Estimation and prediction for nonlinear time series
Borovkova, Svetlana Alfredovna

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1998

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 09-04-2019
5 Limit theorems for $U$-statistics and $U$-statistics processes

Recall that the $U$-statistic of degree $m$ corresponding to the kernel function $h : \mathbb{R}^k \rightarrow \mathbb{R}$ is given by

$$U_n = \left( \frac{n}{m} \right)^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} h(X_{i_1}, \ldots, X_{i_m}),$$

where $\{X_n\}_{n \in \mathbb{N}}$ is a stationary sequence of $k$-dimensional random vectors with common distribution $F$. $U$-statistics are unbiased estimators for functionals of the form

$$\theta(F) = \int_{\mathbb{R}^k} h(x_1, \ldots, x_m) dF(x_1) \cdots dF(x_m).$$

In this chapter we consider the asymptotic behaviour of $U$-statistics and empirical processes of $U$-statistics structure (which we shall define later) for absolutely regular sequences and functionals of them. For brevity, we shall state and prove the results for $U$-statistics only in case $m = 2$, i.e. for

$$U_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j),$$

where $h : \mathbb{R}^{2k} \rightarrow \mathbb{R}$ is a symmetric function. Unless stated differently, we shall restrict ourselves to the case $k = 1$. We remark that this is no loss of generality, generalisations to cases with $m > 2$, $k > 1$ are mostly straightforward.

The central limit theorem for $U$-statistics of one-sided functionals (4.1) of absolutely regular processes was shown by Denker and Keller [26]. In the next section we extend their result to the case of two-sided functionals (4.2), under less restrictive conditions on the function $f$. 

91
5.1 CLT for $U$-statistics of functionals of absolutely regular sequences

In this section we consider the $U$-statistic

$$U_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j),$$

for the process $\{X_n\}_{n \in \mathbb{N}}$ which is a functional of an absolutely regular process, and show that the distribution of $U_n$ is asymptotically normal.

Let us assume that the function $h$ is bounded. The case of unbounded $h$ will be partially considered in one of the following sections. Denote

$$h_1(x) = \int_{\mathbb{R}} h(x, y) dF(y).$$

Again, we want both functions $h$ and $h_1$ to preserve in some way the $1$-approximation condition on $\{X_n\}$. For that we shall require, as in Section 4.3, some sort of Lipschitz-continuity condition on $h$ and $h_1$.

We shall say that $h$ satisfies the $p$-Lipschitz condition if for all $X, X', Y$, each having the distribution $F$, there exists a function $\phi(\epsilon) \to 0$ as $\epsilon \to 0$, such that

$$\mathbb{E}[|h(X, Y) - h(X', Y)|^p 1_{|X - X'| \leq \epsilon}] \leq \phi(\epsilon). \quad (5.1)$$

From this condition the $p$-Lipschitz condition (4.46) on the function $h_1$ follows directly, since by taking $Y \sim F$, independent of $X, X'$ and by applying Jensen’s inequality we obtain

$$\mathbb{E} |h_1(X) - h_1(X')|^p 1_{|X - X'| \leq \epsilon} \leq \phi(\epsilon),$$

where $\phi(\epsilon) \to 0$ as $\epsilon \to 0$, since (5.1) is required to hold for all joint distributions of $X, X'$ and $Y$. Note that the $p$-Lipschitz condition (5.1) is with respect to the distribution $F$. We will not express this explicitly, but will keep this in mind.

As above, in place of $\phi(\alpha_1)$ for uniformity we shall always use the new sequence $\phi_1$ defined in (4.49) in terms of $\phi(\alpha_1)$.

First we prove two fundamental lemmas, which are analogous to the lemma of Yoshihara for absolutely regular sequences [86].
Lemma 5.1 Let \( \{X_n\}_{n \in \mathbb{N}} \) be a functional of an absolutely regular process \( \{Z_n\}_{n \in \mathbb{Z}} \) with mixing coefficients \( (\beta_k) \), such that the 1-approximation condition holds, and let \( h(x,y) \) be a 1-Lipschitz continuous function in the sense of (5.1), such that for some \( r > 1 \)

\[
\int_{\mathbb{R}^2} |h(x_0, x_k)| dP_{X_0,X_k}(x_0, x_k) \leq M_1 < \infty
\]

and

\[
\int_{\mathbb{R}^2} |h(x_0, x_k)| dF(x_0) dF(x_k) \leq M_2 < \infty
\]

holds for all \( k \), where \( P_{X_0,X_k} \) is the joint measure induced by \( (X_0, X_k) \), and \( F \) is the distribution of \( X_0 \). Denote \( M = \max(M_1, M_2) \). Then, for \( k \geq 3 \),

\[
\left| \int_{\mathbb{R}^2} h(x_0, x_k) dP_{X_0,X_k}(x_0, x_k) - \int_{\mathbb{R}^2} h(x_0, x_k) dF(x_0) dF(x_k) \right| \leq 4M^{1/r}(\beta_{[k/3]} + \alpha_{[k/3]})^{1/s} + 2\phi_{[k/3]},
\]

where \( \frac{r}{s} + \frac{1}{s} = 1 \).

Proof Let \( \{Z'_n\}_{n \in \mathbb{Z}} \) and \( \{Z''_n\}_{n \in \mathbb{Z}} \) together with the corresponding functional be as in Proposition 4.1. Then the pairs \( (X_0, X_k) \), \( (X'_0, X'_k) \) and \( (X''_0, X''_k) \) have the same distribution, \( (X_0, X_k) \) is independent of \( (X'_0, X'_k) \) and \( (X''_0, X''_k) \) and the properties (3) and (4) hold.

Let \( B \) be the subset of the probability space where \( |X_k - X'_k| \leq \alpha_{[k/3]} \), \( \mathbb{P}(B) = 1 - \beta_{[k/3]} - \alpha_{[k/3]} \), and let \( D \) be the subset of the probability space where \( |X'_0 - X''_0| \leq \alpha_{[k/3]} \), \( \mathbb{P}(D) = 1 - \alpha_{[k/3]} \).

Denote

\[
\mathbb{E}_{F \times F} := \int_{\mathbb{R}^2} h(x_0, x_k) dF(x_0) dF(x_k) \quad \text{and}
\]

\[
\mathbb{E}_{P_{0k}} := \int_{\mathbb{R}^2} h(x_0, x_k) dP_{X_0,X_k}(x_0, x_k).
\]

Then

\[
\left| \mathbb{E}_{F \times F} h(X_0, X_k) - \mathbb{E}_{P_{0k}} h(X_0, X_k) \right| = \left| \mathbb{E}_{P_{0k}}[h(X'_0, X_k) - h(X''_0, X_k)] \right|
\]

\[
\leq \mathbb{E}_{P_{0k}}|h(X'_0, X_k) - h(X''_0, X_k)| + \mathbb{E}_{P_{0k}}|h(X'_0, X_k) - h(X'_0, X'_k)|
\]

\[
\leq \mathbb{E}_{P_{0k}}|h(X'_0, X_k) - h(X''_0, X''_k)|_B + \mathbb{E}_{P_{0k}}|h(X'_0, X_k) - h(X''_0, X''_k)|_D
\]

\[
\leq \phi_{[k/3]} + 2M^{1/r}(\beta_{[k/3]} + \alpha_{[k/3]})^{1/s} + \phi_{[k/3]} + 2M^{1/r} \alpha_{[k/3]}^{1/s}
\]

\[
\leq 2\phi_{[k/3]} + 4M^{1/r}(\beta_{[k/3]} + \alpha_{[k/3]})^{1/s},
\]
by Hölder’s inequality and the 1-Lipschitz condition (5.1). □

The next lemma is a generalisation of the previous one.

Let the process \( \{X_n\}_{n \in \mathbb{Z}} \) be a functional of an absolutely regular process, and let \( i_1 < i_2 < \ldots < i_t \) be arbitrary integers. Let \( P \) denote the joint distribution of \( X_{i_1}, \ldots, X_{i_t} \), and for \( j \in \{1, \ldots, l - 1\} \) let \( P^{(j)}_{(i_1, \ldots, i_j) \times (i_{j+1}, \ldots, i_t)}(E \times G) = P((X_{i_1}, \ldots, X_{i_j}) \in E) \cdot P((X_{i_{j+1}}, \ldots, X_{i_t}) \in G) \),

where \( E \in \mathcal{B}_j, \ G \in \mathcal{B}_{j+1}. \) Let \( \mathbb{E}^{(j)} \) denote the expectation taken with respect to the measure \( P^{(j)}_{(i_1, \ldots, i_j) \times (i_{j+1}, \ldots, i_t)} \) (we shall write just \( \mathbb{E} \) if it is taken with respect to \( P \)). Let \( g : \mathbb{R}^j \rightarrow \mathbb{R} \) be a measurable function. Analogous to (5.1), we say that \( g \) satisfies the \( p \)-Lipschitz condition if for all \( j \in \{1, \ldots, k - 1\} \) a function \( \phi(\epsilon) \) exists such that

\[
\mathbb{E} \left[ \left| g(X_1, \ldots, X_l) - g(X_1, \ldots, X_j, X_j', \ldots, X_l') \right|^p 1_{\sum_{i=1}^t |X_i - X_i'| \leq \epsilon} \right] \leq \phi(\epsilon),
\]

and

\[
\mathbb{E} \left[ \left| g(X_1, \ldots, X_k) - g(X_1, \ldots, X_j', \ldots, X_l') \right|^p 1_{\sum_{i=1}^j |X_i - X_i'| \leq \epsilon} \right] \leq \phi(\epsilon),
\]

where \( \phi(\epsilon) \rightarrow 0 \) as \( \epsilon \rightarrow 0 \) and all \( X_i, X_i' \) have the distribution \( F \).

**Lemma 5.2** Let \( \{X_n\}_{n \in \mathbb{Z}} \) be a functional of an absolutely regular process \( \{Z_n\}_{n \in \mathbb{Z}} \) with mixing coefficients \( (\beta_k) \), such that the 1-approximation condition holds. Let the function \( g \) satisfy the 1-Lipschitz condition in the sense of (5.3). If for some \( r > 1 \) and \( j \in \{1, \ldots, l - 1\} \)

\[
M := \max \left\{ (\mathbb{E}[g(X_{i_1}, \ldots, X_{i_l})]^r)^{1/r}, (\mathbb{E}[g(X_{i_1}, \ldots, X_{i_l})]^r)^{1/r} \right\} < \infty,
\]

then

\[
|\mathbb{E}^{(j)}[g(X_{i_1}, \ldots, X_{i_l}) - g(X_{i_1}, \ldots, X_{i_l})]| \leq 4M(\beta_{|k/3|} + \alpha_{|k/3|})^{1/2} + 2\phi_{|k/3|},
\]

where \( k = |i_{j+1} - i_j| \) and \( \frac{1}{r} + \frac{1}{s} = 1. \)

**Proof** W.l.o.g. we assume that \( i_j = 0 \) and \( i_{j+1} = k \). Let the processes \( \{Z'_n\} \) and \( \{Z''_n\} \), together with the corresponding functionals \( \{X'_n\}_{n \in \mathbb{Z}} \) and \( \{X''_n\}_{n \in \mathbb{Z}} \) be as in Proposition 4.1. Then

\[
(X_{i_1}, \ldots, X_{i_l}) \overset{d}{=} (X'_{i_1}, \ldots, X'_{i_l}) \overset{d}{=} (X''_{i_1}, \ldots, X''_{i_l})
\]
and \((X'_i, \ldots, X'_i)\) is independent of \((X_1, \ldots, X_i)\). Moreover, the properties \((3')\) and \((4')\) hold. Now proceeding in the same way as in Lemma 5.1 and using condition \((5')\), we get

\[
|\mathbf{E} \hat{g}(X_{i_1}, \ldots, X_{i_\ell}) - \mathbf{E} g(X_{i_1}, \ldots, X_{i_\ell})| \\
= |\mathbf{E}[g(X'_{i_1}, \ldots, X'_{i_\ell}, X_{i_{\ell+1}}, \ldots, X_{i_\ell}) - g(X'_{i_1}, \ldots, X'_{i_\ell}, X'_{i_{\ell+1}}, \ldots, X_{i_\ell})]| \\
\leq \mathbf{E}[g(X'_{i_1}, \ldots, X'_{i_\ell}, X'_{i_{\ell+1}}, \ldots, X_{i_\ell}) - g(X'_{i_1}, \ldots, X'_{i_\ell}, X'_{i_{\ell+1}}, \ldots, X'_{i_\ell})] \\
+ \mathbf{E}[g(X'_{i_1}, \ldots, X'_{i_\ell}, X'_{i_{\ell+1}}, \ldots, X_{i_\ell}) - g(X'_{i_1}, \ldots, X'_{i_\ell}, X'_{i_{\ell+1}}, \ldots, X_{i_\ell})] \\
\leq 4M(\beta[k/3] + \alpha[k/3])^{1/s} + 2\phi[k/3].
\]

\[ \Box \]

**Theorem 5.1** Let \(\{X_n\}_{n \in \mathbb{N}}\) be a functional of an absolutely regular process \(\{Z_n\}_{n \in \mathbb{Z}}\) with mixing coefficients \((\beta_k)\), such that the 1-approximation condition holds. Suppose that both \(h\) and \(h_1\) are 1-Lipschitz continuous, and that the sequences \((\beta_k), (\alpha_k)\) and \((\phi_k)\) satisfy the following summability condition:

\[
\sum_{k=1}^{\infty} k^2(\beta_k + \alpha_k + \phi_k) < \infty. \tag{5.5}
\]

Then the series

\[
\sigma^2 = \mathbf{E}[h_1(X_0)]^2 + 2 \sum_{k=1}^{\infty} \text{Cov}[h_1(X_0)h_1(X_k)], \tag{5.6}
\]

converges absolutely, and, if \(\sigma^2 > 0\), then

\[
\sqrt{n}(U_n - \theta) \xrightarrow{d} \mathcal{N}(0, 4\sigma^2).
\]

**Proof** The following decomposition of \(U\)-statistics (Hoeffding decomposition) holds:

\[
U_n = \theta(F) + \frac{2}{n} \sum_{i=1}^{n} [h_1(X_i) - \theta(F)] + R_n,
\]

where

\[
R_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} [h(X_i, X_j) - h_1(X_i) - h_1(X_j) + \theta(F)].
\]
From Theorem 4.2, using the same opening arguments as in Theorem 4.3 and the summability condition (5.5), it follows that

\[ \frac{2}{\sqrt{n}} \sum_{i=1}^{n} [h_1(X_i) - \theta(F)] \xrightarrow{d} \mathcal{N}(0,4\sigma^2), \]

where \( \sigma^2 \) is given by (5.6). Then the statement of the theorem will follow, if we show that \( \sqrt{n}R_n \to 0 \) in probability, as \( n \to \infty \), which, in turn, is implied by \( \mathbb{E}[nR_n^2] \to 0 \) as \( n \to \infty \). This is shown in the following lemma.

**Lemma 5.3** Under the conditions of Theorem 5.1,

\[ \mathbb{E}[nR_n^2] \to 0 \text{ as } n \to \infty. \tag{5.7} \]

**Proof** We have

\[ \mathbb{E}[R_n^2] = \frac{4}{n^2(n-1)^2} \sum_{1 \leq i_1 < j_1 \leq n} \sum_{1 \leq i_2 < j_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})], \tag{5.8} \]

where

\[ J(x, y) = h(x, y) - h_1(x) - h_1(y) + \theta(F). \]

Note that \( J(x, y) \) is degenerate, i.e. \( \int J(x, y)dF(x) = 0 \). Hence if e.g. \( j_2 \neq i_1, i_2, j_1 \), then we get

\[ \int_{\mathbb{R}} J(X_{i_1}, X_{j_1}) J(X_{i_2}, y) dF(y) = 0. \]

Since both \( h \) and \( h_1 \) are bounded and satisfy the 1-Lipschitz conditions (5.1) and (5.2), the function \( g(x_{i_1}, x_{j_1}, x_{i_2}, x_{j_2}) = J(x_{i_1}, x_{j_1}) J(x_{i_2}, x_{j_2}) \) satisfies the 1-Lipschitz condition (5.3). Also, since \( J(X_i, X_j) \) is bounded by 2, the condition (5.4) of Lemma 5.2 is satisfied with \( r = \infty \) (i.e. for \( L_\infty \)-norm of \( g \)). Hence, if \( i_1 \leq i_2 \leq j_1 < j_2 \), Lemma 5.2 implies

\[ \left| \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] - \mathbb{E}_P(1_{i_2 \leq j_1 < j_2}) \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \right| \leq 16(\beta_k + \alpha_k) + 2\phi_k, \]

where \( k = \lfloor |j_2 - j_1|/3 \rfloor \).
Split the sum (5.8) into 2 sums:

\[
\sum_{1 \leq i_1 < j_1, i_2 < j_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1})] J(X_{i_2}, X_{j_2}) \]

(5.9)

= \sum_{1 \leq i_1 < j_1, i_2 < j_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1})] J(X_{i_2}, X_{j_2})

+ \sum_{1 \leq i_1 < j_1, i_2 < j_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1})] J(X_{i_2}, X_{j_2})

In the elements of the first sum at least one index is different from all others, say, \( i_2 \), and suppose \( i_1 \leq i_2 \leq j_1 < j_2 \). Let \( d_i \) be the \( i \)th largest difference between consecutive indices. If \( d_i = j_2 - j_1 \) then

\[
|\mathbb{E}[J(X_{i_1}, X_{j_1})] J(X_{i_2}, X_{j_2})]| \leq 16(\beta_{d_1/3} + \alpha_{d_1/3}) + 2\phi_{d_1/3}.
\]

Then

\[
\sum_{1 \leq i_1 < j_1, d_1 = j_2 - j_1} |\mathbb{E}[J(X_{i_1}, X_{j_1})] J(X_{i_2}, X_{j_2})]| \leq 16 \sum_{1 \leq i_1 < j_1, d_1 = j_2 - j_1} (\beta_{d_1/3} + \alpha_{d_1/3} + \phi_{d_1/3})
\]

\[
\leq 16n \sum_{k=1}^{\infty} k^2(\beta_k + \alpha_k + \phi_k).
\]

If \( d_i \) is not \( j_2 - j_1 \) we apply Lemma 5.2 twice to obtain

\[
\mathbb{E}[J(X_{i_1}, X_{j_1})] J(X_{i_2}, X_{j_2}) \leq 16(\beta_{d_1/3} + \alpha_{d_1/3}) + 2\phi_{d_1/3} + 16(\beta_{d_2/3} + \alpha_{d_2/3}) + 2\phi_{d_2/3}
\]

and then

\[
\sum_{1 \leq i_1 < j_1, d_1 = j_2 - j_1} \mathbb{E}[J(X_{i_1}, X_{j_1})] J(X_{i_2}, X_{j_2})]
\]

\[
\leq 16 \sum_{1 \leq i_1 < j_1, d_1 = j_2 - j_1} (\beta_{d_1/3} + \alpha_{d_1/3} + \phi_{d_1/3})
\]

\[
+ (\beta_{d_2/3} + \alpha_{d_2/3} + \phi_{d_2/3})
\]

\[
\leq 32n^2 \sum_{k=1}^{\infty} k(\beta_k + \alpha_k + \phi_k).
\]
Estimating the sums in the other cases in the same way we get that the first sum is bounded by

$$ \sum_{\substack{1 \leq i_1 < j_1, \ i_2 < j_2 \leq n \\ n \neq i_2 \text{ or } j_1 \neq j_2}} \mathbb{E}[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \leq C n^2 \sum_{k=1}^{n} k(\beta_k + \alpha_k + \phi_k). $$

In the second sum there are at most $n^2$ terms, all are bounded, hence

$$ \sum_{1 \leq i_1 \neq j_1 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \leq C n^2. $$

Combining all the estimates above with the summability conditions on $\beta_k, \alpha_k$ and $\phi_k$, we get that

$$ \sum_{1 \leq i_1 < j_1 \leq n} \sum_{1 \leq i_2 < j_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \leq C n^2, $$

and so, $\mathbb{E}[n R_n^2] \leq \frac{C}{n} \rightarrow 0$ as $n \rightarrow \infty$.

5.2 Weak convergence of the empirical process of $U$-statistics structure

When estimating the correlation dimension by the Grassberger-Proccacia method, we study the dependence of the sample correlation integral $C_n(r)$ on the distance $r$ for different values of $r$. The correlation integral is estimated in the points $(r_1, r_2, \ldots, r_l)$ and then a least-squares regression of $C_n(r_i)$ vs. $r_i$ is fitted. In that case, if the observed sequence $X_1, X_2, \ldots, X_n$ is absolutely regular, or is the functional of the absolutely regular process, the central limit theorem by Denker and Keller [26] implies that $(C_n(r_i))_{i=1,\ldots,l}$ has a multivariate normal distribution with mean zero and some given covariance structure (see Chapter 2). However, in the definition of the correlation dimension (2.2) the distance parameter $r$ changes continuously on some interval close to 0. So it is interesting to study the sample correlation integral $C_n(r)$ not as a random variable, or a collection of them, but as a function of $r$. In the following sections we shall address this problem in more general context.

In the case of the sample correlation integral the kernel function $h(x,y) = h(x,y,r) = 1(\|x-y\| \leq r)$ depends also on a real parameter $r : \ r \in (0, r_0]$,
5.2. Weak convergence of the empirical process of U-statistics structure

for some \( r_0 > 0 \). This occurs also in many other applications, for instance, in the analysis of the archaeological data, so-called “ley hunting”. Suppose \( X_1, \ldots, X_n \) are observations from an unknown distribution \( G \) on \( \mathbb{R}^2 \), and we are interested in testing randomness against presence of some collinearities in the data. For this Broadbent and Heaton (see Silverman and Brown [69]) suggested the following approach: denote \( \alpha(x, y, z) = \)”the largest angle of the triangle \( xyz \” \), and study the statistics

\[
T_n(\epsilon) = \left( \frac{n}{3} \right)^{-1} \sum_{1 \leq i < j < k \leq n} 1(\alpha(X_i, X_j, X_k) > \pi - \epsilon)
\]

which is a U-statistic of degree 3 estimating

\[
\theta_G(\epsilon) = \mathbb{P}(\alpha(X, Y, Z) > \pi - \epsilon),
\]

where \( X, Y, Z \) are chosen independently according to the distribution \( G \). In this case we also are interested in the behaviour of \( T_n(\epsilon) \) not for a fixed \( \epsilon \), but for \( \epsilon \) on some interval \( (0, \epsilon_0] \).

In these examples the U-statistic

\[
U_n = U_n(t) = \left( \frac{n}{m} \right)^{-1} \sum_{1 \leq i_1 < \ldots < i_m \leq n} h(X_{i_1}, \ldots, X_{i_m}, t)
\]

is a random function of the real parameter \( t \), rather than a random variable. Denote

\[
U(t) = \int_{\mathbb{R}^m} h(x_1, \ldots, x_m, t) dF(x_1) \ldots dF(x_n),
\]

and consider the following process

\[
W_n(t) = \sqrt{n}(U_n(t) - U(t)),
\]

which we call the empirical process of U-statistics structure.

\( W_n(t) \) can be regarded as a generalisation of the classical empirical process. Construct the empirical distribution \( G_n \) from the set of all random vectors

\[
\{(X_{i_1}, \ldots, X_{i_m}) : 1 \leq i_1 < \ldots < i_m \leq n\},
\]

i.e.

\[
G_n := \left( \frac{n}{m} \right)^{-1} \sum_{1 \leq i_1 < \ldots < i_m \leq n} \delta_{X_{i_1}, \ldots, X_{i_m}},
\]
where $\delta_{X_{1},...,X_{m}}$ puts mass 1 on the $m$-tuple $(X_{1},...,X_{m})$. Let $G = F \times F \times \ldots \times F$. Then the empirical process of $U$-statistics structure is the processes

$$W_{n}(t) = \sqrt{n} \int_{\mathbb{R}^{m \times m}} h(x_{1},...,x_{m},t)d(G_{n} - G)$$

$$= \sqrt{n} \left[ \binom{n}{m}^{-1} \sum_{1 \leq i_{1},...,i_{m} \leq n} h(X_{i_{1}},...,X_{i_{m}},t) - U(t) \right].$$

If $X_{t} \in \mathbb{R}$ and $h(x,t) = 1(x \leq t)$, then $W_{n}(t)$ is the ordinary empirical process.

The weak convergence of the empirical processes of $U$-statistics structure to a Gaussian process was shown in Silverman [68] for the case of i.i.d. random variables, and further such processes were studied by Dehling, Denker and Philipp [23], Serfling [65], Helmers, Janssen and Serfling [39], Nolan and Pollard [53] and others.

In this section we want to study the behaviour of the process $W_{n}(t)$ when the sequence $X_{1},X_{2},...$ is stationary and absolute regular. We prove weak convergence of the empirical process of $U$-statistics structure for absolutely regular sequences to a Gaussian process under certain conditions on the class of kernel functions (Theorem 5.2). The next section contains some lemmas concerning the remainder term of the Hoeffding decomposition of the $U$-statistics. In Section 5.4 we shall consider the empirical process of $U$-statistics structure for functionals of absolutely regular sequences.

Let $\{X_{n}\}_{n \in \mathbb{N}}$ be a stationary sequence of $k$-dimensional random vectors ($k \geq 1$) and let $F$ be the distribution of $X_{0}$. We restrict ourselves to the case of $U$-statistics with $m = 2$, i.e. we consider

$$U_{n}(t) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_{i},X_{j},t)$$

and the empirical process of $U$-statistics structure

$$W_{n}(t) = \sqrt{n}(U_{n}(t) - U(t)),$$

where

$$U(t) = \int_{\mathbb{R}^{2k}} h(x,y,t)dF(x)dF(y).$$
Suppose the function $U(t)$ is continuous on the interval $[0, t_0]$ for some $t_0 > 0$ (w.l.o.g. we may take $t_0 = 1$) and satisfies the Lipschitz condition of the form:

$$|U(t) - U(s)| \leq C|t - s|$$  \hspace{1cm} (5.10)

for all $t, s \in [0, 1]$, where $C$ is a constant independent on $t, s$.

Introduce the class $\mathcal{H}$ of all kernel functions $h : \mathbb{R}^{2k} \times [0, 1] \rightarrow \mathbb{R}$ satisfying for all $x, y; t \in [0, 1]$

(A1) $0 \leq h(x, y, t) \leq 1$, $h(x, y, 0) = 0$ and $h(x, y, t)$ is increasing in $t$.

Denote

$$h_1(x, t) = \int_{\mathbb{R}^k} h(x, y, t) dF(y).$$

Note that, if the kernel function $h$ satisfies the conditions (A1) and (A2), then the function $h_1$ satisfies the analogous conditions, i.e. $0 \leq h_1(x, t) \leq 1$, $h_1(x, 0) = 0$ and it is increasing in $t$.

We shall also need the following condition, analogous to (4.30). Suppose that for all $t, s$ with $0 \leq s \leq t \leq 1$ and all $k \geq 1$

$$|\mathbb{E}(h(X_0, X_k, t) - h(X_0, X_k, s))| \leq C|t - s|,$$  \hspace{1cm} (5.11)

where the constant $C$ does not depend on $t, s, k$.

Now we formulate the result on weak convergence of $W_n$.

**Theorem 5.2** Let $\{X_n\}_{n \in \mathbb{N}}$ be a stationary and absolutely regular sequence with mixing coefficients $(\beta_k)$ satisfying $\sum k^2 \beta_k^{1/2 - \tau} < \infty$ for some $\tau \in (0, 1)$. Suppose that $h \in \mathcal{H}$, that $U(t)$ is continuous on $[0, 1]$ and that (5.10) and (5.11) hold. Then the process $W_n$ converges weakly in $D[0,1]$ to the mean-zero Gaussian process $W$ with covariance structure

$$\mathbb{E}\{W(s)W(t)\} = 4 \text{Cov}[h_1(X_1, s)h_1(X_1, t)] + 4 \sum_{k=1}^{\infty} \text{Cov}[h_1(X_1, s)h_1(X_{k+1}, t)] + 4 \sum_{k=1}^{\infty} \text{Cov}[h_1(X_1, t)h_1(X_{k+1}, s)].$$  \hspace{1cm} (5.12)

Moreover, the series (5.12) converge absolutely, and the limit process $W$ has continuous sample paths on $[0, 1]$ with probability 1.
Proof According to Hoeffding’s projection method, $U_n(t)$ can be decomposed as

$$U_n(t) = U(t) + \frac{2}{n} \sum_{i=1}^{n} [h_1(X_i,t) - U(t)] + R_n(t),$$

where

$$R_n(t) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} [h(X_i, X_j,t) - h_1(X_i,t) - h_1(X_j,t) + U(t)]$$

is the remainder of $U_n(t)$. Then the process $W_n(t)$ can be written as

$$W_n(t) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} [h_1(X_i,t) - U(t)] + \sqrt{n}R_n(t) = W'_n(t) + \sqrt{n}R_n(t),$$

where

$$W'_n(t) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} [h_1(X_i,t) - U(t)].$$

(5.13)

If we are able to show that $W'_n$ converges weakly on $[0,1]$ to a Gaussian process $Z$, and the remainder term $\sqrt{n}R_n(t)$ converges to 0 in probability uniformly on $[0,1]$, then the Slutsky argument proves weak convergence on $[0,1]$ of $W_n$ to the process $W$.

Note that the process $W'_n$ has the form of an empirical process indexed by functions $h_1(x,t)$. It satisfies all the conditions of Theorem 4.3 for strong mixing and, hence, also for absolutely regular sequences. Then, by this theorem, $W'_n$ converges weakly in $D[0,1]$ to the mean-zero Gaussian process $W$ with covariance structure given by (4.31), which is equivalent to (5.12) up to the factor $4$ (due to the factor $2$ on the r.h.s. of (5.13)).

Convergence of the remainder term to 0 is essential for the case of $U$-statistics. We present the proof of it in the following section.

5.3 Bounds on the remainder $R_n(t)$

The first lemma establishes conditions under which an arbitrary sequence of processes converges to zero in probability uniformly over an interval. Since it does not specify the form of the process or any dependence structure, this lemma can be applied in a more general context and may be of independent interest.
In this lemma we shall use the ordinary Lipschitz-continuity condition, so we shall first remind the reader of the definition of a Lipschitz-continuous function on $[0, 1]$.

**Definition 5.1** A function $f$ on $[0, 1]$ is called Lipschitz-continuous if for all $s, t \in [0, 1]$ there is a constant $C$ such that

$$|f(t) - f(s)| \leq C|t - s|.$$ 

In what follows $C, C_1, C_2, \ldots$ denote some positive constants.

**Lemma 5.4** Let $\{X_n(t), t \in [0, 1]\}_{n \in \mathbb{N}}$ be a sequence of $\mathbb{R}$-valued stochastic processes with $X_n(0) = 0$ a.s. Suppose that for some $\alpha, \beta, \gamma > 0$, $\alpha \leq 1$, $\gamma < 1$.

(i) There exists a Lipschitz-continuous function $f$ on $[0, 1]$ such that

$$\mathbb{E}|X_n(t) - X_n(s)|^2 \leq \frac{|f(t) - f(s)|^\gamma}{n^\beta}, \; \forall s, t \in [0, 1].$$

(ii) There exist a Lipschitz-continuous function $g$, a monotone Lipschitz-continuous function $\Lambda$ and a process $Y_n$ on $[0, 1]$ with

$$\mathbb{E}|Y_n(t) - Y_n(s)|^r \leq C_1[(g(t) - g(s))^{1+h} + \frac{1}{n}(g(t) - g(s))^{1+h}], \; \forall s, t \in [0, 1]$$

for some $r > 0$, $0 < h < 1$; such that for all $s, t, \delta$ with $0 \leq s \leq t < s + \delta \leq 1$:

$$|X_n(t) - X_n(s)| \leq |X_n(s + \delta) - X_n(s)| + |Y_n(s + \delta) - Y_n(s)| + |\Lambda(s + \delta) - \Lambda(s)| n^\alpha.$$ 

Then

$$\sup_{t \in [0, 1]} |X_n(t)| \to 0 \text{ in probability},$$

whenever $\alpha(1 - \gamma) < \beta$.

**Remark:** The particular usefulness of this lemma is in the fact that for uniform convergence of a process to 0 on an interval we do not need bounds on the moments of increments higher than 2 (which are necessary in most similar applications) due to the factor $n^\delta$ in the denominator of (5.14).

**Proof** For $k = 0, 1, 2, \ldots, K$ (we shall specify $K$ later) define refining
partitions of $[0, 1]$ into $2^k$ subintervals: $0 = s_0^{(k)} < s_1^{(k)} < \cdots < s_{2^k}^{(k)} = 1$, such that for any $i = 1, 2, \ldots, 2^k$

$$|s_i^{(k)} - s_{i-1}^{(k)}| \leq 2^{-k}, \quad (5.17)$$

Then, due to Lipschitz-continuity of $\Lambda$, also

$$|\Lambda(s_i^{(k)}) - \Lambda(s_{i-1}^{(k)})| \leq C \cdot 2^{-k}. \quad (5.18)$$

For $t \in [0, 1]$ we denote by $i_k(t)$ the index of that point of the $k$th partition that is closest to $t$ from the left, i.e.

$$s_{i_k(t)}^{(k)} \leq t < s_{i_k(t)+1}^{(k)}.$$

In this way we have defined a "chain":

$$0 = s_0^{(0)} \leq s_1^{(1)} \leq \cdots \leq s_k^{(k)} \leq t \leq s_{k+1}^{(k)},$$

and we can write

$$X_n(t) = \sum_{k=1}^{K} [X_n(s_{i_k(t)}^{(k)}) - X_n(s_{i_{k-1}(t)}^{(k)})] + [X_n(t) - X_n(s_{i_k(t)}^{(k)})].$$

Then

$$\sup_{t \in [0, 1]} |X_n(t)| \leq \max_{i_1(t)} |X_n(s_{i_1(t)}^{(1)}) - X_n(s_{i_0(t)}^{(0)})|$$

$$+ \max_{i_2(t)} |X_n(s_{i_2(t)}^{(2)}) - X_n(s_{i_1(t)}^{(1)})| + \cdots$$

$$+ \max_{i_{k-1}(t), i_k(t)} |X_n(s_{i_k(t)}^{(k)}) - X_n(s_{i_{k-1}(t)}^{(k-1)})|$$

$$+ \sup_{s_{i_k(t)}^{(k)} \leq t < s_{i_{k+1}(t)}^{(k)}} |X_n(t) - X_n(s_{i_k(t)}^{(k)})|.$$  

By (5.16) and (5.18)

$$\sup_{s_{i_k(t)}^{(k)} \leq t < s_{i_{k+1}(t)}^{(k)}} |X_n(t) - X_n(s_{i_k(t)}^{(k)})|$$

$$\leq \max_{i_k(t)} |X_n(s_{i_k(t)+1}^{(k)}) - X_n(s_{i_k(t)}^{(k)})|$$

$$+ \max_{i_k(t)} |Y_n(s_{i_k(t)+1}^{(k)}) - Y_n(s_{i_k(t)}^{(k)})| + C \cdot 2^{-K} n^\alpha.$$
Now take $K = \lceil \alpha \log_2 n + \log_2 \frac{2C}{\epsilon} \rceil + 1$, so that $C \cdot 2^{-K} n^{\alpha} < \frac{\epsilon}{2}$. Then, since
\[ \sum_{k=1}^{\infty} \frac{1}{(k+2)^r} < \frac{1}{2}, \]

\[
P \left( \sup_{0 \leq t \leq 1} |X_n(t)| > \epsilon \right) \]
\[ \leq \sum_{k=1}^{K} P \left( \max_{i_{k-1}(t), i_k(t)} |X_n(s_{i_k(t)}^{(k)}) - X_n(s_{i_k(t)}^{(k-1)})| > \frac{\epsilon}{(k+2)^2} \right) \]
\[ + P \left( \max_{i_k(t)} |X_n(s_{i_k(t)}^{(k)}) - X_n(s_{i_k(t)}^{(K)})| > \frac{\epsilon}{(K+3)^2} \right) \]
\[ + P \left( \max_{i_k(t)} |Y_n(s_{i_k(t)}^{(K)}) - Y_n(s_{i_k(t)}^{(K)})| > \frac{\epsilon}{(K+4)^2} \right). \quad (5.19) \]

By Chebyshev’s inequality and (5.14)
\[
P \left( \max_{i_{k-1}(t), i_k(t)} |X_n(s_{i_k(t)}^{(k)}) - X_n(s_{i_k(t)}^{(k-1)})| > \frac{\epsilon}{(k+2)^2} \right) \]
\[ \leq \sum_{i=1}^{2^k} P \left( |X_n(s_i^{(k)}) - X_n(s_{i-1}^{(k)})| > \frac{\epsilon}{(k+2)^2} \right) \]
\[ \leq \sum_{i=1}^{2^k} \left| f(s_i^{(k)}) - f(s_{i-1}^{(k)}) \right| \gamma (k+2)^4 \]
\[ \leq \frac{C_1 (k+2)^4}{n^{\beta} \epsilon^2} \sum_{i=1}^{2^k} |s_i^{(k)} - s_{i-1}^{(k)}| \gamma \leq \frac{C_2 (k+2)^4}{n^{\beta} \epsilon^2} \cdot 2^{k(1-\gamma)}. \]

(here we used the fact that we have nested partitions and Lipschitz-continuity of $f$ and $A$ together with (5.17)). In the same way we get
\[
P \left( \max_{i_k(t)} |X_n(s_{i_k(t)}^{(K)}) - X_n(s_{i_k(t)}^{(K)})| > \frac{\epsilon}{(K+3)^2} \right) \leq \frac{C_3 (K+3)^4}{n^{\beta} \epsilon^2} \cdot 2^{K(1-\gamma)}. \quad (5.21) \]

Similarly, by Markov’s inequality and (5.15) we have
\[
P \left( \max_{i_k(t)} |Y_n(s_{i_k(t)}^{(K)}) - Y_n(s_{i_k(t)}^{(K)})| > \frac{\epsilon}{(K+4)^2} \right) \]
\[ \leq \sum_{i=1}^{2^k} P \left( |Y_n(s_i^{(K)}) - Y_n(s_{i-1}^{(K)})| > \frac{\epsilon}{(K+4)^2} \right) \leq \frac{C_4 (K+4)^{2r}}{\epsilon^r} \cdot 2^{-Kh}. \quad (5.22) \]
since for our choice of \( h, K \) and \( \alpha \) we get \( [(g(t) - g(s))^1 + (g(t) - g(s))^2] \leq C_5 \cdot 2^{-K(1+h)} \). Then, combining (5.20), (5.21) and (5.22) with conditions on \( \alpha, \beta, \gamma \) we obtain

\[
P \left( \sup_{0 \leq t \leq 1} |X_n(t)| > \epsilon \right) \\
\leq \frac{C_5 K(K + 3)^4}{n^\beta \epsilon^2} \cdot 2^{K(1-\gamma)} + \frac{C_4 (K + 4)^{2\gamma}}{\epsilon^r} \cdot 2^{-K\epsilon} \\
\leq \frac{C_6 (K + 3)^5}{\epsilon^2} \cdot n^{\alpha(1-\gamma)-\beta} + \frac{C_7 (K + 4)^{2\gamma}}{\epsilon^r} \cdot n^{-\alpha \beta} \rightarrow 0,
\]

because \((K + 3)^p = (\alpha \log n + \log 2 \cdot \frac{2C}{\epsilon} + 4)^p = o(n^\epsilon)\) for any \( \epsilon, p > 0 \).\(\square\)

Now we want to apply Lemma 5.4 to the process \( \sqrt{n}R_n(t) \). In the next lemma we verify condition (5.14).

**Lemma 5.5** Let the conditions of Theorem 5.2 be satisfied. Then

\[
E[R_n(t) - R_n(s)]^2 \leq \frac{C|t - s|^{1/2}}{n^2}
\]

for all \( s, t: 0 \leq s < t \leq 1 \), where \( C \) is some positive constant.

In the proof of this lemma we shall make use of a fundamental lemma by Yoshihara concerning absolutely regular sequences (Lemma 1 in [86]), so we formulate it here.

Let \( X_1, X_2, \ldots \) be a stationary absolutely regular sequence of random vectors in \( \mathbb{R}^k \) and let \( i_1 < i_2 < \cdots < i_m \) be arbitrary integers. As above, by \( \mathcal{P} \) denote the joint distribution of \( X_{i_1}, \ldots, X_{i_m} \) and for any \( j \in \{1, \ldots, m-1\} \) define a measure \( \mathcal{P}^{(j)} = \mathcal{P}^{(j)}_{(X_{i_1}, \ldots, X_{i_j} \times (X_{i_{j+1}}, \ldots, X_{i_m})} \) on \( \mathbb{R}^{k \times m} \) by

\[
\mathcal{P}^{(j)}(E \times G) = \mathcal{P}((X_{i_1}, \ldots, X_{i_j}) \in E) \cdot \mathcal{P}((X_{i_{j+1}}, \ldots, X_{i_m}) \in G),
\]

where \( E \in \sigma(\mathbb{R}^{k \times j}), \ G \in \sigma(\mathbb{R}^{k \times (m-j)}). \)

**Lemma 5.6** (Yoshihara) Let \( j \in \{1, \ldots, m-1\} \) and let \( f(x_1, \ldots, x_m) \) be a measurable function such that

\[
\left( \int_{\mathbb{R}^{k \times m}} |f(x_1, \ldots, x_m)|^r d\mathcal{P}^{(j)} \right)^{1/r} \leq M_1 < \infty
\]
and
\[
\left( \int_{\mathbb{R}^1 \times \ldots \times \mathbb{R}^m} |f(x_1, \ldots, x_m)|^r d\mathbf{P} \right)^{1/r} \leq M_2 < \infty,
\]
where \( r > 1 \). Denote \( M = \max(M_1, M_2) \). Then
\[
|\int_{\mathbb{R}^1 \times \ldots \times \mathbb{R}^m} f(x_1, \ldots, x_m) d\mathbf{P} - \int_{\mathbb{R}^1 \times \ldots \times \mathbb{R}^m} f(x_1, \ldots, x_m) d\mathbf{P}^{(j)}| \leq 4M\beta_d^{1/s},
\]
where \( d = |i_{j+1} - i_j| \) and \( \frac{1}{s} + \frac{1}{r} = 1 \).

Note that Lemma 5.2 is a version of this lemma for the case of nearly regular processes.

**Proof of Lemma 5.5** We have
\[
\mathbb{E} \left[ (R_n(t) - R_n(s))^2 \right] = \frac{4}{n^2(n-1)^2} \sum_{1 \leq i < j \leq n} (h(X_i, X_j, t) - h(X_i, X_j, s) - (h_1(X_i, t) - h_1(X_i, s)) - h_1(X_i, s)) - (h_1(X_j, t) - h_1(X_j, s)) + U(t) - U(s))^2
\]
\[
= \frac{4}{n^2(n-1)^2} \sum_{1 \leq i < j \leq n} \sum_{1 \leq i_j < j} \sum \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})], \quad (5.23)
\]
where we denoted
\[
J(X_i, X_j) = h(X_i, X_j, t) - h(X_i, X_j, s) - (h_1(X_i, t) - h_1(X_i, s)) - (h_1(X_j, t) - h_1(X_j, s)) + U(t) - U(s)
\]
(for simplicity of notation we skip in \( J(X_i, X_j) \) the dependence on \( t, s \).

If in the quadruple of indices \( (i_1, j_1, i_2, j_2) \) there exists an index which is different from the other three, say, \( j_2 \neq i_1, i_2, j_1 \), then
\[
\int_{\mathbb{R}^2} J(X_{i_1}, X_{j_1}) J(X_{i_2}, y) d\mathbf{F}(y) = 0.\]

We want to make use of this by applying Lemma 5.6 to the terms in the sum (5.23). For this, we split the sum (5.23) into 2 sums:
\[
\sum_{1 \leq i_1 < j_1, i_2 < j_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \quad (5.24)
\]
\[
= \sum_{1 \leq i_1 < j_1, i_2 < j_2 \leq n, i_1 \neq i_2 \text{ or } j_1 \neq j_2} \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})]
\]
\[
+ \sum_{1 \leq i_1 < j_1, i_2 < j_2 \leq n, i_1 = i_2 \text{ and } j_1 = j_2} \mathbb{E}[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})].
\]
In the elements of the first sum at least one index is different from all others, say, \( j_1 \), and suppose \( i_1 \leq i_2 \leq j_1 < j_2 \). Let \( d_1 \) be the largest difference between consecutive indices. If \( d_1 = j_2 - j_1 \) then from Lemma 5.6 with \( r = s = 2 \) we have that
\[
E[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \leq 4M^{1/2} \beta_{d_1}^{1/2}.
\]
Here
\[
M = \max \{ E[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})]^2, E[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})]^2 \},
\]
where we denoted
\[
E_1[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})]^2 = \int_{\mathbb{R}^4} [J(x_{i_1}, x_{j_1}) J(x_{i_2}, x_{j_2})]^2 dP(x_{i_1}, x_{i_2}, x_{j_1}, x_{j_2}).
\]
Recall that \( h(x, y, t), h_1(x, t) \) and \( U(t) \) are increasing in \( t \) and take values between 0 and 1. Moreover, \( E h_1(X_1, t) = U(t) \). Then, taking \( J \) to the power 2 and using these properties, we get
\[
E[J(X_{i_1}, X_{j_1})]^2 \leq 2E(h(X_{i_1}, X_{j_1}, t) - h(X_{i_1}, X_{j_1}, s)) + 2E[(h_1(X_{i_1}, t) - h_1(X_{i_1}, s)]^2 + 2(U(t) - U(s))
\]
and
\[
E_{F \times F}[J(X_{i_1}, X_{j_1})]^2 \leq 2E[(h_1(X_{i_1}, t) - h_1(X_{i_1}, s)]^2 + 4(U(t) - U(s))
\]
and
\[
E_{F \times F}[J(X_{i_1}, X_{j_1})]^2 \leq 2E[(h_1(X_{i_1}, t) - h_1(X_{i_1}, s)]^2 + 4(U(t) - U(s))
\]
Since \( |J(X_{i_1}, X_{j_1})| \leq 2 \), we have
\[
M \leq 4\max \left\{ \frac{E[J(X_{i_1}, X_{j_1})]^2, E_{F \times F}[J(X_{i_1}, X_{j_1})]^2] \right\}
\]
and
\[
M_{t, s} \leq C|t - s|.
\]
by Lipschitz-continuity of \( U(t) \) and the condition (5.11). Then
\[
\sum_{1 \leq i_1 \leq i_2 \leq j_1 < j_2 \leq n} E[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \leq 4M_{t, s}^{1/2} \sum_{1 \leq i_1 \leq i_2 \leq j_1 < j_2 \leq n} \beta_{d_1}^{1/2}
\]
and
\[
\sum_{1 \leq i_1 \leq i_2 \leq j_1 < j_2 \leq n} k^2 \beta_{k}^{1/2}.
\]
If $d_1$ is not $j_2 - j_1$ we apply Lemma 5.6 twice to obtain
\[
E[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \leq 4M_{t,s}^{1/2} \beta_{d_1}^{1/2} + 4M_{t,s}^{1/2} \beta_{d_2}^{1/2}
\]
and then
\[
\sum_{1 \leq i_1 \leq i_2 < j_1 < j_2 \leq n \atop d_1 \neq j_2 - j_1} E[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \leq 4M_{t,s}^{1/2} \sum_{1 \leq i_1 < i_2 < j_1 < j_2 \leq n \atop d_1 \neq j_2 - j_1} (\beta_{d_1}^{1/2} + \beta_{d_2}^{1/2}) \leq 8M_{t,s}^{1/2} n^2 \sum_{k=1}^n k\beta_k^{1/2}.
\]

Estimating the sums in the other cases in the same way we get that the first sum is bounded by
\[
\sum_{1 \leq i_1 \leq i_2 < j_1 < j_2 \leq n \atop i_1 \neq i_2 \text{ or } j_1 \neq j_2} E[J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \leq 8M_{t,s}^{1/2} n^2 \sum_{k=1}^n k\beta_k^{1/2}. \tag{5.26}
\]

For the elements of the second sum, analogous to (5.25), we have
\[
E \left[ J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2}) \right] = E[J(X_{i_1}, X_{j_1})]^2 
\leq \left[ 6U(t) - U(s) \right] + 2| Eh(X_{i_1}, X_{j_1}, t) - Eh(X_{i_1}, X_{j_1}, s) | 
= \frac{M_{t,s}}{4} \leq \frac{3}{2} M_{t,s}^{1/2},
\]
and then
\[
\sum_{1 \leq i_1 \neq j_1 \leq n} E[J(X_{i_1}, X_{j_1}) J(X_{i_1}, X_{j_1})] \leq \frac{3}{2} M_{t,s}^{1/2} n^2. \tag{5.27}
\]

So combining (5.24), (5.25), (5.26) and (5.27) we obtain
\[
E[R_n(t) - R_n(s)]^2 \leq \frac{C|t - s|^{1/2}}{(n - 1)^2} \approx \frac{C|t - s|^{1/2}}{n^2}
\]
for $n$ big enough. This completes the proof of the lemma. \qed
Lemma 5.7 Under the conditions of Theorem 5.2

\[ \sup_{0 \leq t \leq 1} \sqrt{n}|R_n(t)| \rightarrow 0 \]

in probability, as \( n \rightarrow \infty \).

Proof According to the previous lemma, the condition (5.14) of Lemma 5.4 holds with \( \gamma = \frac{1}{2} \) and \( \beta = 1 \).

For any \( s, t, \delta > 0 \) with \( 0 \leq s \leq t \leq s + \delta \leq 1 \) we have

\[
R_n(t) - R_n(s) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h(X_i, X_j, t) - h(X_i, X_j, s))
- (h_1(X_i, t) - h_1(X_i, s)) - (h_1(X_j, t) - h_1(X_j, s))
+ (U(t) - U(s))] \\
\leq \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h(X_i, X_j, s + \delta) - h(X_i, X_j, s))
- (h_1(X_i, s + \delta) - h_1(X_i, s))
- (h_1(X_j, s + \delta) - h_1(X_j, s))
+ (U(s + \delta) - U(s))] \\
+ \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h_1(X_i, s + \delta) - h_1(X_i, s))
+ (h_1(X_j, s + \delta) - h_1(X_j, s))] \\
\leq |R_n(s + \delta) - R_n(s)|
+ \frac{2}{n} \sum_{i=1}^{n} [(h_1(X_i, s + \delta) - U(s + \delta))
- (h_1(X_i, s) - U(s))] + 2[U(s + \delta) - U(s)] \\
\leq |R_n(s + \delta) - R_n(s)| + \frac{1}{\sqrt{n}} |W'_n(s + \delta) - W'_n(s)|
+ 2[U(s + \delta) - U(s)]
\]

and

\[
R_n(s) - R_n(t) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [(h(X_i, X_j, s) - h(X_i, X_j, t))
- (h_1(X_i, s) - h_1(X_i, t)) - (h_1(X_j, s) - h_1(X_j, t))
+ (U(s) - U(t))] \\
\]
\begin{align*}
\leq & \ |R_n(s + \delta) - R_n(s)| \\
+ & \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \left[ (h_1(X_i, t) - h_1(X_i, s)) \\
+ & (h_1(X_j, t) - h_1(X_j, s)) + (h_1(X_j, s + \delta) - h_1(X_j, s)) \\
+ & (h_1(X_i, s + \delta) - h_1(X_i, s)) \right] \\
\leq & \frac{4}{n} \sum_{i=1}^{n} [(h_1(X_i, s + \delta) - U(s + \delta)) - (h_1(X_i, s) - U(s))] \\
+ & 4U(s + \delta) - U(s) \\
\leq & \ |R_n(s + \delta) - R_n(s)| + \frac{2}{\sqrt{n}} |W'_n(s + \delta) - W'_n(s)| \\
+ & 4|U(s + \delta) - U(s)|,
\end{align*}

from which it follows that

\begin{equation}
\sqrt{n}[R_n(t) - R_n(s)] \leq \sqrt{n}[R_n(s + \delta) - R_n(s)] + 2|W'_n(s + \delta) - W'_n(s)| \\
+ 4\sqrt{n}|U(s + \delta) - U(s)|. \tag{5.28}
\end{equation}

where

\[ W'_n(t) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} [h_1(X_i, t) - U(t)], \]

as in (5.13). Note that (5.28) is the representation (5.14) of Lemma 5.4 for the process \( \sqrt{n}R_n(t) \). Now we only have to check that the condition (5.15) of Lemma 5.4 holds for \( W'_n(t) \). For this note that (5.15) holds for \( W'_n(t) \) with \( r = 4 \) by Lemma 4.12 if

\[ \mathbf{E}|(h_1(X_1, t) - U(t)) - (h_1(X_1, s) - U(s))|^{2+\delta'} \leq C|t - s| \tag{5.29} \]

for some \( \delta' > 0 \). Using the properties that both \( U(t) \) and \( h_1(x, t) \) are increasing in \( t \) and take values between 0 and 1, and that \( \mathbf{E}h_1(X_1, t) = U(t) \), we get

\begin{align*}
\mathbf{E} \quad & [(h_1(X_1, t) - U(t)) - (h_1(X_1, s) - U(s))]^2 \\
\leq & \mathbf{E}[h_1(X_1, t) - h_1(X_1, s)]^2 + (U(t) - U(s))^2 \\
- & 2(U(t) - U(s))\mathbf{E}[h_1(X_1, t) - h_1(X_1, s)] \\
\leq & \mathbf{E}[h_1(X_1, t) - h_1(X_1, s)] + (U(t) - U(s)) \\
= & 2(U(t) - U(s)) \leq C(t - s). \tag{5.30}
\end{align*}
Limit theorems for $U$-statistics and $U$-statistics processes

Since $|(h_1(X_1,t) - U(t)) - (h_1(X_1,s) - U(s))| \leq 1$, (5.29) follows from (5.30).

So the condition (5.16) of Lemma 5.4 holds with $\alpha = \frac{1}{2}$. Thus, by applying Lemma 5.4 to the process $\sqrt{n}P_n(t)$, the statement of the lemma follows. □

By this lemma we also complete the proof of Theorem 5.2. □

Remark: It seems possible to replace in Theorem 5.2 the boundedness condition on the kernel function $(0 \leq h(x,y,t) \leq 1)$ by the weaker condition of boundedness of its $(2 + \delta)$th moments, for some $\delta > 0$.

5.4 $U$-process for functionals of absolutely regular sequences

Here we shall use the notation of the previous section.

Theorem 5.3 Let $\{X_n\}_{n \in \mathbb{N}}$ be a functional of an absolutely regular process $\{Z_n\}_{n \in \mathbb{Z}}$ with mixing coefficients $(\beta_k)$, such that 1-approximation condition holds. Suppose that $h \in \mathcal{H}$, that for all $t$ both $h$ and $h_1$ are 1-Lipschitz continuous in the sense (5.1) and (5.2), and that the sequences $(\beta_k)$, $(\alpha_k)$ and $(\phi_k)$ satisfy the following summability conditions:

$$\sum_{k=0}^{\infty} k^2(\alpha_k + \beta_k + \phi_k)^{\tau/(2+\tau)} < \infty, \quad (5.31)$$

for some $\tau \in (0,1)$. Suppose, moreover, that $U(t)$ is continuous on $[0,1]$ and that the conditions (5.10) and (5.11) hold. Then the process $\{W_n = \sqrt{n}(U_n(t) - U(t)), t \in [0,1]\}_{n \in \mathbb{N}}$ converges weakly in $D[0,1]$ to the mean-zero Gaussian process $\{W(t), t \in [0,1]\}$ with covariance structure

$$\mathbb{E}[W(s)W(t)] = 4 \text{Cov}[h_1(X_1,s)h_1(X_1,t)]$$
$$+ 4 \sum_{k=1}^{\infty} \text{Cov}[h_1(X_1,s)h_1(X_{k+1},t)]$$
$$+ 4 \sum_{k=1}^{\infty} \text{Cov}[h_1(X_1,t)h_1(X_{k+1},s)]. \quad (5.32)$$

Moreover, the series (5.32) converge absolutely, and the limit process $W$ has continuous sample paths on $[0,1]$ with probability 1.
5.4. $U$-process for functionals of absolutely regular sequences

**Proof** By the Hoeffding decomposition,

$$W_n(t) = \sqrt{n}(U_n(t) - U(t)) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} [h_1(X_i, t) - U(t)] + \sqrt{n}R_n(t).$$

Theorem 4.4 implies that

$$W'_n(t) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} [h_1(X_i, t) - U(t)] \overset{W}{\rightarrow} W(t) \text{ as } n \rightarrow \infty, \ t \in [0, 1],$$

where $W$ is a zero-mean Gaussian process with the covariance structure given by (5.32). The statement of the theorem will follow if we show that

$$\sup_{t \in [0,1]} \sqrt{n}|R_n(t)| \rightarrow 0 \text{ in probability, as } n \rightarrow \infty.$$  

This is shown in the following lemma.

**Lemma 5.8** Let the condition of Theorem 5.3 be satisfied. Then

$$\sup_{t \in [0,1]} \sqrt{n}|R_n(t)| \rightarrow 0 \text{ in probability, as } n \rightarrow \infty.$$  

**Proof** The statement of the lemma can be proved by applying Lemma 5.4 of the previous section to $\sqrt{n}|R_n(t)|$. Note that the condition (iii) of that lemma holds with

$$X_n(t) = \sqrt{n}R_n(t), \ Y_n(t) = W'_n(t), \ A(t) = U(t), \ r = 4, \ \alpha = 1/2,$$

by the same arguments as in Lemma 5.7 and the bound (4.55) on the fourth moment of $(W'_n(t) - W'_n(s))$. What remains to check, is the condition (i), or, rather, the form of it which was used in the proof of Lemma 5.4.

Recall that there we have defined refining partitions of $[0,1]$ into $2^k$ subintervals of the size $C2^{-k}$, $k = 1, 2, ..., K$ with $K = O(\alpha \log_2 n)$, so that the size of the finest partition is $\Delta = Cn^{-\alpha}$. For such partitions we required that

$$\mathbb{E}_n[R_n(t + \Delta) - R_n(t)]^2 \leq \frac{C\Delta^\gamma}{n^\beta} \quad (5.33)$$

for some $\gamma \in (0,1), \beta > 0$, such that $\alpha(1-\gamma) < \beta$. So, rather then verifying the condition (i), we shall check its special case (5.33) with the specified size of partition $\Delta = Cn^{-\alpha}$. For this we shall proceed in the way similar to Lemma 5.5.
Let $t, s \in [0, 1]$ be such that $|t - s| = \Delta$. We have

$$
\mathbb{E} \left[ R_n(t) - R_n(s) \right]^2
\leq \frac{4}{n^2(n - 1)^2} \sum_{1 \leq i_1 < j_1 \leq n} \sum_{1 \leq i_2 < j_2 \leq n} \mathbb{E} |J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})| (5.34)
$$

where

$$
J(X_i, X_j) = h(X_i, X_j, t) - h(X_i, X_j, s) - (h(X_i, t) - h(X_i, s))
- (h(X_j, t) - h(X_j, s)) + U(t) - U(s).
$$

We shall estimate most of the individual terms in the sum (5.34) with the help of Lemma 5.2. For this note again that if, for instance, $i_1 \neq i_2, j_1, j_2$, then

$$
\int_{\mathbb{R}} J(y, X_{j_1}) J(X_{i_2}, X_{j_2}) dF(y) = 0,
$$

and then, if $i_1 < j_1 \leq i_2 \leq j_2$,

$$
\left| \mathbb{E} [J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] - \mathbb{E} P_{(j_1, i_2, j_2) \times i_1} [J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] \right|
\leq 4M^{1/2}(\beta_k + \alpha_k)^{1/2} + 2\phi_k,
$$

where $k = \lfloor |j_1 - i_1|/3 \rfloor$. Here we have denoted

$$
M = \max \left\{ \mathbb{E} |J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})|^2, \mathbb{E} P_{(j_1, i_2, j_2) \times i_1} [J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})|^2] \right\}. (5.35)
$$

Since $J(X_i, X_j)$ is bounded by 2, it is easy to see that, analogous to (5.25),

$$
M \leq 4 \max \left\{ \mathbb{E} |J(X_{i_1}, X_{j_1})|^2, \mathbb{E} P_{(j_1, i_2, j_2) \times i_1} [J(X_{i_1}, X_{j_1})|^2] \right\}
\leq 4 \left[ 2 \mathbb{E} h(X_{i_1}, X_{j_1}, t) - \mathbb{E} h(X_{i_1}, X_{j_1}, s) \right] + 6(U(t) - U(s))
\leq C \Delta,
$$

due to the conditions (5.11) and (5.10).

Consider now $k_0$, such that $\phi_{k_0} \leq \Delta$, and split the sum (5.34) into 3 sums:

$$
\sum_{1 \leq i_1 < j_1, i_2 < j_2 \leq n} \mathbb{E} [J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2})] (5.36)
$$
5.4. U-process for functionals of absolutely regular sequences

\[ \sum_{1\leq i_1 < j_1, i_2 < j_2 \leq n} E[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \]
\[ + \sum_{1\leq i_1 < j_1, i_2 < j_2 \leq n, |i_1-i_2|/3 \geq k_0 \text{ or } |i_1-j_2|/3 \geq k_0} E[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \]
\[ + \sum_{1\leq i_1 < j_1, i_2 < j_2 \leq n, i_1=i_2 \text{ and } j_1=j_2} E[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})]. \]

To estimate the first sum in (5.36), note that

\[ \sum_{1\leq i_1 < j_1, i_2 < j_2 \leq n, |i_1-i_2|/3 \geq k_0 \text{ or } |i_1-j_2|/3 \geq k_0} E[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \]
\[ \leq n \sum_{i,j,k \geq k_0, i+j+k \leq n} E[J(X_0, X_i)J(X_{i+j}, X_{i+j+k})], \]

where \( i, k > 0, -n \leq j \leq n \) and \( i+j \geq 0 \).

Suppose \( |i/3| \geq k_0 \). If \( i > k \) and, if \( j > 0 \), also \( i > j \) (i.e. if \( i \) is the largest gap between indices), then

\[ E[J(X_0, X_i)J(X_{i+j}, X_{i+j+k})] \leq 4C\Delta^{1/2} (\beta_{|i/3|} + \alpha_{|i/3|})^{1/2} + \phi_{|i/3|} \]
\[ \leq C \Delta^{1/2} (\beta_{|i/3|} + \alpha_{|i/3|})^{1/2} + \phi_{|i/3|}^{1/2} \]

and, so,

\[ \sum_{i,j,k \geq k_0, i+j+k \leq n} E[J(X_0, X_i)J(X_{i+j}, X_{i+j+k})] \]
\[ \leq C \Delta^{1/2} \sum_{i=k_0}^{n} \sum_{j=1}^{i} ((\beta_{|i/3|} + \alpha_{|i/3|})^{1/2} + \phi_{|i/3|}^{1/2}) \]
\[ \leq C \Delta^{1/2} \sum_{i=1}^{n} j^2 (\beta_{|i/3|} + \alpha_{|i/3|})^{1/2} + \phi_{|i/3|}^{1/2} \leq C_1 \Delta^{1/2}, \]

due to the summation conditions on \( \beta_k, \alpha_k \) and \( \phi_k \). If \( j < 0 \), or if \( k \) is the largest gap, the same estimate holds.

In the case \( j > i, k \) we apply Lemma 5.2 twice to get

\[ E \left[ J(X_0, X_i)J(X_{i+j}, X_{i+j+k}) \right] \]

where we used the boundedness of $J(X_i, X_j)$ and the fact that $|j/3| \geq k_0$ since $j > i$. Then

$$
\sum_{i,j,k : |i+j+k| \leq n, |j/3| \geq k_0, j > k, i} \mathbf{E}[J(X_0, X_i)J(X_{i+j}, X_{i+j+k})]
\leq C \Delta^{1/2} \sum_{j=k_0}^{n} \sum_{i=1}^{j} \sum_{k=1}^{j} \left[ (\beta_{[j/3]} + \alpha_{[j/3]})^{1/2} + \phi_{[j/3]}^{1/2} \right] + \left[ (\beta_{[i/3]} + \alpha_{[i/3]})^{1/2} + \phi_{[i/3]}^{1/2} \right]
\leq C \Delta^{1/2} \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_j^2 [(\beta + \alpha)^{1/2} + \phi_1^{1/2}] + n \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_j^2 \right] \leq C_1 n \Delta^{1/2},
$$

and in total, the first sum in (5.36) is bounded by

$$
\sum_{1 \leq i_1 < j_1, i_2 < j_2 \leq n, |i_1-i_2|/3 \geq k_0 \text{ or } |j_1-j_2|/3 \geq k_0} \mathbf{E}[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \leq C n^2 \Delta^{1/2}. \quad (5.37)
$$

In the second sum in (5.36) there are at most $k_0^2 n^2$ terms, each is bounded by

$$
\mathbf{E} \left[ J(X_{i_1}, X_{j_1}) J(X_{i_2}, X_{j_2}) \right]
\leq 2[\mathbf{E}h(X_{i_1}, X_{j_1}, t) - \mathbf{E}h(X_{i_1}, X_{j_1}, s) + (U(t) - U(s))]
\leq C \Delta \leq C_1 \Delta^{1/2}. \quad (5.38)
$$

Recall that we must choose $k_0 = k_0(n)$ such that $\phi_{k_0} \leq \Delta = C n^{-\alpha}, \alpha = 1/2$. We can take $k_0(n) \sim 0(n^{1/4})$, since the summation condition on $\phi_k : \sum_{k=1}^{\infty} k^2 \phi_k < \infty$ implies that

$$
\sum_{k=0}^{\infty} k^2 \phi_k = \sum_{n=1}^{\infty} n^{1/2} \phi_k < \infty.
$$
Hence, for this choice of \( k_0 \), \( n^{1/2} \phi_{k_0} \longrightarrow 0 \) and then \( \phi_{k_0} \leq Cn^{-1/2} = \Delta \) for sufficiently large \( n \). Then
\[
\sum_{1 \leq i_1 < i_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \leq C \Delta^{1/2} k_0 n^2 = C \Delta^{1/2} n^{2 + \frac{3}{4}}. \quad (5.39)
\]
The third sum, which consists of at most \( n^2 \) terms, is bounded in the similar way by
\[
\sum_{1 \leq i_1 < i_2 \leq n} \mathbb{E}[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] \leq Cn^2 \Delta^{1/2}, \quad (5.40)
\]
since here
\[
\mathbb{E}[J(X_{i_1}, X_{j_1})J(X_{i_2}, X_{j_2})] = \mathbb{E}[J(X_{i_1}, X_{j_1})]^2 \\
\leq 2|\mathbb{E} h(X_{i_1}, X_{j_1}, t) - \mathbb{E} h(X_{i_1}, X_{j_1}, s)| + 6(U(t) - U(s)) \\
\leq C \Delta \leq C_1 \Delta^{1/2}.
\]
Thus, (5.37), (5.39) and (5.40) together with (5.34) and (5.36) imply that, in total,
\[
\mathbb{E}[R_n(t + \Delta) - R_n(t)]^2 \leq \frac{C \Delta^{1/2}(n^2 + n^{2 + \frac{3}{4}})}{n^3} \leq \frac{C \Delta^{1/2}}{n^{1/2}}.
\]
Now, the condition (5.33) is satisfied with \( \gamma = 1/2 \) and \( \beta = 1/2 \), so that the criteria \( \alpha(1 - \gamma) = \frac{1}{4} < \beta = \frac{1}{2} \) is also satisfied. Thus, Lemma 5.4 then implies that \( \sup_{t \in [0,1]} \sqrt{n} |R_n(t)| \longrightarrow 0 \), which, in turn, proves this lemma as well as the theorem.

5.5 \textit{U-statistics with unbounded kernel function}

Here we want to consider the case of \( U \)-statistics with unbounded kernel function. In Chapter 3 it was shown that the \( U \)-statistic with unbounded kernel is weakly consistent for stationary ergodic and absolutely regular sequences under some additional condition on the kernel function and the distributional properties of the sequence. Here we show that the weak consistency of \( U \)-statistics still holds under the same conditions, even if the sequence of observations is not itself absolutely regular, but is a functional of an absolutely regular process.
Theorem 5.4 Let \( \{X_n\}_{n \in \mathbb{N}} \) be a functional of a stationary and absolutely regular process such that 1-approximation condition holds and \( \alpha_l \to 0 \) as \( l \to \infty \). Let \( F \) be a distribution of \( X_1 \). Suppose, moreover, that \( h : \mathbb{R}^2 \to \mathbb{R} \) is a measurable symmetric function satisfying 1-Lipschitz condition (5.1), that also \( \phi_l \to 0 \) as \( l \to \infty \), and that the family of random variables \( \{h(X_i, X_j) : i, j \geq 1\} \) is uniformly integrable. Then

\[
U_n = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h(X_i, X_j) \to \int_{\mathbb{R}^2} h(x, y) dF(x) dF(y) = \theta(F)
\]

in probability, as \( n \to \infty \).

Proof By Theorem 4.1 the process \( \{X_n\}_{n \in \mathbb{N}} \) is nearly regular, which means that for all \( K, L, N \) there exists a sequence of i.i.d. \( N \)-dimensional vectors \( \{B'_s\}_{s \geq 1} \) such that for \( (K + 2L, N) \)-blocking \( \{B_s\}_{s \geq 1} \) of \( \{X_n\} \), the blocks \( B'_s, s = 1, 2, \ldots \) have the same distribution as \( B_s, s = 1, 2, \ldots \) and

\[
P(\|B_s - B'_s\| \geq \alpha) \leq \beta_K + 2\alpha L.
\]

(5.41)

Uniform integrability of \( \{h(X_i, X_j), i, j \geq 1\} \) in turn implies that given \( \epsilon > 0 \) there is a \( \delta > 0 \) such that

\[
E|h(X_i, X_j)1_B| \leq \epsilon
\]

for all measurable sets \( B \) with \( P(B) < \delta \). Let \( \epsilon > 0 \), and find \( \delta > 0 \) such that (5.42) holds. Take \( K, L \) so big that \( 2\beta_K + 4\alpha L < \delta \) and \( \phi_l < \epsilon \), and fix \( N \) such that \( \frac{K + 2L}{N + K + 2L} < \epsilon \). If \( n \) is the sample size, then the number of blocks of length \( N \) in the sample is \( p = \lfloor \frac{n}{K + 2L} \rfloor \).

Define the new kernel \( H : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} \) by

\[
H(x, y) = \frac{1}{N^2} \sum_{1 \leq i, j \leq N} h(x_i, y_j),
\]

where \( x = (x_1, \ldots, x_N), y = (y_1, \ldots, y_N) \) are vectors in \( \mathbb{R}^N \). For independent vectors \( \mathbf{X}, \mathbf{Y} \) with marginal distribution of coordinates \( F \)

\[
E H(\mathbf{X}, \mathbf{Y}) = E h(X_1, Y_1) = \int_{\mathbb{R}^2} h(x, y) dF(x) dF(y) = \theta(F).
\]

Denote \( I_s \) the set of indices in the block \( B_s \), \( I'_s \) the set of indices between blocks \( B_s \) and \( B_{s+1} \), and \( i_p \) the last index in the block \( B_p \). Decompose the
5.5. $U$-statistics with unbounded kernel function

$U$-statistic as follows:

$$U_n = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h(X_i, X_j)$$

$$= \frac{1}{n(n-1)} \left[ \sum_{1 \leq k \neq l \leq p} \sum_{i \in I_k} \sum_{j \in I_l} h(X_i, X_j) + 2 \sum_{k,l=1}^{p} \sum_{i \in I_k} \sum_{j \in I_l} h(X_i, X_j) + 2 \sum_{k=1}^{p} \sum_{i,j \in I_k, i \neq j} h(X_i, X_j) + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j) \right].$$ (5.43)

The first term in the above sum can be written as

$$\frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h(X_i, X_j) = \frac{N^2}{n(n-1)} \sum_{1 \leq k \neq l \leq p} H(B_k, B_l).$$

For i.i.d. vectors $B'_1, B'_2, \ldots$, by Hoeffding’s theorem we have

$$\frac{1}{p(1-p)} \sum_{1 \leq k \neq l \leq p} H(B'_k, B'_l) \overset{a.s.}{\to} \mathbb{E} H(B'_1, B'_2) = \theta(F)$$ (5.44)

as $p \to \infty$.

If $k, l$ are such that $B_l - B'_l \parallel \alpha_L$ and $B_k - B'_k \parallel \alpha_L$, then by 1-Lipschitz continuity of $h$

$$\mathbb{E} |H(B_k, B_l) - H(B'_k, B'_l)| \leq \frac{1}{N^2} \sum_{i \in I_k} \sum_{j \in I_l} \mathbb{E}|h(X_i, X_j) - h(X'_i, X'_j)|$$

$$\leq \frac{1}{N^2} \sum_{i \in I_k} \sum_{j \in I_l} \mathbb{E}|h(X_i, X_j) - h(X'_i, X'_j) + h(X'_i, X_j) - h(X'_i, X'_j)|$$

$$\leq \frac{1}{N^2} \left( \sum_{i \in I_k} \sum_{j \in I_l} \mathbb{E}|h(X_i, X_j) - h(X'_i, X_j)| + \sum_{i \in I_k} \sum_{j \in I_l} \mathbb{E}|h(X'_i, X_j) - h(X'_i, X'_j)| \right) \leq \frac{2N^2 \phi_L}{N^2} \leq 2\phi_L. \quad (5.45)$$

Theorem 4.1 implies that $\mathbb{P}(\parallel B_k - B'_k \geq \alpha_L \text{ or } \parallel B_k - B'_k \geq \alpha_L) \leq 2/\beta_k + 4\alpha_L < \delta$, and then by (5.42)

$$\mathbb{E}(|H(\xi_k, \xi_l) - H(\xi'_k, \xi'_l)| \mathbf{1}_{\parallel B_k - B'_k \geq \alpha_L \text{ or } \parallel B_k - B'_k \geq \alpha_L}) \leq 2\epsilon,$$
from which it follows that
\[ \mathbb{E} \left| \sum_{1 \leq k \neq l \leq p} H(B_k, B_l) - \sum_{1 \leq k \neq l \leq p} H(B_k', B_l') \right| \leq 2p^2 \varepsilon + 2p^2 \phi_L, \]
also taking into account (5.45). And then
\[
\mathbb{E} \left[ \frac{N^2}{n(n-1)} \sum_{1 \leq k \neq l \leq p} H(B_k, B_l) - \frac{1}{p_n(p_n - 1)} \sum_{1 \leq k \neq l \leq p} H(B_k', B_l') \right] \\
\leq 2 \phi_L + 2 \varepsilon < 4 \varepsilon. \tag{5.46}
\]
The expectation of the absolute value of the other terms in the decomposition (5.43) is small. This is the consequence the condition on \(N, L, K\) and the fact that \(\sup_{i,j} \mathbb{E}[|h(X_i, X_j)|] < \infty\), which follows from uniform integrability of \(\{h(X_i, X_j), i, j \geq 1\}:
\[
\frac{2}{n(n-1)} \mathbb{E} \left| \sum_{k,l=1}^{p} \sum_{i \in I_k, j \in I'_l} h(X_i, X_j) \right| \\
\leq \frac{2Cp^2 N(K + 2L)}{n(n-1)} \leq \frac{2C(K + 2L)}{N + K + 2L} < 2C \varepsilon
\]
\[
\frac{2}{n(n-1)} \mathbb{E} \left| \sum_{k,l=1}^{p} \sum_{i \in I_k', j \in I'_l'} h(X_i, X_j) \right| \\
\leq \frac{2Cp^2 (K + 2L)^2}{n(n-1)} \leq \frac{2C(K + 2L)^2}{(N + K + 2L)^2} < 2C \varepsilon^2
\]
\[
\frac{1}{n(n-1)} \mathbb{E} \left| \sum_{k=1}^{p} \sum_{i \in I_k} \sum_{j \in I_k', i \neq j} h(X_i, X_j) \right| \\
\leq \frac{CpN^2}{n(n-1)} \leq \frac{CN}{n} \leq C \varepsilon
\]
for \(n\) big enough, and
\[
\frac{2}{n(n-1)} \mathbb{E} \left| \sum_{i=i_p+1}^{n} \sum_{j=1}^{n} h(X_i, X_j) \right| \\
\leq \frac{2Cn(N + K + 2L)}{n(n-1)} \leq \frac{2C(N + K + 2L)}{(n-1)} < 2C \varepsilon
\]
for \(n\) big enough. Hence, (5.44), (5.46) and (5.47) together imply the statement of the theorem. \(\square\)
5.6 Applications to the dimension estimation

Here we shall study applications of Theorems 5.2 and 5.3 to dimension estimation and try to translate the conditions of these theorems into distributional conditions of the data sequence. Also below we shall give a couple of examples of dynamical systems for which some conditions can be verified directly.

Suppose that the sequence of data \( \{X_n\}_{n \in \mathbb{N}} \) is coming from a weak Bernoulli dynamical system, and assume that \( X_i \in \mathcal{A} \subset \mathbb{R}^d \), where \( \mathcal{A} \) is a compact subset of the space (here we assume for simplicity that \( d = 1 \)). Let \( F \) be the distribution of \( X_1 \).

When estimating the correlation dimension by the Grassberger-Procaccia method, we are studying

\[
U_n(t) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} 1\{|X_i - X_j| \leq t\},
\]

which is the sample correlation integral. Here the kernel function is

\[
h(x, y, t) = 1\{|x - y| \leq t\}
\]

and

\[
h_1(x, t) = \int_{\mathcal{A}} 1\{|x - y| \leq t\} dF(y) = \mathbb{P}_Y(|x - Y| \leq t),
\]

\[
U(t) = \int_{\mathcal{A}} 1\{|x - y| \leq t\} dF(x)dF(y).
\]

First we check the 1-Lipschitz conditions (5.1) and (5.2) on the functions \( h \) and \( h_1 \) given above. The condition (5.1) requires that for all \( X_0, X'_0, X_k \) having the distribution \( F \), all \( t \in [0,1] \)

\[
\mathbb{E}[|h(X_0, X_k) - h(X'_0, X_k)| 1_{|X_0 - X'_0| \leq \epsilon}] \leq \phi(\epsilon),
\]

where \( \phi(\epsilon) \rightarrow 0 \) as \( \epsilon \rightarrow 0 \). Let \( t \in [0,1] \) be fixed and take \( X_0, X'_0, X+k \sim F \). In the case of the indicator function \( h \) (5.48) we have

\[
\mathbb{E}[|1\{|X_0 - X_k| \leq t\} - 1\{|X'_0 - X_k| \leq t\}| 1_{|X_0 - X'_0| \leq \epsilon}] = \mathbb{E}[|1\{(|X_0 - X_k| \leq t \& |X'_0 - X_k| > t)\} - 1\{(|X_0 - X_k| > t \& |X'_0 - X_k| \leq t)\}| 1_{|X_0 - X'_0| \leq \epsilon}] \\
\leq \mathbb{P}_{X_0}\{(|X_0 - X_k| \leq t \& |X'_0 - X_k| > t)\} 1_{|X_0 - X'_0| \leq \epsilon} \\
\leq F(t + \epsilon) - F(t - \epsilon),
\]
and the condition (5.1) is satisfied in the following cases:

\begin{itemize}
\item[(s)\,] \( F(t + \epsilon) - F(t - \epsilon) \leq \phi(\epsilon) \to 0 \) as \( \epsilon \to 0 \) if \( F \) is continuous, or
\item[(s)\,] \( F(t + \epsilon) - F(t - \epsilon) \leq 2M\epsilon \) if \( F \) has a bounded (by \( M \)) density, or
\item[(s)\,] \( F(t + \epsilon) - F(t - \epsilon) \leq 2C\epsilon \) if \( F \) is Lipschitz-continuous.
\end{itemize}

The same arguments show that the 1-Lipschitz condition (5.2) holds also in these cases for \( h_1 \) given by (5.49).

Next, we study the Lipschitz condition (5.10) on \( U(t) \). Theorem 5.3 requires that for all \( t, s \in [0, 1] \)

\[ |U(t) - U(s)| \leq C|t - s|, \]

where \( C \) is a positive constant independent on \( t, s \). If the distribution function \( F \) has a bounded density \( f_x \), then, using the formula for transformation of densities, one can show that the density of \( \rho_{xy} = |X - Y| \) also exists (let us call it \( f_{x-y} \)), and is bounded, by, say, \( M < \infty \). Hence, we have

\[ |U(t) - U(s)| = \int_s^t f_{x-y}(\tau) \, d\tau \leq M|t - s|, \]

and, so, the Lipschitz condition on \( U(t) \) is satisfied.

On the other hand, recall that in the case of dimension estimation \( U(t) \) is the correlation integral, which, we assume, obeys the scaling law

\[ C(t) \sim c \cdot t^\alpha \quad \text{as} \quad t \to 0, \]

where \( \alpha \) is the correlation dimension. Assuming that the exact scaling law holds on some interval near 0, say, \([0, 1]\), we have for \( 0 \leq s < t \leq 1 \) and some \( c \in (s, t) \)

\[ |C(t) - C(s)| = t^{\alpha} - s^{\alpha} = (t - s)\alpha c^{\alpha-1} \leq \alpha(t - s) \]

if \( \alpha \geq 1 \), so the Lipschitz condition holds in this case as well.

Finally, we study the condition (5.11), i.e. that for all \( t, s \in [0, 1] \), \( k > 1 \)

\[ |\mathbf{E}h(X_0, X_k, t) - \mathbf{E}h(X_0, X_k, s)| \leq C|t - s|. \quad (5.50) \]

In our case it is equivalent to

\[ \mathbf{E}[\mathbb{1}_{\{|X_0 - X_k| \leq t\}} - \mathbb{1}_{\{|X_0 - X_k| \leq s\}}] = \mathbb{P}(s \leq |X_0 - X_k| \leq t) \leq C|t - s|. \]
If $X_i$'s are $d$-dimensional vectors $(d > 1)$, then we require
\[ P \left( s \leq \| X_0 - X_k \| \leq t \right) \leq C|t - s|, \]
where $\| \cdot \|$ is a norm in $\mathbb{R}^d$ defined in some way (usually the max norm is taken).

If the distribution of the distances $\rho_{0k} = |X_0 - X_k|$ has a density $p(x)$, which is bounded (by $M < \infty$), then the condition (5.30) is satisfied, since then, again,
\[ E = \int_s^t p(\tau)d\tau \leq M|t - s|. \]
On the other hand, if $(X_0, X_k)$ have joint density $f_{0k}$ which is bounded, then, as above, the formula for density transformation implies that the distribution of the distances $\rho_{0k} = |X_0 - X_k|$ also has a density which is bounded, and, hence, the condition (5.30) is satisfied as well.

For many dynamical systems the joint density of pairs $(X_0, X_k)$ does not exist, but the condition (5.30) can still hold. For some dynamical systems the condition (5.30) can be verified directly. Consider first the doubling map on $[0, 1]$ $T_x = 2x \mod 1$ from the example 1 in Chapter 1. This map has the Lebesgue measure on $[0, 1]$ as its invariant measure. Points in $[0, 1]$ can be represented as their binary expansions, i.e. $X_0 = 0.a_1a_2\ldots a_k a_{k+1}$. Then the doubling map $T$ is just a shift by one position to the left, i.e. $X_1 = TX_0 = 0.a_2a_3\ldots$, and $T^k$ is a shift by $k$ positions to the left, i.e. $X_k = T^kX_0 = 0.a_{k+1}a_{k+2}\ldots$. First consider $k = 1$, i.e. the difference $s_1 = X_1 - X_0$. Note that
\[ s_1 = \frac{X_1}{2} - \frac{a_1}{2}, \]
a_1 = 0 or 1. $X_1$ has the uniform distribution on $[0, 1]$, so $\frac{X_1}{2}$ is distributed uniformly on $[0, \frac{1}{2}]$. Now, depending on the value of $a_1$, $s_1$ is distributed uniformly either on $[0, \frac{1}{2}]$ (for $a_1 = 0$), or on $[-\frac{1}{2}, 0]$ (for $a_1 = 1$), and, keeping in mind that these values occur with equal probability $\frac{1}{2}$, and “gluing” the ends of the subintervals $[-\frac{1}{2}, 0]$ and $[0, \frac{1}{2}]$ together, we get that $s_1$ is distributed uniformly on $S^1$ (one-dimensional unit circle). This means that
\[ P\left(s \leq |X_1 - X_0| \leq t\right) = |t - s|. \]
For an arbitrary $k > 1$ note that
\[ s_k = X_k - X_0 = X_k(1 - 2^{-k}) - 0.a_1a_2\ldots a_k. \]
Limit theorems for $U$-statistics and $U$-statistics processes

$X_k$ has the uniform distribution on $[0,1]$, so $X_k(1-2^{-k})$ is distributed uniformly on $[0,1-2^{-k}]$. $0.a_1a_2...a_k$ can take $2^k$ values with equal probability. For each of these values the distribution of $s_k$ is also uniform, but on shifted (by $2^{-k}$) interval of length $1-2^{-k}$. Again, gluing these intervals together, we get that the total distribution of $s_k$ is uniform on $S^1$, and so,

$$\mathbf{P}(s \leq |X_k-X_0| \leq t) = |t-s|,$$

and the condition (5.50) is satisfied for this example with equality and $C = 1$.

A more complicated example for which the condition (5.50) can also be verified directly, is the torus automorphism on the 2-dimensional torus $\tau^2 = S^1 \times S^1$, given by

$$Tx = Ax \mod 1, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{with} \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Let $X_0 = \begin{pmatrix} X_0^1 \\ X_0^2 \end{pmatrix}$ be uniformly distributed on $\tau^2$, i.e. $X_0^{1,2} \sim Un[0,1]$. Then

$$X_k = A^k X_0 \mod 1 = \begin{pmatrix} f_k & f_k-1 \\ f_k-1 & f_k-2 \end{pmatrix} \begin{pmatrix} X_0^1 \\ X_0^2 \end{pmatrix} \mod 1,$$

where $f_i$ are Fibonacci numbers, and

$$X_k - X_0 = (A^k - I) X_0 \mod 1 = \begin{pmatrix} f_k-1 & f_k-1 \\ f_k-1 & f_k-2 - 1 \end{pmatrix} \begin{pmatrix} X_0^1 \\ X_0^2 \end{pmatrix} \mod 1$$

$$= \begin{pmatrix} (f_k-1)X_0^1 + f_k-1X_0^2 \\ f_k-1X_0^1 + (f_k-2-1)X_0^2 \end{pmatrix} \mod 1. \quad (5.51)$$

Both $X_0^1$ and $X_0^2$ are distributed uniformly on $[0,1]$, and, since all entries of the matrix in (5.51) are integers and everything is taken modulo 1, both coordinates of the vector $X_k - X_0$ are also distributed uniformly on $[0,1]$. Suppose for norm in $\mathbf{R}^2$ is taken maximum norm (which is the usual choice in these applications), or either of the coordinates. Then, for all $t,s \in [0,1]$

$$\mathbf{P}(s \leq \|X_k - X_0\| \leq t) = |t-s|$$

and the condition (5.50) is satisfied in this example as well with equality and $C = 1$. 

In conclusion we mention that for the application of Theorem 5.4 to the
Takens estimator we refer the reader to Section 3 of Chapter 3, since the
conditions of Theorem 5.4 for functionals of absolutely regular sequences
are equivalent to those of Theorem 3.2 for absolutely regular sequences.