INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus
Mathématique

Herold Dehling, Davide Giraudo and Dalibor Volný

Some remarks on the ergodic theorem for U-statistics

Volume 361 (2023), p. 1511-1519

https://doi.org/10.5802/crmath.494

Les Comptes Rendus. Mathématique *sont membres du*
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org
e-ISSN : 1778-3569

Comptes Rendus
Mathématique
2023, Vol. 361, p. 1511-1519
https://doi.org/10.5802/crmath.494

Probability theory, Statistics / *Probabilités, Statistiques*

Some remarks on the ergodic theorem for U-statistics

Quelques remarques sur le théorème ergodique pour les U-statistiques

Herold Dehling [a], Davide Giraudo [*, b] and Dalibor Volný [c]

[a] Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany

[b] Institut de Recherche Mathématique Avancée UMR 7501, Université de Strasbourg and CNRS
7 rue René Descartes 67000 Strasbourg, France

[c] University de Rouen, LMRS and CNRS UMR 6085.

E-mails: herold.dehling@ruhr-uni-bochum.de, dgiraudo@unistra.fr, dalibor.volny@univ-rouen.fr

Abstract. In this note, we investigate the convergence of a U-statistic of order two having stationary ergodic data. We will find sufficient conditions for the almost sure and L^1 convergence and present some counter-examples showing that the U-statistic itself might fail to converge: centering is needed as well as finiteness of $\sup_{j\geq 2}\mathbb{E}[|h(X_1,X_j)|]$.

Résumé. Dans cette note, nous étudions le théorème ergodique pour des U-statisques d'ordre 2 dont les données sont issues d'une suite strictement stationnaire. Nous présentons des conditions suffisantes pour la convergence presque sûre et dans L^1 ainsi que des contre-exemples montrant que la U-statistique seule peut ne pas converger: un terme de centrage est requis ainsi que la finitude de $\sup_{j\geq 2}\mathbb{E}[|h(X_1,X_j)|]$.

2020 Mathematics Subject Classification. 37A30, 60F05.

Manuscript received 9 February 2023, accepted 31 March 2023.

1. Introduction

In this note, we investigate the validity of the U-statistics ergodic theorem, i.e. the almost sure convergence

$$\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n} h(X_i,X_j)\longrightarrow \iint h(x,y)\mathrm{d}F(x)\mathrm{d}F(y), \tag{1}$$

where $(X_i)_{i\geq 1}$ is a stationary ergodic process with marginal distribution F, and $h(x,y)$ is a symmetric kernel that is $F\times F$ integrable. Birkhoff's ergodic theorem establishes the analogous

* Corresponding author.

result for the time averages $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$, while Hoeffding [6] established (1) for i.i.d. processes $(X_i)_{i\geq 1}$. These two classical results naturally lead to the conjecture that (1) should hold without further assumptions, i.e. for all stationary ergodic processes $(X_i)_{i\geq 1}$ and all $L_1(F\times F)$ functions $h(x,y)$. Aaronson et al. [1] proved a partial result in this direction, namely showing that (1) holds for all $F\times F$ almost everywhere continuous and bounded kernels $h(x,y)$. At the same time, they presented counterexamples showing that (1) does not hold in full generality. One of their counterexamples is a bounded kernel where the set of discontinuities has positive $F\times F$ measure, while the other counterexample is an $F\times F$ almost everywhere continuous, but unbounded kernel.

The U-statistic ergodic theorem has subsequently been addressed by various authors, e.g. Arcones [2], Borovkova, Burton and Dehling [4]; see also the review paper by Borovkova, Burton and Dehling [5]. These papers provide both sufficient conditions for (1) to hold, as well as further counterexamples, both for stationary ergodic processes as well as under stronger mixing assumptions. Most of the positive results also address other forms of convergence in (1) such as convergence in probability and L^1-convergence. Arcones [2] proved the ergodic theorem for absolutely regular processes under some moment assumptions. Borovkova, Burton and Dehling [5] investigated convergence in probability in (1), with a special focus on the kernel $h(x,y)=\log(|x-y|)$, which arises in connection with the Takens estimator for the correlation dimension.

A common feature of all these examples is that they satisfy a modified version of the U-statistics ergodic theorem, namely

$$\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n}\left(h(X_i,X_j)-\mathbb{E}\left[h\left(X_i,X_j\right)\right]\right)\longrightarrow 0, \tag{2}$$

assuming that $\mathbb{E}\left[\left|h\left(X_i,X_j\right)\right|\right]<\infty$ for all i,j.

It might thus seem natural to conjecture that (2) holds without further assumptions. In this note, we present a counterexample that disproves this conjecture. In addition, we will give a short proof of the U-statistics ergodic theorem for bounded $F\times F$-almost everywhere continuous kernels, and give a new condition for L^1-convergence.

2. A short proof of the ergodic theorem for U-statistics

In this section, we present a short proof of the U-statistics ergodic theorem that was first established in Aaronson et al. [1]. For the special case, when the process has values in $\mathbb{R}^k$, this proof is contained in Borovkova, Burton and Dehling [5]. Here, we give the proof for processes with values in an arbitrary separable metric space.

Theorem 1. *Let $(X_k)_{k\geq 0}$ be a stationary ergodic process with values in the separable metric space S and marginal distribution F, and let $h: S\times S\to\mathbb{R}$ be a symmetric kernel that is bounded and $F\times F$-almost everywhere continuous. Then, as $n\to\infty$*

$$\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n} h(X_i,X_j)\longrightarrow\iint h(x,y)\mathrm{d}F(x)\mathrm{d}F(y),$$

almost surely.

Proof. We define the empirical distribution of the first n random variables

$$F_n=\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i},$$

where δ_x denotes the Dirac delta measure in x. For any $L_1(F)$-function $f\colon S\to\mathbb{R}$, we obtain by Birkhoff's ergodic theorem

$$\int_S f(x)\,\mathrm{d}F_n(x)=\frac{1}{n}\sum_{i=1}^{n} f(X_i)\to\int_S f(x)\mathrm{d}F(x),$$

almost surely. This convergence holds in particular for any bounded measurable function $f \in C_b(S)$. Since S is separable, there exists a countably family of functions $f_i \in C_b(S)$, $i \geq 1$, that is convergence determining, i.e. that convergence of the integrals $\int f_i(x) d\mu_n(x) \to \int f_i(x) d\mu(x)$, for all $i \geq 1$, implies weak convergence of the probability measures μ_n to μ. Now, up to a set of measure 0, we get

$$\int_S f_i(x)\,\mathrm{d}F_n(x) = \frac{1}{n}\sum_{j=1}^{n} f_i\left(X_j\right) \to \int_S f_i(x)\mathrm{d}F(x),$$

for all $i \geq 1$, and thus $F_n \Rightarrow F$ weakly. This is in fact Varadarajan's argument [8] for the fact that the empirical distribution of i.i.d. data $X_1, \ldots, X_n$ converges weakly almost surely to the true distribution F.

By Theorem 3.2 of Billingsley [3, p. 21], we obtain convergence of the empirical product measure

$$F_n \times F_n \Rightarrow F \times F,$$

except on a set of measure 0. Thus, for any bounded $F \times F$-a.e. continuous function $h \colon S \times S \to \mathbb{R}$, we obtain by the portmanteau theorem

$$\frac{1}{n^2}\sum_{1 \leq i,j \leq n} h(X_i, X_j) = \iint h(x,y)\mathrm{d}F_n(x)\,\mathrm{d}F_n(y) \to \iint h(x,y)\mathrm{d}F(x)\,\mathrm{d}F(y),$$

almost surely. Since h is bounded, we obtain $\frac{1}{n^2}\sum_{i=1}^{n} h(X_i, X_i) \to 0$, and thus

$$\frac{1}{n^2}\sum_{1 \leq i \neq j \leq n} h(X_i, X_j) \to \iint h(x,y)\mathrm{d}F(x)\,\mathrm{d}F(y),$$

almost surely. □

3. Convergence in L^1 in the ergodic theorem for U-statistics

In this section, we present two sufficient conditions for the convergence in L^1 of a U-statistic to $\iint h\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)$, where F denotes the distribution of X_0. The first sufficient condition imposes a restriction on the continuity points of the kernel combined with a uniform integrability assumption. The second sufficient condition imposes a restriction on the joint distribution of vectors $\left(X_0, X_k\right), k \geq 1$, but no other assumption is required for the kernel h.

Theorem 2. *Let $(X_i)_{i \geq 1}$ be a stationary ergodic sequence taking values in $\mathbb{R}^d$ and let $h \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a measurable function such that the family $\left\{h\left(X_1, X_j\right), j \geq 1\right\}$ is uniformly integrable. Let F be the distribution of X_1. Assume that one of the following assumptions is satisfied:*

- (A.1) *the function h is $F \times F$ almost everywhere continuous and symmetric.*
- (A.2) *$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|h\left(x,y\right)\right|\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)$ is finite, the random variable X_0 has a bounded density with respect to the Lebesgue measure on $\mathbb{R}^d$ and for each $k \geq 1$, the vector $\left(X_0, X_k\right)$ has a density f_k with respect to the Lebesgue measure on $\mathbb{R}^d \times \mathbb{R}^d$ and $\sup_{k \geq 1}\sup_{s,t \in \mathbb{R}^d} f_k\left(s,t\right)$ is finite.*

Then

$$\lim_{n \to \infty}\mathbb{E}\left[\left|\frac{1}{\binom{n}{2}}\sum_{1 \leq i < j \leq n} h\left(X_i, X_j\right) - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)\right|\right] = 0. \tag{3}$$

Proof. Let us prove Theorem 2 under assumption (A.1). By Theorem 1 in [4], we know that $\frac{1}{\binom{n}{2}}\sum_{1 \leq i < j \leq n} h\left(X_i, X_j\right) \to \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)$ in probability. Then it suffices to notice that uniform integrability of $\left\{h\left(X_1, X_j\right), j \geq 1\right\}$ implies that of $\left\{\frac{1}{\binom{n}{2}}\sum_{1 \leq i < j \leq n} h\left(X_i, X_j\right), n \geq 2\right\}$.

We will prove Theorem 2 under assumption (A.2) in three steps: first we will show that (3) holds when h is a product of indicator functions of Borel subsets of $\mathbb{R}^d$. Then we will show the result by approximating the map $\left(x,y\right) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto h\left(x,y\right)\mathbb{1}_{[-R,R]^d}\left(x\right)\mathbb{1}_{[-R,R]^d}\left(y\right)\mathbb{1}_{|h(x,y)| \leq R}$

in $L^1\left(\mathbb{P}_{(X_0,X_k)}\right)$ uniformly with respect to k by a linear combination of products of indicator functions. Finally we will conclude by uniform integrability.

First step. Assume that $h\left(x,y\right)=\mathbb{1}_A\left(x\right)\mathbb{1}_B\left(y\right)$, where A and B are Borel subsets of $\mathbb{R}^d$. Observe that

$$\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n}h\left(X_i,X_j\right)=\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n}\mathbb{1}_A\left(X_i\right)\mathbb{1}_B\left(X_j\right) \tag{4}$$

$$=\frac{1}{\binom{n}{2}}\sum_{j=2}^{n}\mathbb{1}_B\left(X_j\right)\sum_{i=1}^{j-1}\mathbb{1}_A\left(X_i\right) \tag{5}$$

$$=\frac{1}{\binom{n}{2}}\sum_{j=2}^{n}\left(j-1\right)\mathbb{1}_B\left(X_j\right)Y_j, \tag{6}$$

where

$$Y_j=\frac{1}{j-1}\sum_{i=1}^{j-1}\mathbb{1}_A\left(X_i\right). \tag{7}$$

Therefore, the following decomposition takes place:

$$\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n}h\left(X_i,X_j\right)$$
$$=\frac{1}{\binom{n}{2}}\sum_{j=2}^{n}\left(j-1\right)\mathbb{1}_B\left(X_j\right)\left(Y_j-\mathbb{P}\left(X_0\in A\right)\right)+\mathbb{P}\left(X_0\in A\right)\frac{1}{\binom{n}{2}}\sum_{j=2}^{n}\left(j-1\right)\mathbb{1}_B\left(X_j\right). \tag{8}$$

Observe that by the ergodic theorem and the Lebesgue dominated convergence theorem, the first term of the right hand side of (8) converges to 0 in L^1. Moreover, by the ergodic theorem and a summation by parts,

$$\mathbb{E}\left[\left|\frac{1}{\binom{n}{2}}\sum_{j=2}^{n}\left(j-1\right)\mathbb{1}_B\left(X_j\right)-\mathbb{P}\left(X_0\in B\right)\right|\right]\to 0, \tag{9}$$

hence we derive that

$$\lim_{n\to\infty}\mathbb{E}\left[\left|\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n}\mathbb{1}_A\left(X_i\right)\mathbb{1}_B\left(X_j\right)-\mathbb{P}\left(X_0\in A\right)\mathbb{P}\left(X_0\in B\right)\right|\right]=0 \tag{10}$$

where $\mathbb{P}\left(X_0\in A\right)\mathbb{P}\left(X_0\in B\right)=\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}h\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)$.

Second step. Let $R>0$ be fixed and define

$$h^{(R)}\left(x,y\right)=h\left(x,y\right)\mathbb{1}_{[-R,R]^d}\left(x\right)\mathbb{1}_{[-R,R]^d}\left(y\right)\mathbb{1}_{|h(x,y)|\leq R}, \tag{11}$$

which is integrable. By a standard result in measure theory, we know that for each positive ε, there exists an integer N, constants $c_1,\dots,c_N$ and sets $A_{\varepsilon,\ell},B_{\varepsilon,\ell}, 1\leq\ell\leq N$, such that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\left|h^{(R)}\left(x,y\right)-h_\varepsilon\left(x,y\right)\right|\mathrm{d}\lambda_d\left(x\right)\mathrm{d}\lambda_d\left(y\right)\leq\varepsilon, \tag{12}$$

where

$$h_\varepsilon\left(x,y\right)=\sum_{\ell=1}^{N}c_\ell\mathbb{1}_{A_{\varepsilon,\ell}}\left(x\right)\mathbb{1}_{B_{\varepsilon,\ell}}\left(y\right). \tag{13}$$

Therefore, using stationarity and the fact that $\left(X_i,X_j\right)$ has a density f_{j-i} which is bounded by a constant M independent of $\left(i,j\right)$,

$$\mathbb{E}\left[\left|h^{(R)}\left(X_i,X_j\right)-h_\varepsilon\left(X_i,X_j\right)\right|\right]=\mathbb{E}\left[\left|h^{(R)}\left(X_0,X_{j-i}\right)-h_\varepsilon\left(X_0,X_{j-i}\right)\right|\right]$$
$$=\int_{\mathbb{R}^d\times\mathbb{R}^d}\left|h^{(R)}\left(x,y\right)-h_\varepsilon\left(x,y\right)\right|f_{j-i}\left(x,y\right)\mathrm{d}\lambda_d\left(x\right)\mathrm{d}\lambda_d\left(y\right)\leq M\varepsilon$$

and

$$\mathbb{E}\left[\left|\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h^{(R)}\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)-\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h_\varepsilon\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)\right|\right]$$
$$\leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|h^{(R)}\left(x,y\right)-h_\varepsilon\left(x,y\right)\right| f_{X_0}\left(x\right) f_{X_0}\left(y\right)\mathrm{d}\lambda_d(x)\mathrm{d}\lambda_d(y)\leq \sup_{t\in\mathbb{R}^d} f_{X_0}\left(t\right)\varepsilon. \quad (14)$$

Consequently,

$$\mathbb{E}\left[\left|\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n} h^{(R)}\left(X_i,X_j\right)-\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h^{(R)}\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)\right|\right]$$
$$\leq \mathbb{E}\left[\left|\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n} h_\varepsilon\left(X_i,X_j\right)-\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h_\varepsilon\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)\right|\right]+\left(M+\sup_{t\in\mathbb{R}^d} f_{X_0}\left(t\right)\right)\varepsilon. \quad (15)$$

By the first step and the triangle inequality, we deduce that for each positive ε,

$$\limsup_{n\to\infty}\mathbb{E}\left[\left|\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n} h^{(R)}\left(X_i,X_j\right)-\int_{\mathbb{R}^d\times\mathbb{R}^d} h^{(R)}\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)\right|\right]$$
$$\leq \left(M+\sup_{t\in\mathbb{R}^d} f_{X_0}\left(t\right)\right)\varepsilon, \quad (16)$$

and hence (3) holds with h replaced by h_R.

Third step. By uniform integrability, for each positive ε, there exists δ such that for each A satisfying $\mathbb{P}(A)<\delta$, $\sup_{1\leq i<j}\mathbb{E}\left[\left|h\left(X_i,X_j\right)\right|\mathbb{1}_A\right]<\varepsilon$. Let R be such that $\mathbb{P}\left(X_1\notin[-R,R]^d\right)<\delta$, $\sup_{j\geq 2}\mathbb{E}\left[\left|h\left(X_1,X_j\right)\right|\mathbb{1}_{\{|h(X_1,X_j)|>R\}}\right]<\varepsilon$ and $\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|h\left(x,y\right)-h^{(R)}\left(x,y\right)\right|\mathrm{d}F(x)\mathrm{d}F(y)<\varepsilon$. Then for $h^{(R)}$ defined as in (11),

$$\mathbb{E}\left[\left|h\left(X_i,X_j\right)-h^{(R)}\left(X_i,X_j\right)\right|\right]$$
$$\leq \mathbb{E}\left[\left|h\left(X_i,X_j\right)\right|\left(\mathbb{1}_{\{X_i\notin[-R,R]^d\}}+\mathbb{1}_{\{X_j\notin[-R,R]^d\}}+\mathbb{1}_{\{|h(X_1,X_j)|>R\}}\right)\right]\leq 3\varepsilon \quad (17)$$

and it follows that

$$\mathbb{E}\left[\left|\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n} h\left(X_i,X_j\right)-\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)\right|\right]$$
$$\leq \mathbb{E}\left[\left|\frac{1}{\binom{n}{2}}\sum_{1\leq i<j\leq n} h^{(R)}\left(X_i,X_j\right)-\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h^{(R)}\left(x,y\right)\mathrm{d}F\left(x\right)\mathrm{d}F\left(y\right)\right|\right]+4\varepsilon, \quad (18)$$

and we conclude by the second step. This ends the proof of Theorem 2. □

4. Examples of failure of the convergence of U-statistics

Example 4.1 given in [1] shows that there exists a stationary ergodic sequence $(X_i)_{i\geq 1}$ and a bounded measurable function for which $\left(\binom{n}{2}^{-1}\sum_{1\leq i<j\leq n} h\left(X_i,X_j\right)\right)_{n\geq 2}$ converges, but not to the integral of $h\left(x,y\right)$ with respect to the product of the law of X_1.

In a similar setting, we are able to formulate two examples, the first showing that the sequence $\left(\binom{n}{2}^{-1}\sum_{1\leq i<j\leq n} h\left(X_i,X_j\right)\right)_{n\geq 2}$ may fail to converge in probability even if $\left|h\left(X_i,X_j\right)\right|$ is bounded by 1, and the second one showing that a centered U-statistic $\left(\binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}\left(h\left(X_i,X_j\right)-\mathbb{E}\left[h\left(X_i,X_j\right)\right]\right)\right)_{n\geq 2}$ may also fail to converge in probability.

We consider the transformation $Tx=2x \bmod 1$ of the unit interval $[0,1)$ equipped with the Borel sigma field $\mathscr{B}$ and Lebesgue measure λ. We define $X_0(x)=x$, $X_k(x)=T^kx$ and $U\colon L^1\to L^1$ by $UY=Y\circ T$, $Y\in L^1$.

4.1. *Example 1: non-convergence of the U-statistics*

Proposition 3. *There exists a strictly stationary ergodic sequence $(X_i)_{i\geq 1}$ and a bounded measurable symmetric function $h\colon \mathbb{R}^2\to\mathbb{R}$ such that the sequence $\left(\binom{n}{2}^{-1}\sum_{1\leq i<j\leq n} h\left(X_i,X_j\right)\right)_{n\geq 2}$ does not converge in probability.*

Proof. Let $(N_\ell)_{\ell\geq 1}$ and $\left(N'_\ell\right)_{\ell\geq 0}$ be sequences of positive integers such that $N'_0=1$ and for $\ell\geq 1$, $N_\ell<N'_\ell<N_{\ell+1}$ and

$$N'_\ell/N_\ell\geq \ell,\quad N_{\ell+1}/N'_\ell\to\infty. \tag{19}$$

We define

$$I=\bigcup_{\ell\geq 0} I_\ell,\quad I_\ell:=\left\{k\in\mathbb{N}: N'_\ell<k\leq N_{\ell+1}\right\} \tag{20}$$

$$G=\bigcup_{k\in I}\left\{\left(x,T^kx\right): x\in[0,1)\right\} \tag{21}$$

and for $x,y\in[0,1)$,

$$h(x,y)=\mathbb{1}_G(x,y)+\mathbb{1}_G(y,x).$$

Since for $i<j$ and $k\geq 1$, the equality $T^ix=T^{k+j}x$ can hold only for a countable set of x (namely, the dyadic rationals), we obtain for $1\leq i<j$ the identity $h\left(X_i,X_j\right)=\mathbb{1}_G\left(X_i,X_j\right)$ almost surely. Moreover, by definition, $\left(X_i,X_j\right)\in G$ if and only if $T^{i+k}x=T^jx$ for some $k\in I$. Almost surely, the latter identity holds if and only if $k=j-i$, and thus

$$h\left(X_i,X_j\right)=\begin{cases}1 & \text{if } j-i\in I,\\ 0 & \text{if } j-i\in\mathbb{N}\setminus I.\end{cases}$$

In particular, $\left|h\left(X_i,X_j\right)\right|\leq 1$. By (19) we have

$$\frac{1}{N_\ell(N_\ell-1)}\sum_{1\leq i<j\leq N_\ell} h\left(X_i,X_j\right)\to\frac{1}{2},\quad \frac{1}{N'_\ell(N'_\ell-1)}\sum_{1\leq i<j\leq N'_\ell} h\left(X_i,X_j\right)\to 0. \tag{22}$$

Indeed, first observe that for each integer n,

$$\sum_{1\leq i<j\leq n} h\left(X_i,X_j\right)=\sum_{j=2}^{n}\sum_{i=1}^{j-1}\mathbb{1}_{j-i\in I} \tag{23}$$

$$=\sum_{j=2}^{n}\sum_{k=1}^{j-1}\mathbb{1}_{k\in I} \tag{24}$$

$$=\sum_{k=1}^{n-1}\sum_{j=k+1}^{n}\mathbb{1}_{k\in I}=\sum_{k=1}^{n-1}(n-k)\mathbb{1}_{k\in I} \tag{25}$$

hence by definition of I, we get that for $\ell\geq 3$,

$$\frac{1}{N_\ell(N_\ell-1)}\sum_{1\leq i<j\leq N_\ell} h\left(X_i,X_j\right)=\frac{1}{N_\ell(N_\ell-1)}\sum_{u=1}^{\ell-2}\sum_{k\in I_u}(N_\ell-k)+\frac{1}{N_\ell(N_\ell-1)}\sum_{k\in I_{\ell-1}}(N_\ell-k)$$
$$=: A_\ell+B_\ell. \tag{26}$$

Note that bounding for $1\leq u\leq \ell-2$ the term $\sum_{k\in I_u}(N_\ell-k)$ by $N_\ell\operatorname{Card}(I_u)$, and $\operatorname{Card}(I_u)$ by $N_{u+1}-N_u$, we get

$$A_\ell\leq\frac{1}{N_\ell-1}\sum_{u=1}^{\ell-2}(N_{u+1}-N_u)\leq\frac{N_{\ell-1}-N_1}{N_\ell-1} \tag{27}$$

and using (19), we get $A_\ell\to 0$. Moreover,

$$B_\ell=\frac{1}{N_\ell(N_\ell-1)}\sum_{k=N'_{\ell-1}+1}^{N_\ell}(N_\ell-k)=\frac{1}{N_\ell(N_\ell-1)}\sum_{j=0}^{N_\ell-N'_{\ell-1}-1} j\sim\frac{1}{2}\frac{\left(N_\ell-N'_{\ell-1}-1\right)^2}{N_\ell^2} \tag{28}$$

hence $B_\ell \to 1/2$, which proves the first part of (22). The second one follows from the observation that $\{1,\dots,N'_\ell-1\}\cap I_u$ is empty if $u\geq \ell-1$, which gives in view of (25),

$$\begin{aligned}\frac{1}{N'_\ell(N'_\ell-1)}\sum_{1\leq i<j\leq N'_\ell} h\left(X_i,X_j\right) &= \frac{1}{N'_\ell(N'_\ell-1)}\sum_{k=1}^{N'_\ell-1}\left(N'_\ell-k\right)\mathbb{1}_{k\in I}\\ &= \frac{1}{N'_\ell(N'_\ell-1)}\sum_{u=1}^{\ell-2}\sum_{k\in I_u}\left(N'_\ell-k\right)\\ &\leq \frac{1}{N'_\ell}\sum_{u=1}^{\ell-2}\left(N_{u+1}-N'_u\right)\\ &\leq \frac{1}{N'_\ell}\sum_{u=1}^{\ell-2}\left(N_{u+1}-N_u\right)\leq \frac{N'_{\ell-1}}{N'_\ell},\end{aligned}$$

where the second inequality follows from $N_u<N'_u$, and $N'_{\ell-1}/N_\ell$ goes to 0 by (19). □

4.2. *Example 2: non-convergence of a centered U-statistic*

Proposition 4. *There exists a strictly stationary ergodic sequence $(X_i)_{i\geq 1}$ and a symmetric measurable function $h\colon \mathbb{R}^2\to\mathbb{R}$ such that for each $i<j$, $\mathbb{E}\left[\left|h\left(X_i,X_j\right)\right|\right]$ is finite but the sequence $\left(\binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}\left(h\left(X_i,X_j\right)-\mathbb{E}\left[h\left(X_i,X_j\right)\right]\right)\right)_{n\geq 2}$ does not converge in probability.*

Note that in this example, $\sup_{j\geq 2}\mathbb{E}\left[\left|h\left(X_1,X_j\right)\right|\right]$ is infinite. Moreover, the sequence $\left(\binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}\left(h\left(X_i,X_j\right)-\mathbb{E}\left[h\left(X_i,X_j\right)\right]\right)\right)_{n\geq 2}$ converges in distribution to a centered non-degenerated Gaussian random variable.

Proof. We take the same probability space and transformation as above. For $k=1,2,\dots$ define

$$\overline{G}_k = T^{-k}\left([1/2,1)\right),\ G_k=\left\{\left(x,T^kx\right): x\in\overline{G}_k\right\},\ h\left(x,y\right)=\sum_{k=1}^{\infty}a_k\left(\mathbb{1}_{G_k}\left(x,y\right)+\mathbb{1}_{G_k}\left(y,x\right)\right),$$

where

$$a_k = k^{3/2}-(k-1)^{3/2}\text{ for }k\geq 2\text{ and }a_1=1. \tag{29}$$

By similar arguments as in the proof of Proposition 3, the following equality holds almost surely for each $1\leq i<j$:

$$h\left(X_i,X_j\right)=a_{j-i}U^jf,\text{ where }f=1_{[1/2,1)},$$

hence $\mathbb{E}\left[h\left(X_i,X_j\right)\right]=a_{j-i}/2$, and

$$\sum_{1\leq i<j\leq n}h\left(X_i,X_j\right)=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}a_{j-i}U^jf=\sum_{j=2}^{n}\sum_{i=1}^{j-1}a_{j-i}U^jf=\sum_{j=2}^{n}\left(j-1\right)^{3/2}U^jf, \tag{30}$$

where $f=\mathbb{1}_{[1/2,1)}$. In order to have a better understanding of U^jf, we introduce the intervals

$$I_{j,\ell}=\left[\frac{\ell-1}{2^j},\frac{\ell}{2^j}\right),\ j\geq 1, 1\leq \ell\leq 2^j. \tag{31}$$

Lemma 5. *The sequence $\left(U^j\left(f-1/2\right)\right)_{j\geq 1}$ is a martingale difference sequence with respect to the filtration $\left(\mathscr{F}_j\right)_{j\geq 0}$, where $\mathscr{F}_j=\sigma\left(I_{j,\ell},1\leq\ell\leq 2^j\right)$ and $\mathscr{F}_0=\{\emptyset,\Omega\}$.*

Proof. We show by induction on $j\geq 1$ that

$$f\circ T^j(x)=\mathbb{1}_{\bigcup_{\ell=1}^{2^{j-1}}I_{j,2\ell}}(x). \tag{32}$$

For $j=1$, notice that if $x\in[0,1/2)$, then $f\circ T(x)=f(2x)=\mathbb{1}_{[1/2,1]}(2x)=\mathbb{1}_{[1/4,1/2)}(x)$ and if $x\in[1/2,1)$, then $f\circ T(x)=f(2x-1)=\mathbb{1}_{[1/2,1]}(2x-1)=\mathbb{1}_{[3/2,2)}(2x)=\mathbb{1}_{[3/4,1)}(x)$ hence for each $x\in[0,1)$, $f\circ T(x)=\mathbb{1}_{[1/4,1/2)}(x)+\mathbb{1}_{[3/4,1)}(x)$.

Assume now that (32) holds true for some $j \geq 1$ and let us show that

$$f \circ T^{j+1}(x) = \mathbb{1}_{\bigcup_{\ell=1}^{2^j} I_{j+1,2\ell}}(x). \tag{33}$$

By (32) with x replaced by Tx, we derive that

$$f \circ T^{j+1}(x) = \mathbb{1}_{\bigcup_{\ell=1}^{2^{j-1}} I_{j,2\ell}}(Tx). \tag{34}$$

If $x \in [0,1/2)$, then

$$\mathbb{1}_{\bigcup_{\ell=1}^{2^{j-1}} I_{j,2\ell}}(Tx) = \mathbb{1}_{\bigcup_{\ell=1}^{2^{j-1}} I_{j,2\ell}}(2x) = \mathbb{1}_{\bigcup_{\ell=1}^{2^{j-1}} I_{j+1,2\ell}}(x)$$

and (33) holds, and if $x \in [1/2,1)$, then

$$\mathbb{1}_{\bigcup_{j=1}^{2^{j-1}} I_{j,2\ell}}(Tx) = \mathbb{1}_{\bigcup_{\ell=1}^{2^{j-1}} I_{j,2\ell}}(2x-1) = \sum_{\ell=1}^{2^{j-1}} \mathbb{1}_{I_{j,2\ell}}(2x-1) = \sum_{\ell=1}^{2^{j-1}} \mathbb{1}_{I_{j,2\ell+2^j}}(x) = \mathbb{1}_{\bigcup_{\ell=2^{j-1}+1}^{2^j} I_{j+1,2\ell}}(x)$$

hence (33) also holds.

By (32), it is clear that $U^j f$ is $\mathscr{F}_j$-measurable. Moreover,

$$\mathbb{E}\left[U^{j+1}(f-1/2) \,\middle|\, \mathscr{F}_j\right] = \sum_{\ell=1}^{2^j} \mathbb{E}\left[\mathbb{1}_{I_{j+1,2\ell}} - 1/2 \,\middle|\, \mathscr{F}_j\right] = 0, \tag{35}$$

which ends the proof of Lemma 5. □

Notice that $\binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left(h\left(X_i, X_j\right) - \mathbb{E}\left[h\left(X_i, X_j\right)\right]\right) = \sum_{j=1}^n d_{n,j}$, where

$$d_{n,j} = \binom{n}{2}^{-1} (j-1)^{3/2} \left(U^j f - \frac{1}{2}\right), j \geq 2, d_{n,1} = 0. \tag{36}$$

Then $\left(d_{n,j}\right)_{j \geq 1}$ is a martingale difference sequence with respect to the filtration $\left(\mathscr{F}_j\right)_{j \geq 0}$ given as in Lemma 5. Recall that by [7], if $\left(d_{n,j}\right)_{n \geq 1, 1 \leq j \leq n}$ is an array of martingale differences, such that

$$\max_{1 \leq j \leq n} \left|d_{n,j}\right| \to 0 \text{ in probability}, \tag{37}$$

$$\text{there exists } M > 0 \text{ such that } \sup_{n \geq 1} \max_{1 \leq j \leq n} \mathbb{E}\left[d_{n,j}^2\right] \leq M \text{ and} \tag{38}$$

$$\sum_{j=1}^n d_{n,j}^2 \to \sigma^2 \text{ in probability}, \tag{39}$$

then $\sum_{j=1}^n d_{n,j}$ converges in distribution to a centered normal distribution with variance σ^2.

Noticing that $\left|U^j f(x) - 1/2\right| = 1/2$, we can see that (37) and (38) are satisfied as well as (39) with $\sigma^2 = 1/4$.

Letting $Y_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left(h\left(X_i, X_j\right) - \mathbb{E}\left[h\left(X_i, X_j\right)\right]\right)$, we thus get that $Y_n \to N(0, 1/4)$. Expressing $Y_{2n} - Y_n$ as a sum of a martingale difference array, the same argument as above gives that $Y_{2n} - Y_n$ converges in distribution to a non-degenerated normal random variable hence $(Y_n)_{n \geq 1}$ cannot converge in probability. □

Acknowledgements

The authors would like to thank the editor for a suggested improvement of the counterexamples and for having helped us to clarify where the symmetry is needed.

References

[1] J. Aaronson, R. Burton, H. Dehling, D. Gilat, T. Hill, B. Weiss, "Strong laws for L- and U-statistics", *Trans. Am. Math. Soc.* **348** (1996), no. 7, p. 2845-2866.

[2] M. A. Arcones, "The law of large numbers for U-statistics under absolute regularity", *Electron. Commun. Probab.* **3** (1998), p. 13-19.

[3] P. Billingsley, *Convergence of probability measures*, John Wiley & Sons, 1968, xii+253 pages.

[4] S. Borovkova, R. Burton, H. Dehling, "Consistency of the Takens estimator for the correlation dimension", *Ann. Appl. Probab.* **9** (1999), no. 2, p. 376-390.

[5] ———, "From dimension estimation to asymptotics of dependent U-statistics", in *Limit theorems in probability and statistics, Vol. I (Balatonlelle, 1999)*, János Bolyai Mathematical Society, 2002, p. 201-234.

[6] W. Hoeffding, *The strong law of large numbers for U-statistics*, University of North Carolina Institute of Statistics Mimeograph Series, vol. 302, 1961.

[7] D. L. McLeish, "Dependent central limit theorems and invariance principles", *Ann. Probab.* **2** (1974), p. 620-628.

[8] V. S. Varadarajan, "Weak convergence of measures on separable metric spaces", *Sankhyā, Ser. A* **19** (1958), p. 15-22.