Central Limit Theorem and the Bootstrap for U-Statistics of Strongly Mixing Data

Herold Dehling*, Martin Wendler†

December 5, 2008

The asymptotic normality of U-statistics has so far been proved for iid data and under various mixing conditions such as absolute regularity, but not for strong mixing. We use a coupling technique introduced 1983 by Bradley [5] to prove a new generalized covariance inequality similar to Yoshihara's [19]. It follows from the Hoeffding-decomposition and this inequality, that U-statistics of strongly mixing observations converge to a normal limit if the kernel of the U-statistic fulfills some moment and continuity conditions.

The validity of the bootstrap for U-statistics has until now only been established in the case of iid data (see Bickel and Freedman [3]). For mixing data, Politis and Romano [15] proposed the circular block bootstrap, which leads to a consistent estimation of the sample mean's distribution. We extend these results to U-statistics of weakly dependent data and prove a CLT for the circular block bootstrap version of U-statistics under absolute regularity and strong mixing. We also calculate a rate of convergence for the bootstrap variance estimator of a U-statistic.

1 U-Statistic CLT

1.1 U-Statistics under Independence

U-statistics are a broad class of nonlinear functionals, including many well-known examples such as the variance estimator or the Cramer-von Mises-statistic. For simplicity of notation, we concentrate on the case of bivariate U-statistics.

*Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany

†Corresponding Author, Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany, Fax: +49-234-3214039, Email address: Martin.Wendler@rub.de

Definition 1.1. *A U-statistic with a symmetric and measurable kernel $h : \mathbb{R}^2 \to \mathbb{R}$ is defined as*

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

$U_n(h)$ is the uniformly minimum variance estimator of $\theta = E[h(X_1, X_{2.})]$, if $X_1, \dots, X_n$ are iid with an arbitrary absolutely continuous distribution. To prove asymptotic normality of U-statistics, Hoeffding [11] decomposed $U_n(h)$ as follows:

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

with

$$\begin{aligned} h_1(x) &:= Eh(x, X_2) - \theta \\ h_2(x,y) &:= h(x,y) - h_1(x) - h_1(y) - \theta. \end{aligned}$$

$\frac{2}{\sqrt{n}}\sum_{i=1}^{n} h_1(X_i)$ is a sum of iid random variables with a normal limit distribution, $\frac{2}{\sqrt{n}(n-1)}\sum_{1\le i<j\le n} h_2(X_i, X_j)$ is called the degenerate part of the U-statistic and converges to zero in probability, as its parts are uncorrelated, so the U-statistic is asymptotically normal.

1.2 U-Statistics under Absolute Regularity

Under dependence, the summands of the degenerate part can be correlated and this can change the limit distribution. Under the strong assumption of $\star$-mixing, Sen [16] showed that U-statistics are asymptotically normal. Yoshihara assumed $X_1, \dots, X_n$ to be stationary and absolutely regular and proved a CLT for U-Statistics under this weaker condition (for a detailed description of the various mixing conditions see Doukhan [9] and Bradley [6]).

Definition 1.2. *A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is called absolutely regular, if*

$$\beta(m) := \sup\left\{\beta\left((X_1, \dots, X_k), (X_j)_{j\ge k+m}\right) | k \in \mathbb{N}\right\} \xrightarrow{m\to\infty} 0$$

where β is the absolute regularity coefficient defined as

$$\beta(Y, Z) := E\left[\sup_{A\in\sigma(Y)} |P[A|Z] - P[A]|\right].$$

Yoshihara has proved the asymptotic normality of the U-statistic $U_n(h)$ using the Hoeffding-decomposition and generalized covariance inequalities. With increasing distance between the indices i_1, i_2, i_3, i_4, the covariance of $h_2(X_{i_1}, X_{i_2})$ and $h_2(X_{i_3}, X_{i_4})$ becomes smaller and therefore the degenerate part vanishes as in the independent case.

Theorem 1.3 (Yoshihara [19]). *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary, absolutely regular sequence of random variables. If there are $\delta, M > 0$, so that for all $k \in \mathbb{N}_0$*

$$\iint |h(x_1, x_2)|^{2+\delta} \, dF(x_1) \, dF(x_2) \leq M \tag{1}$$

$$\int |h(x_1, x_k)|^{2+\delta} \, dP(x_1, x_k) \leq M \tag{2}$$

and for a $\delta' < \delta$

$$\beta(n) = O\left(n^{-\frac{2+\delta'}{\delta'}}\right) \tag{3}$$

then with $\sigma_\infty^2 = \operatorname{Var}[h_1(X_1)] + 2\sum_{k=1}^{\infty} \operatorname{Cov}[h_1(X_1), h_1(X_{1+k})]$:

$$\sqrt{n}(U_n(h) - \theta) \xrightarrow{\mathcal{D}} N\left(0, 4\sigma_\infty^2\right) \tag{4}$$

Denker and Keller [8] have weakened the mixing assumption to functionals of absolutely regular processes, Borovkova, Burton and Dehling [4] showed convergence of the emperical U-process to a Gaussian process.

1.3 U-Statistics under Strong Mixing

We want to extend Yoshihara's CLT to random variables, which fullfill the strong mixing condition:

Definition 1.4. *A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is called strongly mixing if*

$$\alpha(m) = \sup\left\{\alpha\left((X_1, \ldots, X_k), (X_j)_{j\geq k+m}\right) | k \in \mathbb{N}\right\} \xrightarrow{m\to\infty} 0$$

where α is the strong mixing coefficient defined as

$$\alpha(Y, Z) = \sup_{\substack{A\in\sigma(Y)\\ B\in\sigma(Z)}} |P(A\cap B) - P(A)P(B)|.$$

Strong mixing is weaker than absolute regularity, but absolute regularity and strong mixing are equivalent for random variables, which take their values in a finite set. You can approximate general random variables by such discrete ones. This discretization works only for U-Statistics if the kernel is continuous in some sense:

Definition 1.5. *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process. A kernel h is called $\mathcal{P}$-Lipschitz-continuous if there is a constant $L > 0$ with*

$$E\left[\left|h(X, Y) - h\left(X', Y\right)\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \leq L\epsilon \tag{5}$$

for every $\epsilon > 0$, every pair X and Y with the common distribution $\mathcal{P}_{X_1,X_k}$ for a $k \in \mathbb{N}$ or $\mathcal{P}_{X_1} \times \mathcal{P}_{X_1}$ and X' and Y also with one of these common distributions.

$\mathcal{P}$-Lipschitz-continuity is a special case of p-continuity established by Borovkova, Burton and Dehling [4]. It is clear that every Lipschitz-continuous kernel is $\mathcal{P}$-Lipschitz-continuous. But this definition holds also for many kernels that are not Lipschitz-continuous in the ordinary sense:

Example 1.6 (Variance estimation)**.** Consider stationary random variables with bounded variance and the kernel $h(x,y) = \frac{1}{2}(x-y)^2$. The related U-statistic is the well known variance estimator

$$U_n(h) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

For random variables X, X' and Y as above, we get:

$$\begin{aligned}
&E\left[\left|\frac{1}{2}(X-Y)^2 - \frac{1}{2}\left(X'-Y\right)^2\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \\
\leq& \frac{1}{2}E\left[\left|X^2 - X'^2\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] + E\left[\left|X-X'\right| |Y| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \\
\leq& \epsilon E|X| + \frac{1}{2}\epsilon^2 + \epsilon E|Y| = 2\epsilon E|X| + \frac{1}{2}\epsilon^2
\end{aligned}$$

Furthermore, we have

$$E\left[\left|\frac{1}{2}(X-Y)^2 - \frac{1}{2}\left(X'-Y\right)^2\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \leq 4E\left[X^2\right] < \infty.$$

These two bounds prove the $\mathcal{P}$-Lipschitz-continuity.

Example 1.7 (Dimension estimation)**.** Let $t > 0$. The kernel $h(x,y) = \mathbb{1}_{\{|x-y|<t\}}$ is related to the Grassberger-Procaccia dimension estimator. It is $\mathcal{P}$-Lipschitz-continuous if there is an $L > 0$, such that for all $\epsilon > 0$ and every common distribution of X and Y from definiton 3.1:

$$P\left[t-\epsilon \leq |X-Y| \leq t+\epsilon\right] \leq L\epsilon$$

The difference between $\mathbb{1}_{\{|X-Y|<t\}}$ and $\mathbb{1}_{\{|X'-Y|<t\}}$ is not 0, iff $|X-Y| < t$ and $|X'-Y| \geq t$ or the other way round. As $|X-Y'| \leq \epsilon$, it follows that $t-\epsilon \leq |X-Y| \leq t+\epsilon$. Therefore

$$E\left[\left|\mathbb{1}_{\{|X-Y|<t\}} - \mathbb{1}_{\{|X'-Y|<t\}}\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \leq P\left[t-\epsilon \leq |X-Y| \leq t+\epsilon\right] \leq L\epsilon.$$

Using the Hoeffding decomposition and a new generalized covariance inequality for h_2 under strong mixing, the asymptotic normality of a U-statistic follows. You need more technical conditions than under absolute regularity: A faster decay of the mixing coefficient, some finite moments of X_i and the $\mathcal{P}$-Lipschitz-continuity.

Theorem 1.8. *Let h be a $\mathcal{P}$-Lipschitz-continuous kernel, $(X_n)_{n\in\mathbb{N}}$ a stationary sequence of random variables. If there is $\gamma > 0$ with $E|X_k|^\gamma < \infty$ and $M > 0$, $\delta > 0$, so that for* *all* $k \in \mathbb{N}_0$

$$\iint |h(x_1, x_2)|^{2+\delta} \, dF(x_1) \, dF(x_2) \leq M$$

$$\int |h(x_1, x_k)|^{2+\delta} \, dP(x_1, x_k) \leq M$$

and furthermore for a $\rho > \frac{3\gamma\delta+\delta+5\gamma+2}{2\gamma\delta}$

$$\alpha(n) = O\left(n^{-\rho}\right) \tag{6}$$

then

$$\sqrt{n}\left(U_n(h) - \theta\right) \xrightarrow{\mathcal{D}} N\left(0, 4\sigma_\infty^2\right) \tag{7}$$

with $\sigma_\infty^2 = \mathrm{Var}\left[h_1(X_1)\right] + 2\sum_{k=1}^{\infty} \mathrm{Cov}\left[h_1(X_1) \, h_1(X_{1+k})\right]$.

2 Bootstrap

2.1 Bootstrapping the Sample Mean

Efron [10] proposed a method called bootstrap to estimate the unknown distribution of an estimator $\hat{\theta}(X_1, \ldots, X_n)$: The original sample $X_1, \ldots, X_n$ with unknown distribution is replaced by $X_1^\star, \ldots, X_n^\star$, which have the distribution function $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i < t}$ and are conditionally independent for given $X_1, \ldots, X_n$. Singh [18] mentioned, that this method can fail under dependence.

Kuensch [13] proposed resampling blocks of consecutive observations instead of single observations. We consider the circular block bootstrap introduced by Politis and Romano [15]. Instead of the original sample of n observations with an unknown distribution, construct new samples $X_1^\star, \ldots, X_{bl}^\star$ as follows: Extend the sample $X_1, \ldots, X_n$ periodically by $X_{i+n} = X_i$, choose blocks of l consecutive observations of the sample randomly and repeat that $b = \lfloor \frac{n}{l} \rfloor$ times independently: For $j = 1, \ldots, n$

$$P^\star\left[X_1^\star = X_j, \ldots, X_l^\star = X_{j+l-1}\right] = \frac{1}{n} \tag{8}$$

For strongly mixing stationary processes, the bootstrap version of the sample mean $\bar{X}_n^\star = \frac{1}{bl}\sum_{i=1}^{bl} X_i^\star$ has asymptotically the same normal distribution as the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$:

Theorem 2.1 (Shao, Yu [17])**.** *Let* $(X_n)_{n\in\mathbb{N}}$ *be a stationary sequence of random variables with* $E|X_1|^{2+\delta} < \infty$ *and* $\alpha(n) = O\left(n^{-r}\right)$*, where* $\delta > 0$ *and* $r > \frac{2+\delta}{\delta}$ *and* $l = l_n \xrightarrow{n\to\infty} \infty$*,* $l = O\left(n^{1-\epsilon}\right)$ *for some* $\epsilon > 0$*. Then:*

$$\sup_{x\in\mathbb{R}} \left| P^\star\left[\sqrt{bl}\left(\bar{X}_n^\star - \bar{X}\right) \leq x\right] - P\left[\sqrt{n}\left(\bar{X} - E[X]\right) \leq x\right] \right| \xrightarrow{a.s.} 0 \tag{9}$$

$$\left| \mathrm{Var}^\star\left[\sqrt{bl}\bar{X}_n^\star\right] - \mathrm{Var}\left[\sqrt{n}\bar{X}\right] \right| \xrightarrow{a.s.} 0 \tag{10}$$

Furthermore, we can get a rate of convergence according to the block length l. With increasing block length, the bias becomes smaller and the variance becomes bigger. The optimal choice for l to minimize the mean squared error (MSE) of $\mathrm{Var}^\star\left[\sqrt{bl}\bar{X}_n^\star\right]$ is

$$l^0 = Kn^{\frac{1}{3}} + o\left(n^{\frac{1}{3}}\right) \tag{11}$$

for a constant $K > 0$. For this block length, we get:

Theorem 2.2 (Lahiri [14]). *If $(X_1)_{n\in\mathbb{N}}$ is a stationary, strongly mixing sequence with $E|X_1|^{8+\delta} < \infty$ and $\sum_{n=1}^{\infty} n^7 \alpha^{\frac{\delta}{8+\delta}}(n) < \infty$ for a $\delta > 0$, then*

$$\min_l \mathrm{MSE}\left(\mathrm{Var}^\star\left[\sqrt{bl}\bar{X}_n^\star\right]\right) = O\left(n^{-\frac{2}{3}}\right). \tag{12}$$

2.2 Bootstrapping U-Statistics

To bootstrap a U-statistic one can resample blocks of observations and plug them in:

$$\begin{aligned}U_n^\star(h) = \frac{2}{bl(bl-1)} \sum_{1\le i<j\le bl} h\left(X_i^\star, X_j^\star\right) \\ = \theta + \frac{2}{bl}\sum_{i=1}^{bl} h_1\left(X_i^\star\right) + \frac{2}{bl(bl-1)} \sum_{1\le i<j\le bl} h_2\left(X_i^\star, X_j^\star\right)\end{aligned}$$

We show that the bootstrap version of a U-statistic has the same asymptotic variance and the same normal limit distribution as the U-statistic itself, using the Hoeffding decomposition and the fact that the block bootstrap is consistent for the sample mean.

Theorem 2.3. *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary, mixing process and h a kernel, such that for a $\delta > 0$, $M > 0$:*

$$\begin{aligned}\iint \left|h(x_1, x_2)\right|^{2+\delta} dF(x_1)\, dF(x_2) \le M \\ \forall k \in \mathbb{N}_0: \quad \int \left|h(x_1, x_{1+k})\right|^{2+\delta} dP(x_1, x_{1+k}) \le M\end{aligned}$$

Let l be the block length with $l \xrightarrow{n\to\infty} \infty$ and $l = O\left(n^{1-\epsilon}\right)$ for some $\epsilon > 0$. If one of the following two conditions holds

- *for a $\delta' \in (0, \delta)$: $\beta(n) = O\left(n^{-\frac{2+\delta'}{\delta'}}\right)$*
- *h is $\mathcal{P}$-Lipschitz, $E|X_1|^\gamma < \infty$ for a $\gamma > 0$ and for $\rho > \frac{3\gamma\delta+\delta+5\gamma+2}{2\gamma\delta}$: $\alpha(n) = O\left(n^{-\rho}\right)$*

then

$$\left|\mathrm{Var}^\star\left[\sqrt{bl}U_n^\star(h)\right]-\mathrm{Var}\left[\sqrt{n}U_n(h)\right]\right|\xrightarrow{\mathcal{P}}0 \tag{13}$$

$$\sup_{x\in\mathbb{R}}\left|P^\star\left[\sqrt{bl}\left(U_n^\star(h)-E^\star\left[U_n^\star\right]\right)\le x\right]-P\left[\sqrt{n}\left(U_n(h)-\theta\right)\le x\right]\right|\xrightarrow{\mathcal{P}}0. \tag{14}$$

If we assume the existence of higher moments, we can achieve almost sure convergence:

Theorem 2.4. *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary and absolutely regular process and h a kernel, such that for a $\delta>0$, $M>0$:*

$$\iint\left|h(x_1,x_2)\right|^{4+\delta}dF(x_1)\,dF(x_2)\le M$$

$$\forall k\in\mathbb{N}_0:\quad\int\left|h(x_1,x_{1+k})\right|^{4+\delta}dP(x_1,x_{1+k})\le M$$

and for a $\delta'\in(0,\delta)$ $\beta(n)=O\left(n^{-\frac{3(4+\delta')}{\delta'}}\right)$ and additionally $l\xrightarrow{n\to\infty}\infty$ and $l=O\left(n^{1-\epsilon}\right)$ for some $\epsilon>0$, then

$$\left|\mathrm{Var}^\star\left[\sqrt{bl}U_n^\star(h)\right]-\mathrm{Var}\left[\sqrt{n}U_n(h)\right]\right|\xrightarrow{a.s.}0 \tag{15}$$

$$\sup_{x\in\mathbb{R}}\left|P^\star\left[\sqrt{bl}\left(U_n^\star(h)-E^\star\left[U_n^\star\right]\right)\le x\right]-P\left[\sqrt{n}\left(U_n(h)-\theta\right)\le x\right]\right|\xrightarrow{a.s.}0. \tag{16}$$

The degenerate part of the bootstrapped U-statistic converges to zero with a rate, which does not depend on the block length and is faster than the convergence of the sample mean. Choosing the optimal block length for the linear part $\frac{2}{\sqrt{n}}\sum_{i=1}^n h_1(X_i)$, we can achieve the following rate of convergence:

Corollary 2.5. *If the assumptions of Theorem 2.4 hold and additionally for a $\tilde{\delta}>0$: $E\left|h_1(X_1)\right|^{8+\tilde{\delta}}<\infty$ and $\sum_{n=1}^\infty n^7\alpha^{\frac{\tilde{\delta}}{8+\tilde{\delta}}}(n)<\infty$, the optimal block length l^0 for $\frac{1}{n}\sum h_1(X_i)$ is asymptotically optimal for U_n and the variance estimator converges with the following rate:*

$$\mathrm{MSE}\left(\mathrm{Var}^\star\left[\sqrt{bl}U_n^\star(h)\right]\right)=O\left(n^{-\frac{2}{3}}\right) \tag{17}$$

3 Auxiliary Results

3.1 Generalized Covariance Inequalities

Yoshihara has proved the asymptotic normality of the U-statistic $U_n(h)$ with the help of the Hoeffding-decomposition and the following generalized covariance inequality:

Lemma 3.1 (Yoshihara [19])**.** *If there are $\delta,M>0$, so that for all $k\in\mathbb{N}_0$*

$$\iint\left|h(x_1,x_2)\right|^{2+\delta}dF(x_1)\,dF(x_2)\le M$$

$$\int\left|h(x_1,x_k)\right|^{2+\delta}dP(x_1,x_k)\le M$$

then there is a constant K, such that for $m = \max\left\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\right\}$, where $i_{(1)} \leq i_{(2)} \leq i_{(3)} \leq i_{(4)}$ the following inequality holds:

$$\left|E\left[h_2\left(X_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right)\right]\right| \leq K\beta^{\frac{\delta}{2+\delta}}(m) \tag{18}$$

To prove Lemma 3.1 under absolute regularity, you can use coupling techniques (see Berbee [1] and Berkes, Philipp [2]): For dependent random variables X and Y, one can find a random variable X', such that

- X' has the same distribution as X,
- X' and Y are independent,
- $P\left[X \neq X'\right] = \beta(X, Y)$.

Such a coupling is impossible under strong mixing, as can be seen e.g. from the results of Dehling [7]. Bradley [5], however, was able to establish a weaker type of coupling for strongly mixing random variables, using the fact that absolute regularity and strong mixing are equivalent for random variables taking their values in a finite set and approximating general random variables by such discrete ones:

Lemma 3.2 (Bradley [5])**.** *Let X, Y be random variables, X real-valued with $E|X|^\gamma \leq \infty$. Let $0 < \epsilon \leq \|X\|_\gamma$. Then there exists (after replacing the underlying probability space by a bigger one if necessary) a random variable X' such that*

- *X' has the same distribution as X,*
- *X' and Y are independent,*
- $$P\left[\left|X - X'\right| \geq \epsilon\right] \leq 18 \frac{\|X\|_\gamma^{\frac{\gamma}{2+\gamma}}}{\epsilon^{\frac{\gamma}{2+\gamma}}} \alpha^{\frac{2\gamma}{2+\gamma}}(X, Y). \tag{19}$$

As this coupling under strong mixing allows small differences between X and X' (while X and X' are equal with high probability in the case of absolute regularity), we need the $\mathcal{P}$-Lipschitz-continuity of the kernel.

Lemma 3.3. *Let h be a $\mathcal{P}$-Lipschitz-continuous kernel with constant L, $(X_n)_{n\in\mathbb{N}}$ a stationary sequence of random variables. If there is a $\gamma > 0$ with $E|X_k|^\gamma < \infty$ and $M > 0$, $\delta > 0$, so that for all $k \in \mathbb{N}_0$*

$$\iint \left|h(x_1, x_2)\right|^{2+\delta} dF(x_1)\, dF(x_2) \leq M$$
$$\int \left|h(x_1, x_k)\right|^{2+\delta} dP(x_1, x_k) \leq M$$

then there exists a constant $K = K\left(\gamma, \|X_1\|_\gamma, \delta, M, L\right)$, such that the following inequality holds with $m = \max\left\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\right\}$:

$$\left|E\left[h_2\left(X_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right)\right]\right| \leq K\alpha^{\frac{2\gamma\delta}{3\gamma\delta+\delta+5\gamma+2}}(m) \tag{20}$$

3.2 Bounds for the Degenerate Part of a U-Statistic

With the covariance inequalities one can show that the covariance of the summands $h(X_i, X_j)$ is small if the gap between the indices is big enough. Therefore, the degenerate part decreases fast enough, so that it does not disturb the asymptotic normality of $\frac{1}{\sqrt{n}} \sum h_1(X_i)$.

Lemma 3.4 (Yoshihara [19]). *If the assumptions of Lemma 3.1 hold and furthermore for a $\delta' < \delta$*

$$\beta(n) = O\left(n^{-\frac{2+\delta'}{\delta'}}\right)$$

then for $U_n(h_2)$:

$$\begin{aligned} E\left[nU_n^2(h_2)\right] &\leq \frac{4}{n(n-1)^2} \sum_{1\leq i_1<i_2\leq n} \sum_{1\leq i_3<i_4\leq n} \left|E\left[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})\right]\right| \\ &\leq \frac{4}{n^3} \sum_{i_1,i_2,i_3,i_4=1}^{n} \left|E\left[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})\right]\right| = O\left(n^{-\eta}\right) \end{aligned} \tag{21}$$

with $\eta = \min\left\{2\frac{\delta-\delta'}{\delta'(2+\delta)}, 1\right\}$.

So $\sqrt{n}U_n(h_2)$ vanishes as n increases. For one of our later results, we also need another one of Yoshihara's lemmas (Our assumptions and result differ slightly from the lemma in [19], as we believe there is a misprint):

Lemma 3.5 (Yoshihara [19]). *If*

$$\begin{aligned} &\iint \left|h(x_1, x_2)\right|^{4+\delta} dF(x_1)\, dF(x_2) \leq M \\ \forall k \in \mathbb{N}_0: \quad &\int \left|h(x_1, x_{1+k})\right|^{4+\delta} dP(x_1, x_{1+k}) \leq M \end{aligned}$$

and for a $\delta' \in (0, \delta)$ $\beta(n) = O\left(n^{-\frac{3(4+\delta')}{\delta'}}\right)$, then for $\eta' = \min\left\{12\frac{\delta-\delta'}{\delta'(4+\delta)}, 1\right\}$

$$\begin{aligned} &E\left[n^2 U_n^4(h_2)\right] \\ &\quad\leq \frac{16}{n^6} \sum_{i_1,\ldots,i_8=1}^{n} \left|E\left[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4}) h_2(X_{i_5}, X_{i_6}) h_2(X_{i_7}, X_{i_8})\right]\right| \\ &\quad= O\left(n^{-1-\eta'}\right). \end{aligned} \tag{22}$$

Now we show a result analogous to the Lemma 3.4 under strong mixing:

Lemma 3.6. *If the assumptions of Lemma 3.3 hold and for a $\rho > \frac{3\gamma\delta+\delta+5\gamma+2}{2\gamma\delta}$*

$$\alpha(n) = O\left(n^{-\rho}\right)$$

then for $U_n(h_2)$*:*

$$E\left[nU_n^2(h_2)\right] \le \frac{4}{n(n-1)^2} \sum_{1\le i_1<i_2\le n} \sum_{1\le i_3<i_4\le n} \left|E\left[h_2(X_{i_1},X_{i_2})\,h_2(X_{i_3},X_{i_4})\right]\right|$$
$$\le \frac{4}{n^3} \sum_{i_1,i_2,i_3,i_4=1}^{n} \left|E\left[h_2(X_{i_1},X_{i_2})\,h_2(X_{i_3},X_{i_4})\right]\right| = O\left(n^{-\eta}\right) \quad (23)$$

with $\eta = \min\left\{\rho\frac{2\gamma\delta}{3\gamma\delta+\delta+5\gamma+2} - 1, 1\right\} > 0$.

We need a bound for $U_n^\star(h_2) = \frac{2}{bl(bl-1)}\sum_{1\le i<j\le n} h_2\left(X_i^\star, X_j^\star\right)$. Using Yoshihara's inequality for the second moment respectively Lemma 3.6 and the fact that the bootstrap expectation of a U-statistic is similar to a von Mises-statistic, we get:

Lemma 3.7. *Let* $(X_n)_{n\in\mathbb{N}}$ *be a stationary, mixing process and* h *a kernel, such that for a* $\delta > 0$, $M > 0$*:*

$$\iint |h(x_1,x_2)|^{2+\delta}\, dF(x_1)\, dF(x_2) \le M$$
$$\forall k \in \mathbb{N}_0: \quad \int |h(x_1,x_{1+k})|^{2+\delta}\, dP(x_1,x_{1+k}) \le M$$

If one of the following two conditions holds

- *for a* $\delta' \in (0,\delta)$*:* $\beta(n) = O\left(n^{-\frac{2+\delta'}{\delta'}}\right)$
- h *is* $\mathcal{P}$*-Lipschitz,* $E|X_1|^\gamma < \infty$ *for a* $\gamma > 0$ *and for* $\rho > \frac{3\gamma\delta+\delta+5\gamma+2}{2\gamma\delta}$*:* $\alpha(n) = O(n^{-\rho})$

then for $\eta = \min\left\{2\frac{\delta-\delta'}{\delta'(2+\delta)}, 1\right\}$ *respectively* $\eta = \min\left\{\rho\frac{2\gamma\delta}{3\gamma\delta+\delta+5\gamma+2} - 1, 1\right\}$*:*

$$E\left[E^\star\left[blU_n^{\star 2}(h_2)\right]\right] = O\left(n^{-\eta}\right). \quad (24)$$

With the inequality for the fourth moment, we can calculate a faster rate of convergence. Note that this rate does not depend on the block length.

Lemma 3.8. *If*

$$\iint |h(x_1,x_2)|^{4+\delta}\, dF(x_1)\, dF(x_2) \le M$$
$$\forall k \in \mathbb{N}_0: \quad \int |h(x_1,x_{1+k})|^{4+\delta}\, dP(x_1,x_{1+k}) \le M$$

and for a $\delta' \in (0,\delta)$ $\beta(n) = O\left(n^{-\frac{3(4+\delta')}{\delta'}}\right)$*, then for* $\eta' = \min\left\{12\frac{\delta-\delta'}{\delta'(4+\delta)}, 1\right\}$

$$E\left[E^\star\left[(bl)^2 U_n^{\star 4}(h_2)\right]\right] = O\left(n^{-1-\eta'}\right). \quad (25)$$

4 Proofs

We will first prove the auxiliary results and after that the CLT and the theorems about the bootstrap.

4.1 Auxiliary Results

Proof of Lemma 3.3: For simplicity, we consider only the case $i_1 < i_2 < i_3 < i_4$ and $i_2 - i_1 \geq i_4 - i_3$. Let $\epsilon > 0$, $K > 0$ and define:

$$h_{2,K}(x,y) = \begin{cases} h_2(x,y) & \text{if } |h_2(x,y)| \leq \sqrt{K} \\ \sqrt{K} & \text{if } h_2(x,y) > \sqrt{K} \\ -\sqrt{K} & \text{if } h_2(x,y) < -\sqrt{K} \end{cases}$$

h_2 is $\mathcal{P}$-Lipschitz-continuous with constant $2L$, as for all X, X', Y as in defintion 1.5 and Y' with the same distribution as Y and independent of X and X':

$$\begin{aligned} &E\left[\left|h_2(X,Y) - h_2\left(X',Y\right)\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \\ \leq &E\left[\left|h(X,Y) - h\left(X',Y\right)\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] + E\left[\left|h_1(X) - h_1\left(X'\right)\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \\ \leq &E\left[\left|h(X,Y) - h\left(X',Y\right)\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \\ &+ E\left[\left|h\left(X,Y'\right) - h\left(X',Y'\right)\right| \mathbb{1}_{\{|X-X'|\leq\epsilon\}}\right] \leq 2L\epsilon. \end{aligned}$$

Obviously, $h_{2,K}$ is $\mathcal{P}$-Lipschitz-continuous with the same constant $2L$ as h_2. With Lemma 3.2, choose a random variable X'_{i_1} independent of $X_{i_2}, X_{i_3}, X_{i_4}$ with

$$P\left[\left|X_{i_1} - X'_{i_1}\right| \geq \epsilon\right] \leq 18 \frac{\|X\|_\gamma^{\frac{\gamma}{2+\gamma}}}{\epsilon^{\frac{\gamma}{2+\gamma}}} \alpha^{\frac{2\gamma}{2+\gamma}}(m).$$

As h_2 is a degenerate kernel, we have

$$E\left[h_2\left(X'_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right)\right] = 0.$$

Therefore, we get:

$$\begin{aligned} &\left|E\left[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})\right]\right| \\ = &\left|E\left[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})\right] - E\left[h_2\left(X'_{i_1}, X_{i_2}\right) h_2(X_{i_3}, X_{i_4})\right]\right| \\ = &\left|E\left[\left(h_2(X_{i_1}, X_{i_2}) - h_2\left(X'_{i_1}, X_{i_2}\right)\right) h_2(X_{i_3}, X_{i_4})\right]\right| \\ \leq &E\left[\left|\left(h_{2,K}(X_{i_1}, X_{i_2}) - h_{2,K}\left(X'_{i_1}, X_{i_2}\right)\right) h_{2,K}(X_{i_3}, X_{i_4})\right| \mathbb{1}_{\left\{\left|X_{i_1} - X'_{i_1}\right| \leq \epsilon\right\}}\right] \\ &+ E\left[\left|\left(h_{2,K}(X_{i_1}, X_{i_2}) - h_{2,K}\left(X'_{i_1}, X_{i_2}\right)\right) h_{2,K}(X_{i_3}, X_{i_4})\right| \mathbb{1}_{\left\{\left|X_{i_1} - X'_{i_1}\right| > \epsilon\right\}}\right] \\ &+ E\left[\left|h_{2,K}(X_{i_1}, X_{i_2}) h_{2,K}(X_{i_3}, X_{i_4}) - h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})\right|\right] \\ &+ E\left[\left|h_{2,K}\left(X'_{i_1}, X_{i_2}\right) h_{2,K}(X_{i_3}, X_{i_4}) - h_2\left(X'_{i_1}, X_{i_2}\right) h_2(X_{i_3}, X_{i_4})\right|\right] \end{aligned}$$

Because of the $\mathcal{P}$-Lipschitz-continuity and $|h_{2,K}(X_3,X_4)| \leq \sqrt{K}$, the first summand is smaller than $2L\epsilon\sqrt{K}$. In consequence of Lemma 3.2, the second term is bounded by

$$P\left[\left|X_{i_1} - X'_{i_1}\right| \geq \epsilon\right] 2K \leq 36 \frac{\|X\|_\gamma^{\frac{\gamma}{2+\gamma}}}{\epsilon^{\frac{\gamma}{2+\gamma}}} \alpha^{\frac{2\gamma}{2+\gamma}}(m) K.$$

For the third summand, we get:

$$\begin{aligned}
& E\left[\left|h_{2,K}(X_{i_1},X_{i_2}) h_{2,K}(X_{i_3},X_{i_4}) - h_2(X_{i_1},X_{i_2}) h_2(X_{i_3},X_{i_4})\right|\right] \\
\leq\ & E\Big[\left(|h_2(X_{i_1},X_{i_2})| - \sqrt{K}\right)|h_2(X_{i_3},X_{i_4})| \\
& \qquad\qquad\qquad\qquad \mathbb{1}_{\left\{|h_2(X_{i_1},X_{i_2})|>\sqrt{K}, |h_2(X_{i_3},X_{i_4})|\leq\sqrt{K}\right\}}\Big] \\
& + E\Big[|h_2(X_{i_1},X_{i_2})|\left(|h_2(X_{i_3},X_{i_4})| - \sqrt{K}\right) \\
& \qquad\qquad\qquad\qquad \mathbb{1}_{\left\{|h_2(X_{i_1},X_{i_2})|\leq\sqrt{K}, |h_2(X_{i_3},X_{i_4})|>\sqrt{K}\right\}}\Big] \\
& + E\Big[\left|\left(|h_2(X_{i_1},X_{i_2})| - \sqrt{K}\right)\left(|h_2(X_{i_3},X_{i_4})| - \sqrt{K}\right)\right| \\
& \qquad\qquad\qquad\qquad \mathbb{1}_{\left\{|h_2(X_{i_1},X_{i_2})|>\sqrt{K}, |h_2(X_{i_3},X_{i_4})|>\sqrt{K}\right\}}\Big] \\
\leq\ & E\left[\left(|h_2(X_{i_1},X_{i_2})| - \sqrt{K}\right)\sqrt{K}\mathbb{1}_{\left\{|h_2(X_{i_1},X_{i_2})|>\sqrt{K}\right\}}\right] \\
& \qquad + E\left[\left(|h_2(X_{i_3},X_{i_4})| - \sqrt{K}\right)\sqrt{K}\mathbb{1}_{\left\{|h_2(X_{i_3},X_{i_4})|>\sqrt{K}\right\}}\right] \\
& \qquad + \frac{1}{2}E\left[\left(|h_2(X_{i_1},X_{i_2})| - \sqrt{K}\right)^2 \mathbb{1}_{\left\{|h_2(X_{i_1},X_{i_2})|>\sqrt{K}\right\}}\right] \\
& \qquad + \frac{1}{2}E\left[\left(|h_2(X_{i_3},X_{i_4})| - \sqrt{K}\right)^2 \mathbb{1}_{\left\{|h_2(X_{i_3},X_{i_4})|>\sqrt{K}\right\}}\right] \\
\leq\ & \frac{1}{2}E\left[h_2^2(X_{i_1},X_{i_2}) \mathbb{1}_{\left\{|h_2(X_{i_1},X_{i_2})|>\sqrt{K}\right\}}\right] \\
& \qquad\qquad\qquad\qquad + \frac{1}{2}E\left[h_2^2(X_{i_3},X_{i_4}) \mathbb{1}_{\left\{|h_2(X_{i_3},X_{i_4})|>\sqrt{K}\right\}}\right] \\
& \qquad\qquad \leq \frac{1}{2}\frac{E|h_2(X_{i_1},X_{i_2})|^{2+\delta}}{K^{\frac{\delta}{2}}} + \frac{1}{2}\frac{E|h_2(X_{i_3},X_{i_4})|^{2+\delta}}{K^{\frac{\delta}{2}}} \leq \frac{M}{K^{\frac{\delta}{2}}}
\end{aligned}$$

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

$$\begin{aligned}
& |E[h_2(X_{i_1},X_{i_2}) h_2(X_{i_3},X_{i_4})]| \\
& \qquad \leq 2L\epsilon\sqrt{K} + 36\frac{\|X\|_\gamma^{\frac{\gamma}{2+\gamma}}}{\epsilon^{\frac{\gamma}{2+\gamma}}}\alpha^{\frac{2\gamma}{2+\gamma}}(m) K + 2\frac{M}{K^{\frac{\delta}{2}}} =: f(\epsilon, K)
\end{aligned}$$

Setting $\epsilon^0 = \|X_1\|_\gamma^{\frac{\gamma}{3\gamma+1}} L^{-\frac{2\gamma+1}{3\gamma+1}} \alpha(m)^{\frac{2\gamma}{3\gamma+1}} K^{\frac{\gamma+\frac{1}{2}}{3\gamma+1}}$, we obtain:

$$f\left(\epsilon^0, K\right) = 38\|X_1\|_\gamma^{\frac{\gamma}{3\gamma+1}} L^{\frac{\gamma}{3\gamma+1}} K^{\frac{\frac{5}{2}\gamma+1}{3\gamma+1}} \alpha^{\frac{2\gamma}{3\gamma+1}}(m) + 2\frac{M}{K^{\frac{\delta}{2}}}$$

With $K^0 = \|X_1\|_\gamma^{-\frac{2\gamma}{3\gamma\delta+\delta+5\gamma+2}} L^{-\frac{2\gamma}{3\gamma\delta+\delta+5\gamma+2}} \alpha(m)^{-\frac{4\gamma}{3\gamma\delta+\delta+5\gamma+2}} M^{\frac{6\gamma+2}{3\gamma\delta+\delta+5\gamma+2}}$, we finally get the bound:

$$\begin{aligned}
&|E[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})]| \\
&\quad \leq f(\epsilon^0, K^0) = 40 \|X_1\|_\gamma^{\frac{\gamma\delta}{3\gamma\delta+\delta+5\gamma+2}} L^{\frac{\gamma\delta}{3\gamma\delta+\delta+5\gamma+2}} M^{\frac{5\gamma+2}{3\gamma\delta+\delta+5\gamma+2}} \alpha(m)^{\frac{2\gamma\delta}{3\gamma\delta+\delta+5\gamma+2}} \qquad (26)
\end{aligned}$$

□

Proof of Lemma 3.6: The proof is exactly the same as of Yoshihara's Lemma 3.4, using Lemma 3.3 instead of 3.1. Therefore, we concentrate on the case $i_1 < i_2 < i_3 < i_4$ and $i_2 - i_1 \geq i_4 - i_3$. If $i_2 - i_1 = m$, there are at most n possibilities for i_1 and i_3 and m possibilities for i_4:

$$\begin{aligned}
\sum_{\substack{i_1<i_2<i_3<i_4 \\ i_2-i_1\geq i_4-i_3}} |E[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})]| &\leq n^2 \sum_{m=1}^{n} mK\alpha(m)^{\frac{2\gamma\delta}{3\gamma\delta+\delta+5\gamma+2}} \\
&\leq K_2 n^2 \sum_{m=1}^{n} m^{1-\rho\frac{2\gamma\delta}{3\gamma\delta+\delta+5\gamma+2}} = O\left(n^{3-\eta}\right)
\end{aligned}$$

With a similar argument for the other cases, we get

$$\sum_{i_1,i_2,i_3,i_4=1}^{n} |E[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})]| = O\left(n^{3-\eta}\right).$$

□

Proof of Lemma 3.7: The bootstrapped expectation of $h_2\left(X_{i_1}^\star, X_{i_2}^\star\right) h_2\left(X_{i_3}^\star, X_{i_4}^\star\right)$ (conditionally on $(X_n)_{n\in\mathbb{N}}$) depends on the way the indices i_1, i_2, i_3, i_4 are allocated to the different blocks. First consider indices i_1, i_2, i_3, i_4 lying in different blocks (therefore, $X_{i_1}^\star, \ldots, X_{i_4}^\star$ are independent for fixed $(X_n)_{n\in\mathbb{N}}$). Then the bootstrapped expectation of $h_2\left(X_{i_1}^\star, X_{i_2}^\star\right) h_2\left(X_{i_3}^\star, X_{i_4}^\star\right)$ is a von Mises-statistic and we get

$$\begin{aligned}
&\left|E\left[E^\star\left[h_2\left(X_{i_1}^\star, X_{i_2}^\star\right) h_2\left(X_{i_3}^\star, X_{i_4}^\star\right)\right]\right]\right| \\
=&\left|E\left[\frac{1}{n^4} \sum_{i_1,i_2,i_3,i_4=1}^{n} h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})\right]\right| \\
\leq&\frac{1}{n^4} \sum_{i_1,i_2,i_3,i_4=1}^{n} |E[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})]|.
\end{aligned}$$

There are at most n^4 possibilities for the four indices to be in four different blocks, so

$$\begin{aligned}
\sum_{\substack{i_1,i_2,i_3,i_4 \\ \text{4 diff. blocks}}} &\left|E\left[E^\star\left[h_2\left(X_{i_1}^\star, X_{i_2}^\star\right) h_2\left(X_{i_3}^\star, X_{i_4}^\star\right)\right]\right]\right| \\
&\leq \sum_{i_1,i_2,i_3,i_4=1}^{n} |E[h_2(X_{i_1}, X_{i_2}) h_2(X_{i_3}, X_{i_4})]|.
\end{aligned}$$

As an example, let i_1 and i_2 now lie in the same block (write $i_1 \sim i_2$) with $i_2 - i_i = k$, while i_3, i_4 lie in two further blocks. The bootstrapped expectation is no longer a von Mises-statistic, as $X^\star_{i_1}$ and $X^\star_{i_2}$ are dependent. To repair this, add up the expected values for all i_2 in the same block as i_1 and take into acount that there are at most n^3 possibilities for i_1, i_3, i_4:

$$
\begin{aligned}
&\left|E\left[E^\star\left[h_2\left(X^\star_{i_1}, X^\star_{i_2}\right) h_2\left(X^\star_{i_3}, X^\star_{i_4}\right)\right]\right]\right| \\
&\qquad\qquad \leq \frac{1}{n^3} \sum_{i_1,i_3,i_4=1}^{n} \left|E\left[h_2\left(X_{i_1}, X_{i_1+k}\right) h_2\left(X_{i_3}, X_{i_4}\right)\right]\right| \\
\Rightarrow &\sum_{\substack{i_2 \\ i_1 \sim i_2}} \left|E\left[E^\star\left[h_2\left(X^\star_{i_1}, X^\star_{i_2}\right) h_2\left(X^\star_{i_3}, X^\star_{i_4}\right)\right]\right]\right| \\
&\qquad\qquad \leq \frac{1}{n^3} \sum_{i_1,i_2,i_3,i_4=1}^{n} \left|E\left[h_2\left(X_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right)\right]\right| \\
\Rightarrow &\sum_{\substack{i_1,i_2,i_3,i_4 \\ i_1 \sim i_2}} \left|E\left[E^\star\left[h_2\left(X^\star_{i_1}, X^\star_{i_2}\right) h_2\left(X^\star_{i_3}, X^\star_{i_4}\right)\right]\right]\right| \\
&\qquad\qquad\qquad\qquad \leq \sum_{i_1,i_2,i_3,i_4=1}^{n} \left|E\left[h_2\left(X_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right)\right]\right|
\end{aligned}
$$

When the indices are allocated to the blocks in another way, analogous arguments can be used. Totally, we get by Lemma 3.4 or 3.6, keeping in mind that $\frac{bl}{n} \to 1$:

$$
\begin{aligned}
E&\left[E^\star\left[\left(\frac{2}{\sqrt{bl}\,(bl-1)} \sum_{1 \leq i < j \leq bl} h_2\left(X^\star_i, X^\star_j\right)\right)^2\right]\right] \\
&\leq \frac{4}{bl\,(bl-1)^2} \sum_{i_1,i_2,i_3,i_4=1}^{n} \left|E\left[E^\star\left[h_2\left(X^\star_{i_1}, X^\star_{i_2}\right) h_2\left(X^\star_{i_3}, X^\star_{i_4}\right)\right]\right]\right| \\
&\leq \frac{K}{bl\,(bl-1)^2} \sum_{i_1,i_2,i_3,i_4=1}^{n} \left|E\left[h_2\left(X_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right)\right]\right| = O\left(n^{-\eta}\right) \qquad (27)
\end{aligned}
$$

□

Proof of Lemma 3.8: We use similar arguments as above. If $i_1, \ldots, i_8$ are in 8 different blocks, then the bootstrapped expectation is bounded by

$$
\begin{aligned}
&\left|E\left[E^\star\left[h_2\left(X^\star_{i_1}, X^\star_{i_2}\right) h_2\left(X^\star_{i_3}, X^\star_{i_4}\right) h_2\left(X^\star_{i_5}, X^\star_{i_6}\right) h_2\left(X^\star_{i_7}, X^\star_{i_8}\right)\right]\right]\right| \\
&\qquad \leq \frac{1}{n^8} \sum_{i_1,\ldots,i_8=1}^{n} \left|E\left[h_2\left(X_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right) h_2\left(X_{i_5}, X_{i_6}\right) h_2\left(X_{i_7}, X_{i_8}\right)\right]\right|.
\end{aligned}
$$

Let now lie i_1 and i_2 in the same block and the other indices in different blocks. Then add up the expectations for all i_2 in the same block as i_1:

$$\left|\sum_{i_2} E\left[E^\star\left[h_2\left(X^\star_{i_1}, X^\star_{i_2}\right) h_2\left(X^\star_{i_3}, X^\star_{i_4}\right) h_2\left(X^\star_{i_5}, X^\star_{i_6}\right) h_2\left(X^\star_{i_7}, X^\star_{i_8}\right)\right]\right]\right|$$
$$\leq \frac{1}{n^7} \sum_{i_1,\ldots,i_8=1}^{n} \left|E\left[h_2\left(X_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right) h_2\left(X_{i_5}, X_{i_6}\right) h_2\left(X_{i_7}, X_{i_8}\right)\right]\right|$$

Treating the other cases in the same way to obtain by Lemma 3.5

$$\begin{aligned} &E\left[E^\star\left[(bl)^2 U_n^{\star 4}\left(h_2\right)\right]\right] \\ &\leq \frac{K}{(bl)^2\,(bl-1)^4} \\ &\qquad \sum_{i_1,\ldots,i_8=1}^{n} \left|E\left[h_2\left(X_{i_1}, X_{i_2}\right) h_2\left(X_{i_3}, X_{i_4}\right) h_2\left(X_{i_5}, X_{i_6}\right) h_2\left(X_{i_7}, X_{i_8}\right)\right]\right| \\ &\qquad = O\left(n^{-1-\eta'}\right). \end{aligned} \tag{28}$$

□

4.2 U-Statistics under Strong Mixing

Proof of Theorem 1.8: Use the Hoeffding-decomposition:

$$\sqrt{n}\left(U_n(h) - \theta\right) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} h_1\left(X_i\right) + \sqrt{n} U_n\left(h_2\right)$$

The first summand has a normal limit with variance $4\sigma^2_\infty$ by Theorem 1.7 of Ibragimov [12]. The second summand converges in probability to zero because of Lemma 3.6. The theorem follows with the Lemma of Slutzky. □

4.3 Bootstrapping U-Statistics

Proof of Theorem 2.3: Use the Hoeffding-decomposition

$$U_n^\star(h) = \theta + \frac{2}{bl} \sum_{i=1}^{bl} h_1\left(X_i^\star\right) + U_n^\star\left(h_2\right)$$

By Theorem 2.1:

$$\left|\operatorname{Var}^\star\left[\frac{2}{\sqrt{bl}} \sum_{i=1}^{bl} h_1\left(X_i^\star\right)\right] - \operatorname{Var}\left[\frac{2}{\sqrt{n}} \sum_{i=1}^{n} h_1\left(X_i\right)\right]\right| \xrightarrow{a.s.} 0$$

By the Lemmas 3.4 or 3.6 and 3.7:

$$\operatorname{Var}\left[\sqrt{n}U_n\left(h_2\right)\right] \xrightarrow{n\to\infty} 0 \tag{29}$$

$$\operatorname{Var}^\star\left[\sqrt{bl}U_n^\star\left(h_2\right)\right] \xrightarrow{\mathcal{P}} 0 \tag{30}$$

This together proves line (13). To prove line (14), note that for every subsequence of $\left(\operatorname{Var}^\star\left[\sqrt{bl}U_n^\star\left(h_2\right)\right]\right)_{n\in\mathbb{N}}$, there exists another almost sure convergent subsequence $(n_k)_{k\in\mathbb{N}}$, and by the Lemma of Slutzky

$$\sup_{x\in\mathbb{R}}\left|P^\star\left[\sqrt{b_{n_k}l_{n_k}}\left(U_{n_k}^\star\left(h\right)-E^\star\left[U_{n_k}^\star\left(h\right)\right]\right)\leq x\right]\right.$$
$$\left.-P^\star\left[\frac{2}{\sqrt{b_{n_k}l_{n_k}}}\sum_{i=1}^{b_{n_k}l_{n_k}}\left(h_1\left(X_i^\star\right)-E^\star\left[h_1\left(X_1^\star\right)\right]\right)\leq x\right]\right| \xrightarrow{a.s.} 0. \tag{31}$$

From Lemma 3.4 or 3.6 and the Lemma of Slutzky follows:

$$\sup_{x\in\mathbb{R}}\left|P\left[\sqrt{n}\left(U_n\left(h\right)-\theta\right)\leq x\right]-P\left[\frac{2}{n}\sum_{i=1}^{n}h_1\left(X_i\right)\leq x\right]\right| \xrightarrow{n\to\infty} 0 \tag{32}$$

With Theorem 2.1 and the triangle inequality, (14) holds for the subsequence $(n_k)_{k\in\mathbb{N}}$ almost surely. Since the subsequence is arbitrary, (14) holds in probability. □

Proof of Theorem 2.4: We get from Lemma 3.8 and the Chebyshev inequality

$$P\left[\operatorname{Var}^\star\left[\sqrt{bl}U_n^\star\left(h_2\right)\right]>\epsilon\right]\leq\frac{1}{\epsilon^2}E\left[n^2U_n^4\left(h_2\right)\right]=O\left(n^{-1-\eta'}\right). \tag{33}$$

As these probabilities are summable, the convergence in line (30) holds almost surely under this conditions. □

Proof of Corollary 2.5: By Theorem 2.2, the rate of convergence follows for the variance of $\frac{2}{\sqrt{bl}}\sum_{i=1}^{bl}h_1\left(X_i^\star\right)$. The faster convergence to zero of $(bl)^2\,U_n^{\star 4}\left(h_2\right)$ (Lemma 3.8) completes the proof. □

References

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

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

[3] P.J. Bickel, D.A. Freedman, Some asymptotic theory for the bootstrap, *Ann. Stat.* **9** (1981) 1196-1217.

[4] 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.

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

[6] R.C. Bradley, *Introduction to strong mixing conditions*, volume 1-3, Kendrick Press, 2007.

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

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

[9] P. Doukhan, *Mixing*, Springer, New York, 1994.

[10] B. Efron, Bootstrap methods: another look at the jackknife, *Ann. Stat.* **7** (1979) 1-26.

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

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

[13] H.R. Kuensch, The jackknife and the bootstrap for general stationary observations, *Ann. Stat.* **17** (1989) 1217-1241.

[14] S.N. Lahiri, Theoretical comparisons of block bootstrap methods, *Ann. Stat.* **27** (1999) 386-404.

[15] D.N. Politis, J.P. Romano, A circular block resampling procedure for stationary data, in: R. Lepage. L. Billard, (Eds.) *Exploring the Limits of Bootstrap*, Wiley, New York, 1992, pp. 263-270.

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

[17] Q. Shao, H. Yu, Bootstrapping the sample means for stationary mixing sequences, *Stochastic Process. Appl.* **48** (1993) 175-190.

[18] K. Singh, On the asymptotic accuracy of Efron's bootstrap, *Ann. Stat.* **9** (1981) 1187-1195.

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