thus

$$\partial E_k(\varepsilon) \subseteq \bigcup_{ij} [\{(x_1,x_2) \mid |H(x_1, x_2) - H_{i,j}| = \varepsilon \} \cup \partial\{A_{k,i} \times A_{k,j}\}].$$

Now in view of continuity of H there exists an $\{\varepsilon_s\} \downarrow 0$ as $s \to \infty$ which is independent of i and j such that

$$P_0\{(x_1,x_2) \mid H(x_1, x_2) = H_{i,j} \pm \varepsilon_s \} = 0.$$

Noticing

$$P_0(\partial\{A_{k,i} \times A_{k,j}\}) = 0,$$

we have

$$P_0(\partial E_k(\varepsilon_s)) = \{(x_1,x_2) \mid H_k^*(x_1, x_2) \mid = \varepsilon_s\} = 0.$$

This completes the proof for h continuous.

When h is a.e continuous , denote the set of all non-continuous points by D, then the result can be derived from

$$\partial E_k(\varepsilon) \subseteq \bigcup_{ij} [\{(x_1,x_2) \mid |H(x_1, x_2) - H_{i,j}| = \varepsilon\} \cup \partial\{A_{k,i} \times A_{k,j}\} \cup D]. \qquad (12)$$

The proof is complete.

Proposition 3 : Let h be a bounded kernel; if h is a.e. continuous w.r.to Lebesgue measure on $\mathbf{R}^m$ and F is absolutely continuous w.r.to Lebesgue measure on $\mathbf{R}$, then there exists a n_k corresponding k such that

$$\sup_{n \geq n_k} E \mid R_n(H_k^*) \mid] \to 0 \text{ ,when } k \to \infty.$$

PROOF: Since

$$R_n(H_k^*) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i_1 < i_2 \leq n} H_k^*(X_{i_1}, X_{i_2})$$

$$= \frac{1}{\binom{n}{2}} \sum_{1 \leq i_1 \leq n} \ldots \sum_t H_k^*(X_{i_1}, X_{i_1+t}),$$

using the stationarity of the strong mixing process $\{X_n\}$

$$E \mid R_n(H_k^*) \mid \leq \frac{1}{\binom{n}{2}} \sum_{1 \leq i_1 \leq n} \ldots \sum_t E \mid H_k^*(X_{i_1}, X_{i_1+t}) \mid$$

$$\leq \frac{C}{n} \sum_t \sup_{1 \leq i_1 \leq n} E \mid H_k^*(X_{i_1}, X_{i_1+t}) \mid$$

$$\leq \frac{C}{n} \sum_t E\left| H_k^*(X_{i_1}, X_{i_1+t}) \right| .$$

Since h is bounded , it follows that

$$\frac{C}{n} \sum_t E\left| H_k^*(X_{i_1}, X_{i_1+t}) \right|$$

$$\leq \frac{C}{n}\{ \sum_t [\varepsilon P\{ \left| H_k^*(X_{i_1}, X_{i_1+t}) \right| \leq \varepsilon\}$$

$$+P\{ \left| H_k^*(X_{i_1}, X_{i_1+t}) \right| > \varepsilon\}]\}$$

$$\leq C\varepsilon + \frac{C}{n} \sum_t P_t\{E_k(\varepsilon)^c\}$$

$$\leq C\varepsilon + \frac{C}{n} \sum_t P_0\{E_k(\varepsilon)^c\} + \frac{C}{n} \sum_t (P_t\{E_k(\varepsilon)^c\} - P_0\{E_k(\varepsilon)^c\})$$

$$\leq C\varepsilon + CP_0\{E_k(\varepsilon)^c\} + \frac{C}{n} \sum_t (P_t\{E_k(\varepsilon)^c\} - P_0\{E_k(\varepsilon)^c\}) . \qquad (13)$$

Since P_t weakly converge to P_0 (by strong mixing property),

$$P_t\{E_k(\varepsilon)^c\} \to P_0\{E_k(\varepsilon)^c\}, \qquad t \to \infty,$$

in view of the above lemma (see the definition of weak convergence in Billingsely [2]) ,Thus the last term of (13)→ 0 when $n \to \infty$. Noticing $H_k^* \to 0$ (a.s. P_0)as $k \to \infty$ (see (10) in [4.3]), we have $P_0\{E_k(\varepsilon)^c\} \to 0$ when $k \to \infty$. Hence it follows from (13) that

$$\sup_{n \geq n_k} E\left| R_n (H_k^*) \right|] \to 0$$

when $k \to \infty$. The proof is complete.

PROOF OF THEOREM D : At first we consider h is bounded. Since

$$U_n(h)-\theta = \widehat{U}_n(h)+R_n(H)$$

$$= \widehat{U}_n(h) + R_n(H_k) + R_n(H_k^*)$$

and (as in the proof of Theorem A [4.1])

$$\widehat{U}_n(h) \to 0 \text{(in Prob.)},$$

we only need to consider the last two terms. Using

$$E\{R_n(H_k)\}^2 \to 0$$

(see Proposition 1[3.1])as $n \to \infty$ for each fixed k, and (see Proposition 3 in [4.4])

$$\sup_{n \geq n_k} E|R_n(H_k^*)| \;] \to 0 \text{ as } k \to \infty,$$

we get

$$R_n(H_k) + R_n(H_k^*) \to 0 \text{ (in Prob.) as } n \to \infty$$

So

$$U_n(h)-\theta \to 0 \text{ (in Prob.) as } n \to \infty.$$

Now we consider h is unbounded. Define H_k^+ as

$$H_k^+(x_1,x_2) = H(x_1,x_2) \text{ when } |x_i| \leq k \; (i=1,2),$$

$$H_k(x_1,x_2) = 0 \qquad \text{otherwise.}$$

Since, setting $H_k^{+*} = H - H_k^+$,

$$U_n(h)-\theta = \widehat{U}_n(h)+R_n(H)$$

$$= \widehat{U}_n(h) + R_n(H_k^+) + R_n(H_k^{+*})$$

and for h unbounded

$$\hat{U}_n(h) \to 0 \text{ (in Prob.) as } n \to \infty,$$

we now consider $R_n(H_k^+)$ and $R_n(H_k^{+*})$. It is obvious that H_k^+ is a.s. bounded for each k and a.e. continuous w.r. to Lebesgue measure on R^m, so $R_n(H_k^+) \to 0$ (in Prob.) as $n \to \infty$ for each fixed k. Also we will show below that

$$\sup_{n \geq n_k} E\,|\,R_n(H_k^{+*})\,| \to 0 \text{ (in Prob.) as } k \to \infty.$$

Then combining the last result with the preceding assertion, the proof would be complete. Now Denote E_k by

$$E_k = \{(x_1, x_2) \mid |H(x_1, x_2)| \leq k\}, \tag{14}$$

then $\partial E_k(\varepsilon) \subseteq \{(x_1, x_2) \mid |H(x_1, x_2)| = k\} \cup D$, thus there exists $\{k\} \uparrow \infty$ such that E_k is a P_0 –continuity set. Since, again in view of stationarity of $\{X_n\}$ and the finite moment condition,

$$E\,|\,R_n(H_k^{+*})\,| \leq \frac{1}{\binom{n}{2}} \sum_{1 \leq i_1 \leq n} \cdots \sum_t E\,|\,H_k^{+*}(X_{i_1}, X_{i_1+t})\,|$$

$$\leq \frac{C}{n} \sum_t \sup_{1 \leq i_1 \leq n} E\,|\,H_k^{+*}(X_{i_1}, X_{i_1+t})\,|$$

$$\leq \frac{C}{n} \sum_t E\,|\,H_k^{+*}(X_{i_1}, X_{i_1+t})\,|$$

$$\leq \frac{C}{n} \sum_t E[\,|\,H(X_{i_1}, X_{i_1+t})\,|\, I(E_k^c)]$$

$$\leq \frac{C}{n} \sum_t P_t\{E_k^c\}^{1/2} \{E[H(X_{i_1}, X_{i_1+t})]^2\}^{1/2}$$

$$\leq \frac{C}{n} \sum_t P_t\{E_k^c\}^{1/2}$$

$$\leq \frac{C}{n} \sum_t P_0(E_k^c)^{1/2} + \frac{C}{n} \sum_t [P_t(E_k^c)^{1/2} - P_0(E_k^c)^{1/2}]$$

$$\leq CP_0(E_k^c)^{1/2} + \frac{C}{n} \sum_t [P_t(E_k^c)^{1/2} - P_0(E_k^c)^{1/2}] \quad . \tag{15}$$

Since P_t weakly converge to P_0 (by strong mixing property),

$$P_t(E_k^c) \rightarrow P_0(E_k^c), \qquad t \rightarrow \infty,$$

Thus the last term of (15)$\rightarrow$ 0 when $n \rightarrow \infty$. Noticing $E_0 | h | < \infty$,we have $P_0(E_k^c) \rightarrow 0$ when $k \rightarrow \infty$. Hence it follows from (15) that

$$\sup_{n \geq n_k} E | R_n (H_k^{+*}) |] \rightarrow 0 .$$

The proof is complete.

PROOF OF THEOREM C : From the proof of theorem A we know that U_n - $\theta \rightarrow$ 0(in prob.) for a bounded function h. Note that when h is unbounded but continuous we can define a continuous and bounded function H_k for each k such that

$$H_k^+ (x_1, x_2) = H(x_1, x_2) \text{ when } |x_i| \leq k \ (i=1,2). \tag{16}$$

Thus, by using the fact that H_k^+ is bounded and continuous we can finish the proof in the same way as the unbounded part of Theorem D.

REFERENCES

[1] Athreya, K. B. and Pantula, S. G. :(1984) Mixing properties of Harris chains. *Institute of Statistics Mimeograph Series* No. 1649, North Carolina State University.

[2] Billingsley, P. :(1968) *Convergence of Probability Measures*. Wiley, New York.

[3] Chanda, K.C. :(1974) .Strong mixing properties of linear stochastic processes. *J. Appl. Prob.* **11**, 401-408.

[4] Chernoff, H. and Savage , I. R. :(1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29**, 972-994.

[5] Dehling, H.,Philipp,W. :(1982) Almost sure invariance principles for weakly dependent vector-valued random variable. *Ann. Prob.* **10**, 689-701.

[6] Dehling, H.: (1983) Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. Verw. Geb.* **63**, 393-432.

[7] Denker,M., and Keller, G. :(1983). On U-statistics and von Mises functionals for weakly dependent processes. *Z. Wahrsch. Verw. Geb.* **64**, 505-522.

[8] Doob, J.L.: (1953) Stochastic Processes. Wiley, New York.

[9] Fears, T.R. and Mehra, K.L. :(1974)Weak convergence of a two-sample empirical process and a Chernoff-Savage Theorem for ϕ-moxing sequences. *Ann. Statist.* **2**,586-596.

[10] Gorodetskii, V.V. :(1977). On the strong mixing property for linear sequences. *Theor. Probab. Appl.* **22**,411-413.

[11] Hàjek, J.: (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39**, 325-346.

[12] Harel, M.,and Puri, M. L.: (1989). Limiting behavior U-statistics, V-statistics and one sample rank order statistics for nonstationary absolutely regular processes. *J. Multivariate anal.* **29**.

[13] Hoeffding, W. :(1948) A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.

[14] Ibragimov, I. A. and Linnik, Yu. V. :(1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen, Netherlands.

[15] Kolmogorov, A. N. and Rozanov, Yu. A. :(1960). On the strong mixing conditions of a stationary Gaussian process. *Thor. Veroyatnost. i Primenen* **5**, 222-227.

[16] Mehra, K.L.:(1975) Weak convergence and one-sample rank statistics under ϕ-mixing. *Canad. Math. Bull.* **18**, 555-565.

[17] Narasimhan, R. :(1973) *Analysis on Real and Complex Manifolds*. North-Holland, Inc..

[18] Philipp, W. :(1969). The central limit theory for mixing sequences of random variables. *Z. Warsch. Verw. Geb.* **12**, 155-171.

[19] Rosenblatt, M. :(1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci., U.S.A.* **42**, 43-47.

[20] Rosenblatt, M. :(1971) *Markov Processes, Structure and Asymptotic Behavior* . Springer , New York.

[21] Rozanov, Yu.A., Volkonski, V.A.:(1959). Some limit theorems for random functions I. *Theor. Probab. Appl.* **4**, 178-197.

[22] Rozanov, Yu.A., Volkonski, V.A.:(1961). Some limit theorems for random functions II *Theor. Probab. Appl.* **6**, 186-198.

[23] Sen, P. K.:(1972). Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. *Z. Wahrsch. Verw. Geb.* **25**, 71-82.

[24] Serfling, R. J.: (1968). Contributions to central limit theory for dependent variables. *Ann. Math. Statist.* **39** , 1158-1195.

[25] Serfling, R. J. :(1968). The Wilcoxon two-sample statistic on strongly mixing processes. *Ann. Math. Statist.* **9** , 1202-1209.

[26] Serfling, R. J. :(1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

[27] Stout, W.F. :(1974) *Almost Sure Convergence*.Academic Press, Inc.

[28] Withers, C.S. :(1981). Conditions for linear processes to be strong-mixing. *Z. Wahrsch. Verw. Geb.* **57**, 477-480.

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

[30] Yoshihara, K.:(1978). Limiting behavior of one sample rank order statistics for absolutely regular processes. *Z. Wahrsch. Verw. Geb.* **43**, 107-127.

[31] Yosida,K.: (1980). *Functional Analysis*. Springer, New York.