Empirical U-Quantiles of Dependent Data

Empirische U-Quantile abhängiger Daten

Martin Wendler

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
an der Fakultät für Mathematik der Ruhr-Universität Bochum

2011

Poetic Abstract

Ein Fürst und doch keines Fürsten Sohn,
Gesetze der Erdmutter waren der Lohn
Für den, der vermaß das Land, das Trigon.

Gebrannt und gehängt, geschleudert, geschlagen,
So hören wir täglich ihr Rufen und Klagen,
doch keiner der Frommen wird sich erbarmen.[1]

Die zwei nun ergeben das Höchste von vielen,
beweisen sie nur genug eigenen Willen.
Dies woll'n wir bei Paaren im Spiegel erzielen.

[1] Second verse freely adapted from by Bentzien [12], page 34.

Abstract

The aim of this thesis is the investigation of the asymptotic behaviour of empirical U-quantiles under dependence. U-quantiles are a generalization of order statistcs and are applicated in robust statistics. Examples are the Hodges-Lehmann estimator of location or the Q_n estimator of scale. Furthermore, we want to study generalized linear statistics, which are linear combinations of U-quantiles. An important tool for the analysis of the asymptotic distribution of quantiles is the Bahadur representation, which gives a relation to the empirical distribution function. U-quantiles can be approximated by the empirical U-distribution function in this way. For this reason, we will develop some new results for U-statistics under different mixing assumptions. We will give pointwise and functional central limit theorems, laws of the iterated logarithm and strong invariance principles for U-quantiles.

Ziel dieser Arbeit ist es, das asymptotische Verhalten von empirischen U-Quantilen unter Abhängigkeit zu analysieren. U-Quantile sind eine Verallgemeinerung von Ordnungsstatistiken und haben Anwendung in der robusten Statistik, so z. B. der Hodges-Lehmann-Schätzer für die Lage einer Verteilung oder der Q_n Skalenschätzer. Weiterhin wollen wir auch verallgemeinerte lineare Statistiken untersuchen, also Linearkombinationen von U-Quantilen. Ein wichtiges Hilfsmittel für die Analyse der asymptotischen Verteilung von Quantilen ist die Bahadur-Darstellung, mit der ein Zusammenhang zur empirischen Verteilungsfunktion hergestellt wird. U-Quantile lassen sich auf diese Weise durch die empirische U-Verteilungsfunktion approximieren. Aus diesem Grund werden wir auch einige neue Resultate für U-Statistiken unter verschiedenen Mischungsbedingungen entwickeln. Wir werden sowohl punktweise also auch funktionale Versionen des zentralen Grenzwertsatzes, des Gesetzes des iterierten Logarithmus und des starken Invarianzprinzips für U-Quantile beweisen.

Preface

During the period from May 2008 until June 2011, I worked on the topic of empirical U-quantiles, hoping that the results of my research will be interesting for statisticians as well as for probabilists. I also hope that I have found a balanced approach between theory and applicaton. Empirical U-quantiles have application in robust statistics, i.e. the analysis of data that might be contaminated by extreme outliers. We extend different classic theorems of probability theory to U-statistics and U-quantiles of dependent data. Our technical conditions include series from mathematics like continued fraction as well as from application like GARCH models used for financial data.

This thesis is based on a series of four papers: Dehling and Wendler [32], [33] on U-statistics, and Wendler [93], [94] on U-quantiles. However, instead of simply glueing these four articles together, we restructured the material, adjusted the notation and improved the technical details in different places.

After an introduction, the second chapter will be a short summary of concepts of measuring dependence and properties of random variables that are weak dependent in the sense of strong mixing or absolute regularity. Chapter 3 and 4 will give an introduction to U-statistics, U-quantiles are dealt with in Chapter 5 and 6. First, we will investigate the pointwise asymptotic behaviour and then the functional limit theorems. The four main Chapters 3 to 6 are divided in three sections each: Definitions and examples in the first part, technical lemmas in the second and main results in the third. I tried to give the reader all the necessary information to understand the ideas and the technical method used without boring him with an excessive repetition of well-known facts.

This research would not have been possible without the support of numerous people. I want to express my deepest gratitude to Herold Dehling, who proposed this interesting topic to me, discussed it with me many times and gave me advice whenever I needed it. I learned a lot from him during the time I wrote my diploma thesis and my PhD-thesis.

I am thankful to Thomas Kott, Aeneas Rooch, Daniel Vogel, Ting Zhang and anonymous referees who read my papers carefully and helped me to improve them with their critical comments. I owe thanks to Julia Tullius who helped me reduce my mistakes concerning English grammar and spelling. I am very grateful that Wei Biao

Wu gave me the opportunity to visit him at the University of Chicago for half a year. I appreciate the financial aid of the German Academic Foundation (Studienstiftung des deutschen Volkes) and the collaborative research center on dynamical structures (SFB 823 Dynamische Strukturen) of the German Research Foundation (DFG).

Last, but not least, I want to thank Bettina and Detlef Wendler from all my heart for being such good parents. Thank you for your encouragement and support during my studies.

Contents

List of Symbols

a_l	approximation constant, $A_l := \sqrt{2\sum_{n=l}^{\infty} a_n}$	p. 19
$\alpha(k)$	strong mixing coefficient	p. 17
$\beta(k)$	absolute regularity coefficient	p. 18
C	constant, which may change its value from line to line	
Cov	covariance of two random variables	
E	expectation	
g_1, g_2	terms of Hoeffding decomposition of kernel g	p. 29
h_1, h_2	terns of Hoeffding decomposition of kernel function h	p. 50
$\mathcal{K}$	reproducing kernel Hilbert space	p. 52
L	constant in variation conditions	p. 20, 30, 50
$\mathbb{N}$	natural numbers $0, 1, 2, \dots$	
$N(\mu, \sigma^2)$	normal distribution with expectation μ and variance σ^2	
P	probability	
$\mathbb{R}$	real numbers	
$R_n(p)$	remainder term in Bahadur representation	p. 80
$t_p = U^{-1}(p)$	U-quantile	p. 65
θ	expectation of a U-statistic	p. 29
$\mathcal{U}_K$	unit ball of the reproducing kernel Hilbert space	p. 52
$U_n(g)$	U-statistic with kernel g	p. 27
$U_n(t)$	empirical U-distribution function	p. 50
$U(t)$	U-distribution function	p. 50
Var	variance of a random variable	
$\mathbb{Z}$	integers	
$\Vert \cdot \Vert$	Euclidean norm on $\mathbb{R}^d$	
$\Vert \cdot \Vert_p$	L^p-norm for random variables	
$\lfloor \cdot \rfloor$	integer part	
$\lceil \cdot \rceil$	ceiling function	

1 Introduction

1.1 Robust Estimation

Let us start with a classic statistical problem. Assume that $(X_n)_{n\in\mathbb{N}}$ is a sequence of independent random variables with a $N(\mu, \sigma^2)$ distribution, which is a standard normal distribution shifted by the location parameter μ and stretched by the scale parameter σ. As the normal distribution has all moments, we have that $\mu = EX_i$ is the expectation and $\sigma^2 = \operatorname{Var} X_i$ is the variance of the random variables. The task is to find good estimators $\hat{\mu}_n = \hat{\mu}(X_1, \ldots, X_n)$ for the location and $\hat{\sigma}_n^2 = \hat{\sigma}^2(X_1, \ldots, X_n)$ for the variance of X_i. Good estimators should have a low bias and a low variance.

The classic solution to this problem are the sample mean $\hat{\mu}_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and the sample variance $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}\right)^2$. There are many justifications for this estimation procedure, one is the theory of U-statistics introduced by Halmos [45] and Hoeffding [49]. U-statistics are generalized means. To estimate a parameter θ which can be expressed as $\theta = E[g(X_i, X_j)]$ for a symmetric measurable function, the U-statistic

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

is (under independence) an unbiased estimator, as indicated by the U in its name. If we investigate the model of independently distributed random variables with an arbitrary density, then $U_n(g)$ is even the uniformly minimum variance unbiased estimator.

In this thesis, we will deal with stationary sequences $(X_n)_{n\in\mathbb{N}}$ of random variables that do not behave as nicely; instead of independence, we will allow different forms of short range dependence including examples such as linear processes of the form $X_n = \sum_{i=1}^\infty a_i Z_{n-i}$, GARCH processes, which are used to model volatility clustering in financial data, and observations from dynamical systems, where $X_{n+1} = T(X_n)$ for an expanding map T. If moments high enough exist (more than second moments for the sample mean, more than fourth moments for the sample variance), we can still use U-statistics. They might not be unbiased, but the bias will vanish asymptotically. The analysis of the asymptotic behaviour will be much more complicated than in the independent case and will be discussed in Chapter 3.

An important object to characterize dynamical systems with $X_{n+1} = T(X_n)$ is the correlation integral

$$C(r) = P(\|(X_{n+1}, X_n) - (Y_{n+1}, Y_n)\| \leq r)$$

where $(Y_n)_{n\in\mathbb{N}}$ is a independent copy of $(X_n)_{n\in\mathbb{N}}$. This can be stimated by the empirical U-distribution function, which is a family of U-statistics indexed by $r \in \mathbb{R}$ (which also could be described as a U-statistc with values in a function space). Chapter 4 is about the empirical U-distribution function.

In addition to dependence, we do not assume that the random variables are normally distributed, the distribution might have much heavier tails, second moments might not exists. Then the sample variance is not a consistent estimator for the scale, as it will not converge. An alternative to estimators based on means or generalized means (U-statistics) are estimators based on empirical quantiles or generalized empirical quantiles (U-quantiles). Let $X_{(1)} \leq X_{(2)} \leq \ldots X_{(n)}$ be the order statistic, i.e. the ordered sample $X_1, \ldots, X_n$. Then the empirical p-quantile is defined as $X_{(\lceil pn \rceil)}$ and the p-U-quantile as the p-quantile of the sample $(g(X_i, X_j))_{1\leq i<j\leq n}$.

Some robust estimators of location that can deal better with heavy tailed data are the median $X_{(\lceil 0.5n \rceil)}$, the trimmed mean $\bar{X}_{0.25} := \frac{2}{n}\sum_{i=\frac{1}{4}n}^{\frac{3}{4}n} X_{(i)}$, which is a linear statistic (a linear combination of quantiles), or the Hodges-Lehmann estimator, which is the 0.5-U-quantile for the kernel function $g(x, y) = \frac{1}{2}(x+y)$. To estimate the scale, one could use the inter quartile distance $T_n = X_{(\lceil 0.75 \rceil)} - X_{(\lceil 0.25 \rceil)}$, the Q_n estimator, which is the 0.25-U-quantile for the kernel function $g(x, y) = |x-y|$ or the winsorized variance, which is the mean of the values of $(X_i - X_j)^2$, where the biggest 25% are replaced by the 0.75-U-quantile for the kernel function $g(x, y) = (x - y)^2$. This is a generalized linear statistic. The asymptotic behaviour of these types of statistics will be investigated in Chapters 5 and 6.

1.2 Four Steps of Linearization

The best studied object in probability theory is the partial sum $\sum_{i=1}^n X_i$ for independent and identically distributed random variables. The U-quantiles and generalized linear statistics we investigate in this thesis are more complicated objects, so we will try to approximate them by such partial sums of independent random variables to investigate their asymptotic behaviour. This will be done in four steps:

Bahadur [11] showed that quantiles are close to the empirical distribution function $F_n(t) = \sum_{i=1}^n \mathbb{1}_{\{X_i \leq t\}}$. The idea is that the empirical quantile will converge to the inverse $F^{-1}(p)$ of the distribution function F. The quantile can be expressed as the generalized inverse $F_n^{-1}(p)$ of empirical distribution function. So if F_n is close to F, there is hope that the slope of F_n is close to the slope of F, which would lead to

$$\frac{F_n(F^{-1}(p)) - p}{F^{-1}(p) - F_n^{-1}(p)} \approx f(F^{-1}(p)).$$

This relation can indeed be used to derive the asymptotic behaviour of quantiles and has been generalized to U-quantiles by Geertsema [42] (but only for independent random variables). We will establish a generalized Bahadur representation under dependence in Theorem 5.3.1 and Theorem 6.3.1.

For U-quantiles, the Bahadur representation gives a relation to the empirical U-distribution function, which is still not a simple partial sum of random variables. Hoeffding [49] found a way to decompose U-statistics into a partial sum and a so-called degenerate part, which is a U-statistic with uncorrelated summands. For dependent data, the summands of the degenerate part might be correlated, but the correlation can be bounded with the help of generalized covariance inequalities, a method first used by Yoshihara [98]. We will investigate the rate of almost sure convergence for the degenerate part in Proposition 3.3.2 and Proposition 4.2.3.

The first two steps (Bahadur representation and Hoeffding decomposition) lead to a partial sum, but as we do not assume that the random variables are independent, there is still a way to go. In this thesis, we will consider sequences $(X_n)_{n\in\mathbb{N}}$ that are near epoch dependent on an underlying process $(Z_n)_{n\in\mathbb{N}}$. Near epoch dependence roughly means that the random variables are close to a function of finitely many of the underlying random variables, $X_n \approx h(Z_{n-l}, \ldots, Z_{n+l})$. If $(Z_n)_{n\in\mathbb{N}}$ is short range dependent (for example in the sense of absolute regularity), then $(h(Z_{n-l}, \ldots, Z_{n+l}))_{n\in\mathbb{N}}$ is also short range dependent.

So the last step is to show that the partial sum of short range dependent random variables have the same behaviour as the ones of independent random variables. Two techniques are useful: blocking and coupling. If big blocks are seperated by small blocks, the depedence vanishes as the size of the small blocks increases. Furthermore, absolute regular observation can be replaced by a independent ones by coupling methods, such that the replacement differs from the original random variables only with a small probability. More details about near epoch dependence and short range dependence in the sense of strong mixing or absolute regularity can be found in Chapter 2.

1.3 Main Limit Theorems of this Thesis

With the four steps described above, we can approximate U-quantiles of dependent data by a partial sum of independent data. What asymptotic behaviour can be expected for such an object? Note that the sample mean $\bar{X}_n$ is a consistent estimator for the mean, that means $\frac{1}{n}\sum_{i=1}^n(X_i-EX_i)$ converges to zero for independent random variables. But there are different types of convergence in probability theory: The weak law of large numbers gives convergence in probability, the strong law of large numbers almost sure convergence (which implies convergence in probability).

The question arises what other norming sequences than $\frac{1}{n}$ we could use and still get convergence. It turns out that the answer is different for the two types of convergence. For convergence in probability, the borderline case is the norming sequence $\frac{1}{\sqrt{n}}$: $\frac{1}{\sqrt{n}}\sum_{i=1}^n(X_i-EX_i)$ does not converge to 0 in probability, but weakly to a normal distribution. This was first proved by de Moivre for Bernoulli distributed random variables [34]. We will show that this holds for U-statistics of dependent data in Theorem 3.3.1 and for U-quantiles in Corollary 5.3.3.

For the almost sure convergence, the interesting borderline case with a nontrivial limit behaviour is the norming sequence $\frac{1}{\sqrt{n\log\log n}}$. We obtain the following asymptotic behaviour:

$$\limsup_{n\to\infty}\pm\sqrt{\frac{1}{2\operatorname{Var}[X_1]n\log\log n}}\sum_{i=1}^n(X_i-EX_i)=1$$

almost surely. This law of the iterated logarithm was originally established for partial sums of independent and bounded random variables by Khintchine in 1927 [61] and has been extended to dependent random variables by many authors. We will give a version for U-statistics of dependent data in Theorem 3.3.3 and for U-quantiles in Corollary 5.3.4.

As we want to investigate not only single U-quantiles, but also generalized linear statistics, we are interested in functional limit theorems. The empirical process $(\frac{1}{\sqrt{n}}\sum_{i=1}^n(\mathbb{1}_{\{X_i\leq t\}}-P(X_i\leq t)))_{t\in[0,1]}$ converges weakly to a Brownian Bridge, as was proved by Donsker [36], we will give a functional central limit theorem for the empirical U-process in Corollary 4.3.2 and for the empirical U-quantile process in Corollary 6.3.3. To prove this, we will establish an almost sure invariance principle for the empirical U-process in Theorem 4.3.1, i.e. an almost sure approximation by a Gaussian process, similar to the invariance principle Kiefer [60] gave for the empirical process. Theorem 6.3.2 says that a strong invariance principle also holds for the empirical U-quantile process.

The functional law of the iterated logarithm established by Finkelstein [41] says that $((\frac{1}{\sqrt{2n\log\log n}}\sum_{i=1}^n(\mathbb{1}_{\{X_i\leq t\}}-P(X_i\leq t)))_{t\in[0,1]})_{n\in\mathbb{N}}$ is almost surely a relatively compact sequence of functions. As functional laws of the iterated logarithm hold for Gaussian sequences, we will conclude that the functional law of the iterated logarithm is valid for the empirical U-process (Corollary 4.3.3) and for the empirical U-quantile process (Corollary 6.3.4).

2 Short Range Dependent Data

2.1 Measuring Dependence

Many classic results in probability theory like the central limit theorem or the law of the iterated logarithm start with the assumption that $(X_n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed random variables. In this thesis, we want to avoid the condition of independence, but still assume that $(X_n)_{n\in\mathbb{N}}$ is stationary. If the dependence is too strong, we cannot expect a similar limit behaviour. For example, if $X_n = X_1$, then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = X_1$ and the law of large numbers fails. There are different methods to measure dependence and to define short range dependence. For independent random variables X_n, X_m and all measurable sets A, B, we have $P(X_n \in A, X_m \in B) - P(A)P(B) = 0$. For dependent random variables, this equality will not hold for all sets A, B. There are a bunch of so-called mixing assumption (like strong mixing introduced by Rosenblatt [78], absolute regularity, uniform mixing) which say that the this difference converges to 0 as $|n-m| \to \infty$ (uniformly for all sets A and B in some sense).

On the other hand, under independence we have $\mathrm{Cov}(f(X_n), g(X_m)) = 0$ for all functions f and g for which this covariance exists. Doukhan and Louhichi [39] established a concept of weak dependence taking the maximum of the covariance $\mathrm{Cov}(f(X_n), g(X_m))$ for f and g in some class of test functions. Many of the well-known mixing coefficients can be rewritten as bounds for such covariances. A third concept of measuring dependence uses the fact that many processes can be represented as functionals of underlying sequences of random variables, for example linear processes or GARCH processes. If the underlying sequence $(\xi_n)_{n\in\mathbb{N}}$ is independent and $X_n = f(\xi_{n-l}, \ldots, \xi_{n+l})$ is a function of a finite part of the underlying sequence, then the process $(X_n)_{n\in\mathbb{N}}$ is m-dependent. More general, if the sequence $(\xi_n)_{n\in\mathbb{N}}$ is short range dependent and X_n can be approximated by a function of a finite part of the underlying sequence, then $(X_n)_{n\in\mathbb{N}}$ is short range dependent. Such processes X_n are called approximating functionals of $(\xi_n)_{n\in\mathbb{N}}$ or near epoch dependent on $(\xi_n)_{n\in\mathbb{N}}$.

In this thesis, we will concentrate on the first concept (mixing) and on the third concept (near epoch dependence) and the combination of them. Let us now introduce the strong mixing coefficient:

Definition 2.1.1. 1. Let $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ be two σ-fields on the probability space $(\Omega, \mathcal{F}, P)$. Then the strong mixing coefficient of $\mathcal{A}$ and $\mathcal{B}$ is given by

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup \left\{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{A}, B \in \mathcal{B} \right\}.$$

2. Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process. Then the strong mixing coefficients of (X_n) are given by

$$\alpha(k) = \sup_{n\in\mathbb{N}} \alpha\left(\mathcal{F}_1^n, \mathcal{F}_{n+k}^\infty\right),$$

where $\mathcal{F}_a^l$ is the σ-field generated by random variables $X_a, \ldots, X_l$, and $(X_n)_{n\in\mathbb{N}}$ is called strongly mixing if $\alpha(k) \to 0$ as $k \to \infty$.

Strong mixing in the sense of α-mixing is the weakest of the well-known strong mixing conditions (see Bradley [20] and Doukhan [38] for details about all mixing conditions) and it is a very common assumption and covers some examples, like linear processes if some regularity condition on the density of the innovations hold, as described in Withers [95]. If the innovations have a discrete distribution, the strong mixing assumption might not hold, see Andrews [2]. Strong mixing also excludes data from expanding dynamical systems with $X_{n+1} = T(X_n)$ for a piecewise smooth and expanding map $T : [0,1] \to [0,1]$:

Example 2.1.2. Let $(Z_n)_{n\in\mathbb{N}}$ be independent r.v.'s with $P[Z_n = 1] = P[Z_n = 0] = \frac{1}{2}$ and

$$X_n = \sum_{k=n}^{\infty} \frac{1}{2^{k-n+1}} Z_k.$$

Then $X_{n+1} = 2X_n$ [mod 1] and $(X_n)_{n\in\mathbb{N}}$ is not strong mixing, as

$$\left| P\left[X_1 \in \bigcup_{i=1}^{2^{(k-1)}} \left[(2i-2)2^{-k}, (2i-1)2^{-k}\right],\ X_k \in \left[0, \frac{1}{2}\right] \right] \right.$$
$$\left. - P\left[X_1 \in \bigcup_{i=1}^{2^{(k-1)}} \left[(2i-2)2^{-k}, (2i-1)2^{-k}\right] \right] P\left[X_k \in \left[0, \frac{1}{2}\right] \right] \right| = \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

For this reason, we will study sequences which are near epoch dependent on absolutely regular processes. Absolute regularity was introduced by Volkonskii and Rozanov [91], [92].

Definition 2.1.3. 1. Let $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ be two σ-fields on the probability space $(\Omega, \mathcal{F}, P)$. The absolute regularity coefficient of $\mathcal{A}$ and $\mathcal{B}$ is given by

$$\beta(\mathcal{A}, \mathcal{B}) = E \sup_{A \in \mathcal{A}} |P(A|\mathcal{B}) - P(A)|.$$

2. Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process. Then the absolute regularity coefficients of $(X_n)_{n\in\mathbb{N}}$ are given by
$$\beta(k) = \sup_{n\in\mathbb{N}} \beta\left(\mathcal{F}_1^n, \mathcal{F}_{n+k}^\infty\right),$$
and $(X_n)_{n\in\mathbb{N}}$ is called absolutely regular, if $\beta(k) \to 0$ as $k \to \infty$.

For all $A \in \mathcal{F}_1^n$, $B \in \mathcal{F}_{n+k}^\infty$, we have that
$$|P(A\cap B)-P(A)P(B)| \leq \frac{1}{2}\left(|P(A\cap B) - P(A)P(B)| + |P(A\cap B^c) - P(A)P(B^c)|\right)$$
$$= \frac{1}{2}\left(|P(A|B) - P(A)|P(B) + |P(A|B^c) - P(A)|P(B^c)\right)$$
$$= \frac{1}{2}E|P(A|\sigma(\{B\})) - P(A)| \leq \frac{1}{2}E \sup_{A\in\mathcal{F}_{-\infty}^n} |P(A|\sigma(\{B\})) - P(A)| \leq \frac{1}{2}\beta(k)$$
so that $\alpha(k) \leq \frac{1}{2}\beta(k)$ (absolute regularity implies strong mixing). But we will not study absolutely regular sequences themself, but near epoch dependent functionals:

Definition 2.1.4. Let $((X_n, Z_n))_{n\in\mathbb{Z}}$ be a stationary process. We say that $(X_n)_{n\in\mathbb{N}}$ is L^1 near epoch dependent on the process $(Z_n)_{n\in\mathbb{Z}}$ with approximation constants $(a_l)_{l\in\mathbb{N}}$, if
$$E\left|X_1 - E(X_1|\mathcal{G}_{-l}^l)\right| \leq a_l \qquad l = 0, 1, 2\ldots$$
where $\lim_{l\to\infty} a_l = 0$ and $\mathcal{G}_{-l}^l$ is the σ-field generated by $Z_{-l}, \ldots, Z_l$.

Near epoch dependent processes are often called approximating functionals, for example in Borovkova et al [18]. In the literature one often finds L^2 near epoch dependence (where the L^1 norm in the definition is replaced by the L^2 norm), but this requires second moments and we are interested in robust estimation. So we want to allow heavier tails and consider L^1 near epoch dependence. Furthermore, we do not require that the underlying process is independent, it only has to be weakly dependent in the sense of absolute regularity. ARMA and GARCH processes are near epoch dependent, see Hansen [46]. GARCH processes play an important role in financial mathematics, they are used to model volatility clustering in financial data. The simplest GARCH model is the following: Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent standard normal random variables and $X_n = \sigma_n\xi_n$, where $(\sigma_n)_{n\in\mathbb{N}}$ is a random sequence with $\sigma_n^2 = \alpha_0 + \alpha_1 X_{n-1}^2 + \alpha_2\sigma_{n-1}^2$.

Linear process are near epoch dependent, regardless if the innovations have a density or not. However, for later application, we will need the random variables $(X_n)_{n\in\mathbb{N}}$ to have a density and it is not clear that this holds for discrete innovations. But in many examples, the density exists. Solomyak [87] proved that if $(Z_n)_{n\in\mathbb{N}}$ are independent identically distributed and take the values $1, -1$, then for almost all $\lambda \in (\frac{1}{2}, 1)$, the distribution of $\sum_{n=1}^\infty \lambda^n Z_n$ has a density.

This class of near epoch dependent sequences also covers data from dynamical systems, which are deterministic except for the initial value and not covered by strong mixing or near epoch dependence on an independent sequence. Let $T:[0,1]\to[0,1]$ be a piecewise smooth and expanding map such that $\inf_{x\in[0,1]}|T'(x)|>1$. Then there is a stationary process $(X_n)_{n\in\mathbb{N}}$ such that $X_{n+1}=T(X_n)$ which can be represented as a functional of an absolutely regular process, for details see Hofbauer and Keller [52]. The map $T(x)=\frac{1}{x}-\lfloor\frac{1}{x}\rfloor$ is related to the continued fraction

$$X_n=f\left((Z_{n+k})_{k\in\mathbb{N}}\right)=\cfrac{1}{Z_n+\cfrac{1}{Z_{n+1}+\cfrac{1}{Z_{n+2}+\cdots}}}$$

where $(Z_n)_{n\in\mathbb{N}}$ is a stationary, absolutely regular process (even uniformly mixing, see Billingsley [16], p. 50) taking values in $\mathbb{N}$ if the distribution of X_0 is the Gauss measure given by the density $f(x)=\frac{1}{\log 2}\frac{1}{1+x}$ (Note that this map T is not uniformly expanding).

The conditional expectation $E(X_1|\mathcal{G}_{-l}^{l})$ is not easy to deal with. Borovkova et al. [18] gave the following characterization of near epoch dependence: By our assumption, $X_n=E(X_n|\mathcal{G}_{-\infty}^{\infty})$ almost surely, so we can write $X_n=f((Z_{n+k})_{k\in\mathbb{Z}})$.

Lemma 2.1.5. *[Borovkova et al. [18]] Let $(Z_n)_{n\in\mathbb{Z}}$ be a stationary process and $X_n=f((Z_{n+k})_{k\in\mathbb{Z}})$.*

1. *Let $(X_n)_{n\in\mathbb{N}}$ be near epoch dependent on $(Z_n)_{n\in\mathbb{Z}}$ with approximation constants $(a_n)_{n\in\mathbb{N}}$ and $(Z'_n)_{n\in\mathbb{Z}}$ be a copy of $(Z_n)_{n\in\mathbb{Z}}$ such that $Z_{-l}=Z'_{-l},Z_{-l+1}=Z'_{-l+1},\ldots,Z_l=Z'_l$, then for $X'_n=f((Z'_{n+k})_{k\in\mathbb{N}})$*

$$E|X_0-X'_0|\leq 2a_l.$$

2. *If for all copies of $(Z_n)_{n\in\mathbb{Z}}$ with $Z_{-l}=Z'_{-l},Z_{-l+1}=Z'_{-l+1},\ldots,Z_l=Z'_l$ we have $E|X_0-X'_0|\leq a_l$, then $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent on $(Z_n)_{n\in\mathbb{Z}}$ with approximation constants $(a_n)_{n\in\mathbb{N}}$.*

In our proofs, we will investigate not the near epoch dependent process $(X_n)_{n\in\mathbb{N}}$ itself, but some function $(g(X_n))_{n\in\mathbb{N}}$. For general g, it is not clear that the near epoch dependence holds for this transformed process, we will need an additional continuity condition to guarantee near epoch dependence.

Definition 2.1.6. Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary process. A function $g:\mathbb{R}\to\mathbb{R}$ satisfies the variation condition (with respect to the distribution of X_1), if there is a constant L such that

$$E\left[\sup_{\|x-X_0\|\leq\epsilon}|g(x)-g(X_0)|\right]\leq L\epsilon.$$

This condition was introduced by Denker and Keller [35] and is slightly different from the continuity condition used by Borovkova et al. [18]. Note that the variation of g might hold for one distribution, but not for a different one. Now we can give conditions for the near epoch dependence of $(g(X_n))_{n\in\mathbb{N}}$ similar to Proposition 2.11 of Borovkova et al. [18]:

Lemma 2.1.7. *Let $(X_n)_{n\in\mathbb{N}}$ be L^1 near epoch dependent on the process $(Z_n)_{n\in\mathbb{Z}}$ with approximation constants $(a_l)_{l\in\mathbb{N}}$ and g satisfy the variation condition with constant L.*

1. *If $E|g(X_1)|^{1+\delta}<\infty$, then $(g(X_n))_{n\in\mathbb{N}}$ is near epoch dependent on $(Z_n)_{n\in\mathbb{Z}}$ with approximation constants*
$$a_l' = (L + 2^{\frac{1+2\delta}{1+\delta}}\|g(X_1)\|_{1+\delta})a_l^{\frac{\delta}{1+2\delta}}.$$

2. *If $g(X_1)$ is bounded, then $(g(X_n))_{n\in\mathbb{N}}$ is near epoch dependent on $(Z_n)_{n\in\mathbb{Z}}$ with approximation constants*
$$a_l' = (L + 4\|g(X_1)\|_\infty)a_l^{\frac{1}{2}}.$$

Proof. 1. Let $(Z_n')_{n\in\mathbb{Z}}$ be a copy of $(Z_n)_{n\in\mathbb{Z}}$ as in Lemma 2.1.5, then $E|X_0-X_0'|\leq 2a_l$ and by the Markov inequality $P(|X_0-X_0'|>a_l^{\frac{\delta}{1+2\delta}})\leq 2a_l^{\frac{1+\delta}{1+2\delta}}$. So by the variation condition and by the Hölder inequality

$$\begin{aligned}
&E|g(X_0)-g(X_0')|\\
=&E|g(X_0)-g(X_0')|\mathbb{1}_{\{|X_0-X_0'|\leq a^{\frac{\delta}{1+2\delta}}\}} + E|g(X_0)-g(X_0')|\mathbb{1}_{\{|X_0-X_0'|> a^{\frac{\delta}{1+2\delta}}\}}\\
\leq& La^{\frac{\delta}{1+2\delta}} + 2\|g(X_1)\|_{1+\delta}(P(|X_0-X_0'|>a^{\frac{\delta}{1+2\delta}}))^{\frac{\delta}{1+\delta}}\\
\leq&(L+2^{\frac{1+2\delta}{1+\delta}}\|g(X_1)\|_{1+\delta})a_l^{\frac{\delta}{1+2\delta}}
\end{aligned}$$

and the statement follows with Lemma 2.1.5.

2. With X_0' as above, we have by the Markov inequality $P(|X_0-X_0'|>\sqrt{a_l})\leq 2\sqrt{a_l}$. So by the variation condition

$$\begin{aligned}
&E|g(X_0)-g(X_0')|\\
=&E|g(X_0)-g(X_0')|\mathbb{1}_{\{|X_0-X_0'|\leq\sqrt{a_l}\}} + E|g(X_0)-g(X_0')|\mathbb{1}_{\{|X_0-X_0'|>\sqrt{a_l}\}}\\
\leq& L\sqrt{a_l} + 2\|g(X_1)\|_\infty(P(|X_0-X_0'|>\sqrt{a_l}))\\
\leq&(L+4\|g(X_1)\|_\infty)a_l^{\frac{1}{2}}
\end{aligned}$$

and the statement follows with Lemma 2.1.5. □

2.2 Covariance Inequalities

An important tool in the analysis of weakly dependent random variables are covariance inequalities. We can write $\alpha(n) = \sup |E[Y_1Y_2] - E[Y_1]E[Y_2]|$, where Y_1 and Y_2 random variables such that Y_1 is measurable with respect to $\mathcal{F}_1^k$ and Y_2 with respect to $\mathcal{F}_{k+n}^\infty$ and take only the values 0 and 1. For more general random variables, we have

Lemma 2.2.1. *[Davydov [25]] Let be Y_1 and Y_2 random variables such that Y_1 is measurable with respect to $\mathcal{F}_1^k$ and Y_2 with respect to $\mathcal{F}_{k+n}^\infty$ for some $k \in \mathbb{N}$, then*

$$|E[Y_1Y_2] - E[Y_1]E[Y_2]| \leq 10 \|Y_1\|_{p_1} \|Y_2\|_{p_2} \alpha^{\frac{1}{p_3}}(n)$$

for all $p_1, p_2, p_3 \in [1, \infty]$ with $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = 1$.

In this inequality, the case $p_i = \infty$ is included and should be understood as $\frac{1}{\infty} = 0$. For a stationary and strongly mixing sequence $(X_n)_{n\in\mathbb{N}}$, this inequality is useful to find bounds for the variance. Assume that $\|X_1\|_{2+\delta} < \infty$ and $\sum_{k=0}^\infty \alpha^{\frac{\delta}{2+\delta}}(k) < \infty$, then

$$\begin{aligned}\operatorname{Var}[\bar{X}] &= \operatorname{Var}[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n^2}\sum_{i,j=1}^n \operatorname{Cov}[X_i, X_j] \\ &= \frac{1}{n^2} 2\sum_{k=1}^{n-1}(n-k)\operatorname{Cov}[X_0, X_k] + \frac{1}{n}\operatorname{Var}[X_1] \leq \frac{2}{n}\sum_{k=0}^{n-1}\alpha^{\frac{\delta}{2+\delta}}(k)\|X_1\|_{2+\delta}^2.\end{aligned}$$

By dominated convergence, one obtains

$$\operatorname{Var}[\sqrt{n}\bar{X}] \xrightarrow{n\to\infty} \operatorname{Var} X_1 + 2\sum_{k=1}^{n-1}\operatorname{Cov}[X_0, X_k].$$

For treating stochastic processes, fourth moment bounds are a common tool. As a immediate consequence of Lemma 2.2.1 we have the following inequality, which will help us with higher moments:

Lemma 2.2.2. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary strongly mixing sequence of random variables, then for $i \leq j \leq k \leq l$*

$$|E[X_iX_jX_kX_l] - E[X_i]E[X_jX_kX_l]| \leq 10 \|X_i\|_{p_1} \|X_jX_kX_l\|_{p_2} \alpha^{\frac{1}{p_3}}(j-i)$$

and

$$|E[X_iX_jX_kX_l] - E[X_iX_j]E[X_kX_l]| \leq 10 \|X_iX_j\|_{p_1} \|X_kX_l\|_{p_2} \alpha^{\frac{1}{p_3}}(k-j)$$

for all $p_1, p_2, p_3 \in [1, \infty]$ with $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = 1$.

With easy calculations or as a special case of Theorem 2 of Yokoyama [97], a fourth moment inequality follows

Lemma 2.2.3. *[Yokoyama [97]] Let $(X_n)_{n\in\mathbb{N}}$ be a strongly mixing and bounded sequence with $\sum_{k=1}^{\infty} k\alpha(k) < \infty$, then there exists a constant C such that*

$$E\left(\sum_{i=1}^{n} X_i\right)^4 \leq Cn^2.$$

For near epoch dependent sequences, Borovkova et al. [18] proved similar covariance inequalities and derived moment bounds:

Lemma 2.2.4. *[Borovkova et al. [18]] Let $(X_n)_{n\in\mathbb{Z}}$ be a near epoch dependent sequence with approximation constants $(a_l)_{l\in\mathbb{N}}$ on an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$ and $\|X_0\|_{2+\delta} < \infty$ for some $\delta \in (0,\infty]$. Then*

$$|E[X_iX_{i+k}] - (EX_i)(EX_{i+k})| \leq 2\|X_0\|_{2+\delta}^2 \left(\beta\left(\lfloor\frac{k}{3}\rfloor\right)\right)^{\frac{\delta}{2+\delta}} + 4\|X_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} a_{\lfloor\frac{k}{3}\rfloor}^{\frac{\delta}{1+\delta}}.$$

Lemma 2.2.5. *[Borovkova et al. [18]] Let $(X_n)_{n\in\mathbb{Z}}$ be a bounded near epoch dependent sequence with approximation constants $(a_l)_{l\in\mathbb{N}}$ on an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$. Then*

$$\begin{aligned}&|E[X_iX_jX_kX_l] - E[X_i]E[X_jX_kX_l]|\\ &\qquad\leq \left(6\|X_0\|_{2+\delta}^2\left(\beta\left(\lfloor\frac{j-i}{3}\rfloor\right)\right)^{\frac{\delta}{2+\delta}} + 8\|X_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} a_{\lfloor\frac{j-i}{3}\rfloor}^{\frac{\delta}{1+\delta}}\right)\|X_0\|_\infty^2\end{aligned}$$

and

$$\begin{aligned}&|E[X_iX_jX_kX_l] - E[X_iX_j]E[X_kX_l]|\\ &\qquad\leq \left(6\|X_0\|_{2+\delta}^2\left(\beta\left(\lfloor\frac{k-j}{3}\rfloor\right)\right)^{\frac{\delta}{2+\delta}} + 8\|X_0\|_{2+\delta}^{\frac{2+\delta}{1+\delta}} a_{\lfloor\frac{k-j}{3}\rfloor}^{\frac{\delta}{1+\delta}}\right)\|X_0\|_\infty^2.\end{aligned}$$

Lemma 2.2.6. *[Borovkova et al. [18]] Let $(X_n)_{n\in\mathbb{Z}}$ be a bounded near epoch dependent sequence with approximation constants $(a_l)_{l\in\mathbb{N}}$ on an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$. Assume that $\sum_{k=1}^{\infty} k^2(a_k + \beta(k)) < \infty$, then*

$$E\left(\sum_{i=1}^{n} X_i\right)^4 \leq Cn^2.$$

2.3 Coupling

A very powerful tool for proving limit theorems for dependent data are coupling techniques, especially if one deals with nonlinear functionals (where usual covariances are not enough). The idea is to replace the dependent random variabls by independent random variables with the smallest error possible. Berkes and Philipp [14] and [15] introduced this kind of approximation by independent random variables. The mixing assumption of absolute regularity is very suitable for establishing such an approximation, as Berbee [13] noticed.

Lemma 2.3.1. *[Berbee [13]] Let Y be a random variable taking values in $\mathbb{R}^d$ and $\mathcal{A} \subset \mathcal{F}$ a σ-field. Suppose that on the probability space $(\Omega, \mathcal{F}, P)$ there exist a random variable U which is uniformly distributed on $[0,1]$ and independent of the σ-field generated by Y and $\mathcal{A}$. Then there exist a random variable Y' such that*

1. *the random variables Y and Y' have the same distribution,*
2. *Y' is independent of the σ-field $\mathcal{A}$,*
3. *$P(Y \neq Y') = \beta(\sigma(Y), \mathcal{A})$.*

We will only give a brief sketch of the proof, following the arguments of Bryc [21]. Suppose that Y has a finite support $t_1, \ldots, t_n$ and the σ-field $\mathcal{A}$ is generated by a random variable X. One can construct a random variable Y' with the same distribution as Y and

$$P\left[Y = Y' = t_i | X\right] = \min\left\{P\left[Y = t_i | X\right], P\left[Y = t_i\right]\right\}$$
$$P\left[Y' = t_i | X\right] = P\left[Y = t_i\right].$$

Then Y' is independent of X and

$$\begin{aligned}
P\left[Y \neq Y' | X\right] &= 1 - \sum_{i=1}^{n} P\left[Y = Y' = t_i | X\right] \\
&= 1 - \sum_{i=1}^{n} \min\left\{P\left[Y = t_i | X\right], P\left[Y = t_i\right]\right\} \\
&= \sum_{i=1}^{n} \left(P\left[Y = t_i\right] - \min\left\{P\left[Y = t_i | X\right], P\left[Y = t_i\right]\right\}\right) \\
&= \sum_{i:\, P[Y=t_i] > P[Y=t_i|X]} \left(P\left[Y = t_i\right] - P\left[Y = t_i | X\right]\right) \\
&= \sup_{A \subset \{t_1, \ldots, t_n\}} \left|P\left[Y \in A\right] - P\left[Y \in A | X\right]\right|
\end{aligned}$$

and by iterated expectation

$$P[Y \neq Y'] = E[P[Y \neq Y'|X]] \\ = E\left[\sup_{A\subset\{t_1,\ldots,t_n\}} |P[Y\in A] - P[Y\in A|X]|\right] = \beta(X,Y).$$

For the proof the general Theorem, one has to approximate random variables by discrete ones. Such a coupling is impossible under strong mixing, as can be seen e.g. from the results of Dehling [27]. Bradley [19], however, was able to establish a weaker type of coupling for strongly mixing random variables, using the fact that absolute regularity coefficient and the strong mixing coefficient are equivalent for random variables taking their values in a finite set and approximating general random variables by such discrete ones. We will use a later version of this coupling by Rio [77]:

Lemma 2.3.2. *[Rio [77]] Let Y be a bounded, real-valued random variable and $\mathcal{A} \subset \mathcal{F}$ a σ-field. Suppose that on the probability space $(\Omega, \mathcal{F}, P)$ there exist a random variable U which is uniformly distributed on $[0,1]$ and independent of the σ-field generated by Y and $\mathcal{A}$. Then there exist a random variable Y' such that*

1. *the random variables Y and Y' have the same distribution,*
2. *Y' is independent of the σ-field $\mathcal{A}$,*
3. *$E|Y-Y'| \leq 4\|Y\|_\infty \alpha(\sigma(Y), \mathcal{A})$.*

Unlike Berbee's theorem for absolute regularity, this lemma under strong mixing does not give equality with high probability, Y and Y' are only close in L^1 distance. This will force us to impose additional continuity conditions on a kernel g when we investigate U-statistics later. If the random variables are not bounded, we will need the following coupling lemma:

Lemma 2.3.3. *Let Y be a real-valued random variable with $E|Y|^\rho$ for a $\rho > 0$ and $\mathcal{A} \subset \mathcal{F}$ a σ-field. Suppose that on the probability space $(\Omega, \mathcal{F}, P)$ there exist two independent random variables U_1, U_2 which are uniformly distributed on $[0,1]$ and independent of the σ-field generated by Y and $\mathcal{A}$. Then there exist a random variable Y' such that for every $\epsilon > 0$*

1. *the random variables Y and Y' have the same distribution,*
2. *Y' is independent of the σ-field $\mathcal{A}$,*
3. *$P(|Y-Y'| \geq \epsilon) \leq 6\frac{\|Y\|_\rho^{\frac{\rho}{1+\rho}} \alpha^{\frac{\rho}{1+\rho}}(\sigma(Y),\mathcal{A})}{\epsilon^{\frac{\rho}{1+\rho}}}$.*

Proof. Let $B = \|Y\|_\rho^{\frac{\rho}{1+\rho}} \epsilon^{\frac{1}{1+\rho}} \alpha^{-\frac{1}{1+\rho}}(\sigma(Y), \mathcal{A})$ and $Y_B = Y\mathbb{1}_{|Y|\leq B}$. Then there is a Y'_B independent of $\mathcal{A}$ with the same distribution as Y_B such that $E|Y_B - Y'_B| \leq 4B\alpha(\sigma(Y_B), \mathcal{A}) \leq 4B\alpha(\sigma(Y), \mathcal{A})$ and by the Markov inequality $P(|Y_B - Y'_B| \geq \epsilon) \leq 4\frac{B\alpha(\sigma(Y),\mathcal{A})}{\epsilon}$. Furthermore, there exists a random variable Y'' with the same distribution as $Y - Y_B$, independent of Y, Y'_B and $\mathcal{A}$. We set

$$Y' = \begin{cases} Y'_B & \text{if } Y'_B \neq 0 \\ Y'' & \text{if } Y'_B = 0 \end{cases},$$

so that $Y'' = Y' - Y'_B$ and Y' is independent of $\mathcal{A}$. Then

$$\begin{aligned} P(|Y - Y'| \geq \epsilon) &\leq P(|Y_B - Y'_B| \geq \epsilon) + P(|Y| > B) + P(|Y'| > B) \\ &\leq 4\frac{B\alpha(\sigma(Y), \mathcal{A})}{\epsilon} + 2\frac{\|Y\|_\rho^\rho}{B^\rho} = 6\frac{\|Y\|_\rho^{\frac{\rho}{1+\rho}} \alpha^{\frac{\rho}{1+\rho}}(\sigma(Y), \mathcal{A})}{\epsilon^{\frac{\rho}{1+\rho}}}. \end{aligned}$$

□

Such coupling results exists also for near epoch dependent sequences, developed by Borovkova et al. [18], using the approximation of near epoch dependent sequences by functions of a finite part of the underlying process and the coupling under absolute regularity.

Lemma 2.3.4. *[Borovkova et al. [18]] Let $(X_n)_{n\in\mathbb{N}}$ be a near epoch dependent sequence with approximation constants $(a_k)_{k\in\mathbb{N}}$ on an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$ with mixing coefficients $(\beta(k))_{k\in\mathbb{N}}$. Given an integer m, there exist random sequences $(X'_n)_{n\in\mathbb{N}}$, $(X''_n)_{n\in\mathbb{N}}$, such that*

1. *$(X'_n)_{n\in\mathbb{N}}$, $(X''_n)_{n\in\mathbb{N}}$ have the same distribution as $(X_n)_{n\in\mathbb{N}}$,*
2. *$(X'_n)_{n\in\mathbb{N}}$ is independent of $(X''_n)_{n\in\mathbb{N}}$,*
3. *$P\left[\sum_{i=m}^\infty |X_i - X'_i| > A_{\lfloor \frac{m}{3} \rfloor}\right] \leq A_{\lfloor \frac{m}{3} \rfloor} + \beta\left(\lfloor \frac{m}{3} \rfloor\right)$,*
4. *$P\left[\sum_{i=0}^\infty \left|X'_{-i} - X''_{-i}\right| > A_{\lfloor \frac{m}{3} \rfloor}\right] \leq A_{\lfloor \frac{m}{3} \rfloor}$,*

where

$$A_l := \sqrt{2\sum_{n=l}^\infty a_n}.$$

3 U-Statistics

3.1 Definition and Applications

U-statistics are a class of nonlinear statistics and can be described as generalized means and were introduced independently in the second half of the 1940s by Halmos [45] and Hoeffding [49]. Van Mises [68] investigated the similar statistics named after him. Many sample statistics can be written as a U-statistic, at least asymptotically, e.g. the sample variance or the Cramer-von Mises-statistic. For an overview on U-staistics, we recommend the book of Koroljuk and Borovsikich [62]. For simplicity of notation, we concentrate on the case of bivariate U-statistics.

Definition 3.1.1. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a measurable and symmetric function.

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

is called U-statistic with kernel g.

If $(X_n)_{n \in \mathbb{N}}$ is a sequence of independently identically distributed random variables, $U_n(g)$ is an unbiased estimator for $\theta := E[g(X_1, X_2)]$. If we investigate the model of independently distributed random variables with an arbitrary density, then $U_n(g)$ is a uniformly minimum variance unbiased estimator. In this case, the order statistics $X_{(1)}, \ldots, X_{(n)}$ (the ordered sample with $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ and $\{X_{(1)}, \ldots, X_{(n)}\} = \{X_1, \ldots, X_n\}$) is sufficient and complete. Hence the expection

$$U_n(g) = E[g(X_1, X_2) | X_{(1)}, \ldots, X_{(n)}]$$

conditionalized on the order statistic has uniformly the lowest variance.

Let us now give some examples:

Example 3.1.2 (Sample Mean). Let $g\left(x, y\right) = \frac{1}{2}\left(x + y\right)$. Then $\theta = EX_1$ and the related U-statistic is the sample mean:

$$U_n(g) = \frac{1}{n\left(n-1\right)} \sum_{1 \leq i < j \leq n} \left(X_i + X_j\right) = \frac{1}{n} \sum_{i=1}^{n} X_i =: \bar{X}.$$

Example 3.1.3 (Sample Variance)**.** Let $g(x,y) = \frac{1}{2}(x-y)^2$. It follows that

$$\begin{aligned}\theta &= \frac{1}{2}E\left[X_1^2 - 2X_1X_2 + X_2^2\right] \\ &= \frac{1}{2}\left(\mathrm{Var}\left[X_1\right] + E^2\left[X_1\right] + \mathrm{Var}\left[X_2\right] + E^2\left[X_2\right] - 2E\left[X_1\right]E\left[X_2\right]\right) = \mathrm{Var}\left[X_1\right].\end{aligned}$$

The related U-statistic is then the well-known sample variance:

$$\begin{aligned}U_n(g) &= \frac{1}{n(n-1)}\sum_{1\le i<j\le n}\left(X_i - X_j\right)^2 = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(X_i - X_j\right)^2 \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - \bar{X}\sum_{i=1}^{n}X_i\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.\end{aligned}$$

Example 3.1.4 (Gini's mean difference)**.** Let $g(x,y) = |x-y|$. Then the corresponding U-statistic is

$$U_n(g) = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\left|X_i - X_j\right|,$$

known as Gini's mean difference, which is an estimator of scale which does not require second moments.

Example 3.1.5 (Cramer-von Mises-statistic)**.** Let

$$g(x,y) = \int_0^1\left(\mathbb{1}_{\{x\le t\}} - t\right)\left(\mathbb{1}_{\{y\le t\}} - t\right)dt.$$

This leads to the following U-statistic:

$$\begin{aligned}U_n(g) &= \frac{1}{n(n-1)}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^1\left(\mathbb{1}_{\{X_i\le t\}} - t\right)\left(\mathbb{1}_{\{X_j\le t\}} - t\right)dt - \sum_{i=1}^{n}g\left(X_i, X_i\right)\right) \\ &= \frac{n}{n-1}\int_0^1\left(\hat{F}_n(t) - t\right)^2 dt - \frac{1}{n(n-1)}\sum_{i=1}^{n}g\left(X_i, X_i\right) \\ &:= \frac{n}{n-1}V_n - \frac{1}{n(n-1)}\sum_{i=1}^{n}g\left(X_i, X_i\right).\end{aligned}$$

V_n is called Cramer-von Mises-statistic and can be used for testing the hypothesis that X_n has a uniform distribution on $[0,1]$ as an alternative to the Kolmogorow-Smirnoff-statistic $K_n := \sup_{t\in[0,1]}\left|\hat{F}_n(t) - t\right|$ (also called discrepancy).

Example 3.1.6 (χ^2 goodness of fit test)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process such that X_1 can take only the values $t_1, \ldots, t_k$. Furthermore, let be $p_1, \ldots, p_k > 0$ with $\sum_{i=1}^n p_i = 1$ and

$$g(x, y) = \sum_{i=1}^k \frac{1}{p_i} \left(\mathbb{1}_{\{x=t_i\}} - p_i\right) \left(\mathbb{1}_{\{y=t_i\}} - p_i\right).$$

The related U-statistic is

$$\begin{aligned} U_n(g) =& \frac{1}{n(n-1)} \sum_{l=1}^k \left(\frac{1}{p_l} \left(\sum_{i=1}^n \left(\mathbb{1}_{\{X_i=t_l\}} - p_l\right) \right)^2 - \frac{1}{p_l} \sum_{i=1}^n \left(\mathbb{1}_{\{X_i=t_l\}} - p_l\right)^2 \right) \\ =& \frac{1}{n-1}\chi^2 - \frac{1}{n(n-1)} \sum_{l=1}^k \frac{1}{p_l} \sum_{i=1}^n \left(\mathbb{1}_{\{X_i=t_l\}} - p_l\right)^2. \end{aligned}$$

χ^2 is used for testing the hypothesis that $P[X_1 = t_l] = p_i$ for $l = 1, \ldots, k$.

The limit behaviour of partial sums is well understood for indepenedent data as well as for many types of dependent sequences. As U-statistics have a more complex structure, the key tool in the investigation of U-statistics is their approximation by a linear part, the so-called Hoeffding-decomposition [49]:

Definition 3.1.7. We can write

$$U_n(g) = \theta + \frac{2}{n} \sum_{i=1}^n g_1(X_i) + \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} g_2(X_i, X_j)$$

where

$$\begin{aligned} \theta &:= Eg(X, Y) \\ g_1(x) &:= g(x, Y) - \theta \\ g_2(x, y) &:= g(x, y) - g_1(x) - g_1(y) - \theta \end{aligned}$$

for some independent copies of X, Y of X_1. $\frac{2}{n}\sum_{i=1}^n g_1(X_i)$ is called linear part and $U_n(g_2) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} g_2(X_i, X_j)$ degenerate part of the U-statistic $U_n(g)$.

It might be surprising that this decomposition is helpful, because we have to deal with the degenerate part $U_n(g_2)$, which is a U-statistic itself. But it has the special degeneracy property: For independent data, we have that

$$\begin{aligned} E[g_2(X_1, X_2) | X_2] &= E[g(X_1, X_2) - \theta - g_1(X_1) - g_1(X_2) | X_2] \\ &= E[g(X_1, X_2) | X_2] - \theta - E[g_1(X_1)] - g_1(X_2) \\ &= g_1(X_2) - E[g_1(X_1)] - g_1(X_2) = -E[g_1(X_1)] = 0. \end{aligned}$$

It follows that the summands of $U_n(g_2)$ are uncorrelated, as for three different indices i, j, k

$$\begin{aligned} E\left[g_2\left(X_i, X_j\right) g_2\left(X_j, X_k\right)\right] &= E\left[E\left[g_2\left(X_i, X_j\right) g_2\left(X_j, X_k\right) | X_j, X_k\right]\right] \\ &= E\left[E\left[g_2\left(X_i, X_j\right) | X_j\right] g_2\left(X_j, X_k\right)\right] \\ &= 0 = E\left[g_2\left(X_i, X_j\right)\right] E\left[g_2\left(X_j, X_k\right)\right]. \end{aligned}$$

So we have that $\mathrm{Var}[U_n(g_2)] = \frac{4}{n^2(n-1^2)} \sum_{1\leq i<j\leq n} \mathrm{Var}\, g(X_i, X_j) = O(n^{-2})$. Unfortunately, the summands of the degenerate part $U_n(g_2)$ can be correlated for dependent data. Yoshihara [98] established generalized covariance inequalities under absolutely regularity and concluded that the variance of $U_n(g_2)$ is still decreasing fast enough. Such inequalities can be proved with the help of coupling techniques, where dependent random variables are replaced by independent ones. The independent random variables are close to the original ones with high probablity. As we want to conclude that the values of the kernel g_2 are not changed much, we have to introduce a continuity property, proposed by Denker and Keller [35]:

Definition 3.1.8. A kernel g satisfies the variation condition if there exists constant L and $\epsilon_0 > 0$, such that for all $\epsilon \in (0, \epsilon_0)$

$$E\left[\sup_{\|(x,y)-(X,Y)\|\leq\epsilon} \left|g\left(x,y\right) - g\left(X,Y\right)\right|\right] \leq L\epsilon,$$

where X, Y are independent with the same distribution as X_1 and $\|\cdot\|$ denotes the Euclidean norm.

This condition could be described as Lipschitz-continuity in mean. To investigate quantiles, we will have to deal with indicator functions in Chapters 4 to 6, but we will see that the variation condition also holds in these cases. Let us now have a look at the examples above:

Example 3.1.9 (Sample Mean)**.** The kernel $g\left(x,y\right) = \frac{1}{2}\left(x+y\right)$ satisfies the variation condition as it is Lipschitz-continuous.

Example 3.1.10 (Sample Variance)**.** Let $g\left(x,y\right) = \frac{1}{2}\left(x-y\right)^2$. Then

$$\begin{aligned} &E \sup_{\|(x,y)-(X,Y)\|\leq\epsilon} \left|g\left(x,y\right) - g\left(X,Y\right)\right| \\ =&E \sup_{\|(x,y)-(X,Y)\|\leq\epsilon} \left|((x-y)-(X-Y))((x-y)+(X-Y))\right| \\ =&E \sup_{\|(x,y)-(X,Y)\|\leq\epsilon} \left|((x-X)+(y-Y))(x+X-y-Y)\right| \leq \epsilon\left(4E|X|+2\epsilon\right), \end{aligned}$$

so the variation condition holds as long as X_1 has finite first moment, which is a mild assumption for a variance estimator.

Example 3.1.11 (Gini's mean difference)**.** The kernel $g(x,y) = |x-y|$ is Lipschitz-continuous, therefore it satisfies the variation condition.

Example 3.1.12 (Cramer-von Mises-statistic)**.** Let

$$g(x,y) = \int_0^1 \left(\mathbb{1}_{\{x\le t\}} - t\right)\left(\mathbb{1}_{\{y\le t\}} - t\right) dt.$$

Then

$$\begin{aligned}
&\sup_{\|(x,y)-(X,Y)\|\le\epsilon} |g(x,y) - g(X,Y)| \\
\le &\sup_{\|(x,y)-(X,Y)\|\le\epsilon} \left|\int_0^1 (\mathbb{1}_{\{x\le t\}} - \mathbb{1}_{\{X\le t\}})\mathbb{1}_{\{y\le t\}} dt\right| \\
&\quad + \sup_{\|(x,y)-(X,Y)\|\le\epsilon} \left|\int_0^1 \mathbb{1}_{\{X\le t\}}(\mathbb{1}_{\{y\le t\}} - \mathbb{1}_{\{Y\le t\}}) dt\right| \\
\le &\sup_{\|(x,y)-(X,Y)\|\le\epsilon} \int_0^1 |\mathbb{1}_{\{x\le t\}} - \mathbb{1}_{\{X\le t\}}| dt + \sup_{\|(x,y)-(X,Y)\|\le\epsilon} \int_0^1 |\mathbb{1}_{\{y\le t\}} - \mathbb{1}_{\{Y\le t\}}| dt \\
\le &\sup_{\|(x,y)-(X,Y)\|} |x-X| + \sup_{\|(x,y)-(X,Y)\|} |y-Y| = 2\epsilon,
\end{aligned}$$

so the variation condition holds.

Example 3.1.13 (χ^2 goodness of fit test)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process such that X_1 can take only the values $t_1, \ldots, t_k$. Furthermore, let be $p_1, \ldots, p_k > 0$ with $\sum_{i=1}^n p_i = 1$ and

$$g(x,y) = \sum_{i=1}^k \frac{1}{p_i} \left(\mathbb{1}_{\{x=t_i\}} - p_i\right)\left(\mathbb{1}_{\{y=t_i\}} - p_i\right).$$

If $\epsilon < \min\left\{|t_i - t_j| \middle| i,j \in 1\ldots,k\right\}$, then

$$\sup_{\|(x,y)-(X,Y)\|\le\epsilon} |g(x,y) - g(X,Y)| = 0,$$

so the variation condition holds.

We will need the variation conditions for the function g_1 of the linear part and of the kernel g_2 of the degenerate part. As this two functions are unknown in many applications, we check the variation condition for g and then use the following:

Lemma 3.1.14. *Let the variation condition hold for a kernel g.*

1. *The variation condition also holds for g_1.*

2. The variation condition also holds for g_2.

Proof. 1. Recall that $g_1(x) = E[g(x, X_1)] - \theta$, so

$$\begin{aligned}
&E\left[\sup_{\|(x,y)-(X,Y)\|\leq\epsilon,\ \|(x',y')-(X,Y)\|\leq\epsilon} |g_1(x) - g_1(x')|\right] \\
=&E\left[\sup_{|x-X|\leq\epsilon,\ |x'-X|\leq\epsilon} |E[g(x,Y)] - E[g(x',Y)]|\right] \\
\leq&E\left[\sup_{|x-X|\leq\epsilon,\ |x'-X|\leq\epsilon} E|g(x,Y) - g(x',Y)|\right] \\
\leq&E\left[\sup_{|x-X|\leq\epsilon,\ |x'-X|\leq\epsilon} |g(x,Y) - g(x',Y)|\right] \\
\leq&E\left[\sup_{\|(x,y)-(X,Y)\|\leq\epsilon,\ \|(x',y')-(X,Y)\|\leq\epsilon} |g(x,y) - g(x',y')|\right] \leq L\epsilon.
\end{aligned}$$

2. The set of kernels satisfying the variation condition is obviously a vector space. As $g_2(x,y) = g(x,y) - g_1(x) - g_1(y) - \theta$, the condition follows directly from part 1. of this lemma. □

3.2 Generalized Covariance and Moment Inequalities

To show in the dependent case that the degenerate part of a U-statistic is negligible, we need bounds for the covariance. To establish covariance inequalities, we can use coupling techniques. Assume that g is a bounded kernel (then g_2 is also bounded by some constant C) and that $(X_n)_{n\in\mathbb{N}}$ is an absolutely regular process. For $i \leq j \leq k \leq l$, choose with Lemma 2.3.1 a random variable X_i' with the same distribution as X_i, independent of X_j, X_k, X_l, such that $P(X_i \neq X_i') \leq \beta(j-i)$. Then

$$\begin{aligned}
E[g_2(X_i', X_j)\, g_2(X_k, X_l)] &= E[E[g_2(X_i', X_j)\, g_2(X_k, X_l)\,|X_j, X_k, X_l]] \\
&= E[E[g_2(X_i', X_j)\,|X_j]\, g_2(X_k, X_l)] \\
&= 0 = E[g_2(X_i', X_j)]\, E[g_2(X_k, X_l)]
\end{aligned}$$

as the distribution of X_i' does not change conditionalized on X_i, X_j, X_k. Additionally, we have

$$\begin{aligned}
&|E[g_2(X_i', X_j)\, g_2(X_k, X_l)] - E[g_2(X_i, X_j)\, g_2(X_k, X_l)]| \\
&\qquad\qquad\qquad\qquad \leq 2CP(X_i \neq X_i^\star) = 2C\beta(j-i)
\end{aligned}$$

and arrive at

$$|E[g_2(X_i', X_j) g_2(X_k, X_l)]| \leq 2CP(X_i \neq X_i^\star) = 2C\beta(j-i).$$

If the kernel g is not bounded, we need some restriction on the moments. For independent data, second moments of the kernel are required, but in the dependent case, one needs slightly more:

Definition 3.2.1. Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process. A kernel g has uniform $(2+\delta)$-moments if for all $k \in \mathbb{N}_0$

$$E|g(X_1, X_k)|^{2+\delta} \leq M,$$
$$E|g(X, Y)|^{2+\delta} \leq M,$$

where X, Y are independent copies of X_1.

If g has uniform $(2+\delta)$-moments, then

$$E|g_1(X_1)|^{2+\delta} = E|E[g(X,Y)|Y] - Eg(X,Y)|^{2+\delta} \leq 2^{2+\delta}M$$

and as g_2 is defined as $g_2(x,y) = g(x,y) - g_1(x) - g_2(y) - \theta$, also the degenerate kernel g_2 has uniform $(2+\delta)$-moments. Now by the arguments above and some trimming method, we arrive at the the following lemma:

Lemma 3.2.2. *[Yoshihara [98]] Assume that $(X_n)_{n\in\mathbb{N}}$ is an absolutely regular process. Let $m = \max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\}$, where $\{i_1, i_2, i_3, i_4\} = \{i_{(1)}, i_{(2)}, i_{(3)}, i_{(4)}\}$ and $i_{(1)} \leq i_{(2)} \leq i_{(3)} \leq i_{(4)}$.*

1. *If g is a bounded kernel, then there is a constant C such that*

$$|E[g_2(X_{i_1}, X_{i_2}) g_2(X_{i_3}, X_{i_4})]| \leq C\beta(m).$$

2. *If g is a kernel with uniform $(2+\delta)$-moments for a $\delta > 0$, then there is a constant C such that*

$$|E[g_2(X_{i_1}, X_{i_2}) g_2(X_{i_3}, X_{i_4})]| \leq C\beta^{\frac{\delta}{2+\delta}}(m).$$

Using the coupling technique for strongly mixing data, we can prove similar inequalities:

Lemma 3.2.3. *Assume that $(X_n)_{n\in\mathbb{N}}$ is a strongly mixing sequence with $E|X_1|^\rho < \infty$ for a $\rho > 0$ and that the kernel g satisfies the variation condition. Let $m = \max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\}$, where $\{i_1, i_2, i_3, i_4\} = \{i_{(1)}, i_{(2)}, i_{(3)}, i_{(4)}\}$ and $i_{(1)} \leq i_{(2)} \leq i_{(3)} \leq i_{(4)}$.*

1. *If g is a bounded kernel, then there is a constant C such that*

$$\left|E\left[g_2\left(X_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|\leq C\alpha^{\frac{\rho}{2\rho+1}}(m).$$

2. *If g is a kernel with uniform $(2+\delta)$-moments for a $\delta>0$, then there is a constant C such that*

$$\left|E\left[g_2\left(X_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|\leq C\alpha^{\frac{\rho\delta}{2\rho\delta+\delta+3\rho+2}}(m).$$

Proof. 1. For simplicity, we consider only the case $i_1<i_2<i_3<i_4$ and $i_2-i_1\geq i_4-i_3$. After enlarging the probability space if necessary, we can assume that there exists a random variable with an uniform distribution on $[0,1]$ and independent of $(X_n)_{n\in\mathbb{N}}$. With Lemma 2.3.3, choose a random variable X'_{i_1} independent of X_{i_2},X_{i_3},X_{i_4} with the same distribution as X_{i_1} and

$$P\left[\left|X_{i_1}-X'_{i_1}\right|\geq\epsilon\right]\leq 6\frac{\|X_1\|_\rho^{\frac{\rho}{1+\rho}}\alpha^{\frac{\rho}{1+\rho}}(m)}{\epsilon^{\frac{\rho}{1+\rho}}}.$$

As g_2 is a degenerate kernel, we have

$$\begin{aligned}E\left[g_2\left(X'_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]&=E\left[E\left[g_2\left(X'_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)|X_{i_2},X_{i_3},X_{i_4}\right]\right]\\&=E\left[E\left[g_2\left(X'_{i_1},X_{i_2}\right)|X_{i_2}\right]g_2\left(X_{i_3},X_{i_4}\right)\right]=0.\end{aligned}$$

If g_2 is bounded by M and satisfies the variation condition with constant L (which holds by our assumptions and Lemma 3.1.14), then by the variation condition for every ϵ

$$\begin{aligned}&\left|E\left[g_2\left(X_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|\\=&\left|E\left[g_2\left(X_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]-E\left[g_2\left(X'_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|\\=&\left|E\left[\left(g_2\left(X_{i_1},X_{i_2}\right)-g_2\left(X'_{i_1},X_{i_2}\right)\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|\\\leq&M\left(\left|E\left[\left(g_2\left(X_{i_1},X_{i_2}\right)-g_2\left(X'_{i_1},X_{i_2}\right)\right)\mathbb{1}_{\{|X_{i_1}-X'_{i_1}|\geq\epsilon\}}\right]\right|\right.\\&\left.+\left|E\left[\left(g_2\left(X_{i_1},X_{i_2}\right)-g_2\left(X'_{i_1},X_{i_2}\right)\right)\mathbb{1}_{\{|X_{i_1}-X'_{i_1}|<\epsilon\}}\right]\right|\right)\\\leq&2M^2P\left[\left|X_{i_1}-X'_{i_1}\right|\geq\epsilon\right]+L\epsilon M=12M^2\frac{\|X_1\|_\rho^{\frac{\rho}{1+\rho}}\alpha^{\frac{\rho}{1+\rho}}(m)}{\epsilon^{\frac{\rho}{1+\rho}}}+L\epsilon M.\end{aligned}$$

Setting $\epsilon=\|X_1\|_\rho^{\frac{\rho}{2\rho+1}}\alpha(m)^{\frac{\rho}{2\rho+1}}$, we arrive at

$$\left|E\left[g_2\left(X_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|\leq\left(12M^2+LM\right)\|X_1\|_\rho^{\frac{\rho}{2\rho+1}}\alpha^{\frac{\rho}{2\rho+1}}(m).$$

2. Let $\epsilon > 0$, $K > 0$ and define:

$$g_{2,K}(x,y) = \begin{cases} g_2(x,y) & \text{if } |g_2(x,y)| \le \sqrt{K} \\ \sqrt{K} & \text{if } g_2(x,y) > \sqrt{K} \\ -\sqrt{K} & \text{if } g_2(x,y) < -\sqrt{K} \end{cases}$$

Obviously, $g_{2,K}$ satisfies the variation condition with the same constant L as g_2. Let X'_{i_1} be a random variable as above independent of $X_{i_2}, X_{i_3}, X_{i_4}$ with

$$P\left[\left|X_{i_1} - X'_{i_1}\right| \ge \epsilon\right] \le 6\frac{\|X_1\|_\rho^{\frac{\rho}{1+\rho}} \alpha^{\frac{\rho}{1+\rho}}(m)}{\epsilon^{\frac{\rho}{1+\rho}}}.$$

As g_2 is a degenerate kernel, we have

$$E\left[g_2\left(X'_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right] = 0.$$

Therefore, we get:

$$\begin{aligned}
&\left|E\left[g_2\left(X_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right]\right| \\
=&\left|E\left[g_2\left(X_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right] - E\left[g_2\left(X'_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right]\right| \\
=&\left|E\left[\left(g_2\left(X_{i_1}, X_{i_2}\right) - g_2\left(X'_{i_1}, X_{i_2}\right)\right) g_2\left(X_{i_3}, X_{i_4}\right)\right]\right| \\
\le& E\left[\left|\left(g_{2,K}\left(X_{i_1}, X_{i_2}\right) - g_{2,K}\left(X'_{i_1}, X_{i_2}\right)\right) g_{2,K}\left(X_{i_3}, X_{i_4}\right)\right| \mathbb{1}_{\left\{\left|X_{i_1} - X'_{i_1}\right| \le \epsilon\right\}}\right] \\
&+ E\left[\left|\left(g_{2,K}\left(X_{i_1}, X_{i_2}\right) - g_{2,K}\left(X'_{i_1}, X_{i_2}\right)\right) g_{2,K}\left(X_{i_3}, X_{i_4}\right)\right| \mathbb{1}_{\left\{\left|X_{i_1} - X'_{i_1}\right| > \epsilon\right\}}\right] \\
&+ E\left[\left|g_{2,K}\left(X_{i_1}, X_{i_2}\right) g_{2,K}\left(X_{i_3}, X_{i_4}\right) - g_2\left(X_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right|\right] \\
&+ E\left[\left|g_{2,K}\left(X'_{i_1}, X_{i_2}\right) g_{2,K}\left(X_{i_3}, X_{i_4}\right) - g_2\left(X'_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right|\right].
\end{aligned}$$

Because of the variation condition and $|g_{2,K}(X_3, X_4)| \le \sqrt{K}$, the first summand is smaller than $L\epsilon\sqrt{K}$. In consequence of Lemma 2.3.1, the second term is bounded by

$$P\left[\left|X_{i_1} - X'_{i_1}\right| \ge \epsilon\right] 2K \le 12\frac{\|X_1\|_\rho^{\frac{\rho}{1+\rho}} \alpha^{\frac{\rho}{1+\rho}}(m)}{\epsilon^{\frac{\rho}{1+\rho}}} K.$$

Let the $(2+\delta)$-moments of g_2 be uniformly bounded by M. For the third summand, we get:

$$\begin{aligned}
&E\left[\left|g_{2,K}\left(X_{i_1}, X_{i_2}\right) g_{2,K}\left(X_{i_3}, X_{i_4}\right) - g_2\left(X_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right|\right] \\
\le& E\left[\left(\left|g_2\left(X_{i_1}, X_{i_2}\right)\right| - \sqrt{K}\right) \left|g_2\left(X_{i_3}, X_{i_4}\right)\right| \mathbb{1}_{\left\{\left|g_2\left(X_{i_1}, X_{i_2}\right)\right| > \sqrt{K}, \left|g_2\left(X_{i_3}, X_{i_4}\right)\right| \le \sqrt{K}\right\}}\right] \\
&+ E\left[\left|g_2\left(X_{i_1}, X_{i_2}\right)\right| \left(\left|g_2\left(X_{i_3}, X_{i_4}\right)\right| - \sqrt{K}\right) \mathbb{1}_{\left\{\left|g_2\left(X_{i_1}, X_{i_2}\right)\right| \le \sqrt{K}, \left|g_2\left(X_{i_3}, X_{i_4}\right)\right| > \sqrt{K}\right\}}\right] \\
&+ E\left[\left|\left(\left|g_2\left(X_{i_1}, X_{i_2}\right)\right| - \sqrt{K}\right) \left(\left|g_2\left(X_{i_3}, X_{i_4}\right)\right| - \sqrt{K}\right)\right| \right. \\
&\qquad\qquad\qquad \left. \mathbb{1}_{\left\{\left|g_2\left(X_{i_1}, X_{i_2}\right)\right| > \sqrt{K}, \left|g_2\left(X_{i_3}, X_{i_4}\right)\right| > \sqrt{K}\right\}}\right]
\end{aligned}$$

$$\leq E\left[\left(|g_2(X_{i_1},X_{i_2})|-\sqrt{K}\right)\sqrt{K}\mathbb{1}_{\left\{|g_2(X_{i_1},X_{i_2})|>\sqrt{K}\right\}}\right]$$
$$+E\left[\left(|g_2(X_{i_3},X_{i_4})|-\sqrt{K}\right)\sqrt{K}\mathbb{1}_{\left\{|g_2(X_{i_3},X_{i_4})|>\sqrt{K}\right\}}\right]$$
$$+\frac{1}{2}E\left[\left(|g_2(X_{i_1},X_{i_2})|-\sqrt{K}\right)^2\mathbb{1}_{\left\{|g_2(X_{i_1},X_{i_2})|>\sqrt{K}\right\}}\right]$$
$$+\frac{1}{2}E\left[\left(|g_2(X_{i_3},X_{i_4})|-\sqrt{K}\right)^2\mathbb{1}_{\left\{|g_2(X_{i_3},X_{i_4})|>\sqrt{K}\right\}}\right]$$
$$\leq\frac{1}{2}E\left[g_2^2(X_{i_1},X_{i_2})\mathbb{1}_{\left\{|g_2(X_{i_1},X_{i_2})|>\sqrt{K}\right\}}\right]+\frac{1}{2}E\left[g_2^2(X_{i_3},X_{i_4})\mathbb{1}_{\left\{|g_2(X_{i_3},X_{i_4})|>\sqrt{K}\right\}}\right]$$
$$\leq\frac{1}{2}\frac{E|g_2(X_{i_1},X_{i_2})|^{2+\delta}}{K^{\frac{\delta}{2}}}+\frac{1}{2}\frac{E|g_2(X_{i_3},X_{i_4})|^{2+\delta}}{K^{\frac{\delta}{2}}}\leq\frac{M}{K^{\frac{\delta}{2}}}.$$

After treating the fourth summand in the same way, we totally get:

$$|E[g_2(X_{i_1},X_{i_2})g_2(X_{i_3},X_{i_4})]|$$
$$\leq L\epsilon\sqrt{K}+12\frac{\|X_1\|_\rho^{\frac{\rho}{1+\rho}}\alpha^{\frac{\rho}{1+\rho}}(m)}{\epsilon^{\frac{\rho}{1+\rho}}}K+2\frac{M}{K^{\frac{\delta}{2}}}=:f(\epsilon,K).$$

Setting $\epsilon^0=\|X_1\|_\rho^{\frac{\rho}{2\rho+1}}\alpha(m)^{\frac{\rho}{2\rho+1}}K^{\frac{\rho+1}{4\rho+2}}$, we obtain:

$$f\left(\epsilon^0,K\right)=(L+12)\|X_1\|_\rho^{\frac{\rho}{2\rho+1}}\alpha(m)^{\frac{\rho}{2\rho+1}}K^{\frac{3\rho+2}{4\rho+2}}+2\frac{M}{K^{\frac{\delta}{2}}}.$$

With $K^0=\|X_1\|_\rho^{-\frac{2\rho}{2\rho\delta+\delta+3\rho+2}}\alpha(m)^{-\frac{2\rho}{2\rho\delta+\delta+3\rho+2}}$, we get the bound:

$$|E[g_2(X_{i_1},X_{i_2})g_2(X_{i_3},X_{i_4})]|$$
$$\leq f\left(\epsilon^0,K^0\right)=(L+12+2M)\|X_1\|_\rho^{\frac{\rho\delta}{2\rho\delta+\delta+3\rho+2}}\alpha^{\frac{\rho\delta}{2\rho\delta+\delta+3\rho+2}}(m).$$

□

The next lemma is very similar to the generalized covariance inequality given by Borovkova et al. [18], but with a different continuity condition.

Lemma 3.2.4. *Let $(X_n)_{n\in\mathbb{N}}$ be near epoch dependent on an absolutely regular process (Z_n) with constants a_l. Define $A_L=\sqrt{2\sum_{i=L}^{\infty}a_i}$ and $\beta(j)$ as the mixing coefficient of (Z_n). Let g satisfy the variation condition.*

1. *If g is bounded, then there exists a constant C, such that*

$$|E[g_2(X_{i_1},X_{i_2})g_2(X_{i_3},X_{i_4})]|\leq C\beta\left(\lfloor\frac{m}{3}\rfloor\right)+CA_{\lfloor\frac{m}{3}\rfloor}.$$

2. If g has uniform $(2+\delta)$-moments, then there exists a constant C, such that

$$\left|E\left[g_2\left(X_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right| \leq C\beta^{\frac{\delta}{2+\delta}}\left(\lfloor\frac{m}{3}\rfloor\right)+CA_{\lfloor\frac{m}{3}\rfloor}^{\frac{\delta}{2+\delta}}.$$

Proof. 1. For simplicity, we consider only the case $0=i_1<i_2<i_3<i_4$ and $m=i_2-i_1\geq i_4-i_3$. With Lemma 2.3.4, there exist sequences $(X'_n)_{n\in\mathbb{Z}}$ and $(X''_n)_{n\in\mathbb{Z}}$ with the same distribution as $(X_n)_{n\in\mathbb{Z}}$, such that

1. $(X''_n)_{n\in\mathbb{Z}}$ is independent of $(X_n)_{n\in\mathbb{Z}}$,
2. $P\left[\sum_{i=m}^{\infty}|X_i-X'_i|>A_{\lfloor\frac{m}{3}\rfloor}\right]\leq A_{\lfloor\frac{m}{3}\rfloor}+\beta\left(\lfloor\frac{m}{3}\rfloor\right)$,
3. $P\left[\sum_{i=0}^{\infty}\left|X'_{-i}-X''_{-i}\right|>A_{\lfloor\frac{m}{3}\rfloor}\right]\leq A_{\lfloor\frac{m}{3}\rfloor}$.

By the degeneracy of g_2, we have that $g_2\left(X''_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)=0$ and by construction

$$E\left[g_2\left(X_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]=E\left[g_2\left(X'_{i_1},X'_{i_2}\right)g_2\left(X'_{i_3},X'_{i_4}\right)\right].$$

With the triangular inequality it follows that

$$\begin{aligned}
&\left|E\left[g_2\left(X_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|\\
=&\left|E\left[g_2\left(X'_{i_1},X'_{i_2}\right)g_2\left(X'_{i_3},X'_{i_4}\right)\right]-E\left[g_2\left(X''_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|\\
\leq&\left|E\left[g_2\left(X'_{i_1},X'_{i_2}\right)g_2\left(X'_{i_3},X'_{i_4}\right)-g_2\left(X''_{i_1},X_{i_2}\right)g_2\left(X'_{i_3},X'_{i_4}\right)\right]\right|\\
&+\left|E\left[g_2\left(X''_{i_1},X_{i_2}\right)g_2\left(X'_{i_3},X'_{i_4}\right)-g_2\left(X''_{i_1},X_{i_2}\right)g_2\left(X_{i_3},X_{i_4}\right)\right]\right|.
\end{aligned}$$

In order to keep this proof short, we treat only the first summand. As g is bounded and satisfies the variation condition, g_2 is bounded by some M and the variation condition holds with constant L, then by the variation condition and the coupling as in Lemma 2.3.4

$$\begin{aligned}
&\left|E\left[\left(g_2\left(X'_{i_1},X'_{i_2}\right)-g_2\left(X''_{i_1},X_{i_2}\right)\right)g_2\left(X'_{i_3},X'_{i_4}\right)\right]\right|\\
\leq&M\left(\left|E\left[\left(g_2\left(X''_{i_1},X_{i_2}\right)-g_2\left(X'_{i_1},X'_{i_2}\right)\right)\mathbb{1}_{\{|X''_{i_1}-X'_{i_1}|\leq A_{\lfloor\frac{m}{3}\rfloor},|X_{i_2}-X'_{i_2}|\leq A_{\lfloor\frac{m}{3}\rfloor}\}}\right]\right|\right.\\
&\left.+\left|E\left[\left(g_2\left(X_{i_1},X_{i_2}\right)-g_2\left(X'_{i_1},X_{i_2}\right)\right)\mathbb{1}_{\{|X''_{i_1}-X'_{i_1}|\leq A_{\lfloor\frac{m}{3}\rfloor},|X_{i_2}-X'_{i_2}|\leq A_{\lfloor\frac{m}{3}\rfloor}\}^C}\right]\right|\right)\\
\leq&2M^2P\left[\left|X_{i_1}-X''_{i_1}\right|\geq A_{\lfloor\frac{m}{3}\rfloor}\right]+2M^2P\left[\left|X_{i_2}-X'_{i_2}\right|\geq A_{\lfloor\frac{m}{3}\rfloor}\right]+L\sqrt{2}A_{\lfloor\frac{m}{3}\rfloor}M\\
\leq&(4M^2+\sqrt{2}LM)\left(\beta\left(\lfloor\frac{m}{3}\rfloor\right)+A_{\lfloor\frac{m}{3}\rfloor}\right).
\end{aligned}$$

2. As above, choose sequences $(X'_n)_{n\in\mathbb{Z}}$ and $(X''_n)_{n\in\mathbb{Z}}$, such that $(X''_n)_{n\in\mathbb{Z}}$ is independent of $(X_n)_{n\in\mathbb{Z}}$ with

$$P\left[\sum_{i=m}^{\infty}|X_i - X'_i| > A_{\lfloor\frac{m}{3}\rfloor}\right] \leq A_{\lfloor\frac{m}{3}\rfloor} + \beta\left(\lfloor\frac{m}{3}\rfloor\right)$$
$$P\left[\sum_{i=0}^{\infty}\left|X'_{-i} - X''_{-i}\right| > A_{\lfloor\frac{m}{3}\rfloor}\right] \leq A_{\lfloor\frac{m}{3}\rfloor}.$$

Then

$$\begin{aligned}&|E\left[g_2\left(X_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right]|\\ \leq&\left|E\left[g_2\left(X'_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right) - g_2\left(X''_{i_1}, X_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right)\right]\right|\\ &+\left|E\left[g_2\left(X''_{i_1}, X_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right) - g_2\left(X''_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right]\right|.\end{aligned}$$

We will again concentrate on the first summand. As in the proof of Lemma 3.2.3, part 2., define

$$g_{2,K}(x,y) = \begin{cases} g_2(x,y) & \text{if } |g_2(x,y)| \leq \sqrt{K} \\ \sqrt{K} & \text{if } g_2(x,y) > \sqrt{K} \\ -\sqrt{K} & \text{if } g_2(x,y) < -\sqrt{K} \end{cases}$$

It is clear that $g_{2,K}$ satisfies the variation condition with the same constant L as g_2 (which satisfies the variation condition because of Lemma 3.1.14). We get that

$$\begin{aligned}&\left|E\left[g_2\left(X'_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right) - g_2\left(X''_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right)\right]\right]\\ =&\left|E\left[\left(g_2\left(X'_{i_1}, X'_{i_2}\right) - g_2\left(X''_{i_1}, X'_{i_2}\right)\right) g_2\left(X'_{i_3}, X'_{i_4}\right)\right]\right|\\ \leq& E\left[\left|\left(g_{2,K}\left(X'_{i_1}, X'_{i_2}\right) - g_{2,K}\left(X''_{i_1}, X'_{i_2}\right)\right) g_{2,K}\left(X'_{i_3}, X'_{i_4}\right)\right|\right]\\ &+ E\left[\left|g_{2,K}\left(X'_{i_1}, X'_{i_2}\right) g_{2,K}\left(X'_{i_3}, X'_{i_4}\right) - g_2\left(X'_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right)\right|\right]\\ &+ E\left[\left|g_{2,K}\left(X''_{i_1}, X'_{i_2}\right) g_{2,K}\left(X'_{i_3}, X'_{i_4}\right) - g_2\left(X''_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right)\right|\right].\end{aligned}$$

As in the proof of the first part of this lemma, we have for the first summand:

$$\begin{aligned}&\left|E\left[\left(g_2\left(X'_{i_1}, X'_{i_2}\right) - g_2\left(X''_{i_1}, X'_{i_2}\right)\right) g_2\left(X'_{i_3}, X'_{i_4}\right)\right]\right|\\ \leq&(4K + \sqrt{2}L\sqrt{K})\left(\beta\left(\lfloor\frac{m}{3}\rfloor\right) + A_{\lfloor\frac{m}{3}\rfloor}\right).\end{aligned}$$

As $g_2\left(X'_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right)$ and $g_2\left(X''_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right)$ are random variables with $(1+\frac{\delta}{2})$-moments smaller than M from the definition of the uniform $(2+\delta)$-moments, the second and the third summand are bounded by $\frac{M}{K^{\frac{\delta}{2}}}$. Totally, we get

$$\begin{aligned}&\left|E\left[g_2\left(X'_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right) - g_2\left(X''_{i_1}, X'_{i_2}\right) g_2\left(X'_{i_3}, X'_{i_4}\right)\right]\right|\\ &\qquad\leq (4K + \sqrt{2}L\sqrt{K})\left(\beta\left(\lfloor\frac{m}{3}\rfloor\right) + A_{\lfloor\frac{m}{3}\rfloor}\right) + 2\frac{M}{K^{\frac{\delta}{2}}}.\end{aligned}$$

Setting $K = \left(A_{\lfloor \frac{k}{3} \rfloor} + \beta\left(\lfloor \frac{k}{3} \rfloor\right)\right)^{-\frac{2}{2+\delta}} M^{\frac{2}{2+\delta}}$, keeping in mind that K is non-decreasing and treating the summand in the same way, one easily obtains

$$|E\left[g_2\left(X_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right]| \leq CM^{\frac{2}{2+\delta}} \left(\beta^{\frac{\delta}{2+\delta}}\left(\lfloor \frac{m}{3} \rfloor\right) + A_{\lfloor \frac{m}{3} \rfloor}^{\frac{\delta}{2+\delta}}\right)$$

for a constant C, which proves the lemma. □

Now we can derive a second moment inequality for the degenerate part of a U-statistic, following the arguments of Yoshihara [98]. In the optimal case $\tau = 0$ (the series in the lemma are summable), the order of the variance is n^2, which is the same as in the independent case.

Lemma 3.2.5. *Let g be a kernel that satisfies the variation condition. Let be $\tau \geq 0$ such that one of the following conditions hold for the process $(X_n)_{n\in\mathbb{N}}$:*

1. *g is bounded, $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, $E|X_1|^\rho < \infty$ for a $\rho > 0$ and $\sum_{k=0}^n k\alpha^{\frac{\rho}{2\rho+1}}(k) = O(n^\tau)$.*

2. *g has uniform $(2+\delta)$-moments for a $\delta > 0$, $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, $E|X_1|^\rho < \infty$ for a $\rho > 0$ and $\sum_{k=0}^n k\alpha^{\frac{\rho\delta}{2\rho\delta+\delta+3\rho+2}}(k) = O(n^\tau)$.*

3. *g is bounded, $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent with approximation constants $(a_k)_{k\in\mathbb{N}}$ on an absolutely regular process with mixing coefficients $(\beta(k))_{k\in\mathbb{N}}$ and $\sum_{k=0}^n k(\beta(k) + A_k) = O(n^\tau)$ for $A_L = \sqrt{2\sum_{i=L}^\infty a_i}$.*

4. *g has uniform $(2+\delta)$-moments for a $\delta > 0$, $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent with approximation constants $(a_k)_{k\in\mathbb{N}}$ on an absolutely regular process with mixing coefficients $(\beta(k))_{k\in\mathbb{N}}$ and $\sum_{k=0}^n k(\beta^{\frac{\delta}{2+\delta}}(k) + A_k^{\frac{\delta}{2+\delta}}) = O(n^\tau)$ for $A_L = \sqrt{2\sum_{i=L}^\infty a_i}$.*

Then

$$\sum_{i_1,i_2,i_3,i_4=1}^{n} |E\left[g_2\left(X_{i_1}, X_{i_2}\right) g_2\left(X_{i_3}, X_{i_4}\right)\right]| = O\left(n^{2+\tau}\right).$$

Proof. The proofs of the four parts of this lemmas are exactly the same, using the differents parts of Lemma 3.2.3 respectively Lemma 3.2.4, so we will show only part 1. Let $\{i_1, i_2, i_3, i_4\} = \left\{i_{(1)}, i_{(2)}, i_{(3)}, i_{(4)}\right\}$ and $i_{(1)} \leq i_{(2)} \leq i_{(3)} \leq i_{(4)}$, then we can rewrite the sum in the statement of our lemma as

$$\sum_{i_1,i_2,i_3,i_4=1}^{n} |E[g_2(X_{i_1},X_{i_2})\, g_2(X_{i_3},X_{i_4})]|$$
$$= \sum_{k=0}^{n} \sum_{\substack{i_1,i_2,i_3,i_4 \\ \max\{i_{(2)}-i_{(1)},i_{(4)}-i_{(3)}\}=k}} |E[g_2(X_{i_1},X_{i_2})\, g_2(X_{i_3},X_{i_4})]|$$
$$\leq C \sum_{k=0}^{n} \sum_{\substack{i_1,i_2,i_3,i_4 \\ \max\{i_{(2)}-i_{(1)},i_{(4)}-i_{(3)}\}=k}} \alpha^{\frac{\rho}{2\rho+1}}(k).$$

We have to calculate the number of quadrubles (i_1, i_2, i_3, i_4) such that $\max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\} = k$. First note that there are at most 6 quadrubles wich lead to the same ordered numbers $i_{(1)}, i_{(2)}, i_{(3)}, i_{(4)}$. There are at most n^2 possibilities to choose $i_{(1)}$ and $i_{(4)}$. If $i_{(2)} - i_{(1)} = \max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\} = k$, then $i_{(2)}$ is already fixed and there are k possibilities $i_{(3)}$. The same argument applies if $i_{(4)} - i_{(3)} = \max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\} = k$, so we arrive at

$$\sum_{i_1,i_2,i_3,i_4=1}^{n} |E[g_2(X_{i_1},X_{i_2})\, g_2(X_{i_3},X_{i_4})]| \leq Cn^2 \sum_{k=0}^{n} k\alpha^{\frac{\rho}{2\rho+1}}(k) = O\left(n^{2+\tau}\right).$$

□

3.3 Central Limit Theorem and Law of the Iterated Logarithm

We are now prepared to give versions of the classic limit theorems in probability theory for U-statistics under dependence: the central limit theorem and the law of the iterated logarithm. Similar results can be found in Dehling, Wendler [32], [33]. The proofs will all make use of the Hoeffding decomposition

$$U_n(g) = \theta + \frac{2}{n}\sum_{i=1}^{n} g_1(X_i) + \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} g_2(X_i,X_j).$$

There already exist different results for the linear part and the moment inequalities of the previous section will allow us to show that the degenerate part will converge to 0. After the central limit theorem under independence by Hoeffding [49], there are many results under dependence: Under the strong assumption of m-dependence or $\star$-mixing, Sen [80], [82] proved asymptotic normality. Yoshihara [98] assumed $(X_n)_{n\in\mathbb{N}}$ to be stationary and absolutely regular and proved a central limit theorem for U-Statistics under this weaker condition. Denker and Keller [35] have relaxed the mixing assumption to sequences which are near epoch dependent on absolutely regular processes, Borovkova et al. [18] showed convergence of the empirical U-process to a Gaussian process. Recently, Hsing and Wu [53] proved a central limit theorem for weighted U-statistics of processes that are functionals of an underlying independent i.i.d. process with some physical dependence measure (a more refined version of near epoch dependence).

Theorem 3.3.1. *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process and g be a kernel that satisfies the variation condition and let one of the following four conditions hold:*

1. *g is bounded, $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, $E|X_1|^\rho < \infty$ for a $\rho > 0$ and $\alpha(k) = O(k^{-\alpha})$ for an $\alpha > \frac{2\rho+1}{\rho}$.*
2. *g has uniform $(2+\delta)$-moments for a $\delta > 0$, $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, $E|X_1|^\rho < \infty$ for a $\rho > 0$ and $\alpha(k) = O(k^{-\alpha})$ for an $\alpha > \frac{2\rho\delta+\delta+3\rho+2}{\rho\delta}$.*
3. *g is bounded, $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent with approximation constants $(a_k)_{k\in\mathbb{N}}$ on an absolutely regular process with mixing coefficients $(\beta(k))_{k\in\mathbb{N}}$ and there is a $\beta > 1$ such that $\beta(k) = O(k^{-\beta})$ and $a_k = O(k^{-\beta-2})$.*
4. *g has uniform $(2+\delta)$-moments for a $\delta > 0$, $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent with approximation constants $(a_k)_{k\in\mathbb{N}}$ on an absolutely regular process with mixing coefficients $(\beta(k))_{k\in\mathbb{N}}$ and there is a $\beta > \frac{2+\delta}{\delta}$ such that $\beta(k) = O(k^{-\beta})$ and $a_k = O(k^{-2\beta-1})$.*

Then

$$\sqrt{n}\left(U_n(g) - \theta\right) \xrightarrow{\mathcal{D}} N\left(0, 4\sigma^2\right)$$

with

$$\sigma^2 = \operatorname{Var}\left[g_1(X_1)\right] + 2\sum_{k=1}^{\infty} \operatorname{Cov}\left[g_1(X_1)\, g_1(X_{1+k})\right].$$

If $\sigma^2 = 0$, when the statement of the theorem should be read as convergence to 0.

Proof. We use the decomposition

$$\sqrt{n}\left(U_n(g) - \theta\right) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n} g_1(X_i) + \sqrt{n}\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} g_2(X_i, X_j).$$

For the four different conditions of the theorem, we will proof that $\frac{2}{\sqrt{n}}\sum_{i=1}^{n} g_1(X_i)$ is asymptotically normal, that $\sigma^2 < \infty$ and that the degenerate part

$$\sqrt{n}\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} g_2(X_i, X_j) \xrightarrow{n\to\infty} 0$$

in probability. The statement of the theorem will then follow with the help of Slutzky's theorem.

1. First note that if $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, then the same holds for the sequence $(g(X_n))_{n\in\mathbb{N}}$ and the mixing coefficients are smaller or equal to the mixing coefficients $(\alpha(k))_{k\in\mathbb{N}}$. So by Theorem 1.6 of Ibragimov [55], we have that $\sigma^2 < \infty$ and that $\frac{2}{\sqrt{n}}\sum_{i=1}^n g_1(X_i)$ is asymptotically normal. Moreover, we have $\sum_{k=0}^n k\alpha^{\frac{\rho}{2\rho+1}}(k) \leq \sum_{k=0}^n k^{1-\alpha\frac{\rho}{2\rho+1}} = O(n^\tau)$ for a $\tau < 1$ and by the triangular inequality and the first part of Lemma 3.2.5

$$\begin{aligned}\operatorname{Var}&\left[\sqrt{n}\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} g_2(X_i,X_j)\right]\\&\leq \frac{1}{n^3}\sum_{i_1,i_2,i_3,i_4=1}^{n}\left|E\left[g_2(X_{i_1},X_{i_2})\,g_2(X_{i_3},X_{i_4})\right]\right| = O(\frac{1}{n^3}n^{2+\tau}) = o(1),\end{aligned}$$

the Chebyshev inequality completes the proof.

2. With Theorem 1.7 of Ibragimov [55] $\sigma^2 < \infty$ and $\frac{2}{\sqrt{n}}\sum_{i=1}^n g_1(X_i)$ is asymptotically normal. Moreover, we have $\sum_{k=0}^n k\alpha^{\frac{\rho\delta}{2\rho\delta+\delta+3\rho+2}}(k) = O(n^\tau)$ for a $\tau < 1$ and by second part of Lemma 3.2.5

$$\operatorname{Var}\left[\sqrt{n}\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} g_2(X_i,X_j)\right] = O(\frac{1}{n^3}n^{2+\tau}) = o(1).$$

3. Note that $g_1(X_i)$ is bounded and that it is near epoch dependent with approximation constants $a'_l = a_l^{\frac{1}{2}}$ by Lemma 2.1.7. As $\sum_{k=1}^\infty \beta(k) < \infty$ and $\sum_{k=1}^\infty a'_k < \infty$, it follows by Thereom 2.3 of Ibragimov [55] that $\sigma^2 < \infty$ and $\frac{2}{\sqrt{n}}\sum_{i=1}^n g_1(X_i)$ is asymptotically normal. Moreover, we have $A_L = (2\sum_{i=L}^\infty a_i)^{\frac{1}{2}} = O(n^{-\frac{\beta+1}{2}})$ and $\sum_{k=0}^\infty k(\beta(k)+A_k) = O(n^\tau) < \infty$ for a $\tau < 1$, so by the third part of Lemma 3.2.5

$$\operatorname{Var}\left[\sqrt{n}\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} g_2(X_i,X_j)\right] = O(\frac{1}{n^3}n^{2+\tau}) = o(1).$$

4. $g_1(X_i)$ have finite $(2+\delta)$-moments and are near epoch dependent with approxi-

mation constants $a_l' = Ca_l^{\frac{1+\delta}{3+2\delta}}$ by Lemma 2.1.7. So we can conclude that

$$\begin{aligned}
&E\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^{\frac{2+\delta}{1+\delta}} \\
&= E\left[\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^{\frac{2+\delta}{1+\delta}} \mathbb{1}_{\{|g_1(X_0)-E[g_1(X_0)|Z_{-l},\ldots,Z_l]|\leq a_l'^{-\frac{1}{1+\delta}}\}}\right] \\
&\quad + E\left[\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^{\frac{2+\delta}{1+\delta}} \mathbb{1}_{\{|g_1(X_0)-E[g_1(X_0)|Z_{-l},\ldots,Z_l]|> a_l'^{-\frac{1}{1+\delta}}\}}\right] \\
&\leq a_l'^{-\frac{1}{(1+\delta)^2}} E\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right| \\
&\qquad + a_l'^{\frac{\delta(2+\delta)}{(1+\delta)^2}} E\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^{2+\delta} \leq C a_l'^{\frac{\delta(2+\delta)}{(1+\delta)^2}} \leq C a_l^{\frac{\delta(2+\delta)}{(1+\delta)(3+2\delta)}}.
\end{aligned}$$

So we have that

$$\sum_{k=1}^{\infty}\left(E\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^{\frac{2+\delta}{1+\delta}}\right)^{\frac{1+\delta}{2+\delta}} \leq C\sum_{k=1}^{\infty} a_l^{\frac{\delta}{3+2\delta}} < \infty,$$

as $a_k = O(n^{-2\beta-1})$ and $2\beta + 1 > \frac{4+3\delta}{\delta} > \frac{3+2\delta}{\delta}$. Furthermore, $\sum_{k=1}^{\infty} \beta^{\frac{\delta}{2+\delta}}(k) < \infty$, so we can apply Theorem 2.2 of Ibragimov [55] to obtain asymptotic normality of $\frac{2}{\sqrt{n}}\sum_{i=1}^{n} g_1(X_i)$ and $\sigma^2 < \infty$. Moreover, $A_k = (2\sum_{i=k}^{\infty} a_i)^{\frac{1}{2}} = O(k^{-\beta})$, so $\sum_{k=0}^{n} k(\beta^{\frac{\delta}{2+\delta}}(k) + A_k^{\frac{\delta}{2+\delta}}) = O(n^{\tau})$ for a $\tau < 1$ and with the fourth part of Lemma 3.2.5

$$\operatorname{Var}\left[\sqrt{n}\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} g_2(X_i, X_j)\right] = O(\frac{1}{n}) = O(\frac{1}{n^3}n^{2+\tau}) = o(1).$$

□

Part 2. of this theorem is similar to the central limit theorems for U-statistics in Dehling and Wendler [32], but the continuity condition differs. The fourth part avoids the condition of $(4+\delta)$-moments required in Borovkova et al. [18]. We are not only interested in the weak convergence, but also in the strong behaviour of U-statistics. Let us first investigate the degenerate part.

Proposition 3.3.2. *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process and g be a kernel that satisfies the variation condition. Let be $\tau \geq 0$ such that one of the following conditions holds:*

1. *g is bounded, $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, $E|X_1|^{\rho} < \infty$ for a $\rho > 0$ and $\sum_{k=0}^{n} k\alpha^{\frac{\rho}{2\rho+1}}(k) = O(n^{\tau})$.*
2. *g has uniform $(2+\delta)$-moments for a $\delta > 0$, $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, $E|X_1|^{\rho} < \infty$ for a $\rho > 0$ and $\sum_{k=0}^{n} k\alpha^{\frac{\rho\delta}{2\rho\delta+\delta+3\rho+2}}(k) = O(n^{\tau})$.*

3. *g is bounded,* $(X_n)_{n\in\mathbb{N}}$ *is near epoch dependent with approximation constants* $(a_k)_{k\in\mathbb{N}}$ *on an absolutely regular process with mixing coefficients* $(\beta(k))_{k\in\mathbb{N}}$. *For* $A_L = \sqrt{2\sum_{i=L}^{\infty} a_i}$: $\sum_{k=0}^{n} k\left(\beta\left(k\right)+A_k\right) = O\left(n^{\tau}\right)$.

4. *g has uniform* $(2+\delta)$*-moments for a* $\delta > 0$, $(X_n)_{n\in\mathbb{N}}$ *is near epoch dependent with approximation constants* $(a_k)_{k\in\mathbb{N}}$ *on an absolutely regular process with mixing coefficients* $(\beta(k))_{k\in\mathbb{N}}$. *For* $A_L = \sqrt{2\sum_{i=L}^{\infty} a_i}$: $\sum_{k=0}^{n} k(\beta^{\frac{\delta}{2+\delta}}\left(k\right)+A_k^{\frac{\delta}{2+\delta}}) = O\left(n^{\tau}\right)$.

Then almost surely

$$\frac{n^{1-\frac{\tau}{2}}}{\log^{\frac{3}{2}} n \log\log n} U_n\left(g_2\right) \to 0.$$

Proof. We define

$$Q_n = \sum_{1\leq i_1<i_2\leq n} g_2\left(X_{i_1}, X_{i_2}\right)$$
$$c_n = \frac{1}{n^{1+\frac{\tau}{2}} \log^{\frac{3}{2}} n \log\log n}.$$

With the method of subsequences, it suffices to show that

$$c_{2^l} Q_{2^l}\left(g_2\right) \to 0$$
$$\max_{2^{l-1}\leq n<2^l} \left|c_n Q_n - c_{2^{l-1}} Q_{2^{l-1}}\right| \to 0$$

almost surely as $l\to\infty$. We use the Chebyshev inequality and Lemma 3.2.5 to prove the first line. For every $\epsilon > 0$:

$$\sum_{l=1}^{\infty} P\left[\left|c_{2^l} Q_{2^l}\left(g_2\right)\right| > \epsilon\right] \leq \frac{1}{\epsilon^2}\sum_{l=1}^{\infty} c_{2^l}^2 E\left[Q_{2^l}^2\left(g_2\right)\right] \leq C\frac{1}{\epsilon^2}\sum_{l=1}^{\infty}\frac{1}{l^3\log^2 l} < \infty$$

so $c_{2^l} Q_{2^l}\left(g_2\right) \to 0$ follows with the Borel-Cantelli Lemma. To prove the convergence of the maxima to 0, we first have to find a bound for the second moments, using a well-known chaining technique. For example, by the triangle inequality we have

$$\left|c_{15}Q_{15} - c_8 Q_8\right| \leq \left|c_{15}Q_{15} - c_{14}Q_{14}\right| + \left|c_{14}Q_{14} - c_{12}Q_{12}\right| + \left|c_{12}Q_{12} - c_8 Q_8\right|.$$

Using such a decomposition for all n with $2^{l-1}\leq n < 2^l$, we conclude that

$$\max_{2^{l-1}\leq n<2^l} \left|c_n Q_n - c_{2^{l-1}} Q_{2^{l-1}}\right|$$
$$\leq \sum_{d=1}^{l} \max_{i=1,\ldots,2^{l-d}} \left|c_{2^{l-1}+i2^{d-1}} Q_{2^{l-1}+i2^{d-1}} - c_{2^{l-1}+(i-1)2^{d-1}} Q_{2^{l-1}+(i-1)2^{d-1}}\right|.$$

As for any random variables $Y_1, \ldots, Y_n$: $E(\max |Y_i|)^2 \leq \sum EY_i^2$, it follows that

$$E\left[\left(\max_{2^{l-1}\leq n<2^l} |c_nQ_n - c_{2^{l-1}}Q_{2^{l-1}}|\right)^2\right]$$

$$\leq l\sum_{d=1}^{l}\sum_{i=1}^{2^{l-d}} E\left[\left(c_{2^{l-1}+i2^{d-1}}Q_{2^{l-1}+i2^{d-1}} - c_{2^{l-1}+(i-1)2^{d-1}}Q_{2^{l-1}+(i-1)2^{d-1}}\right)^2\right]$$

$$\leq l\sum_{d=1}^{l}\sum_{i=1}^{2^{l-d}} E\left[\left(c_{2^{l-1}+i2^{d-1}}\left(Q_{2^{l-1}+i2^{d-1}} - Q_{2^{l-1}+(i-1)2^{d-1}}\right)\right.\right.$$

$$\left.\left.+\left(c_{2^{l-1}+i2^{d-1}} - c_{2^{l-1}+(i-1)2^{d-1}}\right)Q_{2^{l-1}+(i-1)2^{d-1}}\right)^2\right]$$

$$\leq l\sum_{d=1}^{l}\sum_{i=1}^{2^{l-d}} 2c^2_{2^{l-1}+i2^{d-1}}E\left[\left(Q_{2^{l-1}+i2^{d-1}} - Q_{2^{l-1}+(i-1)2^{d-1}}\right)^2\right]$$

$$+l\sum_{d=1}^{l}\sum_{i=1}^{2^{l-d}} 2\left(c_{2^{l-1}+i2^{d-1}} - c_{2^{l-1}+(i-1)2^{d-1}}\right)^2 E\left[Q^2_{2^{l-1}+(i-1)2^{d-1}}\right]$$

$$=\sum_{d=1}^{l} 2c^2_{2^{l-1}+i2^{d-1}}E\left[\sum_{i=1}^{2^{l-d}}\left(Q_{2^{l-1}+i2^{d-1}} - Q_{2^{l-1}+(i-1)2^{d-1}}\right)^2\right]$$

$$+l\sum_{d=1}^{l}\sum_{i=1}^{2^{l-d}}2\left(c_{2^{l-1}+i2^{d-1}}+c_{2^{l-1}+(i-1)2^{d-1}}\right)\left(c_{2^{l-1}+i2^{d-1}}-c_{2^{l-1}+(i-1)2^{d-1}}\right)E\left[Q^2_{2^{l-1}+(i-1)2^{d-1}}\right]$$

$$\leq l^2 6c^2_{2^{l-1}} \sum_{i_1,i_2,i_3,i_4=1}^{2^l} |E\left[g_2\left(X_{i_1}, X_{i_2}\right)g_2\left(X_{i_3}, X_{i_4}\right)\right]| \leq C\frac{1}{l\log^2 l}.$$

In the last line we used the fact that the sequence $(c_n)_{n\in\mathbb{N}}$ is decreasing and Lemma 3.2.5. It now follows for all $\epsilon > 0$ with the Chebyshev inequality

$$\sum_{l=1}^{\infty} P\left[\max_{2^{l-1}\leq n<2^l} |c_nQ_n - c_{2^{l-1}}Q_{2^{l-1}}| > \epsilon\right]$$

$$\leq \frac{1}{\epsilon^2}\sum_{l=1}^{\infty} E\left[\left(\max_{2^{l-1}\leq n<2^l} |c_nQ_n - c_{2^{l-1}}Q_{2^{l-1}}|\right)^2\right] \leq \frac{C}{\epsilon^2}\sum_{l=1}^{\infty}\frac{1}{l\log^2 l} < \infty,$$

the Borel-Cantelli Lemma completes the proof. □

Since $\alpha(k) \leq \frac{1}{4}$, condition 2. in Proposition 3.3.2 is always satisfied with some $\tau \in [0, 2]$. In the extreme case $\tau = 0$, i.e. when the series $\sum_{k=0}^{n} k\alpha^{\frac{\rho\delta}{2\rho\delta+\delta+3\rho+2}}(k)$

converges, the conclusion of Theorem 1 is close to the optimal rate which follows in the independent case from the law of the iterated logarithm for degenerate U-statistics of Dehling, Denker and Philipp [29] and Dehling [28]. Kanagawa and Yoshihara [57] proved such a law of the iterated logarithm under strong mixing. They imposed additional conditions on the summability of the eigenvalues of the function $g(x, y)$ which are not easy to check for many examples. In the other extreme case, the statement of the theorem is rather trivial, as it follows by the strong law of large numbers for U-statistcs. Under independence, this was shown by Hoeffding [50]. Aaronson et al. [1] have shown the following: If $(X_n)_{n\in\mathbb{N}}$ is a stationary ergodic process, $|g(x, y)| \leq f_1(x)f_2(y)$ with $Ef_1(X_1) \leq \infty$, $Ef_2(X_2) \leq \infty$ and one of the three following conditions hold:

1. The distribution of X_1 is discrete,
2. The kernel g is continuous almost everywhere with respect to the distribution of (X, Y), where X, Y are independent with the same distribution as X_1,
3. The sequence $(X_n)_{n\in\mathbb{N}}$ is absolutely regular,

then $U_n(g) \to \theta$ almost surely and consequently $U_n(g_2) \to 0$. The mild assumption $|g(x, y)| \leq f_1(x)f_2(y)$ holds for all our examples. Furthermore, the function $s_n(x, y) = \sup_{\|(x,y)-(x',y')\|\leq\frac{1}{n}} |g(x, y) - g(x', y')|$ in nonincreasing in n and bounded below by 0. So we have monotone convergence and

$$E \lim_{n\to\infty} s_n(X, Y) = \lim_{n\to\infty} Es_n(X, Y) = 0,$$

because of the variation condition, so $\lim_{n\to\infty} s_n(x, y) = 0$ almost everywhere with respect to the distribution of (X, Y), meaning that we have continuity almost everywhere, so the theorem of Aaronson et al. [1] applys. Dehling and Sharipov [31] established a Marcinkiewicz-Zygmund stong law of large numbers under absolute regularity.

We will now proceed with the law of the iterated logarithm for U-statistics. Were are a lot of results for partial sums, but few for U-statistics. The law of the iterated logorithm was originally established for partial sums of independent identically distributed random variables by Khintchine in 1927 [61]. Hartman and Wintner [47] were able to prove Khintchine's result under the optimal condition that the random variables have mean zero and finite second moments. Independently, Philipp [74], Iosifescu [56] and Reznick [76] studied this problem under dependence; Oodaira and Yoshihara [73] weakened their conditions. For partial sums of strongly mixing processes, the sharpest results presently available are due to Rio [77].

Serfling [84] extended the law of the iterated logarithm to U-statistics in the independent case, we will extend this to dependent random variables. The following theorem is similar to Theorem 2 of Dehling and Wendler [33]:

Theorem 3.3.3. *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process and g be a kernel that satisfies the variation condition and let one of the following four conditions hold:*

1. *g is bounded, $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, $E|X_1|^\rho < \infty$ for a $\rho > 0$ and $\alpha(k) = O(k^{-\alpha})$ for an $\alpha > \frac{2\rho+1}{\rho}$.*
2. *g has uniform $(2+\delta)$-moments for a $\delta > 0$, $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, $E|X_1|^\rho < \infty$ for a $\rho > 0$ and $\alpha(k) = O(k^{-\alpha})$ for an $\alpha > \frac{2\rho\delta+\delta+3\rho+2}{\rho\delta}$.*
3. *g is bounded, $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent with approximation constants $(a_k)_{k\in\mathbb{N}}$ on an absolutely regular process with mixing coefficients $(\beta(k))_{k\in\mathbb{N}}$ and there is a $\beta > 1$ such that $\beta(k) = O(k^{-\beta})$ and $a_k = O(k^{-\beta-3})$.*
4. *g has uniform $(2+\delta)$-moments for a $\delta > 0$, $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent with approximation constants $(a_k)_{k\in\mathbb{N}}$ on an absolutely regular process with mixing coefficients $(\beta(k))_{k\in\mathbb{N}}$ and there is a $\beta > \frac{2+\delta}{\delta}$ such that $\beta(k) = O(k^{-\beta})$ and $a_k = O(k^{-3\beta-1})$.*

If $\sigma^2 = \operatorname{Var}[g_1(X_1)] + 2\sum_{k=1}^\infty \operatorname{Cov}[g_1(X_1)\, g_1(X_{1+k})] > 0$ then

$$\limsup_{n\to\infty} \pm\sqrt{\frac{n}{8\sigma^2\log\log n}}\left(U_n(g) - \theta\right) = 1$$

almost surely.

Proof. As shown in the proof of Theorem 3.3.1, the conditions of Proposition 3.3.2 are satisfied for a $\tau < 1$, so

$$U_n(g_2) = o\left(\frac{\log^{\frac{3}{4}} n \log\log n}{n^{1-\frac{\tau}{2}}}\right) = o\left(\sqrt{\frac{\log\log n}{n}}\right)$$

almost surely. It remains to show that almost surely

$$\limsup_{n\to\infty} \pm\sqrt{\frac{1}{2\sigma^2 n\log\log n}}\sum_{i=1}^n g_1(X_i) = 1.$$

1. As $\sum_{k=1}^\infty \alpha(k) < \infty$, this follows by Theorem 2 of Rio [77].

2. As $\sum_{k=1}^\infty k^{\frac{2}{\delta}}\alpha(k) < \infty$, this follows by Theorem 2 of Rio [77].

3. As

$$\begin{aligned} &E\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\dots,Z_l]\right|^2 \\ &\qquad \le CE\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\dots,Z_l]\right| \le C\sqrt{a_l} = O(n^{-\frac{\beta+3}{2}}) \end{aligned}$$

and $\frac{\beta+3}{2} > 2$, we can use Theorem 7 of Oodaira and Yoshihara [73] to obtain the statement of this theorem.

4. We have that $g(X_i)$ have finite $(2+\delta)$-moments and are near epoch dependent with approximation constants $a'_l = Ca_l^{\frac{1+\delta}{3+2\delta}}$, so consequently

$$
\begin{aligned}
&E\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^2 \\
&= E\left[\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^2 \mathbb{1}_{\{|g_1(X_0)-E[g_1(X_0)|Z_{-l},\ldots,Z_l]|\leq a_l'^{-\frac{1}{1+\delta}}\}}\right] \\
&\quad + E\left[\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^2 \mathbb{1}_{\{|g_1(X_0)-E[g_1(X_0)|Z_{-l},\ldots,Z_l]|> a_l'^{-\frac{1}{1+\delta}}\}}\right] \\
&\leq a_l'^{-\frac{1}{1+\delta}} E\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right| \\
&\qquad\qquad + a_l'^{\frac{\delta}{1+\delta}} E\left|g_1(X_0) - E[g_1(X_0)|Z_{-l},\ldots,Z_l]\right|^{2+\delta} \\
&\qquad\qquad\qquad\qquad \leq Ca_l'^{\frac{\delta}{1+\delta}} \leq Ca_l^{\frac{\delta}{3+2\delta}} \leq Cl^{-(3\beta+1)\frac{\delta}{3+2\delta}}
\end{aligned}
$$

and $(3\beta+1)\frac{\delta}{3+2\delta} > \frac{6+4\delta}{\delta}\frac{\delta}{3+2\delta} = 2$. So we can use Theorem 8 of Oodaira and Yoshihara [73] to obtain the statement of this theorem. □

4 U-Processes

4.1 Definition and Applications

Not only U-statistics with fixed kernel g are of interest, but also the empirical U-distribution function $(U_n(t))_{t\in\mathbb{R}}$, which is for fixed t a U-statistic with kernel $h(x,y) := \mathbb{1}_{\{g(x,y)\leq t\}}$. The Grassberger-Procaccia, see [44], and the Takens estimator of the correlation dimension in a dynamical system are based on the empirical U-distribution function. We will shortly describe the idea of dimension estimation, for more details see Borovkova et al. [18].

We have seen that data from a dynamical system $(X_n)_{n\in\mathbb{N}}$ with $X_{n+1} = T(X_n)$ (where T is a piecewise smooth and expanding map) might show a behaviour similar to other weak dependent sequences such as linear processes. But there is a criterion to distinguish this two types of sequences: For a dynamical system, X_{n+1} is a function of X_n, so all vectors $((X_{n+1}, X_n))_{n\in\mathbb{N}}$ are concentrated on the graph of T, while for example for a linear process there the innovations have a density or for a sequence of independent random variables, the vectors $((X_{n+1}, X_n))_{n\in\mathbb{N}}$ also have a density. Let us define the correlation integral

$$C(r) = P(\|(X_{n+1}, X_n) - (Y_{n+1}, Y_n)\| \leq r)$$

where $(Y_n)_{n\in\mathbb{N}}$ is a independent copy of $(X_n)_{n\in\mathbb{N}}$. For dynamical systems we expect for small r a behaviour as $C(r) \approx Cr$, while for independent data, we would have $C(r) \approx Cr^2$, so the exponent (the so-called correlation dimension) tells us about the nature of the process $(X_n)_{n\in\mathbb{N}}$. A natural estimator of $C(r)$ is the U-statistic with kernel $h(x,y,r) := \mathbb{1}_{\{\|x-y\|\leq r\}}$.

Let us note at this point that our theory can be generalized to multidimensional random variables straightforwardly for sequences, that are near epoch dependent on absolutely regular processes, see also Borovkova et al. [18]. In contrast, for strongly mixing sequences, this is not easy, as the coupling lemma 2.3.2 is not dimension free.

The functional central limit theorem for the empirical U-distribution function has been established by Silverman [86], Nolan and Pollard [71] and Arcones and Giné [6] for independent data, by Arcones and Yu [8] and Borovkova [17] for absolutely regular data, and by Borovkova et al. [18] for data, which is near epoch dependent on absolutely regular processes. Lévy-Leduc et al. [66] investigated U-processes of long range dependent data. The functional law of the iterated logarithm for the empirical U-distribution function has been proved by Arcones [3], Arcones and Giné [7] under independence. The Strong invariance principle has been investigated by Dehling et al. [29]. We will show a strong invariance principle under dependence. As a corollary, we will obtain the law of the iterated logarithm to sequences which are strongly mixing or L^1 near epoch dependent on an absolutely regular process and the central limit theorem under conditions which are slightly different from the conditions in Borovkova et al. [18]. Let us now proceed with precise definitions:

Definition 4.1.1. We call a measurable and bounded function $h : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ which is symmetric in the first two arguments and non-decreasing in the third argument a kernel function. We will assume that for all $x, y \in \mathbb{R}$: $\lim_{t\to\infty} h(x,y,t) = 1$, $\lim_{t\to-\infty} h(x,y,t) = 0$. For fixed $t \in \mathbb{R}$, we define

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

and call the process $(U_n(t))_{t\in\mathbb{R}}$ the empirical U-distribution function. We define the U-distribution function as $U(t) := E[h(X,Y,t)]$, where X, Y are independent with the same distribution as X_1, and the empirical U-process as $(\sqrt{n}(U_n(t) - U(t)))_{t\in\mathbb{R}}$.

The main example for this is the empirical distribution function of the sample $(g(X_i, X_j))_{1\leq i<j\leq n}$ which is the empirical U-distibution function with the kernel function $h(x,y,t) := \mathbb{1}_{\{g(x,y)\leq t\}}$. But other kernel function might also be of interest. If we study $h(x,y,t) = \frac{1}{2}(\mathbb{1}_{\{x\leq t\}} + \mathbb{1}_{\{y\leq t\}})$, we obtain the ordinary empirical distribution function. Similar to U-statistics with fixed kernel, we introduce the Hoeffding decomposition of the empirical U-distribution function into a linear and a so-called degenerate part:

$$U_n(t) = U(t) + \frac{2}{n}\sum_{i=1}^{n} h_1(X_i, t) + \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} h_2(X_i, X_j, t)$$

where

$$\begin{aligned} h_1(x,t) &:= Eh(x,Y,t) - U(t) \\ h_2(x,y,t) &:= h(x,y,t) - h_1(x,t) - h_1(y,t) - U(t). \end{aligned}$$

Because we will consider dependent random variables, we need an additional continuity property of the kernel function, a uniform version of Definition 3.1.8:

Definition 4.1.2. h satisfies the uniform variation condition, if is a constant L, such that for all $t \in \mathbb{R}$, $\epsilon > 0$

$$E\left[\sup_{\|(x,y)-(X,Y)\|\leq\epsilon} |h(x,y,t) - h(X,Y,t)|\right] \leq L\epsilon,$$

where X, Y are independent with the same distribution as X_1 and $\|\cdot\|$ denotes the Euclidean norm.

This condition holds for many discontinuous functions, as the jump is averaged out by the expectation.

Example 4.1.3. The kernel function $h(x,y,t) := \mathbb{1}_{\{|x-y|\leq t\}}$ satisfies the uniform variation condition, if X_1 has a bounded densiy. Then $|X-Y|$ also has a bounded density and for every $\epsilon > 0$

$$E\left[\sup_{\|(x,y)-(X,Y)\|\leq\epsilon}\left|\mathbb{1}_{\{|x-y|\leq t\}} - \mathbb{1}_{\{|X-Y|\leq t\}}\right|\right] \leq P\left[t-\sqrt{2}\epsilon < |X-Y| \leq t+\sqrt{2}\epsilon\right] \leq L\epsilon.$$

Note that for this example, the U-distribution function is Lipschitz continuous.

The empirical U-process is a generalization of the empirical process, which is given by $(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbb{1}_{\{X_i\leq t\}} - P(X_i \leq t)))_{t\in[0,1]}$. For uniformly on $[0,1]$ distributed and independent random variables $(X_n)_{n\in\mathbb{N}}$, the empirical process converges weakly to a Brownian Bridge, as was proved by Donsker [36]. The functional law of the iterated logarithm established by Finkelstein [41] says that $((\frac{1}{\sqrt{2n\log\log n}}\sum_{i=1}^{n}(\mathbb{1}_{\{X_i\leq t\}}-P(X_i \leq t)))_{t\in[0,1]})_{n\in\mathbb{N}}$ is almost surely relatively compact and the limit set is given by

$$\{f|f(0)=f(1)=0,\ f \text{ absolutely continuous },\int_0^1 f'^2(t)dt = 1\}.$$

Müller [70] determined the limit distribution of the double indexed empirical process

$$\left(\frac{1}{\sqrt{n}}\sum_{1\leq i\leq sn}(\mathbb{1}_{\{X_i\leq t\}} - t)\right)_{t,s\in[0,1]}.$$

It converges weakly towards a Gaussian process $(K(t,s))_{s,t\in[0,1]}$ with covariance function $EK(t,s)K(t',s') = \min\{s,s'\}(\min\{t,t'\}-tt')$, which is the covariance function of a Brownian bridge in t direction and of a Brownian motion in s direction. Kiefer [60] proved an almost sure invariance principle: After enlarging the probability space, there exists a copy of the Kiefer-Müller process K such that the empirical process and this copy are close together with respect to the supremum norm.

A strong invariance principle is a very powerful asymptotic theorem, as the limit behaviour of Gaussian processes is well understood and it is then possible to conclude that the approximated process has the same asymptotic properties. Note that a Kiefer-Müller processes can be described as a functional Brownian motion, as its increments in s direction are independent Brownian bridges. Kuelbs and Lepage [63] established the law of the iterated logarithm for Brownian motion in Banach spaces.

Berkes and Philipp [14] extended Kiefer's result to dependent random variables. The approximating Gaussian process K has then the covariance function

$$\begin{aligned} EK(t,s)K(t',s') = \min\{s,s'\}\big(4\operatorname{Cov}\left[\mathbb{1}_{\{X_1\le t\}},\mathbb{1}_{\{X_1\le t'\}}\right] \\ + 4\sum_{k=1}^{\infty}\operatorname{Cov}\left[\mathbb{1}_{\{X_1\le t\}},\mathbb{1}_{\{X_{k+1}\le t'\}}\right] + 4\sum_{k=1}^{\infty}\operatorname{Cov}\left[\mathbb{1}_{\{X_{k+1}\le t\}},\mathbb{1}_{\{X_1\le t'\}}\right]\big). \end{aligned}$$

Their proof is based on a special coupling method, measuring dependence by the difference of the characteristic function conditionalized on the distant past and the unconditional one, see Berkes and Philipp [15].

We have the following scaling behaviour: $(\frac{1}{\sqrt{n}}K(t,ns))_{s,t\in[0,1]}$ has the same distribution as $(K(t,s))_{s,t\in[0,1]}$. Furthermore, a functional law of the iterated logarithm holds: The sequence

$$\left((\frac{1}{\sqrt{2n\log\log n}}K(t,ns))_{s,t\in[0,1]}\right)_{n\in\mathbb{N}}$$

is almost surely relatively compact (with respect to the supremum norm). The limit set is the unit ball of the the reproducing kernel Hilbert space, which we will introduce now:

Definition 4.1.4. Let $C:(\mathbb{R}^d)^2\to\mathbb{R}$ a covariance function (that means symmetric and positive semidefinit). We define

$$\mathcal{K}_m := \left\{f:\mathbb{R}^d\to\mathbb{R}\big|f(x)=\sum_{i=1}^m b_iC(x,y_i),\ b_1,\ldots,b_m\in\mathbb{R},y_1,\ldots,y_m\in\mathbb{R}^d\right\}.$$

For $f(x)=\sum_{i=1}^{m_1}b_iC(x,y_i)\in\mathcal{K}_{m_1}$, $g(x)=\sum_{j=1}^{m_2}b_i'C(x,z_i)\in\mathcal{K}_{m_2}$, the inner product of f and g is given by

$$(f,g)=\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}b_ib_j'C(y_i,z_j)$$

and $\sqrt{(f,f)}$ is a norm on every $\mathcal{K}_m$. We call $\mathcal{K}=\overline{\bigcup_{m=1}^{\infty}\mathcal{K}_m}$ (the completion of the union) reproducing kernel Hilbert space associated with covariance function C.

Oodaira [72] noticed that the limit set in the functional law of the iterated logarithm could be described as the unit ball $\mathcal{U}$ of this space. For more information about the reproducing kernel Hilbert space, see Aronszajn [9].

4.2 4th Moment Bounds and Uniform Hoeffding Decomposition

Lemma 4.2.1. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary, strongly mixing sequence with $\alpha(n) = O(n^{-\alpha})$ for some $\alpha > 3$ and $C_1, C_2 > 0$ constants. Then there exists a constant C, such that for all measurable, non-negative functions $g : \mathbb{R} \to \mathbb{R}$ bounded by C_1 and all $n \in \mathbb{N}$*

$$E\left(\sum_{i=1}^{n} g(X_i) - E[g(X_1)]\right)^4 \leq Cn^2(\log n)^2\left(\max\left\{E|g(X_1)|, C_2 n^{-\frac{3}{4}}\right\}\right)^{1+\gamma}$$

with $\gamma = \frac{\alpha-2}{\alpha}$.

Proof. We define the random variables $Y_i = g(X_i) - Eg(X_1)$. Recall that by Lemma 2.2.2 with $p_1 = p_2 = \frac{2\alpha}{\alpha-3}$ and $p_3 = \frac{\alpha}{3}$ we obtain the following three inequalities for all $i, j, k \in \mathbb{N}$:

$$|E[Y_0Y_iY_{i+j}Y_{i+j+k}]| \leq C\alpha^{\frac{3}{\alpha}}(i)\,\|Y_0\|_{\frac{2\alpha}{\alpha-3}}\,\|Y_0Y_jY_{j+k}\|_{\frac{2\alpha}{\alpha-3}},$$
$$|E[Y_0Y_iY_{i+j}Y_{i+j+k}]| \leq C\,|E[Y_0Y_i]|\,|E[Y_0Y_k]| + C\alpha^{\frac{3}{\alpha}}(j)\,\|Y_0Y_i\|_{\frac{2\alpha}{\alpha-3}}\,\|Y_0Y_k\|_{\frac{2\alpha}{\alpha-3}},$$
$$|E[Y_0Y_iY_{i+j}Y_{i+j+k}]| \leq C\alpha^{\frac{3}{\alpha}}(k)\,\|Y_0Y_iY_{i+j}\|_{\frac{2\alpha}{\alpha-3}}\,\|Y_0\|_{\frac{2\alpha}{\alpha-3}}.$$

By Lemma 2.2.1 with $p_1 = p_2 = \frac{2\alpha}{\alpha-1}$ and $p_3 = \alpha$, we get

$$|E[Y_0Y_i]| \leq C\alpha^{\frac{1}{\alpha}}(i)\,\|Y_1\|^2_{\frac{2\alpha}{\alpha-1}}.$$

As Y_n is bounded, we have that

$$\|Y_1\|_{\frac{2\alpha}{\alpha-3}} = \left(E|Y_1|^{\frac{2\alpha}{\alpha-3}}\right)^{\frac{\alpha-3}{2\alpha}} \leq \left(C_1^{\frac{\alpha+3}{\alpha-3}}E|Y_1|^{\frac{2\alpha}{\alpha-3}}\right)^{\frac{\alpha-3}{2\alpha}} \leq C\left(E|Y_1|\right)^{\frac{\alpha-3}{2\alpha}}$$
$$\|Y_1\|_{\frac{2\alpha}{\alpha-1}} \leq \left(C_1^{\frac{\alpha+1}{\alpha-1}}E|Y_1|\right)^{\frac{\alpha-1}{2\alpha}} \leq C\left(E|Y_1|\right)^{\frac{\alpha-1}{2\alpha}}$$
$$\|Y_0Y_jY_{j+k}\|_{\frac{2\alpha}{\alpha-3}} \leq \left\|Y_1^3\right\|_{\frac{2\alpha}{\alpha-3}} \leq C_1^2\,\|Y_1\|_{\frac{2\alpha}{\alpha-3}} \leq C\left(E|Y_1|\right)^{\frac{\alpha-3}{2\alpha}}$$

and it follows that

$$|E[Y_0Y_iY_{i+j}Y_{i+j+k}]| \leq C\alpha^{\frac{1}{\alpha}}(i)\,\alpha^{\frac{1}{\alpha}}(k)\left(E|Y_1|\right)^{\frac{2\alpha-2}{\alpha}} + C\alpha^{\frac{3}{\alpha}}(\max\{i,j,k\})\left(E|Y_1|\right)^{\frac{\alpha-3}{\alpha}}.$$

Now by stationarity of the process and the linearity of the expection the following fourth moment inequality holds:

$$\begin{aligned} E\left(\sum_{i=1}^{n} Y_i\right)^4 &\leq Cn \sum_{i,j,k=1}^{n} \left|E\left[Y_0 Y_i Y_{i+k} Y_{i+k+j}\right]\right| \\ &\leq Cn^2 \sum_{i=1}^{n} \alpha^{\frac{1}{\alpha}}(i) \sum_{k=1}^{n} \alpha^{\frac{1}{\alpha}}(k) \left(E\left|Y_1\right|\right)^{\frac{2\alpha-2}{\alpha}} + Cn \sum_{i=1}^{n} i^2 \alpha^{\frac{3}{\alpha}}(i) \left(E\left|Y_1\right|\right)^{\frac{\alpha-3}{\alpha}}. \end{aligned}$$

As $\max\left\{E\left|g\left(X_1\right)\right|, C_2 n^{-\frac{3}{4}}\right\} \geq C_2 n^{-\frac{\alpha}{\alpha+1}}$, we have that

$$\left(E\left|Y_1\right|\right)^{\frac{\alpha-3}{\alpha}} \leq Cn\left(\max\left\{E\left|g\left(X_1\right)\right|, C_2 n^{-\frac{3}{4}}\right\}\right)^{\frac{2\alpha-2}{\alpha}}$$

and with $\alpha(n) = O\left(n^{-\alpha}\right)$, we arrive at

$$\begin{aligned} &E\left(\sum_{i=1}^{n} Y_i\right)^4 \\ \leq & Cn^2 \sum_{i=1}^{n} \frac{1}{i} \sum_{k=1}^{n} \frac{1}{k} \left(E\left|Y_1\right|\right)^{\frac{2\alpha-2}{\alpha}} + Cn^2 \sum_{i=1}^{n} i^2 \frac{1}{i^3} \left(\max\left\{E\left|g\left(X_1\right)\right|, C_2 n^{-\frac{3}{4}}\right\}\right)^{\frac{2\alpha-2}{\alpha}} \\ \leq & Cn^2 \left(\sum_{i=1}^{n} \frac{1}{i}\right)^2 \left(\max\left\{E\left|g\left(X_1\right)\right|, C_2 n^{-\frac{3}{4}}\right\}\right)^{\frac{2\alpha-2}{\alpha}} \\ = & Cn^2 \left(\log n\right)^2 \left(\max\left\{E\left|g\left(X_1\right)\right|, C_2 n^{-\frac{3}{4}}\right\}\right)^{1+\gamma}. \end{aligned}$$

□

Lemma 4.2.2. *Let $(X_n)_{n\in\mathbb{Z}}$ be a near epoch dependent sequence on an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$ with mixing coefficients $\beta(n) = O\left(n^{-\beta}\right)$ for a $\beta > 3$ and approximation constants $a_n = O\left(n^{-(\beta+3)}\right)$. Let $C_1, C_2, L > 0$ be constants. Then there exists a constant C, such that for all measurable, non-negative functions $g : \mathbb{R} \to \mathbb{R}$ that are bounded by C_1 and satisfy the variation condition with constant L, and all $n \in \mathbb{N}$ we have*

$$E\left(\sum_{i=1}^{n} g\left(X_i\right) - E\left[g\left(X_1\right)\right]\right)^4 \leq Cn^2 \left(\log n\right)^2 \left(\max\{E\left|Y_1\right|, C_2 n^{-\frac{3}{4}}\}\right)^{1+\gamma}$$

with $\gamma = \frac{\beta-3}{\beta+1}$.

Proof. We define the random variables $Y_i = g\left(X_i\right) - Eg\left(X_1\right)$. Then by Lemma 2.1.7, $(Y_n)_{n\in\mathbb{Z}}$ is near epoch with approximation constants $\tilde{a}_n = (L + C_1)\sqrt{a_n} = O\left(n^{-\frac{\beta+3}{2}}\right)$.

Using Lemma 2.2.5 with $\delta = \frac{6}{\beta-3}$, we obtain

$$\begin{aligned}&|EY_0Y_iY_{i+j}Y_{i+j+k}|\\ &\leq C\left(\beta^{\frac{3}{\beta}}\left(\lfloor\frac{\max\{i,j,k\}}{3}\rfloor\right)\|Y_0\|^2_{\frac{2\beta}{\beta-3}} + \tilde{a}^{\frac{6}{\beta+3}}_{\lfloor\frac{\max\{i,j,k\}}{3}\rfloor}\|Y_0\|^{\frac{2\beta}{\beta+3}}_{\frac{2\beta}{\beta-3}}\right) + |E[Y_0Y_i]E[Y_0Y_k]|.\end{aligned}$$

Making use of Lemma 2.2.4 and $\delta = \frac{2}{\beta-1}$, it follows that

$$\begin{aligned}|EY_0Y_iY_{i+j}Y_{i+j+k}| &\leq C\left(\beta^{\frac{3}{\beta}}\left(\lfloor\frac{\max\{i,j,k\}}{3}\rfloor\right)\|Y_0\|^2_{\frac{2\beta}{\beta-3}} + \tilde{a}^{\frac{6}{\beta+3}}_{\lfloor\frac{\max\{i,j,k\}}{3}\rfloor}\|Y_0\|^{\frac{2\beta}{\beta+3}}_{\frac{2\beta}{\beta-3}}\right)\\ &+C\left(\beta^{\frac{1}{\beta}}\left(\lfloor\frac{k}{3}\rfloor\right)\|Y_0\|^2_{\frac{2\beta}{\beta-1}} + \tilde{a}^{\frac{2}{\beta+1}}_{\lfloor\frac{k}{3}\rfloor}\|Y_0\|^{\frac{2\beta}{\beta+1}}_{\frac{2\beta}{\beta-1}}\right)\cdot\left(\beta^{\frac{1}{\beta}}\left(\lfloor\frac{i}{3}\rfloor\right)\|Y_0\|^2_{\frac{2\beta}{\beta-1}} + \tilde{a}^{\frac{2}{\beta+1}}_{\lfloor\frac{i}{3}\rfloor}\|Y_0\|^{\frac{2\beta}{\beta+1}}_{\frac{2\beta}{\beta-1}}\right).\end{aligned}$$

First note that

$$\begin{aligned}\beta^{\frac{1}{\beta}}(n) &= O\left(n^{-1}\right), \quad \tilde{a}^{\frac{2}{\beta+1}} = O\left(n^{-1}\right),\\ \beta^{\frac{3}{\beta}}(n) &= O\left(n^{-3}\right), \quad \tilde{a}^{\frac{6}{\beta+3}} = O\left(n^{-3}\right),\end{aligned}$$

and that

$$\begin{aligned}\|Y_0\|^2_{\frac{2\beta}{\beta-1}} &\leq C\|Y_0\|^{\frac{2\beta}{\beta+1}}_{\frac{2\beta}{\beta-1}} \leq C\|Y_0\|_1^{\frac{\beta-1}{\beta+1}},\\ \|Y_0\|^2_{\frac{2\beta}{\beta-3}} &\leq C\|Y_0\|^{\frac{2\beta}{\beta+3}}_{\frac{2\beta}{\beta-3}} \leq C\|Y_0\|_1^{\frac{\beta-3}{\beta+3}},\end{aligned}$$

as Y_i is bounded. Furthermore, $\|Y_0\|_1^{\frac{\beta-3}{\beta+3}} \leq Cn\left(\max\{\|Y_0\|_1, C_2n^{-\frac{3}{4}}\}\right)^{\frac{2\beta-2}{\beta+1}}$. Now by stationarity

$$\begin{aligned}&E\left(\sum_{i=1}^n Y_i\right)^4 \leq Cn\sum_{i,j,k=1}^n |E[Y_0Y_iY_{i+j}Y_{i+j+k}]|\\ \leq &Cn^2\sum_{i=1}^n \beta^{\frac{1}{\beta}}\left(\lfloor\frac{i}{3}\rfloor\right)\sum_{k=1}^n\beta^{\frac{1}{\beta}}\left(\lfloor\frac{k}{3}\rfloor\right)\|Y_1\|_1^{\frac{2\beta-2}{\beta}} + Cn^2\sum_{i=1}^n\tilde{a}^{\frac{2}{\beta+1}}_{\lfloor\frac{i}{3}\rfloor}\sum_{k=0}^n\tilde{a}^{\frac{2}{\beta+1}}_{\lfloor\frac{k}{3}\rfloor}\|Y_1\|_1^{\frac{2\beta-2}{\beta+1}}\\ &+Cn\sum_{m=1}^n m^2\beta^{\frac{3}{\beta}}\left(\lfloor\frac{m}{3}\rfloor\right)\|Y_0\|_1^{\frac{\beta-3}{\beta}} + Cn\sum_{m=1}^n m^2\tilde{a}^{\frac{6}{\beta+3}}_{\lfloor\frac{m}{3}\rfloor}\|Y_0\|_1^{\frac{\beta-3}{\beta+3}}\\ \leq &Cn^2\sum_{i=1}^n i^{-1}\sum_{k=1}^n k^{-1}\left(\max\{\|Y_0\|_1, C_2n^{-\frac{3}{4}}\}\right)^{\frac{2\beta-2}{\beta+1}}\\ &+Cn^2\sum_{m=1}^n m^2m^{-3}\left(\max\{\|Y_0\|_1, C_2n^{-\frac{3}{4}}\}\right)^{\frac{2\beta-2}{\beta+1}}\\ \leq &Cn^2(\log n)^2\left(\max\{\|Y_0\|_1, C_2n^{-\frac{3}{4}}\}\right)^{\frac{2\beta-2}{\beta+1}} = Cn^2(\log n)^2\left(\max\{\|Y_0\|_1, C_2n^{-\frac{3}{4}}\}\right)^{1+\gamma}.\end{aligned}$$

□

Proposition 4.2.3. *Let h be a kernel function that satisfies the uniform variation condition such that the U-distribution function U is Lipschitz-continuous and one of the following two mixing conditions is satisfied:*

1. *$(X_n)_{n\in\mathbb{N}}$ is strongly mixing with mixing coefficients $\alpha(n) = O(n^{-\alpha})$ for $\alpha \geq 8$ and $E|X_i|^\rho < \infty$ for a $\rho > \frac{1}{4}$.*
2. *$(X_n)_{n\in\mathbb{N}}$ is near epoch dependent on an absolutely regular process with mixing coefficients $\beta(n) = O(n^{-\beta})$ for $\beta \geq 8$ and approximation constants $a_n = O(n^{-a})$ for $a = \max\{\beta+3, 12\}$.*

Then:

$$\sup_{t\in\mathbb{R}} \left| \sum_{1\leq i<j\leq n} h_2(X_i, X_j, t) \right| = o\left(n^{\frac{3}{2}-\frac{\gamma}{8}}\right)$$

almost surely with $\gamma = \frac{\alpha-2}{\alpha}$ respectively $\gamma = \frac{\beta-3}{\beta+1}$.

Proof. We define $Q_n(t) := \sum_{1\leq i<j\leq n} h_2(X_i, X_j, t)$. For $l \in \mathbb{N}$ choose $t_{1,l}, \dots, t_{k-1,l}$ with $k = k_l = O\left(2^{\frac{5}{8}l}\right)$, such that

$$-\infty = t_{0,l} < t_{1,l} < \ldots < t_{k-1,l} < t_{k,l} = \infty,$$

and $2^{-\frac{5}{8}l} \leq |U(t_{r,l}) - U(t_{r-1,l})| \leq 2 \cdot 2^{-\frac{5}{8}l}$. By the Lipschitz-continuity of the U-distribution function $|t_{r,l} - t_{r-1,l}| \geq \frac{C}{2^{\frac{5}{8}l}}$. By our assumptions, h and U are non-decreasing in t, so we have for any $t \in [t_{r-1,l}, t_{r,l}]$ and $2^l \leq n < 2^{l+1}$

$$\begin{aligned}
|Q_n(t)| &= \left| \sum_{1\leq i<j\leq n} (h(X_i, X_j, t) - h_1(X_i, t) - h_1(X_j, t)) - U(t)) \right| \\
&\leq \max\left\{ \left| \sum_{1\leq i<j\leq n} (h(X_i, X_j, t_{r,l}) - h_1(X_i, t) - h_1(X_j, t) - U(t)) \right|, \right. \\
&\qquad \left. \left| \sum_{1\leq i<j\leq n} (h(X_i, X_j, t_{r-1,l}) - h_1(X_i, t) - h_1(X_j, t) - U(t)) \right| \right\} \\
&\leq \max\left\{ \left| \sum_{1\leq i<j\leq n} (h(X_i, X_j, t_{r,l}) - h_1(X_i, t_{r,l}) - h_1(X_j, t_{r,l}) - U(t_{r,l})) \right|, \right. \\
&\qquad \left. \left| \sum_{1\leq i<j\leq n} (h(X_i, X_j, t_{r-1,l}) - h_1(X_i, t_{r-1,l}) - h_1(X_j, t_{r-1,l}) - U(t_{r-1,l})) \right| \right\} \\
&\quad + (n-1) \max\left\{ \left| \sum_{i=1}^n (h_1(X_i, t_{r,l}) - h_1(X_i, t))) \right|, \left| \sum_{i=1}^n (h_1(X_i, t) - h_1(X_i, t_{r-1,l}))) \right| \right\} \\
&\qquad + \frac{n(n-1)}{2} |U(t_{r,l}) - U(t_{r-1,l})|
\end{aligned}$$

$$\leq \max\{|Q_n(t_{r,l})|, |Q_n(t_{r-1,l})|\} + (n-1)\left|\sum_{i=1}^{n}(h_1(X_i,t_{r,l}) - h_1(X_i,t_{r-1,l})))\right| + 2\frac{n(n-1)}{2}|U(t_{r,l}) - U(t_{r-1,l})|.$$

So we have that

$$\sup_{t\in\mathbb{R}}|Q_n(t)| \leq \max_{r=0,\dots,k}|Q_n(t_{r,l})| + \max_{r=0,\dots,k}(n-1)\left|\sum_{i=1}^{n}(h_1(X_i,t_{r,l}) - h_1(X_i,t_{r-1,l})))\right| + \max_{r=0,\dots,k} n(n-1)|U(t_{r,l}) - U(t_{r-1,l})|.$$

We will treat these three summands separately. We have for the last summand that $\max_{r=0,\dots,k} n(n-1)|U(t_{r,l}) - U(t_{r-1,l})| \leq Cn^2 2^{-\frac{5}{8}l} = o\left(n^{\frac{3}{2}-\frac{\gamma}{8}}\right)$ by the choice of $t_1,\dots,t_{k-1}$. For the first summand, we obtain with similar arguments as in the proof of Proposition 3.3.2

$$\begin{aligned}
&E[\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=0,\dots,k}|Q_n(t_{r,l})|^2]\\
\leq&\sum_{r=0}^{k}E\left[\left(\sum_{d=0}^{l}\max_{i=1,\dots,2^{l-d}}\left|Q_{2^l+i2^d}(t_{r,l}) - Q_{2^l+(i-1)2^d}(t_{r,l})\right|\right)^2\right]\\
\leq&\sum_{r=0}^{k}l\sum_{d=0}^{l}\sum_{i=1}^{2^{l-d}}E\left[\left(Q_{2^l+i2^d}(t_{r,l}) - Q_{2^l+(i-1)2^d}(t_{r,l})\right)^2\right]\\
\leq&\sum_{r=0}^{k}l\sum_{d=0}^{l}\sum_{i_1,j_1,i_2,j_2=1}^{2^{l+1}}|E\left[h_2(X_{i_1},X_{j_1},t)h_2(X_{i_2},X_{j_2},t)\right]|\\
\leq&Ckl^2 2^{2(l+1)} \leq Cl^2 2^{(2+\frac{5}{8})l},
\end{aligned}$$

where we used Lemma 3.2.5 in the last line. With the Chebyshev inequality, it follows for every $\epsilon > 0$

$$\sum_{l=1}^{\infty}P\left[\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=0,\dots,k}|Q_n(t_{r,l})| > \epsilon 2^{l(\frac{3}{2}-\frac{\gamma}{8})}\right] \leq \sum_{l=1}^{\infty}\frac{1}{\epsilon^2 2^{l(3-\frac{\gamma}{4})}}E[\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=0,\dots,k}|Q_n(t_{r,l})|^2] \leq \sum_{l=1}^{\infty}\frac{1}{\epsilon^2 2^{l(3-\frac{\gamma}{4})}}l^2 2^{(2+\frac{5}{8})l} < \infty,$$

as $\gamma \leq 1$, so by the Borel Cantelli lemma

$$P\left[\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=0,\dots,k}|Q_n(t_{r,l})| > \epsilon 2^{l(\frac{3}{2}-\frac{\gamma}{8})}\text{ i.o.}\right] = 0$$

(the meaning of the abbreviation i.o. is "infinitely often") and $\max_{r=0,\dots,k}|Q_n(t_{r,l})| = o\left(n^{\frac{3}{2}-\frac{\gamma}{8}}\right)$ almost surely. It remains to show the convergence of the second summand. By Lemma 4.2.1 respectively 4.2.2 for $2^l \leq n < 2^{l+1}$

$$E\left(\sum_{i=1}^{n}(h_1(X_i,t_{r,l}) - h_1(X_i,t_{r-1,l}))\right)^4 \leq Cn^2(\log n)^2 |U(t_{r,l}) - U(t_{r-1,l})|^{1+\gamma}$$

as $|U(t_{r,l}) - U(t_{r-1,l})| \geq 2^{-\frac{5}{8}l} \geq C2^{-\frac{3}{4}l}$ and consequently

$$\begin{aligned}
&E\left(\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=1,\dots,k}(n-1)\left|\sum_{i=1}^{n}(h_1(X_i,t_{r,l}) - h_1(X_i,t_{r-1,l}))\right|\right)^4 \\
&\leq 2^{4(l+1)}\sum_{r=1}^{k}E\left(\max_{n=2^l,\dots,2^{l+1}-1}\left|\sum_{i=1}^{n}(h_1(X_i,t_{r,l}) - h_1(X_i,t_{r-1,l}))\right|\right)^4 \\
&\qquad\qquad \leq C2^{6l}l^2k(\max_{r=1,\dots,k}|U(t_{r,l}) - U(t_{r-1,l})|)^{1+\gamma} \leq Cl^2 2^{(6-\frac{5}{8}\gamma)l},
\end{aligned}$$

where we used Corollary 1 of Móricz to obtain the last line. Remember that by our choice $k = k_l = O(2^{\frac{5}{8}l})$. We conclude that

$$\begin{aligned}
&\sum_{l=0}^{\infty}P\left[\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=1,\dots,k}(n-1)\left|\sum_{i=1}^{n}(h_1(X_i,t_{r,l}) - h_1(X_i,t_{r-1,l}))\right| > \epsilon 2^{(\frac{3}{2}-\frac{\gamma}{8})l}\right] \\
&\leq \sum_{l=0}^{\infty}\frac{C}{\epsilon^4 2^{l(6-\frac{\gamma}{2})}}E\left(\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=1,\dots,k}(n-1)\left|\sum_{i=1}^{n}(h_1(X_i,t_{r,l}) - h_1(X_i,t_{r-1,l}))\right|\right)^4 \\
&\qquad\qquad \leq \sum_{l=0}^{\infty}\frac{C}{\epsilon^4 2^{l(6-\frac{\gamma}{2})}}l^2 2^{(6-\frac{5}{8}\gamma)l} = \sum_{l=0}^{\infty}\frac{Cl^2}{\epsilon^4 2^{\frac{\gamma}{8}l}} < \infty.
\end{aligned}$$

The Borel Cantelli lemma completes the proof. □

4.3 Strong Invariance Principle

The asymptotic theory for the empirical U-process makes use of the Hoeffding decomposition, recall that $h_1(x,t) := E[h(x,Y,t)] - U(t)$. Under the mixing assumptions of the theorem below, the following covariance function converges absolutely and is continuous (compare Theorem 5 of Borovkova et al. [18]):

$$\begin{aligned}
\Gamma(t,t') &= 4\,\mathrm{Cov}\left[h_1(X_1,t), h_1(X_1,t')\right] \\
&\qquad + 4\sum_{k=1}^{\infty}\mathrm{Cov}\left[h_1(X_1,t), h_1(X_{k+1},t')\right] + 4\sum_{k=1}^{\infty}\mathrm{Cov}\left[h_1(X_{k+1},t), h_1(X_1,t')\right].
\end{aligned}$$

The following theorem was first published in Wendler [94]:

Theorem 4.3.1. *Let h be a kernel function that satisfies the uniform variation condition such that the U-distribution function U is Lipschitz-continuous and one of the following two mixing conditions is satisfied:*

1. *$(X_n)_{n\in\mathbb{N}}$ is strongly mixing with mixing coefficients $\alpha(n) = O(n^{-\alpha})$ for $\alpha \geq 8$ and $E|X_i|^\rho < \infty$ for a $\rho > \frac{1}{4}$.*
2. *$(X_n)_{n\in\mathbb{N}}$ is near epoch dependent on an absolutely regular process with mixing coefficients $\beta(n) = O(n^{-\beta})$ for $\beta \geq 8$ with approximation constants $a_n = O(n^{-a})$ for $a = \max\{\beta + 3, 12\}$.*

Then there exists a centered Gaussian process $(K(t,s))_{t,s\in\mathbb{R}}$ (after enlarging the probability space if necessary) with covariance function

$$EK(t,s)K(t',s') = \min\{s,s'\}\,\Gamma(t,t')$$

such that almost surely

$$\sup_{\substack{t\in\mathbb{R}\\ s\in[0,1]}} \frac{1}{\sqrt{n}} \left|\lfloor ns\rfloor (U_{\lfloor ns\rfloor}(t) - U(t)) - K(t,ns)\right| = O(\log^{-\frac{1}{3840}} n).$$

Proof. We use the Hoeffding decomposition

$$U_n(t) = U(t) + \frac{2}{n}\sum_{i=1}^{n} h_1(X_i,t) + \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} h_2(X_i,X_j,t).$$

By Proposition 4.2.3, we have almost surely

$$\sup_{\substack{t\in\mathbb{R}\\ s\in[0,1]}} \frac{1}{n^{\frac{3}{2}}s} \left|\sum_{1\leq i<j\leq ns} h_2(X_i,X_j,t)\right|$$
$$\leq n^{-\frac{\gamma}{8}} \sup_{\substack{t\in\mathbb{R}\\ n'=1,\dots n}} \frac{1}{(n')^{\frac{3}{2}-\frac{\gamma}{8}}} \left|\sum_{1\leq i<j\leq n'} h_2(X_i,X_j,t)\right| = O(n^{-\frac{\gamma}{8}}).$$

So the statement of the theorem follows if we can proof that there exists a centered Gaussian process $(K(t,s))_{t,s\in\mathbb{R}}$

$$\sup_{\substack{t\in\mathbb{R}\\ s\in[0,1]}} \frac{1}{\sqrt{n}} \left|\lfloor ns\rfloor (U_{\lfloor ns\rfloor}(t) - U(t)) - K(t,ns)\right|$$
$$\leq \sup_{\substack{t\in\mathbb{R}\\ s\in[0,1]}} \frac{1}{\sqrt{ns}} \left|\left(2\sum_{1\leq i\leq ns} h_1(X_i,t) - K(t,ns)\right)\right| + \sup_{\substack{t\in\mathbb{R}\\ s\in[0,1]}} \frac{1}{n^{\frac{3}{2}}s} \left|\sum_{1\leq i<j\leq ns} h_2(X_i,X_j,t)\right|$$
$$= O(\log^{-\frac{1}{3840}} n).$$

This proposition is basically Theorem 1 of Berkes and Philipp [14], which we have to generalize in three aspects:

1. Berkes and Philipp assume that the covariance kernel Γ is positive definite, we want to avoid this condition here.

2. Berkes and Philipp consider indicator functions $\mathbb{1}_{\{x\le t\}}$, while in this version of the proposition, we deal with more general functions $Eh(x,Y,t)$.

3. Theorem 1 of Berkes and Philipp is restricted to the distribution function $F(t) = E\mathbb{1}_{\{X_i\le t\}} = t$, we will extend this to a Lipschitz continuous function U, bounded by 0 and 1.

Berkes and Philipp assume that $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent on a strongly mixing process, which holds under both our mixing assumptions.

1. In the proof of their Theorem 1, Berkes and Philipp use the fact that Γ is positive definite only for two steps. Their Proposition 4.1 (page 124) also holds if this is not the case. It is easy to see that the characteristic functions of the finite dimensional distributions then might converge to 1 at some points, but with the required rate. Furthermore, we have to show (page 135) that for all $t_1,\ldots,t_{d_k}\in[0,1]$, $P[\|(K(t_1,1),\ldots,K(t_{d_k},1))\|\ge\frac{1}{4}T_k]\le\delta_k$, where T_k and δ_k are defined in their article. Let $\Gamma_{d_k}=(\Gamma(t_i,t_j))_{1\le i,j\le d_k}$ be the covariance matrix of $K(t_1,1),\ldots,K(t_{d_k},1)$ and λ its biggest eigenvalue. We first consider the case that $\lambda>0$. As Γ_{d_k} is symmetric and positive semidefinite, there exist a matrix $\Gamma_{d_k}^{\frac{1}{2}}$ such that $(\Gamma_{d_k}^{\frac{1}{2}})^t\Gamma_{d_k}^{\frac{1}{2}}=\Gamma_{d_k}$ and the random vector $(K(t_1,1),\ldots,K(t_{d_k},1))$ has the same distribution as $\Gamma_{d_k}^{\frac{1}{2}}(W_1,\ldots,W_{d_k})^t$, where $W_1,\ldots,W_{d_k}$ are independent standard normal random variables. So it follows that

$$
\begin{aligned}
&P[\|(K(t_1,1),\ldots,K(t_{d_k}))\|\ge\frac{1}{4}T_k]=P[\|\Gamma_{d_k}^{\frac{1}{2}}(W_1,\ldots,W_{d_k})\|\ge\frac{1}{4}T_k]\\
&\le P[\sqrt{\lambda}\|(W_1,\ldots,W_{d_k})^t\|\ge\frac{1}{4}T_k]\\
&\qquad\qquad=\frac{1}{(2\pi)^{\frac{1}{2}}d_k}\int_{\|(x_1,\ldots,x_{d_k})\|\ge\frac{1}{4\sqrt{\lambda}}T_k}\exp(-\frac{1}{2}(x_1^2+\ldots+x_{d_k}^2))dx_1\ldots dx_{d_k}.
\end{aligned}
$$

The rest of the proof is then exactly the same as in Berkes and Philipp [14]. In the case $\lambda=0$, we have that $\Gamma=0$, so trivially $P[\|(K(t_1,1),\ldots,K(t_{d_k}))\|\ge\frac{1}{4}T_k]=0\le\delta_k$.

2. The proof uses different properties of the indicator functions. If the process $(X_n)_{n\in\mathbb{N}}$ is near epoch dependent with constants $(a_n)_{n\in\mathbb{N}}$, then as a consequence of Lemma 3.2.1 of Philipp [75] the process $\left(\mathbb{1}_{\{X_n\le t\}}\right)_{n\in\mathbb{N}}$ is near epoch dependent with constants $(\sqrt{a_n})_{n\in\mathbb{N}}$. The same holds for the sequence $(h_1(X_n,t))_{n\in\mathbb{N}}$ by the boundedness of h_1 and Lemma 2.1.7.

Furthermore, h and U are nondecrasing in t. Berkes and Philipp used different moment properties, which we also assume: $h_1(X_n,t)$ is bounded by 1 and $E|h_1(X_n,t) - h_1(X_n,t')| \leq C|t-t'|$ for $t,t' \in \mathbb{R}$, so consequently for $m \in \mathbb{R}$ $\|h_1(X_n,t)\|_m \leq 1$ and $\|h_1(X_n,t) - h_1(X_n,t')\|_m \leq |t-t'|^{\frac{1}{m}}$. So this more general version can be proved along the lines of the proof in Berkes and Philipp [14].

3. If $U(t) = t$ does not hold, note that $Eh_1(X_i,t_p) = U(t_p) = p$ with $t_p = U^{-1}(p) := \inf\{t \in \mathbb{R} | U(t) \geq p\}$, because U is continuous. Clearly, the uniform variation condition holds for $h(x,y,U^{-1}(p))$. Furthermore, notice that if $U(t) = U(s)$, when $h_1(X_i,t) = h_1(X_i,s)$ almost surely by monotonicity of h, so

$$\sum_{i=1}^{n} h_1(X_i,t) = \sum_{i=1}^{n} h_1(X_i,t_{U(t)})$$

almost surely. From the first two parts of the proof, we know that there is a centered Gaussian process $K^\star$ with covariance function

$$E[K^\star(p,s)K^\star(p',s')] = \min\{s,s'\}\,\Gamma(t_p,t_{p'})$$

with

$$\sup_{\substack{p\in[0,1]\\ s\in[0,1]}} \frac{1}{\sqrt{n}} \left| \left(2\sum_{1\leq i\leq ns} h_1(X_i,t_p) - K^\star(p,ns) \right) \right| = O(\log^{-\frac{1}{3840}} n)$$

almost surely. The Gaussian process K with $K(t,s) = K^\star(U(t),s)$ has the required covariance function and

$$\sup_{\substack{t\in\mathbb{R}\\ s\in[0,1]}} \frac{1}{\sqrt{n}} \left| \left(2\sum_{1\leq i\leq ns} h_1(X_i,t) - K(t,ns) \right) \right| =$$
$$\sup_{\substack{t\in\mathbb{R}\\ s\in[0,1]}} \frac{1}{\sqrt{n}} \left| \left(2\sum_{1\leq i\leq ns} h_1(X_i,t_{U(t)}) - K^\star(U(t),ns) \right) \right| = O(\log^{-\frac{1}{3840}} n).$$

□

The rate of convergence to zero in this theorem is very slow, but the same as in Berkes and Philipp [14], as we made use of their results and methods. By the scaling property of the process K, we obtain the asymptotic distribution of $U_{\lfloor ns\rfloor}(t)$, and by Theorem 2.3 of Arcones [4] a functional LIL:

Corollary 4.3.2. *Under the assumptions of Theorem 4.3.1 the empirical U-process*

$$\left(\frac{\lfloor ns\rfloor}{\sqrt{n}} (U_{\lfloor ns\rfloor}(t) - U(t)) \right)_{t\in\mathbb{R}, s\in[0,1]}$$

converges weakly in the space $D(\mathbb{R} \times [0,1])$ *of càdlàg functions (equipped with the supremum norm) to a centered Gaussian process* $(K(t,s))_{t,s\in\mathbb{R}}$ *introduced in Theorem 4.3.1.*

Corollary 4.3.3. *Under the assumptions of Theorem 4.3.1, the sequence*

$$\left(\left(\frac{\lfloor ns\rfloor}{\sqrt{2n\log\log n}}(U_{\lfloor ns\rfloor}(t) - U(t))\right)_{t\in\mathbb{R}, s\in[0,1]}\right)_{n\in\mathbb{N}}$$

is almost surely relatively compact in the space $D(\mathbb{R} \times [0,1])$ *of càdlàg functions (equipped with the supremum norm) and the limit set is the unit ball* $\mathcal{U}_K$ *of the reproducing kernel Hilbert space* $\mathcal{K}$ *associated with the covariance function of the process* K*.*

5 U-Quantiles

5.1 Definition and Applications

Apart from having a small variance and bias, a good estimator should be robust, meaning that extreme outliers which might result from heavy tails or from wrong measurement should not affect the estimation to much. This can be modeled by random variables $(X_n)_{n\in\mathbb{N}}$ having a distribution function $F = (1-\epsilon)F_1+\epsilon F_2$, meaning that with probability ϵ, we have an outlier with distribution F_2, while the good observations have distribution F_1. Robustness can be quantified with the breakdown point, the minimal fraction of observations that can shift the estimation arbitrarily. Suppose we want to estimate a functional $T(F)$ of the distribution by $T(F_n)$, the value of for the empirical distribution function, that $T(F_n)$ is consistent, and that T can take any real value. More precisely, the breakdown point $\epsilon^\star$ is defined as

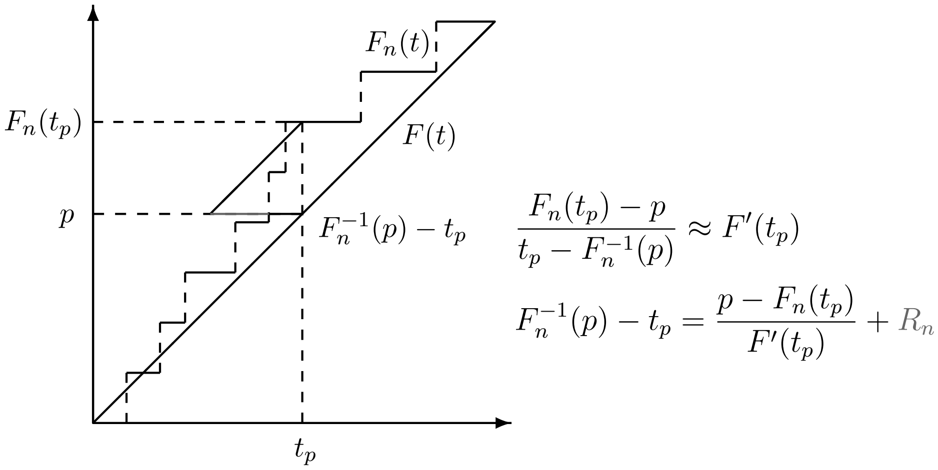
$$\epsilon^\star = \epsilon^\star_{F_1} := \sup\left\{\epsilon > 0 \Big| \sup_{F_2} |T(F_1) - T((1-\epsilon)F_1 + \epsilon F_2)| < \infty\right\}.$$

In many cases, the value of $\epsilon^\star_{F_1}$ does not depend on F_1. For more details on robust estimation see the book of Huber [54]. The sample mean belonging to the functional $T(F) = \int x dF(x)$ is not robust, as its breakdown point is 0. An robust estimator of location is the median ($T(F) = F^{-1}(\frac{1}{2})$) with breakdown point $\epsilon = 0.5$. But the median has a rather low relativ asymptotic efficiency of 64 % for independent normal distributed random variables, meaning that the variance in this situation is increased compared to the sample mean.

The Hodges-Lehmann [51] estimator is another robust estimator of location and is defined as $H_n = \text{median}\left\{\frac{X_i+X_j}{2}\Big| 1 \leq i < j \leq n\right\}$. The breakdown point of this estimator is 29%, while the efficiency is 96% (See Choudhury and Serfling [22]), so the variance is increased only slightly. The Hodges-Lehmann estimator is an example of a U-quantile, i.e. a quantile of the sample $(g(X_i, X_j))_{1\leq i<j\leq n}$, where g is a measurable and symmetric function. In the example of the Hodges-Lehmann estimator, we use $g(x, y) := \frac{1}{2}(x+y)$. To study U-quantiles, we will adopt concepts for ordinary sample quantiles.

Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary sequence of real-valued random variables with distri-

Figure 5.1: Empirical Distribution Function and Remainder in Bahadur Representation



bution function F and $p \in (0,1)$. Then the p-quantile t_p of F is defined as

$$t_p = F^{-1}(p) := \inf \left\{ t \in \mathbb{R} \middle| F(t) \geq t \right\}$$

and can be estimated by the empirical p-quantile, i.e. the $\lceil \frac{n}{p} \rceil$-th order statistic of the sample $X_1 \ldots, X_n$. This also can be expressed as the p-quantile $F_n^{-1}(p)$ of the empirical distribution function $F_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}$. It is clear that $F_n^{-1}(p)$ is greater than t_p iff $F_n(t_p)$ is smaller than p. The relation between the empirical distribution function and the empirical quantile can be refined with the following heuristic argument: If the the function F_n converges to F, one might hope that the slope also converges, so $\dfrac{F_n(t_p) - p}{t_p - F_n^{-1}(p)} \approx f(t_p) := F'(t_p)$. This leads to the Bahadur representation [11]

$$F_n^{-1}(p) = t_p + \frac{p - F_n(t_p)}{f(t_p)} + R_n.$$

Bahadur [11] showed that $R_n = O\left(n^{-\frac{3}{4}} (\log n)^{\frac{1}{2}} (\log \log n)^{\frac{1}{4}}\right)$. This was refined by Kiefer [58] to

$$\limsup_{n \to \infty} \left(\frac{n}{2 \log \log n} \right)^{\frac{3}{4}} R_n = 2^{\frac{1}{2}} 3^{-\frac{3}{4}} p^{\frac{1}{4}} (1-p)^{\frac{1}{4}}.$$

Kiefers proof is very elaborated, a much simpler proof, but with a weaker result (only convergence in probability) was given by Ghosh [43].

The following short calculation shows that R_n is related to the (local) empirical process $\left(F_n(t + t_p) - F_n(t_p) - f(t_p)\, t\right)_t$ centered in $(t_p, F_n(t_p))$ and its inverse denoted by Z_n:

$$\begin{aligned}Z_n(x) &:= \left(F_n(\cdot + t_p) - F_n(t_p)\right)^{-1}(x) - \frac{x}{f(t_p)}\\ &= \inf\left\{s \middle| F_n(s+t_p) - F_n(t_p) \leq x\right\} - \frac{x}{f(t_p)}\\ &= \inf\left\{s \middle| F_n(s) \leq x + F_n(t_p)\right\} - \frac{x}{f(t_p)} - t_p\\ &= F_n^{-1}\left(x + F_n(t_p)\right) - \frac{x}{f(t_p)} - t_p.\end{aligned}$$

So we have

$$Z_n\left(p - F_n(t_p)\right) = F_n^{-1}(p) - t_p + \frac{F_n(t_p) - p}{f(t_p)} = R_n.$$

Deheuvels and Mason [26] used this argument to give a new proof of the precise rate of R_n given by Kiefer [58]. The results under mixing conditions are not as precise. We will prove our result in the following way: The first step of our proof is to show that $\left(F_n(t+t_p) - F_n(t_p) - f(t_p)\,t\right)_{t\in I_n}$ converges to zero at some rate uniformly on intervals $I_1 \supset I_2 \supset I_3 \ldots$ By a theorem of Vervaat [90], $-Z_n$ has the same limit behaviour as the (local) empirical process. We will then conclude that $R_n = Z_n\left(F(t_p) - F_n(t_p)\right)$ converges to zero at the same rate and obtain the central limit theorem and the law of the iterated logarithm as easy corollaries.

There is a broad literature on the Bahadur representation for dependent data beginning with Sen, who studied m-dependent [81] and ϕ-mixing random variables [83]. Babu and Singh [10] proved such a representation under an exponentially fast decay of the strong mixing coefficients, this was weakened by Yoshihara [99] and Sun [88] to a polynomial decay of the strong mixing coefficients. We will obtain a improved version of their results as a special case of our Theorem 5.3.1. Coeurjolly [23] investigated the Bahadur representation for Gaussian processes. Dutta and Sen [40] considered autoregressive processes, Hesse [48], Wu [96] and Kulik [64] established a Bahadur representation for linear processes and Kulik [65] for GARCH processes. Both linear processes and GARCH processes can be treaded as near epoch dependent sequences and are thus included in Theorem 5.3.1.

We are interested in the empirical U-quantile, i.e. the p-quantile of the sample $\left(g(X_i, X_j)\right)_{1\leq i<j\leq n}$ for a measurable, symmetric kernel g, which can be expressed as the generalized inverse of the empirical U-distibution function U_n:

Definition 5.1.1. Let $h : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a kernel function and U the U-distribution function. Then

$$t_p = U^{-1}(p) = \inf\{t | U(t) \geq p\}$$

is called p-U-quantil and

$$U_n^{-1}(p) = \inf\{t | U_n(t) \geq p\}$$

empirical p-U-quantile, where U_n is the empirical U-distibution function.

Let the U-distribution function $U(t) := P[h(X,Y,t)]$ be differentiable in t_p with $u(t_p) := U'(t_p) > 0$. Similarly to a sample quantile, $U_n^{-1}(p)$ can be analyzed with the help of a generalized Bahadur respresentation

$$U_n^{-1}(p) = t_p + \frac{U(t_p) - U_n(t_p)}{u(t_p)} + R_n.$$

For the special case of the Hodges-Lehmann estimator of independent data, Geertsema [42] established a generalized Bahadur representation with $R_n = O(n^{-\frac{3}{4}} \log n)$ almost surely. For general U-quantiles, Dehling et al. [29] and Choudhury and Serfling [22] improved the rate to $R_n = O(n^{-\frac{3}{4}} (\log n)^{\frac{3}{4}})$. Arcones [5] proved the exact order $R_n = O(n^{-\frac{3}{4}} (\log\log n)^{\frac{3}{4}})$ as for sample quantiles. Let us give some examples:

Example 5.1.2 (Hodges-Lehmann estimator)**.** Let $h(x,y,t) = \mathbb{1}_{\left\{\frac{1}{2}(x+y) \leq t\right\}}$. The 0.5-$U$-quantil is the Hodges-Lehmann estimator for location [51]. Note that

$$\sup_{\|(x,y)-(X,Y)\| \leq \epsilon} \left| \mathbb{1}_{\left\{\frac{1}{2}(x+y) \leq t\right\}} - \mathbb{1}_{\left\{\frac{1}{2}(X+Y) \leq t\right\}} \right| = \begin{cases} 1 & \text{if } \frac{X+Y}{2} \in \left(t - \frac{\epsilon}{\sqrt{2}}, t + \frac{\epsilon}{\sqrt{2}}\right] \\ 0 & \text{else} \end{cases}$$

If X_1 has a bounded density, then the density $f_{\frac{1}{2}(X+Y)}$ of $\frac{1}{2}(X+Y)$ is also bounded, so

$$\begin{aligned} E&\left[\sup_{\|(x,y)-(X,Y)\| \leq \epsilon} |h(x,y) - h(X,Y)|\right] \\ &\leq P\left[\frac{X+Y}{2} \in \left(t - \frac{\epsilon}{\sqrt{2}}, t + \frac{\epsilon}{\sqrt{2}}\right]\right] \leq \left(\sqrt{2} \sup_{x \in \mathbb{R}} f_{\frac{1}{2}(X+Y)}(x)\right) \cdot \epsilon \end{aligned}$$

and the kernel function $h(x,y,t) = \mathbb{1}_{\left\{\frac{1}{2}(x+y) \leq t\right\}}$ satisfies the uniform variation condition on $\mathbb{R}$.

Example 5.1.3 (Q_n estimator of scale)**.** Let $h(x,y,t) = \mathbb{1}_{\{|x-y| \leq t\}}$. When the 0.25-$U$-quantile is the Q_n estimator of scale proposed by Rousseeuw and Croux [79]. If X_1 has a bounded density, then with similar arguments as for the Hodges-Lehmann-estimator, the kernel function $h(x,y,t) = \mathbb{1}_{\{|x-y| \leq t\}}$ satisfies the uniform variation condition.

5.2 On the Local Behaviour of the Empirical Distribution Function

Lemma 5.2.1. *Let $h_1 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a non-negative, bounded, measurable function which is non-decreasing in the second argument, let $F(t) := E[h_1(X_1, t)]$ be differentiable in $t_p \in \mathbb{R}$ with $F'(t_p) = f(t_p) > 0$ and*

$$|F(t) - F(t_p) - f(t_p)(t - t_p)| = o\left(|t - t_p|^{\frac{3}{2}}\right) \quad \text{as} \quad t \to t_p.$$

Assume that one of the following two conditions holds:

1. *$(X_n)_{n\in\mathbb{Z}}$ is strongly mixing with $\alpha(n) = O(n^{-\alpha})$ for some $\alpha \geq 3$. Let $\gamma := \frac{\alpha-2}{\alpha}$.*
2. *$(X_n)_{n\in\mathbb{Z}}$ is a near epoch dependent functional on an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$ with mixing coefficients $(\beta(n))_{n\in\mathbb{N}}$ and approximation constants $(a_n)_{n\in\mathbb{N}}$, such that $\beta(n) = \left(n^{-\beta}\right)$ and $a_n = \left(n^{-(\beta+3)}\right)$ for some $\beta > 3$. Let g satisfy the variation condition uniformly in some neighbourhood of t_p and let $\gamma := \frac{\beta-3}{\beta+1}$.*

Then for $F_n(t) := \frac{1}{n}\sum_{i=1}^n h_1(X_i, t)$, $p = F(t_p)$ and any constant $C > 0$

$$\sup_{|t-t_p|\leq C\sqrt{\frac{\log\log n}{n}}} |F_n(t) - F(t) - F_n(t_p) + F(t_p)| = o\left(n^{-\frac{5}{8}-\frac{1}{8}\gamma}(\log n)^{\frac{3}{4}}(\log\log n)^{\frac{1}{2}}\right)$$

almost surely as $n \to \infty$.

Proof. Let $c_n = n^{-\frac{5}{8}-\frac{1}{8}\gamma}(\log n)^{\frac{3}{4}}(\log\log n)^{\frac{1}{2}}$. We first prove that

$$\sum_{l=0}^{\infty} P\left[\max_{2^l\leq n<2^{l+1}} \frac{1}{c_n} \sup_{|t-t_p|\leq C\sqrt{\frac{\log l}{2^l}}} (F_n(t) - F_n(t_p) - F(t) + F(t_p)) > \epsilon\right]$$
$$\leq C\sum_{l=0}^{\infty} \frac{1}{c_{2^l}^4} E\left(\max_{2^l\leq n<2^{l+1}} \sup_{|t-t_p|\leq C\sqrt{\frac{\log l}{2^l}}} (F_n(t) - F_n(t_p) - F(t) + F(t_p))\right)^4 < \infty.$$

The statement of the Lemma will follow by the Borel-Cantelli lemma. We set $d_{2^l} = \left(\frac{\log l}{2^l}\right)^{\frac{3}{4}}$ and $d_n = d_{2^l}$ for $2^l \leq n < 2^{l+1}$. Let $k \in \mathbb{Z}$. As F_n, F are non-decreasing in t, we have for any $t \in [t_p + kd_n, t_p + (k+1)d_n]$ that

$$\begin{aligned}
&|F_n(t) - F_n(t_p) - F(t) + F(t_p)| \\
\leq &\max\{|F_n(t_p + kd_n) - F_n(t_p) - F(t) + F(t_p)|, \\
&\quad |F_n(t_p + (k+1)d_n) - F_n(t_p) - F(t_p) + F(t_p)|\}
\end{aligned}$$

$$\leq \max \{|F_n(t_p + kd_n) - F_n(t_p) - F(t_p + kd_n) + F(t_p)|, \\ |F_n(t_p + (k+1)d_n) - F_n(t_p) - F(t_p + (k+1)d_n) + F(t_p)|\} \\ + |F(t_p + (k+1)d_n) - F(t_p + kd_n)|.$$

It follows that

$$\sup_{|t-t_p|\leq C\sqrt{\frac{\log l}{2^l}}} (F_n(t) - F_n(t_p) - F(t) + F(t_p)) \\ \leq \max_{|k|\leq C\left(2^l \log l\right)^{\frac{1}{4}}} (F_n(t_p + d_n k) - F_n(t_p) - F(t_p + d_n k) + F(t_p)) \\ + \max_{|k|\leq C\left(2^l \log l\right)^{\frac{1}{4}}} |F(t_p + (k+1)d_n) - F(t_p + kd_n)|.$$

From the differentiability condition in our theorem, we conclude that

$$\max_{|k|\leq C\left(2^l \log l\right)^{\frac{1}{4}}} |F(t_p + (k+1)d_n) - F(t_p + kd_n)| \leq f(t_p)d_n + o\left(\frac{\log^{\frac{3}{4}} l}{2^{\frac{3}{4}l}}\right) = o(c_n).$$

Furthermore, we have that for all $k_1, k_2 \leq C\left(2^l \log l\right)^{\frac{1}{4}}$

$$|F(t_p + d_n k_1) - F(t_p + d_n k_2)| = f(t_p)|k_1 - k_2| d_n + o\left(\frac{\log^{\frac{3}{4}} l}{2^{\frac{3}{4}l}}\right) \leq C|k_1 - k_2| d_n.$$

So by Lemma 4.2.1 (under mixing Condition 1.) or Lemma 4.2.2 (under mixing Condition 2.)

$$E(F_n(t_p + d_n k_1) - F_n(t_p + d_n k_2) - F(t_p + d_n k_1) + F(t_p + d_n k_2))^4 \\ \leq C\frac{1}{n^2}(\log n)^2 |k_1 - k_2|^{1+\gamma} d_n^{1+\gamma}.$$

Note that we can represent the differences of the empirical distribution function as a double sum

$$F_n(t_p + d_n k) - F_n(t_p) - F(t_p + d_n k) + F(t_p) \\ = \sum_{i=1}^{n}\sum_{j=1}^{k}(h_1(X_i, t_p + jd_n) - h_1(X_i, t_p + (j-1)d_n) - F(t_p + jd_n) + F(t_p + (j-1)d_n)),$$

so by Corollary 1 of Móricz [69], it then follows that

$$\begin{aligned}
&\frac{1}{c_{2^l}^4} E\left(\max_{2^l \leq n < 2^{l+1}} \max_{|k| \leq C\left(2^l \log l\right)^{\frac{1}{4}}} \left(F_n\left(t_p + d_n k\right) - F_n\left(t_p\right) - F\left(t_p + d_n k\right) + F\left(t_p\right)\right)\right)^4 \\
&\leq C \frac{1}{c_{2^l}^4} E\left(F_n\left(t_p + C\sqrt{\frac{\log\log n}{n}}\right) - F_n\left(t_p - C\sqrt{\frac{\log\log n}{n}}\right)\right. \\
&\qquad\qquad \left. - F\left(t_p + C\sqrt{\frac{\log\log n}{n}}\right) + F\left(t_p - C\sqrt{\frac{\log\log n}{n}}\right)\right)^4 \\
&\leq C \frac{2^{\frac{5+\gamma}{2} l}}{l^3 \left(\log l\right)^2} \frac{l^2}{2^{2l}} \frac{\left(\log l\right)^{\frac{1+\gamma}{2}}}{2^{\frac{1+\gamma}{2} l}} = C \frac{1}{l \left(\log l\right)^{\frac{3-\gamma}{2}}}.
\end{aligned}$$

As $\gamma < 1$, this quantities are summable and by the Markov and Chebyshev inequality, the proof is completed. □

5.3 Central Limit Theorem and Law of the Iterated Logarithm

Before investigating the asymptotic behaviour of U-quantiles, we will investigate the rate of convergence of the remainder term in the Bahadur representation. The following theorem can also be found in Wendler [93]:

Theorem 5.3.1. *Let $h : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a kernel function that satisfies the uniform variation condition in some neighbourhood of t_p. Let $U(t) := E\left[h(X, Y, t)\right]$ be differentiable in $t_p \in \mathbb{R}$ with $U'(t_p) = u(t_p) > 0$ and*

$$\left|U(t) - U(t_p) - u(t_p)(t - t_p)\right| = o\left(\left|t - t_p\right|^{\frac{3}{2}}\right) \quad as \;\; t \to t_p.$$

Assume that one of the following two conditions holds:

1. *$\|X_n\|_1 < \infty$ and $(X_n)_{n\in\mathbb{Z}}$ is strongly mixing and the mixing coefficients satisfy $\alpha(n) = O(n^{-\alpha})$ for some $\alpha \geq 5$. Le $\gamma := \frac{\beta-2}{\beta}$.*

2. *$(X_n)_{n\in\mathbb{Z}}$ is a near epoch dependent functional of an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$ with mixing coefficients $(\beta(n))_{n\in\mathbb{N}}$ and approximation constants $(a_n)_{n\in\mathbb{N}}$, such that $\beta(n) = \left(n^{-\beta}\right)$ and $a_n = \left(n^{-(\beta+3)}\right)$ for some $\beta > 3$. Let $\gamma := \frac{\beta-3}{\beta+1}$.*

Then for $U_n(t) := \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} h(X_i, X_j, t)$, $p = U(t_p)$ *and any constant* $C > 0$

$$\sup_{|t-t_p|\leq C\sqrt{\frac{\log\log n}{n}}} |U_n(t) - U(t) - U_n(t_p) + p| = o\left(n^{-\frac{5}{8}-\frac{1}{8}\gamma}(\log n)^{\frac{3}{4}}(\log\log n)^{\frac{1}{2}}\right)$$

$$R_n := U_n^{-1}(p) - t_p - \frac{p - U_n(t_p)}{u(t_p)} = o\left(n^{-\frac{5}{8}-\frac{1}{8}\gamma}(\log n)^{\frac{3}{4}}(\log\log n)^{\frac{1}{2}}\right)$$

almost surely as $n \to \infty$.

Proof. To prove the first statement of the theorem, we use the Hoeffding decomposition

$$U_n(t) = U(t) + \frac{2}{n}\sum_{i=1}^{n} h_1(X_i, t) + \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} h_2(X_i, X_j, t).$$

As above, we set $c_n = n^{-\frac{5}{8}-\frac{1}{8}\gamma}(\log n)^{\frac{3}{4}}(\log\log n)^{\frac{1}{2}}$, $d_{2^l} = \left(\frac{\log l}{2^l}\right)^{\frac{3}{4}}$ and $d_n = d_{2^l}$ for $2^l \leq n < 2^{l+1}$. We get

$$\begin{aligned}
&\sup_{|t-t_p|\leq C\sqrt{\frac{\log l}{2^l}}} |U_n(t) - U_n(t_p) - U(t) + U(t_p)| \\
\leq &\max_{|k|\leq C\left(2^l \log l\right)^{\frac{1}{4}}} |U_n(t_p + d_n k) - U_n(t_p) - U(t_p + d_n k) + U(t_p)| \\
&+ \max_{|k|\leq C\left(2^l \log l\right)^{\frac{1}{4}}} |U(t_p + d_n(k+1)) - U(t_p + d_n k)|
\end{aligned}$$

and

$$\max_{|k|\leq C\left(2^l \log l\right)^{\frac{1}{4}}} |U(t_p + d_n(k+1)) - U(t_p + d_n k)| = o(c_n).$$

By Lemma 3.1.14 we have that h_1 satisfies the variation condition uniformly in some neighbourhood of t_p. Applying Lemma 5.2.1, we obtain

$$\max_{|k|\leq C\left(2^l \log l\right)^{\frac{1}{4}}} \left|\frac{2}{n}\sum_{i=1}^{n} h_1(X_i, t_p + kd_n) - \frac{2}{n}\sum_{i=1}^{n} h_1(X_i, t_p) - U(t_p + d_n k) + U(t_p)\right| = o(c_n)$$

almost surely. It remains to show that

$$\max_{|k|\leq C\left(2^l \log l\right)^{\frac{1}{4}}} |Q_n(t_p + d_n k) - Q_n(t_p)| = o\left(n^2 c_n\right)$$

almost surely with $Q_n(t) := \sum_{1\leq i<j\leq n} h_2(X_i, X_j, t)$. We first consider mixing assumption 1. (strong mixing) and concentrate on the case $\alpha < 6$. In the case $\alpha \geq 6$, a similar calculation can be done. Recall that for any random variables $Y_1, \dots, Y_m$: $E\left(\max_{i=1,\dots,m} |Y_i|\right)^2 \le \sum_{i=1}^m EY_i^2$ and therefore

$$E\left(\max_{2^{l-1}\le n<2^l} \max_{|k|\le C\left(2^l \log l\right)^{\frac{1}{4}}} \frac{1}{2^{l-1}c_n} \left|Q_n\left(t_p+d_nk\right)-Q_n\left(t_p\right)\right|\right)^2$$

$$\le \frac{1}{2^{2(l-1)}c_{2^l}^2} E\left(\max_{|k|\le C\left(2^l \log l\right)^{\frac{1}{4}}} \sum_{d=1}^{l} \max_{i=1,\dots,2^{l-d}} \left(Q_{2^{(l-1)}+i2^{(d-1)}}\left(t_p+d_nk\right)-Q_{2^{(l-1)}+i2^{(d-1)}}\left(t_p\right)\right)\right)^2$$

$$\le \frac{1}{2^{2(l-1)}c_{2^l}^2} \sum_{|k|\le C\left(2^l \log l\right)^{\frac{1}{4}}} l \sum_{d=1}^{l} \sum_{i=1}^{2^{l-d}} E\left(Q_{2^{(l-1)}+i2^{(d-1)}}\left(t_p+d_nk\right)-Q_{2^{(l-1)}+i2^{(d-1)}}\left(t_p\right)\right)^2$$

$$\le \frac{1}{2^{2(l-1)}c_{2^l}^2} \sum_{|k|\le C\left(2^l \log l\right)^{\frac{1}{4}}} l^2 \sum_{i_1,i_2,i_3,i_4=1}^{2^l}$$

$$\left|E\left[\left(h_2\left(X_{i_1},X_{i_2},t_p+d_nk\right)-h_2\left(X_{i_1},X_{i_2},t_p\right)\right)\left(h_2\left(X_{i_3},X_{i_4},t_p+d_nk\right)-h_2\left(X_{i_3},X_{i_4},t_p\right)\right)\right]\right|,$$

where we used the triangular inequality in the last step. By means of Lemma 3.2.3 and 3.2.5, we arrive at

$$E\left(\max_{2^{l-1}\le n<2^l} \max_{|k|\le C\left(2^l \log l\right)^{\frac{1}{4}}} \frac{1}{2^{l-1}c_n} \left|Q_n\left(t_p+d_nk\right)-Q_n\left(t_p\right)\right|\right)^2$$

$$\le \frac{C}{2^{4l}c_{2^l}^2}\left(\frac{2^l}{\log l}\right)^{\frac{1}{4}} l^2 2^{2l} \sum_{i=1}^{2^l} i\alpha^{\frac{1}{3}}\left(i\right) \le \frac{C2^{l\left(\frac{3}{2}+\frac{1}{4}\gamma\right)}}{2^{4l} l^{\frac{3}{2}}\left(\log l\right)^{\frac{5}{4}}} l^2 2^{l\left(4-\frac{\alpha}{3}\right)} = C\frac{2^{l\left(\frac{3}{2}+\frac{1}{4}\gamma-\frac{\alpha}{3}\right)} l^{\frac{1}{2}}}{\left(\log l\right)^{\frac{5}{4}}}.$$

As $\alpha > 5$, we have that $\frac{3}{2}+\frac{1}{4}\gamma-\frac{1}{3}\alpha = \frac{-4\alpha^2+21\alpha-6}{12\alpha} < 0$ and thus the second moments are summable. The almost sure convergence follows by the Chebyshev inequality and the Borel-Cantelli lemma, so the first statement of the theorem is proved.

Under Condition 2. (near epoch dependence on absolutely regular sequences), we have by Lemma 3.2.4 and $\sum_{i=1}^\infty i\beta(i) < \infty$, $\sum_{i=1}^\infty iA_i < \infty$

$$E\left(\max_{2^{l-1}\le n<2^l} \max_{|k|\le C\left(2^l \log l\right)^{\frac{1}{4}}} \frac{1}{2^{l-1}c_n} \left|Q_n\left(t_p+d_nk\right)-Q_n\left(t_p\right)\right|\right)^2$$

$$\le \frac{C}{2^{4l}c_n^2}\left(\frac{2^l}{\log l}\right)^{\frac{1}{4}} l^2 2^{2l} \sum_{i=1}^{2^l} i\left(\beta(\frac{i}{3})+A_{\frac{i}{3}}\right) \le \frac{C2^{l\left(\frac{3}{2}+\frac{1}{4}\gamma\right)}}{2^{4l} l^{\frac{3}{2}}\left(\log l\right)^{\frac{5}{4}}} l^2 2^{2l} = \frac{Cl^{\frac{1}{2}}}{2^{l\left(\frac{1}{2}-\frac{1}{4}\gamma\right)}\left(\log l\right)^{\frac{5}{4}}}.$$

Since $\gamma \in (0,1)$, we have that $\frac{1}{2}-\frac{1}{4}\gamma > 0$ and the second moments are summable. By the Chebyshev inequality and the Borel-Cantelli lemma, we have now proved that

$$\begin{aligned}&\sup_{|t-t_p|\leq C\sqrt{\frac{\log l}{2^l}}} \left|U_n(t)-U_n(t_p)-U(t)+U(t_p)\right| \\ &= \max_{|k|\leq C\left(2^l\log l\right)^{\frac{1}{4}}} \left|U_n(t_p+d_nk)-U_n(t_p)-U(t_p+d_nk)+U(t_p)\right|+o(c_n)=o(c_n)\end{aligned}$$

almost surely. To prove that

$$R_n := U_n^{-1}(p)-t_p+\frac{U_n(t_p)-U(t_p)}{u(t_p)} = o\left(n^{-\frac{5}{8}-\frac{1}{8}\gamma}(\log n)^{\frac{3}{4}}(\log\log n)^{\frac{1}{2}}\right),$$

let without loss of generality $u(t_p)=1$, otherwise replacing $h(x,y,t)$ by $h\left(x,y,\frac{t}{u(t_p)}\right)$. We represent R_n with the help of the inverse of the local empirical U-process Z_n with

$$\begin{aligned}Z_n(x) &:= \left(U_n(\cdot+t_p)-U_n(t_p)\right)^{-1}(x)-x=\inf\left\{s\middle|U_n(s+t_p)-U_n(t_p)\leq x\right\}-x \\ &= \inf\left\{s\middle|U_n(s)\leq x+U_n(t_p)\right\}-x-t_p=U_n^{-1}(x+U_n(t_p))-x-t_p.\end{aligned}$$

So we have

$$R_n = Z_n\left(U(t_p)-U_n(t_p)\right).$$

By Theorem 3.3.3

$$\limsup_{n\to\infty} \pm\sqrt{\frac{n}{\log\log n}}\left(U_n(t_p)-U(t_p)\right)=C,$$

and by the first statement of the theorem and the differentiability condition we have that

$$\begin{aligned}&\sup_{|x|\leq C\sqrt{\frac{\log\log n}{n}}} \left|U_n(x+t_p)-U_n(t_p)-x\right| \\ &\leq \sup_{|x|\leq C\sqrt{\frac{\log\log n}{n}}} \left|U_n(x+t_p)-U(x+t_p)-U_n(t_p)+U(t_p)\right| \\ &\qquad + \sup_{|x|\leq C\sqrt{\frac{\log\log n}{n}}} \left|U(x+t_p)-U(t_p)-x\right| = o(c_n).\end{aligned}$$

So by Theorem 1 of Vervaat [90]

$$\sup_{|x|\leq C\sqrt{\frac{\log\log n}{n}}} \left|U_n^{-1}(x+t_p)-U_n^{-1}(t_p)-x\right| = o(c_n).$$

$$|R_n| \leq \sup_{|x|\leq C\sqrt{\frac{\log\log n}{n}}} |Z_n(x)| = o(c_n),$$

so the second statement of the Theorem is proved. □

With the kernel function $h(x,y,t) = \frac{1}{2}(\mathbb{1}_{\{x\leq t\}} + \mathbb{1}_{\{x\leq t\}})$, this theorem includes the ordinary distribution function $F_n(t) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{x\leq t\}}$ as a special case. For this kernel function, the degenerate part of the empirical U-distribution function does not exists and if the sequence $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, the variation condition is not needed. Furthermore, the decay of mixing coefficients $\alpha(n) = O(n^{-\alpha})$ for an $\alpha > 3$ is fast enough. This is an improvement of a Theorem by Sun [88], as he assumes a faster decay of the mixing coefficients, namely $\alpha > 10$, and obtains the rate $R_n = o(n^{-\frac{3}{4}+\delta}\log n)$ for any $\delta > \frac{11}{4(\alpha+1)}$. Additional, his proof demands the distribution function to be differentiable twice.

Yoshihara states the rate $R_n = o(n^{-\frac{3}{4}}\log n)$ a.s., but a careful reading shows that there is a mistake in Line (20) of his paper. Instead of

$$E\left|\sum_{j=1}^{n}\sum_{i=1}^{l}\zeta_j\left(\theta+(i-1)q_k, \theta+iq_k\right)\right|^4 \leq C(nlq_k)^{1+\gamma}$$

with $\zeta_i(s,t) = \mathbb{1}_{\{X_i\leq t\}} - \mathbb{1}_{\{X_i\leq s\}} - (F(t)-F(s))$, the inequality should be

$$E\left|\sum_{j=1}^{n}\sum_{i=1}^{l}\zeta_j\left(\theta+(i-1)q_k, \theta+iq_k\right)\right|^4 \leq Cn^2(lq_k)^{1+\gamma}.$$

If this line is corrected, his proof leads to the rate $R_n = o(n^{-\frac{5}{8}-\frac{1}{8}\gamma}(\log n)^{\frac{1}{4}}(\log\log n)^{\frac{1}{2}})$ with $\gamma \leq \frac{1}{5}$, so our theorem is also an improvement.

For a fast decay of the strong mixing coefficient ($\alpha \to \infty$), our rate becomes close to the optimal rate proved by Kiefer [58]. As the empirical distribution function and empirical sample quantiles are included as a special case, we cannot obtain a better rate. However, for some kernel functions and the associated U-quantiles, the empirical U-distribution function might be smoother and the remainder term in the Bahadur representation might converge faster:

Example 5.3.2. We consider again the Hodges-Lehmann estimator associated with the kernel function $h(x,y,t) = \mathbb{1}_{\left\{\frac{1}{2}(x+y)\leq t\right\}}$. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent standard normal random variables. We will show that

$$R_n := U_n^{-1}(p) - t_p - \frac{U(t_p) - U_n(t_p)}{u(t_p)} = o\left(n^{-1}\log^2 n\right)$$

almost surely. We have that $h_1(x,t) = P(Y \leq 2t-x) - P(X+Y\leq 2t) = \Phi(2t-x) - \Phi(\sqrt{2}t)$ is Lipschitz continuous in t uniformly for all x, where Φ is the distribution function of a standard normal random variable, so

$$E\left(\sum_{i=1}^{n} h_1(X_i,s) - h_1(X_i,t) - E\left[h_1(X_1,s) - h_1(X_i,t)\right]\right)^4 \leq Cn^2|s-t|^4.$$

Using this moment inequality instead of Lemma 4.2.1, we can prove Lemma 5.2.1 with $\gamma = 3$ and arrive at

$$\sup_{|t-t_p|\leq C\sqrt{\frac{\log\log n}{n}}} |F_n(t) - F(t) - F_n(t_p) + F(t_p)| = o\left(n^{-3}\log^2 n\right)$$

for $F_n(t) := \frac{1}{n}\sum_{i=1}^n h_1(X_i, t)$ almost surely as $n \to \infty$. Furthermore by the Lipschitz continuity of the U-distribution function, $E\left[h(X_i, X_j, t) - h(X_i, X_j, s)\right] \leq C|s-t|$, so $E\left[h_2(X_i, X_j, t) - h_2(X_i, X_j, s)\right] \leq C|s-t|$ and consequently we have for the degenerate part

$$E\left(Q_n(t_p + \frac{k}{n}) - Q_n(t_p)\right)^2 \leq Cn^2\frac{k}{n}$$

(the summands of $Q_n(t_p + \frac{k}{n}) - Q_n(t_p)$ are unkorrelated). With the same arguments as in the proof of our Theorem 5.3.1, we finally get

$$E\left(\max_{2^{l-1}\leq n<2^l}\max_{|k|\leq C\left(2^l\log l\right)^{\frac{1}{2}}}\frac{1}{n\log^2 n}\left|Q_n\left(t_p + \frac{k}{2^l}\right) - Q_n(t_p)\right|\right)^2$$
$$\leq C\frac{1}{2^{2l}l^4}l^2(2^l\log l)^{\frac{1}{2}}2^{2l}\left(\frac{\log l}{2^l}\right)^{\frac{1}{2}} = C\frac{\log l}{l^2}$$

and as this bounds are summable, we can conclude as before that $R_n = o\left(n^{-1}\log^2 n\right)$ almost surely.

For any kernel functions satisfying the conditions of Theorem 5.3.1, we have that $R_n = o\left(n^{-\frac{1}{2}}\right)$, so if we express the empirical U-quantile as

$$U_n^{-1}(p) = t_p + \frac{U(t_p) - U_n(t_p)}{u(t_p)} + R_n,$$

we can applying Theorem 3.3.1 or 3.3.3 to the U-statistic $U_n(t_p)$ to obtain

Corollary 5.3.3. *Under the assumptions of Theorem 5.3.1 it holds that*

$$\sqrt{n}\left(U_n^{-1}(p) - t_p\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2\right)$$

with

$$\sigma^2 = \frac{4}{u^2(t_p)}\left(\operatorname{Var}\left[h_1(X_1, t_p)\right] + 2\sum_{k=2}^{\infty}\operatorname{Cov}\left[h_1(X_1, t_p), h_1(X_k, t_p)\right]\right).$$

Corollary 5.3.4. *Under the assumptions of Theorem 5.3.1 it holds that*

$$\limsup_{n\to\infty} \pm\sqrt{\frac{n}{2\log\log n}}\left(U_n^{-1}(p) - t_p\right) = \sigma.$$

6 U-Quantile-Processes

6.1 Definition and Applications

In this final chapter, we will study not a single U-quantile, but the empirical U-quantile process $(U_n^{-1}(p))_{p\in I}$ under dependence (where I is some intervall defined later). We will combine methods from Chapter 4 and 5 to do so, developing a uniform version of the generalized Bahadur representation. There are other methods to deduce functional limit theorems for quantile processes from the limit behaviour of empirical processes which also could be generalized to U-quantiles, in particular the functional δ-method, see for example the book of van der Vaart and Wellner [89], page 387. Lévy-Leduc et al. [67] used this method for U-quantiles of long range dependent data. In a very clearly written paper, Vervaat [90] showed that the convergence of inverted processes can be derived with easy analytical arguments, Doss and Gill [37] gave a more general version of his arguments.

The Bahadur representation has the disadvantage that additional calculations are needed, but it gives a deeper inside into the quality of the approximation of the U-quantile process by the empirical U-process. We will examine the rate of convergence of $\sup_{p\in I} R_n(p)$ and use the approximation of the empirical U-process by a Gaussian process. The assumptions on the dependence of the random variables will be the same as in Chapter 4. However, as we divide by u in the Bahadur representation, we have to assume that this derivative behaves nicely. The densitiy of a random variable can not be bounded away from 0 on the whole real line, as it must integrate to 1, and the same problem occurs for the derivative of the U-distribution function, so we limit our investigation to some intervall where the this derivative is bounded away from 0.

The Bahadur representation for the sample quantile process goes back to Kiefer [58] under independence, Babu and Singh [10] proved such an representation for mixing data and Kulik [64] and Wu [96] for linear processes, but there seem to be no such results for the U-quantile process. Csörgő and Révész [24] established a strong invariance principle for the sample quantile process under independence, we will give a strong invariance principle for the U-quantile process under dependence. Additionally to the empirical U-quantile process, we are interested in linear combination of U-quantiles, which can be expressed as linear functional of the empirical U-quantile-process.

Definition 6.1.1. Let be $p_1, \ldots, p_d \in I$, $b_1, \ldots, b_d \in \mathbb{R}$ and J a bounded function that is continuous a.e. and vanishes outside of I. We call a statistic of the form

$$T_n = T\left(U_n^{-1}\right) := \int_I J(p)\, U_n^{-1}(p) dp + \sum_{j=1}^{d} b_j U_n^{-1}(p_j)$$
$$= \sum_{i=1}^{\frac{n(n-1)}{2}} \int_{\frac{2(i-1)}{n(n-1)}}^{\frac{2i}{n(n-1)}} J(t)\, dt \cdot U_n^{-1}\left(\frac{2i}{n(n-1)}\right) + \sum_{j=1}^{d} b_j U_n^{-1}(p_j)$$

generalized linear statistic (GL-statistic).

This generalization of L-statistics was introduced by Serfling [85]. U-statistics, U-quantiles and L-statistics can be written as GL-statistics (though this might be somewhat artificial). For a U-statistic, just take $h(x,y,t) = \mathbb{1}_{\{g(x,y)\le t\}}$ and $J = 1$ (this only works if we can consider the interval $I = [0,1]$). The following example shows how to deal with an ordinary L-statistic.

Example 6.1.2. Let $h(x,y,t) := \frac{1}{2}\left(\mathbb{1}_{\{x\le t\}} + \mathbb{1}_{\{y\le t\}}\right)$, $p_1 = 0.25$, $p_2 = 0.75$, $b_1 = -1$, $b_2 = 1$, and $J = 0$. Then a short calculation shows that the related GL-statistic is

$$T_n = F_n^{-1}(0.75) - F_n^{-1}(0.25),$$

where F_n^{-1} denotes the empirical sample quantile function. This is the well-known inter quartile distance, a robust estimator of scale with 25% breakdown point.

Example 6.1.3. Let $h(x,y,t) := \frac{1}{2}\left(\mathbb{1}_{\{x\le t\}} + \mathbb{1}_{\{y\le t\}}\right)$ and $J(x) = \mathbb{1}_{\{x\in[0.25,0.75]\}}$. This leads to the 25%-trimmed mean as a GL-statistic.

Example 6.1.4. Let $h(x,y,t) := \mathbb{1}_{\left\{\frac{1}{2}(x-y)^2\le t\right\}}$, $p_1 = 0.75$, $b_1 = 0.25$ and $J(x) = \mathbb{1}_{\{x\in[0,0.75]\}}$. The related GL-statistic is called winsorized variance, a robust estimator of scale with 13% breakdown point.

The uniform variation condition also holds for this example, because $h(x,y,t) = \mathbb{1}_{\left\{\frac{1}{2}(x-y)^2\le t\right\}} = \mathbb{1}_{\left\{|x-y|\le\sqrt{2t}\right\}}$ and this is the kernel function of Example 5.1.3.

6.2 On the Continuity of the Empirical U-Process

Lemma 6.2.1. *Let be F a non-decreasing function, $c, l > 0$ constants and $[C_1, C_2] \subset \mathbb{R}$. If for all $t, t' \in [C_1, C_2]$ with $|t - t'| \le l + 2c$*

$$|F(t) - F(t') - (t - t')| \le c,$$

then for all $p, p' \in \mathbb{R}$ with $|p - p'| \le l$ and $F^{-1}(p), F^{-1}(p') \in (C_1 + 2c + l, C_2 - 2c - l)$

$$|F^{-1}(p) - F^{-1}(p') - (p - p')| \le c$$

where $F^{-1}(p) := \inf\left\{t \middle| F(t) \ge p\right\}$ is the generalized inverse.

Proof. Without loss of generality we assume that $p < p'$. Let be $\epsilon \in (0, c)$. By our assumptions

$$F\left(F^{-1}(p) + (p' - p) + c + \epsilon\right) \geq F\left(F^{-1}(p) + \epsilon\right) + (p' - p) + c - c \\ \geq p + (p' - p) = p'.$$

By the definition of F^{-1}, it follows that

$$F^{-1}(p') = \inf\left\{t \middle| F(t) \geq p'\right\} \leq F^{-1}(p) + (p' - p) + c + \epsilon.$$

So taking the limit $\epsilon \to 0$, we obtain

$$F^{-1}(p') \leq F^{-1}(p) + (p' - p) + c.$$

On the other hand

$$F\left(F^{-1}(p) + (p' - p) - c - \epsilon\right) \leq F\left(F^{-1}(p) - \epsilon\right) + (p' - p) - c + c \\ \leq p + (p' - p) = p'.$$

So we have that

$$F^{-1}(p') \geq F^{-1}(p) + (p' - p) - c - \epsilon,$$

and hence $F^{-1}(p') \geq F^{-1}(p) + (p' - p) - c$. Combining the upper and lower inequality for $F^{-1}(p')$, we conclude that $|F^{-1}(p) - F^{-1}(p') - (p - p')| \leq c$. □

Lemma 6.2.2. *Let h be a kernel function that satisfies the uniform variation condition such that U differentiable on an interval $[C_1, C_2]$ with $0 < \inf_{t \in [C_1, C_2]} u(t) \leq \sup_{t \in [C_1, C_2]} u(t) < \infty$ ($u(t) = U'(t)$) and*

$$\sup_{t, t' \in [C_1, C_2]: \ |t - t'| \leq x} |U(t) - U(t') - u(t)(t - t')| = O\left(x^{\frac{5}{4}}\right)$$

and one of the following two mixing conditions is satisfied:

1. *$(X_n)_{n \in \mathbb{N}}$ is strongly mixing with mixing coefficients $\alpha(n) = O(n^{-\alpha})$ for $\alpha \geq 8$ and $E|X_i|^\rho < \infty$ for a $\rho > \frac{1}{4}$.*
2. *$(X_n)_{n \in \mathbb{N}}$ is near epoch dependent on an absolutely regular process with mixing coefficients $\beta(n) = O(n^{-\beta})$ for $\beta \geq 8$ with approximation constants $a_n = O(n^{-a})$ for $a = \max\{\beta + 3, 12\}$.*

Then for any constant $C > 0$

$$\sup_{\substack{t, t' \in [C_1, C_2]: \\ |t - t'| \leq C\sqrt{\frac{\log \log n}{n}}}} |U_n(t) - U_n(t') - u(t)(t - t')| = o(n^{-\frac{1}{2} - \frac{\gamma}{8}} \log n)$$

with γ as in Lemma 4.2.1 respectively 4.2.2.

Proof. As a consequence of the differentiability assumption and $\gamma < 1$

$$\sup_{\substack{t,t' \in [C_1, C_2]: \\ |t-t'| \le C\sqrt{\frac{\log\log n}{n}}}} |U(t) - U(t') - u(t)(t-t')| = o(n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n),$$

so it suffices to show that

$$K_n := \sup_{\substack{t,t' \in [C_1, C_2]: \\ |t-t'| \le C\sqrt{\frac{\log\log n}{n}}}} |U_n(t) - U_n(t') - (U(t) - U(t'))| = o(n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n).$$

For $l \in \mathbb{N}$ choose $t_{1,l}, \dots, t_{k-1,l}$ with $k = k_l = O\left(\sqrt{\frac{2^l}{\log l}}\right)$, $C_1 = t_{0,l} < t_{1,l} < \dots < t_{k-1,l} < t_{k,l} = C_2$ and $\sqrt{\frac{\log l}{2^l}} \le U(t_{r,l}) - U(t_{r-1,l}) \le 2\sqrt{\frac{\log l}{2^l}}$. Clearly

$$\begin{aligned} K_n &\le 2 \max_{r=1,\dots,k} \sup_{t,t' \in [t_{r-1,l}, t_{r,l}]} |U_n(t) - U_n(t') - (U(t) - U(t'))| \\ &\le 4 \max_{r=1,\dots,k} \sup_{t \in [t_{r-1,l}, t_{r,l}]} |U_n(t) - U_n(t_{r-1,l}) - (U(t) - U(t_{r-1,l}))| . \end{aligned}$$

Now choose $m = m_l \in \mathbb{N}$ and for $r = 1, \dots, k$ and $r^\star = 1, \dots, m-1$ real numbers $t^\star_{r^\star,r,l}$, such that $t_{r-1,l} = t^\star_{0,r,l} < t^\star_{1,r,l} < \dots < t^\star_{m-1,r,l} < t^\star_{m,r,l} = t_{r,l}$ and $2^{-(\frac{1}{2}+\frac{\gamma}{8})l} \le U(t^\star_{r^\star,r,l}) - U(t^\star_{r^\star-1,r,l}) \le 2 \cdot 2^{-(\frac{1}{2}+\frac{\gamma}{8})l}$. As U_n and U are non-decreasing, we have for $t \in (t^\star_{r^\star-1,r,l}, t^\star_{r^\star,r,l})$ and $n = 2^l, \dots, 2^{l+1}-1$

$$\begin{aligned} &|U_n(t) - U_n(t_{r-1,l}) - (U(t) - U(t_{r-1,l}))| \\ &\le \max\left\{ \left|U_n(t^\star_{r^\star,r,l}) - U_n(t_{r-1,l}) - (U(t) - U(t_{r-1,l}))\right| ,\right. \\ &\qquad \left. \left|U_n(t^\star_{r^\star-1,r,l}) - U_n(t_{r-1,l}) - (U(t) - U(t_{r-1,l}))\right| \right\} \\ &\le \max\left\{ \left|U_n(t^\star_{r^\star,r,l}) - U_n(t_{r-1,l}) - (U(t^\star_{r^\star,r,l}) - U(t_{r-1,l}))\right| ,\right. \\ &\qquad \left. \left|U_n(t^\star_{r^\star-1,r,l}) - U_n(t_{r-1,l}) - (U(t^\star_{r^\star-1,r,l}) - U(t_{r-1,l}))\right| \right\} + |U(t^\star_{r^\star,r,l}) - U(t^\star_{r^\star-1,r,l})|, \end{aligned}$$

and consequently

$$\begin{aligned} K_n &\le 4 \max_{r=1,\dots,k} \max_{r^\star=1,\dots,m} \left|U_n(t^\star_{r^\star,r,l}) - U_n(t_{r-1,l}) - (U(t^\star_{r^\star,r,l}) - U(t_{r-1,l}))\right| \\ &\qquad + 4 \max_{r=1,\dots,k} \max_{r^\star=1,\dots,m} |U(t^\star_{r^\star,r,l}) - U(t^\star_{r^\star-1,r,l})| \\ &\le 8 \max_{r=1,\dots,k} \max_{r^\star=1,\dots,m} \left| \frac{1}{n} \sum_{1 \le i \le n} h_1(X_i, t^\star_{r^\star,r,l}) - \frac{1}{n} \sum_{1 \le i \le n} h_1(X_i, t_{r-1,l}) \right| \\ &\quad + 4 \max_{r=1,\dots,k} \max_{r^\star=1,\dots,m} \left| \frac{2}{n(n-1)} \left(\sum_{1 \le i < j \le n} h_2(X_i, X_j, t^\star_{r^\star,r,l}) - \sum_{1 \le i < j \le n} h_2(X_i, X_j, t_{r-1,l}) \right) \right| \\ &\qquad + 4 \max_{r=1,\dots,k} \max_{r^\star=1,\dots,m} |U(t^\star_{r^\star,r,l}) - U(t^\star_{r^\star-1,r,l})|. \end{aligned}$$

We will treat these three summands seperately. From our choice of $t^\star_{r^\star,r,l}$, we obtain

$$\max_{r=1,\dots,k}\max_{r^\star=1,\dots,m}|U(t^\star_{r^\star,r,l})-U(t^\star_{r^\star-1,r,l})|\leq 2\cdot 2^{-(\frac{1}{2}+\frac{\gamma}{8})l}=o(n^{-\frac{1}{2}-\frac{\gamma}{8}}\log n).$$

With the help of Proposition 4.2.3, it follows for the degenerate part that

$$\max_{r=1,\dots,k}\max_{r^\star=1,\dots,m}\left|\frac{2}{n(n-1)}\left(\sum_{1\leq i<j\leq n}h_2(X_i,X_j,t^\star_{r^\star,r,l})-\sum_{1\leq i<j\leq n}h_2(X_i,X_j,t_{r-1,l})\right)\right|$$
$$\leq\frac{4}{n(n-1)}\sup_{t\in\mathbb{R}}\left|\sum_{1\leq i<j\leq n}h_2\left(X_i,X_j,t\right)\right|=o\left(n^{-\frac{1}{2}-\frac{\gamma}{8}}\right).$$

Furthermore, we have for the linear part by Lemma 4.2.1 respecitvely 4.2.2 and Corollary 1 of Móricz [69] (which gives moment bounds for the maximum other multidimensional partial sums)

$$E\left[\left(\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=1,\dots,k}\max_{r^\star=1,\dots,m}\left|\sum_{i=1}^n h_1(X_i,t^\star_{r^\star-1,r,l})-\sum_{i=1}^n h_1(X_i,t_{r-1,l})\right|\right)^4\right]$$
$$\leq\sum_{r=1}^k E\left[\left(\max_{n=2^l,\dots,2^{l+1}-1}\max_{m_1=1,\dots,m}\left|\sum_{i=1}^n\sum_{r^\star=1}^{m_1}\left(h_1(X_i,t^\star_{r^\star,r,l})-h_1(X_i,t^\star_{r^\star-1,r,l})\right)\right|\right)^4\right]$$
$$\leq Ck2^{2l}l^2\left(\sqrt{\frac{\log l}{2^l}}\right)^{1+\gamma}\leq Cl^2(\log l)^{\frac{\gamma}{2}}2^{(2-\frac{\gamma}{2})l},$$

as $k=O(\sqrt{\frac{2^l}{\log l}})$. So we can conclude that for any $\epsilon>0$

$$\sum_{l=1}^\infty P\left[\max_{n=2^l,\dots,2^{l+1}-1}\max_{r=1,\dots,k}\max_{r^\star=1,\dots,m}\left|\sum_{i=1}^n\left(h_1(X_i,t^\star_{r^\star-1,r,l})-h_1(X_i,t_{r-1,l})\right)\right|\geq\epsilon 2^{(\frac{1}{2}-\frac{\gamma}{8})l}l\right]$$
$$\leq C\sum_{l=1}^\infty\frac{2^{(2-\frac{\gamma}{2})l}l^2(\log l)^{\frac{\gamma}{2}}}{\epsilon^4 l^4 2^{(2-\frac{\gamma}{2})l}}=C\sum_{l=1}^\infty\frac{(\log l)^{\frac{\gamma}{2}}}{l^2}<\infty.$$

With the Borel Cantelli lemma, it follows that

$$\max_{r=1,\dots,k}\max_{r^\star=1,\dots,m}\left|\sum_{1\leq i\leq n}h_1(X_i,t^\star_{r^\star,r,l})-\sum_{1\leq i\leq n}h_1(X_i,t_{r-1,l})\right|=o(n^{\frac{1}{2}-\frac{\gamma}{8}}\log n)$$

almost surely and finally

$$\max_{r=1,\dots,k}\max_{r^\star=1,\dots,m}\left|\frac{1}{n}\sum_{1\leq i\leq n}h_1(X_i,t^\star_{r^\star,r,l})-\frac{1}{n}\sum_{1\leq i\leq n}h_1(X_i,t_{r-1,l})\right|=o(n^{-\frac{1}{2}-\frac{\gamma}{8}}\log n),$$

so all three summands converge with the required rate and the proof is completed. □

6.3 Strong Invariance Principles

Recall that the remainder term in the generalized Bahadur representation is defined as

$$R_n(p) = U_n^{-1}(p) - t_p - \frac{p - U_n(t_p)}{u(t_p)}$$

and that we write $t_p := U^{-1}(p)$. We set $U_0^{-1}(p) := 0$ as it is not possible to find a generalized inverse of $U_0 = 0$.

Theorem 6.3.1. *Let h be a kernel function that satisfies the uniform variation condition such that U differentiable on an interval $[C_1, C_2]$ with $0 < \inf_{t\in[C_1,C_2]} u(t) \leq \sup_{t\in[C_1,C_2]} u(t) < \infty$ ($u(t) = U'(t)$) and*

$$\sup_{t,t'\in[C_1,C_2]:\ |t-t'|\leq x} |U(t) - U(t') - u(t)(t-t')| = O\left(x^{\frac{5}{4}}\right)$$

and one of the following two mixing conditions is satisfied:

1. *$(X_n)_{n\in\mathbb{N}}$ is strongly mixing with mixing coefficients $\alpha(n) = O(n^{-\alpha})$ for $\alpha \geq 8$ and $E|X_i|^\rho < \infty$ for a $\rho > \frac{1}{4}$.*
2. *$(X_n)_{n\in\mathbb{N}}$ is near epoch dependent on an absolutely regular process with mixing coefficients $\beta(n) = O(n^{-\beta})$ for $\beta \geq 8$ with approximation constants $a_n = O(n^{-a})$ for $a = \max\{\beta + 3, 12\}$.*

Then

$$\sup_{\substack{p\in I\\ s\in[0,1]}} \frac{\lfloor ns\rfloor}{\sqrt{n}} |R_{\lfloor ns\rfloor}(p)| = o(n^{-\frac{\gamma}{8}} \log n)$$

almost surely with $I = [\tilde{C}_1, \tilde{C}_2]$ with $U(C_1) < \tilde{C}_1 < \tilde{C}_2 < U(C_2)$, $\gamma := \frac{\alpha-2}{\alpha}$ (under strong mixing) respectively $\gamma := \frac{\beta-3}{\beta+1}$ (under near epoch dependence on an absolutely regular process).

Note that for a fast decay of the mixing coefficients, the rate becomes close to $n^{-\frac{1}{8}}$, while the optimal rate for sample quantile process of independent data is $n^{-\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log\log n)^{\frac{1}{4}}$, see Kiefer [59].

Proof. To simplify the notation, we will, without loss of generality, assume that $U(p) = p = t_p$ on the interval I. In the general case, one has to change the function $h(x, y, t)$ to $h(x, y, U^{-1}(t))$, as $Eh(X, Y, U^{-1}(p)) = U(U^{-1}(p)) = p$. For related empirical U-process $U_n \circ U^{-1}$, we have

$$\begin{aligned} R_n(p) &= U_n^{-1}(p) - U^{-1}(p) - \frac{p - U_n(U^{-1}(p))}{u(t_p)} \\ &= \frac{1}{u(t_p)}\left((U_n \circ U^{-1})^{-1}(p) - p - (p - U_n \circ U^{-1}(p))\right) + o((U_n^{-1}(p) - U^{-1}(p))^{\frac{5}{4}}), \end{aligned}$$

so $R_n(p)$ is only blown up by a constant because of this transformation. If $U(p) = p = t_p$, then we can write $R_n(p)$ as

$$\begin{aligned} R_n(p) &= U_n^{-1}(p) - t_p + U_n(t_p) - p \\ &= \left(U_n^{-1}(p) - U_n^{-1}(U_n(t_p)) + U_n(t_p) - p\right) + \left(U_n^{-1}(U_n(t_p)) - t_p\right). \end{aligned}$$

Applying Lemma 6.2.2 and Lemma 6.2.1 with $F = U_n$, $c = n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n$ and $l = C\sqrt{\frac{\log\log n}{n}}$, we obtain

$$\sup_{\substack{p,p' \in I: \\ |p-p'| \leq C\sqrt{\frac{\log\log n}{n}}}} \left|U_n^{-1}(p) - U_n^{-1}(p') - (p - p')\right| = o(n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n)$$

almost surely. By Corollary 4.3.3 we have that $\sup_{t\in[C_1,C_2]} (U_n(t_p) - p) \leq C\sqrt{\frac{\log\log n}{n}}$ almost surely, it follows that

$$\begin{aligned} &\sup_{p\in I} \left|U_n^{-1}(p) - U_n^{-1}(U_n(t_p)) + U_n(t_p) - p\right| \\ &\qquad \leq \sup_{\substack{p,p' \in I: \\ |p-p'| \leq C\sqrt{\frac{\log\log n}{n}}}} \left|U_n^{-1}(p) - U_n^{-1}(p') - (p - p')\right| = o(n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n) \end{aligned}$$

almost surely. It remains to show the convergence of $U_n^{-1}(U_n(t_p)) - t_p$. For every $\epsilon > 0$ by the definition of the generalized inverse, $U_n^{-1}(U_n(t_p)) - t_p > \epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n$ only if $U_n(t_p + \epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n) < U_n(t_p)$ and $U_n^{-1}(U_n(t_p)) - t_p \leq -\epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n$ only if $U_n(t_p - \epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n) \geq U_n(t_p)$. So we can conclude that

$$\begin{aligned} &P\left[\sup_{p\in I} |U_n^{-1}(U_n(t_p)) - t_p| > \epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n \ \text{ i.o.}\right] \\ &\leq P\left[\sup_{t\in[C_1, C_2 - \epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n]} U_n(t + \epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n) - U_n(t) \leq 0 \ \text{ i.o.}\right] \\ &\leq P\left[\sup_{\substack{t,t' \in [C_1,C_2] \\ |t-t'| = \epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n}} |U_n(t) - U_n(t') - (U(t) - U(t'))| \geq |U(t) - U(t')| \ \text{ i.o.}\right] \\ &\leq P\left[\sup_{\substack{t,t' \in [C_1,C_2] \\ |t-t'| \leq \epsilon n^{-\frac{1}{2}-\frac{\gamma}{8}} \log n}} |U_n(t) - U_n(t') - (U(t) - U(t'))| \geq \frac{\epsilon \log n}{n^{\frac{1}{2}+\frac{\gamma}{8}} \inf_{t\in[C_1,C_2]} u(t)} \ \text{ i.o.}\right] \\ &= 0, \end{aligned}$$

where the last line is a consequence of Lemma 6.2.2. Now we have proved that $\sup_{p\in I}|R_n(p)| = o(n^{-\frac{1}{2}-\frac{\gamma}{8}}\log n)$, and can finally conclude that

$$\begin{aligned}
&\frac{n^{\frac{\gamma}{8}}}{\log n}\sup_{\substack{p\in I\\ s\in[0,1]}}\frac{\lfloor ns\rfloor}{\sqrt{n}}|R_{\lfloor ns\rfloor}(p)|\\
&\leq \sup_{n'\leq\sqrt{n}}(\frac{n'}{n})^{\frac{1}{2}-\frac{\gamma}{8}}\frac{\log n'}{\log n}\frac{n'^{\frac{1}{2}+\frac{\gamma}{8}}}{\log n'}\sup_{p\in I}|R_{n'}(p)| + \sup_{\sqrt{n}\leq n'\leq n}\frac{n'^{\frac{1}{2}+\frac{\gamma}{8}}}{\log n'}\sup_{p\in I}|R_{n'}(p)|\\
&\qquad\leq Cn^{-\frac{1}{4}+\frac{\gamma}{16}}\sup_{n'\in\mathbb{N}}\sup_{p\in I}|R_{n'}(p)| + \sup_{n'\geq\sqrt{n}}\frac{n'^{\frac{1}{2}+\frac{\gamma}{8}}}{\log n'}\sup_{p\in I}|R_{n'}(p)| \to 0.
\end{aligned}$$

□

Using the Bahadur representation, we can deduce the asymptotic behaviour of the empirical U-quantile process from Theorem 4.3.1.

Theorem 6.3.2. *Let h be a kernel function that satisfies the uniform variation condition such that U differentiable on an interval $[C_1, C_2]$ with $0 < \inf_{t\in[C_1,C_2]} u(t) \leq \sup_{t\in[C_1,C_2]} u(t) < \infty$ ($u(t) = U'(t)$) and*

$$\sup_{t,t'\in[C_1,C_2]:\ \ |t-t'|\leq x}|U(t) - U(t') - u(t)(t-t')| = O\left(x^{\frac{5}{4}}\right)$$

and one of the following two mixing conditions is satisfied:

1. *$(X_n)_{n\in\mathbb{N}}$ is strongly mixing with mixing coefficients $\alpha(n) = O(n^{-\alpha})$ for $\alpha \geq 8$ and $E|X_i|^\rho < \infty$ for a $\rho > \frac{1}{4}$.*
2. *$(X_n)_{n\in\mathbb{N}}$ is near epoch dependent on an absolutely regular process with mixing coefficients $\beta(n) = O(n^{-\beta})$ for $\beta \geq 8$ with approximation constants $a_n = O(n^{-a})$ for $a = \max\{\beta + 3, 12\}$.*

Then there exists a centered Gaussian process $(K'(p,s))_{p\in I,s\in\mathbb{R}}$ (after enlarging the probability space if necessary), where I is the interval introduced in Theorem 6.3.1, with covariance function

$$EK'(p,s)K'(p',s') = \min\{s,s'\}\,\frac{1}{u(t_p)u(t_{p'})}\Gamma(t_p,t_{p'})$$

such that

$$\sup_{\substack{p\in I\\ s\in[0,1]}}\frac{1}{\sqrt{n}}\left|\lfloor ns\rfloor(U^{-1}_{\lfloor ns\rfloor}(p) - t_p) - K'(p,ns)\right| = O(\log^{-\frac{1}{3840}} n).$$

Proof. Define $K'(p,s) := -\frac{1}{u(t_p)}K(t_p,s)$, there K is the Gaussian process introduced in Theorem 4.3.1. K' is then a Gaussian process with covariance function

$$EK'(p,s)K'(p',s') = \min\{s,s'\}\frac{1}{u(t_p)u(t_{p'})}\Gamma(t_p,t_{p'})$$

and by Theorem 4.3.1 and Theorem 6.3.1

$$\begin{aligned}
&\sup_{\substack{p\in I\\ s\in[0,1]}} \frac{1}{\sqrt{n}}\left|\lfloor ns\rfloor(U^{-1}_{\lfloor ns\rfloor}(p)-t_p)-K'(p,ns)\right| \\
\le &\sup_{\substack{p\in I\\ s\in[0,1]}} \frac{1}{\sqrt{n}}\left|\lfloor ns\rfloor(U^{-1}_{\lfloor ns\rfloor}(p)-t_p-\frac{p-U_n(t_p)}{u(t_p)}\right| \\
&+\sup_{\substack{p\in I\\ s\in[0,1]}} \frac{1}{\sqrt{n}}\frac{1}{u(t_p)}\left|\lfloor ns\rfloor(U_{\lfloor ns\rfloor}(t_p)-p)-K(t_p,ns)\right| \\
\le &\sup_{\substack{p\in I\\ s\in[0,1]}} \frac{\lfloor ns\rfloor}{\sqrt{n}}|R_{\lfloor ns\rfloor}(p)| + \frac{1}{\inf_{p\in I}u(t_p)}\sup_{\substack{p\in I\\ s\in[0,1]}}\frac{1}{\sqrt{n}}\left|\lfloor ns\rfloor(U_{\lfloor ns\rfloor}(t_p)-p)-K(t_p,ns)\right| \\
=&O(\log^{-\frac{1}{3840}} n)
\end{aligned}$$

almost surely.

□

K' is a Gaussian process with independent increments in s direction, so we have the following consequences:

Corollary 6.3.3. *Under the assumptions of Theorem 6.3.2*

$$\left(\frac{\lfloor ns\rfloor}{\sqrt{n}}(U^{-1}_{\lfloor ns\rfloor}(p)-t_p)\right)_{t\in\mathbb{R},s\in[0,1]}$$

converges weakly in the space $D(\mathbb{R}\times[0,1])$ of càdlàg functions (equipped with the supremum norm) to the centered Gaussian process $(K'(p,s))_{p\in I,s\in\mathbb{R}}$ introduced in Theorem 6.3.2.

Corollary 6.3.4. *Under the assumptions of Theorem 6.3.2, the sequence*

$$\left(\left(\frac{\lfloor ns\rfloor}{\sqrt{2n\log\log n}}(U^{-1}_{\lfloor ns\rfloor}(p)-t_p)\right)_{p\in I,s\in[0,1]}\right)_{n\in\mathbb{N}}$$

is almost surely relatively compact in the space $D(\mathbb{R}\times[0,1])$ of càdlàg functions (equipped with the supremum norm) and the limit set is the unit ball $\mathcal{U}_{K'}$ of the reproducing kernel Hilbert space $\mathcal{K}'$ associated with the covariance function of the process K'.

As GL-statistics are linear functionals of the empirical U-quantile process, we get an approximation for T_n:

Theorem 6.3.5. *Let be $p_1, \ldots, p_d \in I$ and J a bounded function that is continuous a.e. and vanishes outside of I. Let h be a kernel function that satisfies the uniform variation condition such that U differentiable on an interval $[C_1, C_2]$ with $0 < \inf_{t\in[C_1,C_2]} u(t) \le \sup_{t\in[C_1,C_2]} u(t) < \infty$ ($u(t) = U'(t)$) and*

$$\sup_{t,t'\in[C_1,C_2]:\quad |t-t'|\le x} |U(t) - U(t') - u(t)(t-t')| = O\left(x^{\frac{5}{4}}\right).$$

and one of the following two mixing conditions is satisfied:

1. *$(X_n)_{n\in\mathbb{N}}$ is strongly mixing with mixing coefficients $\alpha(n) = O(n^{-\alpha})$ for $\alpha \ge 8$ and $E|X_i|^\rho < \infty$ for a $\rho > \frac{1}{4}$.*

2. *$(X_n)_{n\in\mathbb{N}}$ is near epoch dependent on an absolutely regular process with mixing coefficients $\beta(n) = O(n^{-\beta})$ for $\beta \ge 8$ with approximation constants $a_n = O(n^{-a})$ for $a = \max\{\beta+3, 12\}$,*

then there exists (after enlarging the probability space if necessary) a Brownian motion B, such that for T_n defined in Definition 6.1.1 and

$$\sigma^2 = \int_{\tilde{C}_1}^{\tilde{C}_2} \int_{\tilde{C}_1}^{\tilde{C}_2} \frac{\Gamma(t_p, t_q)}{u(t_p)u(t_q)} J(p)J(q)dpdq \\ + 2\sum_{j=1}^d b_j \int_{\tilde{C}_1}^{\tilde{C}_2} \frac{\Gamma(t_{p_j}, t_p)}{u(t_{p_j})u(t_p)} J(p)dp + \sum_{i,j=1}^d b_i b_j \frac{\Gamma(t_{p_i}, t_{p_j})}{u(t_{p_i})u(t_{p_j})}$$

we have that

$$\sup_{s\in[0,1]} \frac{1}{\sqrt{n}} \left| \lfloor ns \rfloor (T_{\lfloor ns \rfloor} - T(U^{-1})) - \sigma B(ns) \right| = O(\log^{-\frac{1}{3840}} n)$$

almost surely.

Proof. If $\sigma^2 > 0$, set

$$B(s) = \frac{1}{\sigma} T(K'(\cdot, s)) = \int_I J(p)K'(p,s)dp + \sum_{j=1}^d b_j U_n(p_j).$$

In the case $\sigma^2 = 0$, B may be an arbitrary Brownian motion. As J is a bounded function, T is a linear and Lipschitz continuous functional (with respect to the supremum norm), so

$$\sup_{s\in[0,1]} \frac{1}{\sqrt{n}} \left| \lfloor ns \rfloor (T(U^{-1}_{\lfloor ns \rfloor}) - T(U^{-1})) - \sigma B(ns) \right|$$
$$= \sup_{s\in[0,1]} \frac{1}{\sqrt{n}} \left| T\left(\lfloor ns \rfloor (U^{-1}_{\lfloor ns \rfloor} - U^{-1}) - K'(\cdot, ns) \right) \right|$$
$$\leq C \sup_{\substack{p\in I \\ s\in[0,1]}} \frac{1}{\sqrt{n}} \left| \lfloor ns \rfloor (U^{-1}_{\lfloor ns \rfloor}(p) - t_p) - K'(p, ns) \right| = O(\log^{-\frac{1}{3840}} n).$$

It remains to show that B is a Brownian motion. Cleary, $EB(s) = 0$ for every $s \geq 0$. By the linearity of T, B is a Gaussian process with stationary independent increments. Furthermore

$$E[B^2(s)] = \frac{1}{\sigma^2} \int_{\tilde{C}_1}^{\tilde{C}_2} \int_{\tilde{C}_1}^{\tilde{C}_2} \frac{E[K(t_p, s)K(t_q, s)]}{u(t_p)u(t_q)} J(p)J(q)dpdq$$
$$+ \frac{1}{\sigma^2} 2 \sum_{j=1}^{d} b_j \int_{\tilde{C}_1}^{\tilde{C}_2} \frac{E[K(t_{p_j}, s)K(t_q, s)]}{u(t_{p_j})u(t_p)} J(p)dp + \frac{1}{\sigma^2} \sum_{i,j=1}^{d} b_i b_j \frac{E[K(t_{p_i}, s)K(t_{p_j}, s)]}{u(t_{p_i})u(t_{p_j})}$$
$$= s.$$

□

By the well-known properties of Brownian motions, we have:

Corollary 6.3.6. *Let be $p_1, \ldots, p_d \in I$ and J a bounded function. Under the assumptions of Theorem 6.3.5 for T_n defined in Definition 6.1.1:*

$$\frac{\lfloor ns \rfloor}{\sqrt{n}} (T_{\lfloor ns \rfloor} - T(U^{-1}))$$

converges to the Brownian motion $\sigma B(s)$ with σ^2 as in Theorem 6.3.5.

Corollary 6.3.7. *Let be $p_1, \ldots, p_d \in I$ and J a bounded function. Under the assumptions of Theorem 6.3.5 for T_n defined in Definition 6.1.1:*

$$\left(\frac{\lfloor ns \rfloor}{\sqrt{2n \log \log n}} (T_{\lfloor ns \rfloor} - T(U^{-1}))_{s\in[0,1]} \right)_{n\in\mathbb{N}}$$

is almost surely relatively compact in the space $D[0,1]$ (equipped with the supremum norm) and the limit set is

$$\left\{ f : [0,1] \to \mathbb{R} \middle| f(0) = 0, \int_0^1 f'^2(s)ds \leq \sigma^2 \right\}$$

with σ^2 as in Theorem 6.3.5.

Bibliography

[1] J. Aaronson, R. Burton, H. Dehling, D. Gilat, T. Hill, B. Weiss, Strong laws for L- and U-statistcs, *Trans. Amer. Math. Soc.* **348** (1996) 2845-2866.

[2] D.W.K. Andrews, Non-strong mixing autoregressive processes, *J. Appl. Probab.* **21** (1984) 930-934.

[3] M.A. Arcones, The law of the iterated logarithm for U-processes, *J. Multivariate Anal.* **47** (1993) 139-151.

[4] M.A. Arcones, On the law of the iterated logarithm for Gaussian processes, *J. Theoret. Probab.* **8** (1995) 877-903.

[5] M.A. Arcones, The Bahadur-Kiefer representation for U-quantiles, *Ann. Stat.* **24** (1996) 1400-1422.

[6] M.A. Arcones, E. Giné, Limit Theorems for U-processes, *Ann. Prob.* **21** (1993) 1494-1542.

[7] M.A. Arcones, E. Giné, On the law of the iterated logarithm for canonical U-statistics and processes, *Stochastic Process. Appl.* **58** (1995) 217-245.

[8] M.A. Arcones, B. Yu, Central limit theorem for empirical and U-processes of stationary mixing sequences, *J. Theoret. Probab.* **7** (1997) 47-53.

[9] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **3** (1950) 337-404.

[10] G.J. Babu, K. Singh, On deviations between empirical and quantile processes for mixing random variables, *J. Multivariate Anal.* **8** (1978) 532-549.

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

[12] U. Bentzien, *Der Rätselkasten*, Anaconda, Köln (2007).

[13] H.C.P. Berbee, Random walks with stationary increments and renewal theory, *Mathematisch Centrum* **118** (1979).

[14] I. Berkes, W. Philipp, An almost sure invariance principle for the empirical distribution function of mixing random variables, *Probab. Theory Related Fields* **41** (1977) 115-137.

[15] I. Berkes, W. Philipp, Approximation theorems for independent and weakly dependent random vectors, *Ann. Prob.* **7** (1979) 29-54.

[16] P. Billingsley, *Ergodic theory and information*, Robert E. Krieger Publishing Company, Huntington (1978).

[17] S. Borovkova, Weak convergence of the empirical process of U-statistics structure for dependent observations, *Theory Stoch. Process.* **2** (1996) 114-123.

[18] S. Borovkova, R. Burton, H. Dehling, Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation, *Trans. Amer. Math. Soc.* **353** (2001) 4261–4318.

[19] R.C. Bradley, Approximation theorems for strongly mixing random variables, *Michigan Math. J.* **30** (1983) 69-81.

[20] R.C. Bradley, *Introduction to strong mixing conditions*, volume 1-3, Kendrick Press, Heber City (2007).

[21] W. Bryc, On the approximation theorem of I. Berkes and W. Philipp, *Demonstratio Math.* **15** (1982) 807-816.

[22] J. Choudhury, R.J. Serfling, Generalized order statistics, Bahadur representations, and sequential nonparametric fixed-width confidence intervals, *J. Statist. Plann. Inference* **19** (1988) 269-282.

[23] J.-F. Coeurjolly, Bahadur representation of sample quantiles for functional of Gaussian dependent sequences under a minimal assumption, *Statist. Probab. Lett.* **78** (2008) 2485-2489.

[24] M. Csörgő, P. Révész Strong approximations of the quantile process, *Ann. Stat.* **4** (1978) 882-894.

[25] Yu.A. Davydov, The invariance principle for stationary processes, *Theory of Probab. Appl.* **15** (1970) 487-498.

[26] P.Deheuvels, D.M. Mason, A functional LIL approach to pointwise Bahadur-Kiefer theorems, *Probability in Banach spaces*, **8** 255-266, Progr. Probab., 30, Birkhäuser Boston, Boston, (1992).

[27] H. Dehling, A note on a theorem of Berkes and Philipp, *Z. Wahrsch. verw. Gebiete* **62** (1983) 39-42.

[28] H. Dehling, The functional law of the iterated logarithm for von Mises functionals and multiple Wiener integrals, *J. Multivariate Anal.* **28** (1989) 177-189.

[29] H. Dehling, M. Denker, W. Philipp, The almost sure invariance principle for the empirical process of U-statistic structure, *Annales de l'I.H.P.* **23** (1987) 121-134.

[30] H. Dehling, M. Denker, W. Philipp, Invariance principles for von Mises and U-statistics, *Z. Wahrsch. verw. Gebiete* **67** (1994) 139-167.

[31] H. Dehling, O.S. Sharipov, Marcinkiewicz-Zygmund strong laws for U-statistics of weakly dependent observations, *Statist. Probab. Lett.* **79** (2009) 2028-2036.

[32] H. Dehling, M. Wendler, Central limit theorem and the bootstrap for U-statistics of strongly mixing data, *J. Multivariate Anal.* **101** (2010) 126-137.

[33] H. Dehling, M. Wendler, Law of the iterated logarithm for U-statistics of weakly dependent observations, in: Berkes, Bradley, Dehling, Peligrad, Tichy (Eds): *Dependence in Probability, Analysis and Number Theory*, Kendrick Press, Heber City (2010).

[34] A. de Moivre,*The doctrine of chances*, London, (1738).

[35] M. Denker, G. Keller, Rigorous statistical procedures for data from dynamical systems, *J. Stat. Phys.* **44** (1986) 67-93.

[36] M.D Donsker, Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems, *Ann. Math. Statist.* **23** (1952) 277-281.

[37] H. Doss, R.D. Gill, An elementary approach to weak convergence for quantile processes, with applications to censored survival data, *J. Amer. Stat. Soc.* **87** (1992) 869-877.

[38] P. Doukhan, *Mixing*, Springer, New York, (1995).

[39] P. Doukhan, S. Louhichi, A new weak dependence condition and applications to moment inequalities, *Stoch. Processes and their Appl.* **84** (1999) 313-342.

[40] K. Dutta, P.K. Sen, On the Bahadur representation of sample quantiles in some stationary multivariate autoregressive processes, *J. Multivariate Anal.* **1** (1971) 186-198.

[41] H. Finkelstein, The law of the iterated logarithm for empirical distributions, *Ann. Math. Statist.* **42** (1971) 607-615.

[42] J.C. Geertsema, Sequential confidence intervals based on rank test, *Ann. Math. Stat.* **41** (1970) 1016-1026.

[43] J.K. Ghosh, A new proof of the Bahadur representation of quantiles and an application, *Ann. Math. Statist.* **42** (1971) 1957-1961.

[44] P. Grassberger, I. Procaccia, Characterization of strange attractors, *Phys. Rev. Lett.* **50** (1983) 346-349.

[45] P.R. Halmos, The theory of unbiased estimation, *Ann. Math. Stat.* **17** (1947) 34-43.

[46] B.E. Hansen, GARCH(1,1) processes are near epoch dependent, *Econom. Lett.* **36** (1991) 181-186.

[47] P. Hartman, A. Wintner, On the law of the iterated logarithm, *American Journal of Mathematics* **63** (1941) 169-176.

[48] C.H. Hesse, A Bahadur-type representation for empirical quantiles of a large class of stationary, possibly infinite-variance, linear processes, *Ann. Stat.* **18** (1990) 1188-1202.

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

[50] W. Hoeffding, The strong law of large numbers for U-statistics, *Statistics, Univ. North Carolina, Mimeo Series* (1961).

[51] J.L. Hodges, E.L. Lehmann, Estimates of location based on rank tests, *Ann. Math. Stat.* **34** (1963) 598-611.

[52] F. Hofbauer, G. Keller, Ergodic properties of invariant measures for piecewise monotonic transformations, *Math. Z.* **180** (1982) 119-142.

[53] T. Hsing, W.B. Wu, On weighted U-statistics for stationary processes, *Ann. Prob.* **32** (2004) 1600-1631.

[54] P.J. Huber, *Robust statistics*, Wiley, New York (1981).

[55] I.A. Ibragimov, Some limit theorems for stationary processes, *Theory Prob. Appl.* **7** (1962) 349-382.

[56] M. Iosifescu, The law of the iterated logarithm for a class of dependent random variables, *Theory Prob. Appl.* **13** (1968) 304-313.

[57] S. Kanagawa, K. Yoshihara, The almost sure invariance principles of degenerate U-statistics of degree two for stationary random variables, *Stoch. Processes and their Appl.* **49** (1994) 347-356.

[58] J. Kiefer, On Bahadur's representation of sample quantiles, *Ann. Math. Stat.* **38** (1967) 1323-1342.

[59] J. Kiefer, Deviations between the sample quantile process and the sample df, in: M.L. Puri (Ed): *Nonparametric Techniques in Statistical Inference* (1970).

[60] J. Kiefer, Skorohod embedding of multivariate RV's, and the Sample DF, *Probab. Theory Related Fields* **24** (1972) 1-35.

[61] A. Khintchine, Über das Gesetz der großen Zahlen, *Mathematische Annalen* **96** (1927) 152-168.

[62] V.S. Koroljuk, Y.V. Borovsikich, *Theory of U-Statistics*, Kluwer Academic Publishers, Dorbrecht-Boston-London, (1994).

[63] J. Kuelbs, R. Lepage, The law of the iterated logarithm for Brownian motion in a Banach space, *Trans. Amer. Math. Soc.* **185** (1973) 253-264.

[64] R. Kulik, Bahadur-Kiefer theory for sample quantiles of weakly dependent linear processes, *Bernoulli* **13** (2007) 1071-1090.

[65] R. Kulik, Optimal rates in the Bahadur-Kiefer representation for GARCH sequences, *preprint* arXiv: 0605283 (2008).

[66] C. Lévy-Leduc, H. Boistard, E. Moulines, M.S. Taqqu, V.A. Reisen, Asymptotic properties of U-processes under long-range dependence, *preprint* arXiv:0912.4688.

[67] C. Lévy-Leduc, H. Boistard, E. Moulines, M.S. Taqqu, V.A. Reisen, Robust estimation of scale and of the autocovariance function of Gaussian short and long-range dependent processes *preprint* arXiv:0912.4686.

[68] R.v. Mises, On the asymptotic distribution of differentiable statistical functions, *Ann. Math. Statistics* **18** (1947) 309-348.

[69] F. Móricz, A general moment inequality for the maximum of the rectangular partial sums of multiple series, *Acta Math. Hung.* **43** (1983) 337-346.

[70] D.W. Müller, On Glivenko-Cantelli convergence, *Probab. Theory Related Fields* **16** (1970) 195-210.

[71] D. Nolan, D. Pollard, Functional limit theorems for U-processes, *Ann. Prop.* **16** (1988) 1291-1298.

[72] H. Oodaira, On Strassen's version of the law of the iterated logarithm for Gaussian processes, *Z. Wahrsch. verw. Gebiete* **21** (1972) 289-299.

[73] H. Oodaira, K. Yoshihara, The law of the iterated logarithm for stationary processes satisfying mixing conditions, *Kodai Math. Sem. Rep.* **23** (1971) 311-334.

[74] W. Philipp, Das Gesetz vom iterierten Logarithmus für stark mischende stationäre Prozesse, *Z. Wahrsch. verw. Gebiete* **8** (1967) 204-209.

[75] W. Philipp, A functional law of the iterated logarithm for empirical functions of weakly dependent random variables, *Ann. Prob.* **5** (1977) 319-350.

[76] M. Kh. Reznik, The law of the iterated logarithm for some classes of stationary processes, *Theor. Probability Appl.* **8** (1968) 606-621.

[77] E. Rio, The functional law of the iterated logarithm for stationary, strongly mixing processes, *Ann. Prob.* **23** (1995) 1188-1203.

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

[79] P.J. Rousseeuw, C. Croux, Alternatives to the median absolute deviation, *J. Amer. Stat. Soc.* **88** (1993) 1273-1283.

[80] P.K. Sen, On the properties of U-statistics when the observations are not independent, *Cal. Statist. Assoc. Bull.* **12** (1963) 69-82.

[81] P.K. Sen, Asymptotic normality of sample quantiles for m-dependent processes, *Ann. Math. Statist.* **39** (1968) 1724-1730.

[82] P.K. Sen, Limiting behavior of regular functionals of empirical distributions for stationary $\star$-mixing processes, *Z. Wahrsch. verw. Gebiete* **25** (1972) 71-82.

[83] P.K. Sen, On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables, *J. Multivariate Anal.* **2** (1972) 77-95.

[84] R.J. Serfling, The law of the iterated logarithm for U-statistics and related von Mises statistics, *Ann. Math. Statist.* **42** (1971) 1794.

[85] R.J. Serfling, Generalized L-, M-, and R-statistics, *Ann. Prob.* **12** (1984) 76-86.

[86] B.W. Silverman, Convergence of a class of empirical distribution functions of dependent random variables, *Ann. Prob.* **11** (1983) 745-751.

[87] B. Solomyak, On the random series $\sum \pm\lambda^n$ (an Erdös problem), *Ann. Math.* **142** (1995) 611-625.

[88] S. Sun, The Bahadur representation for sample quantiles under weak dependence, *Statist. Probab. Letters* **76** (2006) 1238-1244.

[89] A.W. van der Vaart, J.A. Wellner, *Weak convergence of empirical processes with applications to statistics*, Springer, New York (1996).

[90] W. Vervaat, Functional central limit theorems for processes with positive drift and their inverses, *Z. Wahrsch. verw. Gebiete* **23** (1972) 245-253.

[91] V.A. Volkonskii, Y.A. Rozanov, Some limit theorems for random functions. I, *Theory of Probab. Appl.* **4** (1959) 178-197.

[92] V.A. Volkonskii, Y.A. Rozanov: Some limit theorems for random functions. II, *Theory of Probab. Appl.* **6** (1961) 186-198.

[93] M. Wendler, Bahadur representation for U-quantiles of dependent data, *J. Multivariate Anal.* **102** (2011) 1064-1079.

[94] M. Wendler, U-processes, U-quantile processes and generalized linear statistics of dependent data, *preprint*, arXiv:1009.5337.

[95] C.S. Withers, Conditions for linear processes to be strong-mixing, *Probab. Theory Related Fields* **57** (1981) 477-480.

[96] W.B. Wu, On the Bahadur representation of sample quantiles for dependent sequences, *Ann. Stat.* **33** (2005) 1934-1963.

[97] R. Yokoyama, Moment Bounds for Stationary Mixing Sequences, *Probab. Theory Related Fields* **52** (1980) 45-57.

[98] K. Yoshihara, Limiting behavior of U-statistics for stationary, absolutely regular processes, *Probab. Theory Related Fields* **35** (1976) 237-252.

[99] K. Yoshihara, The Bahadur representation of sample quantiles for sequences of strongly mixing random variables, *Statist. Probab. Letters* **24** (1995) 299-304.