Some notes on ergodic theorem for U-statistics of order m for stationary and not necessarily ergodic sequences

Davide Giraudo

SOME NOTES ON ERGODIC THEOREM FOR U-STATISTICS OF ORDER m FOR STATIONARY AND NOT NECESSARILY ERGODIC SEQUENCES

DAVIDE GIRAUDO

Abstract. In this note, we give sufficient conditions for the almost sure and the convergence in $\mathbb{L}^p$ of a U-statistic of order m built on a strictly stationary but not necessarily ergodic sequence.

1. Introduction and main results

Hoeffding (1948) introduced the concept of U-statistics of order $m \in \mathbb{N}^*$, defined as follows: if $(X_i)_{i\geqslant 1}$ is strictly stationary sequence taking values in a measurable space $(S,\mathcal{S})$ and $h\colon S^m \to \mathbb{R}$, the U-statistic of kernel h is given by

$$U_{m,n,h} := \frac{1}{\binom{m}{n}} \sum_{(i_\ell)_{\ell\in[\![1,m]\!]}\in \mathrm{Inc}_n^m} h\left(X_{i_1},\dots,X_{i_m}\right), \tag{1.1}$$

where $[\![1,m]\!] = \{k\in\mathbb{N}, 1\leqslant k\leqslant m\}$ and $\mathrm{Inc}_n^m = \left\{(i_\ell)_{\ell\in[\![1,m]\!]}, 1\leqslant i_1<i_2<\cdots<i_m\leqslant n\right\}$. If $(X_i)_{i\geqslant 1}$ is i.i.d. and $\mathbb{E}\left[\left|h\left(X_1,\dots,X_m\right)\right|\right]$ is finite, then $U_{m,n,h}\to\mathbb{E}\left[h\left(X_1,\dots,X_m\right)\right]$ a.s. and in $\mathbb{L}^1$. A natural question is whether for a strictly stationary sequence $(X_i)_{i\geqslant 1}$, the sequence $(U_{m,n,h})_{n\geqslant m}$ converges almost surely or in $\mathbb{L}^1$ to some random variable. Assume first that $(X_i)_{i\geqslant 1}$ is ergodic. It is shown in Aaronson et al. (1996) that if $S=\mathbb{R}$, $(X_i)_{i\geqslant 1}$ has common distribution $\mathbb{P}_{X_0}$, h is bounded and $\mathbb{P}_{X_0}\times\cdots\times\mathbb{P}_{X_0}$ almost everywhere continuous, then

$$U_{m,n,h}\to\int h\left(x_1,\dots,x_m\right)d\mathbb{P}_{X_0}\left(x_1\right)\dots d\mathbb{P}_{X_0}\left(x_m\right) \text{ a.s..} \tag{1.2}$$

Convergence in probability was also investigated in Borovkova et al. (2002). A proof of (1.2) in the context of absolutely regular sequences has been given in Arcones (1998). Moreover, Marcinkievicz law of large numbers for U-statistics of order two has been established in Dehling and Sharipov (2009) for absolutely regular sequences and Giraudo (2021) for sequences expressable as functions of an independent sequence.

It is worth pointing out that in general, the sequence $(U_{2,n,h})_{n\geqslant 2}$ may fail to converge. For instance, Aaronson et al. (1996), Example 4.5, found a non-bounded kernel h and a strictly stationary sequence such that $\limsup_{n\to\infty} U_{2,n,h}=\infty$. Moreover, the example given in Proposition 3 of Dehling et al. (2023) shows the existence of a bounded kernel h and a stationary ergodic sequence $(X_i)_{i\geqslant 1}$ such that a subsequence of $(U_{2,n,h})_{n\geqslant 2}$ converges to 0 almost surely and an other subsequence of $(U_{2,n,h})_{n\geqslant 2}$ converges to 1 almost surely. Also, as Proposition 4 shows, boundedness in $\mathbb{L}^1$ of $(h(X_1,X_j))_{j\geqslant 2}$ plays a key role, otherwise, we can find a kernel h and a strictly stationary sequence $(X_i)_{i\geqslant 1}$ for which the sequence $(U_{2,n,h}-\mathbb{E}\left[U_{2,n,h}\right])_{n\geqslant 2}$ converges to a non-degenerate normal distribution.

Some results have been established in Dehling et al. (2023), assuming that the strictly stationary sequence $(X_i)_{i\geqslant 1}$ is ergodic.

Date: February 16, 2024.

Key words and phrases. U-statistics, ergodic theorem, stationary sequences.

(1) If S is a separable metric space, $h\colon S\times S\to\mathbb{R}$ is a symmetric kernel that is bounded and $\mathbb{P}_{X_0}\times\mathbb{P}_{X_0}$-almost everywhere continuous, then, as $n\to\infty$, $U_{2,n,h}\to\int h(x,y)\,d\mathbb{P}_{X_0}(x)\,d\mathbb{P}_{X_0}(y)$ almost surely.
(2) If $S=\mathbb{R}^d$, the family $\{h(X_1,X_j),j\geqslant 1\}$ is uniformly integrable, h is $\mathbb{P}_{X_0}\times\cdots\times\mathbb{P}_{X_0}$ almost everywhere continuous and symmetric, then

$$\lim_{n\to\infty}\mathbb{E}\left[\left|U_{2,n,h}-\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}h(x,y)\,d\mathbb{P}_{X_0}(x)\,d\mathbb{P}_{X_0}(y)\right|\right]=0. \tag{1.3}$$

(3) If $S=\mathbb{R}^d$, the family $\{h(X_1,X_j),j\geqslant 1\}$ is uniformly integrable, $\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|h(x,y)|\,d\mathbb{P}_{X_0}(x)\,d\mathbb{P}_{X_0}(y)$ is finite, the random variable X_0 has a bounded density with respect to the Lebesgue measure on $\mathbb{R}^d$ and for each $k\geqslant 1$, the vector (X_0,X_k) has a density f_k with respect to the Lebesgue measure of $\mathbb{R}^d\times\mathbb{R}^d$ and $\sup_{k\geqslant 1}\sup_{s,t\in\mathbb{R}^d}f_k(s,t)$ is finite, then (1.3) holds.

Such results lead us to consider the following extensions. The case of U-statistics of order two has been adressed and we may want to extend these results to U-statistics of arbitrary order, especially because such mathematical object is widely used in statistics, for instance in Lyons (2013) for the distance covariance and stochastic geometry (see Lachièze-Rey and Reitzner (2016)). Moreover, it is a natural question to see what happens in the non-ergodic case. It is natural to consider a decomposition of Ω into ergodic components and use the results of the ergodic case to each of them. However, the multiple integral expression of the limit does not give a simple expression. Moreover, the assumptions of the ergodic case for each ergodic component, namely, almost everywhere continuity (for the product law of the marginal distribution) of the kernel and and assumption on density of the vector $(X_{i_1},\dots,X_{i_m})$ does not seem to give a tractable condition. Instead, we will use the following approach: when h is symmetric and bounded, the convergence of the considered U-statistic is viewed as the convergence of random product measures toward a product of random measures (deterministic measures in the ergodic case). When we make an assumption on the density of $(X_{i_1},\dots,X_{i_m})$, we approximate h by linear combinations of products of indicator functions. This approach has similarities with the one used in Denker and Gordin (2014). The case of products of indicators follows then from an application of the usual ergodic theorem.

We will assume that the strictly stationary sequence is such that $X_i=X_0\circ T^i$, where $T\colon\Omega\to\Omega$ is a measure preserving map. By Breiman (1968), page 107, we know that we can find a random variable X and a measure preserving map T on $\mathbb{R}^{\mathbb{Z}}$ such that the sequences $(X_i)_{i\in\mathbb{Z}}$ and $\left(X\circ T^i\right)_{i\in\mathbb{Z}}$ have the same distribution. We will study almost sure convergence and in $\mathbb{L}^p$ sequences of the form $(g_n(X_1,\dots,X_n))_{n\geqslant 1}$, where $g_n\colon S^n\to\mathbb{R}$, which has the same distribution as $\left(g_n\left(X\circ T^1,\dots,X\circ T^n\right)\right)_{n\geqslant 1}$. Therefore, their almost sure convergence are equivalent, as well as their convergence in $\mathbb{L}^p$. To ease the notations, we will write $X=X_0$.

We will study the almost sure convergence of $(U_{m,n,h})_{n\geqslant m}$ and the convergence in $\mathbb{L}^p$. We will denote by $\|Z\|_p:=\left(\mathbb{E}\left[|Z^p|\right]\right)^{1/p}$ the norm of a real-valued random variable Z.

It turns out that the limit will be expressed as an integral with respect to products of a random measure defined as follows:

$$\mu_\omega(A)=\mathbb{E}\left[\mathbf{1}_{X_0\in A}\mid\mathcal{I}\right](\omega),A\in\mathcal{B}(S), \tag{1.4}$$

where $\mathcal{I}$ denotes the σ-algebra of invariant sets, that is, the sets E such that $T^{-1}E=E$. The limit of U-statistics will be expressed as integral with respect to product measure of μ_ω, which lead us to define

$$I_m(S,h,\omega):=\int_{S^m}h(x_1,\dots,x_m)\,d\mu_\omega(x_1)\dots d\mu_\omega(x_m). \tag{1.5}$$

We will also define as $I_m(S,h,\cdot)$ the random variable given by

$$I_m(S,h,\cdot):\omega\mapsto I_m(S,h,\omega). \tag{1.6}$$

Some assumption will be made on the set of discontinuity points of h, which will be denoted by $D(h)$.

1.1. **Almost sure convergence.** Our first result deals with the almost sure convergence of a U-statistic under the assumption of boundedness of the kernel and negligibility of the set of discontinuity with respect to the product of the marginal law.

Theorem 1.1. *Let (S,d) be a separable metric space, let $(X_i)_{i\in\mathbb{Z}}=\left(X_0\circ T^i\right)_{i\in\mathbb{Z}}$ be a strictly stationary sequence. Suppose that $h\colon S^m\to\mathbb{R}$ satisfies the following assumptions:*

(A.1.1) h is symmetric, that is, $h\left(x_{\sigma(1)},\dots,x_{\sigma(m)}\right)=h(x_1,\dots,x_m)$ for each $x_1,\dots,x_m\in S$ and each bijective $\sigma\colon[\![1,m]\!]\to[\![1,m]\!]$,

(A.1.2) h is bounded and

(A.1.3) for almost every $\omega\in\Omega$, $I_m\left(S,\mathbf{1}_{D(h)},\omega\right)=0$, where $D(h)$ denotes the set of discontinuity points of h.

Then for almost every $\omega\in\Omega$, the following convergence holds:

$$\lim_{n\to\infty}U_{m,n,h}(\omega)=I_m(S,h,\omega), \tag{1.7}$$

where $I_m(S,h,\omega)$ is defined as in (1.5).

This result extends Theorem 1 in Dehling et al. (2023) in two directions: first, the case of U-statistics of arbitrary order are considered. Second, we address here the not necessarity ergodic case.

When $(X_i)_{i\geqslant1}$ is ergodic, the measure μ_ω is simply the distribution of X_0 hence the right hand side of (1.7) can be simply expressed as $\mathbb{E}\left[h\left(X_1^{(1)},\dots,X_1^{(m)}\right)\right]$, where $X_1^{(1)},\dots,X_1^{(m)}$ are independent copies of X_1.

The symmetry assumption is needed in order to relate $U_{m,n,h}$ to a sum over a rectangle and then see the convergence in (1.7) as a convergence in distribution of product of random measures.

1.2. **Convergence in $\mathbb{L}^p$, $p\geqslant1$.** In this subsection, we present sufficient conditions for the convergence in $\mathbb{L}^p$ of $(U_{m,n,h})_{n\geqslant1}$.

We start by mentioning the following consequence of Theorem 1.1.

Corollary 1.2. *Let(S,d) be a separable metric space, let $(X_i)_{i\in\mathbb{Z}}=\left(X_0\circ T^i\right)_{i\in\mathbb{Z}}$ be a strictly stationary sequence and let $p\geqslant1$. Suppose that $h\colon S^m\to\mathbb{R}$ satisfies the following assumptions:*

(A.2.1) h is symmetric, that is, $h\left(x_{\sigma(1)},\dots,x_{\sigma(m)}\right)=h(x_1,\dots,x_m)$ for each $x_1,\dots,x_m\in S$ and each bijective $\sigma\colon[\![1,m]\!]\to[\![1,m]\!]$,

(A.2.2) the family $\{|h(X_{i_1},\dots,X_{i_m})|^p,1\leqslant i_1<\dots<i_m\}$ is uniformly integrable.

(A.2.3) the following integral is finite:

$$\int_\Omega\int_{S^m}|h(x_1,\dots,x_m)|^p\,d\mu_\omega(x_1)\dots d\mu_\omega(x_m)\,d\mathbb{P}(\omega). \tag{1.8}$$

(A.2.4) for almost every $\omega\in\Omega$, $I_m\left(S,\mathbf{1}_{D(h)},\omega\right)=0$, where $D(h)$ denotes the set of discontinuity points of h.

Then the following convergence takes place:

$$\lim_{n\to\infty}\left\|U_{m,n,h}-I_m(S,h,\cdot)\right\|_p=0, \tag{1.9}$$

where $I_m(S,h,\cdot)$ is defined as in (1.6).

One can wonder what happens if we remove the symmetry assumption.

Theorem 1.3. *Let (S,d) be a separable metric space, let $(X_i)_{i\in\mathbb{Z}}=\left(X_0\circ T^i\right)_{i\in\mathbb{Z}}$ be a strictly stationary sequence taking values in S and let $p\geqslant 1$. Suppose that $h\colon S^2\to\mathbb{R}$ and $\left(X_0\circ T^i\right)_{i\in\mathbb{Z}}$ satisfy the following assumptions:*

(A.3.1) the collection $\{|h(X_i,X_j)|^p, 1\leqslant i<j\}$ is uniformly integrable.

(A.3.2) for almost every $\omega\in\Omega$, $I_m\left(S,\mathbf{1}_{D(h)},\omega\right)=0$, where $D(h)$ denotes the set of discontinuity points of h.

(A.3.3) the following integral is finite:

$$\int_\Omega\int_{S^2}|h(x,y)|^p\,d\mu_\omega(x)\,d\mu_\omega(y)\,d\mathbb{P}(\omega). \tag{1.10}$$

Then the following convergence takes place:

$$\lim_{n\to\infty}\|U_{2,n,h}-I_2(S,h,\cdot)\|_p=0, \tag{1.11}$$

where $I_2(S,h,\cdot)$ is defined as in (1.6).

This improves Theorem 2 in Dehling et al. (2023) under assumption (A.1) in the paper, since we do not require symmetry of the kernel.

One may wonder why we do not present a similar result for U-statistics of order m. A first idea would be an argument by induction on the dimension. In order to perform the induction step, say from $m=2$ to $m=3$, we would need to show, after a use of the weighted ergodic theorem, the convergence in $\mathbb{L}^p$ of $\binom{n}{2}^{-1}\sum_{1\leqslant i<j\leqslant n}h(X_{-j},X_{-i},X_0)$. Since we assume uniform integrability, it suffices to show the almost sure convergence, which could be established by seeing this almost sure convergence as that of a product of random measures. But without symmetry, we do not know whether the almost sure convergence of the sequence of random measures $\binom{n}{2}^{-1}\sum_{1\leqslant i<j\leqslant n}\delta_{(X_{-j},X_{-i})}$ takes place.

Let us now state a result on the convergence in $\mathbb{L}^p$ without imposing any continuity of the kernel, but making assumptions on the distribution of the vectors $(X_{i_1},\dots,X_{i_m})$.

Theorem 1.4. *Let $(X_i)_{i\in\mathbb{Z}}=\left(X_0\circ T^i\right)_{i\in\mathbb{Z}}$ be a strictly stationary sequence taking values in $\mathbb{R}^d$ and let $p\geqslant 1$. Suppose that $h\colon\left(\mathbb{R}^d\right)^m\to\mathbb{R}$ and $\left(X_0\circ T^i\right)_{i\in\mathbb{Z}}$ satisfy the following assumptions:*

(A.4.1) the collection $\{|h(X_{i_1},\dots,X_{i_m})|^p, 1\leqslant i_1<\dots<i_m\}$ is uniformly integrable.

(A.4.2) for each $(i_\ell)_{\ell\in[\![1,m]\!]}$ such that $1\leqslant i_1<\dots<i_m$, the vector $(X_{i_1},\dots,X_{i_m})$ has a density $f_{i_1,\dots,i_m}$ and there exists a $q_0>1$ such that

$$M_1:=\sup_{(i_\ell)_{\ell\in[\![1,m]\!]}:1\leqslant i_1<\dots<i_m}\int_{\left(\mathbb{R}^d\right)^m}f_{i_1,\dots,i_m}(t_1,\dots,t_m)^{q_0}\,dt_1\dots dt_m<\infty. \tag{1.12}$$

(A.4.3) For almost every ω, the measure μ_ω defined as in (1.4) admits a density f_ω with respect to the Lebesgue measure and there exists a set Ω' having probability one and $q_1>1$ for which

$$M_2:=\sup_{\omega\in\Omega'}\int_{\mathbb{R}^d}f_\omega(t)^{q_1}\,dt<\infty. \tag{1.13}$$

(A.4.4) the following integral is finite:

$$\int_\Omega\int_{\left(\mathbb{R}^d\right)^m}|h(x_1,\dots,x_m)|^p\,d\mu_\omega(x_1)\dots d\mu_\omega(x_m)\,d\mathbb{P}(\omega). \tag{1.14}$$

Then the following convergence hold:

$$\lim_{n\to\infty}\left\|U_{m,n,h}-I_m\left(\mathbb{R}^d,h,\cdot\right)\right\|_p=0, \tag{1.15}$$

where $I_m\left(\mathbb{R}^d, h, \cdot\right)$ *is defined as in* (1.6).

Assumption (A.4.2) is needed in order to approximate h by a linear combination of indicator functions of produts of Borel sets, uniformly with respect to the distribution of $(X_{i_1}, \ldots, X_{i_m})$.

Our Theorem 1.4 improves Theorem 2 in Dehling et al. (2023) under assumption (A.2) in the following directions. First, we provide a result for U-statistics of arbitrary order. Second, the not necessarily ergodic case is addressed. Third, even in the ergodic case, our assumption only require a uniform control on the $\mathbb{L}^{q_1}$ norm of the densities instead of a uniform bound.

1.3. **Examples.** In this Subsection, we give Examples where the conditions of Corollary 1.2, Theorems 1.1 and 1.4 are satisfied.

Corollary 1.5. *Let* $h\colon \mathbb{R}^3 \to \mathbb{R}$ *be the kernel defined as*

$$h\left(x_1, x_2, x_3\right) = \operatorname{sgn}\left(2x_1 - x_2 - x_3\right) + \operatorname{sgn}\left(2x_2 - x_1 - x_3\right) + \operatorname{sgn}\left(2x_3 - x_1 - x_2\right), \tag{1.16}$$

where $\operatorname{sgn}(x)$ *equals* 1 *if* $x > 0$, -1 *if* $x < 0$ *and* $\operatorname{sgn}(0) = 0$. *Let* $(X_i)_{i\in\mathbb{Z}} = \left(X_0 \circ T^i\right)_{i\in\mathbb{Z}}$ *be a strictly stationary sequence of real valued random variables such that for each* $x_0 \in \mathbb{R}$, $\mathbb{P}\left(X_0 = x_0\right) = 0$. *Then for almost every* $\omega \in \Omega$, *the convergence*

$$U_{3,n,h}\left(\omega\right) \to I_3\left(\mathbb{R}, h, \omega\right) \tag{1.17}$$

takes place, where $I_3\left(\mathbb{R}, h, \omega\right)$ *is defined as in* (1.5).

The kernel h defined by (1.16) is used in order to test symmetry of a distribution.
The next example deals with the distance covariance, which is used in order to test independence between two samples.

Corollary 1.6. *Let* (S, d) *be a separable metric space and let* $p \geqslant 1$. *Define* $f\colon S^4 \to \mathbb{R}$ *by*

$$f\left(z_1, z_2, z_3, z_4\right) = d\left(z_1, z_2\right) - d\left(z_1, z_3\right) - d\left(z_2, z_4\right) + d\left(z_3, z_4\right), \tag{1.18}$$

the kernel $g\colon \left(S^2\right)^6 \to \mathbb{R}$ *by*

$$g\left(\left(x_1, y_1\right), \ldots, \left(x_6, y_6\right)\right) = f\left(x_1, x_2, x_3, x_4\right) f\left(y_1, y_2, y_3, y_4\right) \tag{1.19}$$

and its symmetrized version

$$h\left(\left(x_1, y_1\right), \ldots, \left(x_6, y_6\right)\right) = \frac{1}{6!} \sum_{\sigma\in\mathcal{S}_6} g\left(\left(x_{\sigma(1)}, y_{\sigma(1)}\right), \ldots, \left(x_{\sigma(6)}, y_{\sigma(6)}\right)\right). \tag{1.20}$$

Let $\left(\left(X_i, X_i'\right)\right)_{i\in\mathbb{Z}} = \left(\left(X_0, X_0'\right) \circ T^i\right)_{i\in\mathbb{Z}}$ *be a strictly stationary sequence with values in* S^2, *where* $\left(X_0 \circ T^i\right)_{i\in\mathbb{Z}}$ *is independent of* $\left(X_0' \circ T^i\right)_{i\in\mathbb{Z}}$. *Suppose that* $\mathbb{E}\left[d\left(X_0, x_0\right)^p\right] + \mathbb{E}\left[d\left(X_0', x_0\right)^p\right]$ *is finite for some (hence all)* $x_0 \in S$. *Then the following convergence takes place*

$$\lim_{n\to\infty} \left\| U_{6,n,h} - I_6\left(S^2, h, \cdot\right)\right\|_p = 0, \tag{1.21}$$

where $I_6\left(S^2, h, \cdot\right)$ *is defined by*

$$I_6\left(S^2, h, \omega\right) = \int_{S^6} h\left(\left(x_1, y_1\right), \ldots, \left(x_6, y_6\right)\right) d\mu_\omega\left(x_1, y_1\right) \ldots d\mu_\omega\left(x_6, y_6\right) \tag{1.22}$$

and for $A \in \mathcal{B}\left(S^2\right)$,

$$\mu_\omega\left(A\right) = \mathbb{E}\left[\mathbf{1}_{\left(X_0, X_0'\right)\in A} \mid \mathcal{I}\right](\omega). \tag{1.23}$$

Finally, let us give a framework where Theorem 1.4 applies.

Corollary 1.7. *Let $h\colon \mathbb{R}^m \to \mathbb{R}$ be a measurable map and let $p \geqslant 1$. Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary Gaussian sequence. Denote by $\Sigma(i_1,\dots,i_m)$ the covariance matrix of the vector $(X_{i_1},\dots,X_{i_m})$. Suppose that*

$$\inf_{1\leqslant i_1<\dots<i_m} \det\left(\Sigma(i_1,\dots,i_m)\right) > 0, \tag{1.24}$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \left|\mathrm{Cov}\left(X_0, X_i\right)\right| = 0 \text{ and} \tag{1.25}$$

$$\mathbb{E}\left[\left|h\left(X_0^{(1)},\dots,X_0^{(m)}\right)\right|^p\right] < \infty, \tag{1.26}$$

where $X_0^{(1)},\dots,X_0^{(m)}$ are independent copies of X_0 and that (A.4.1) holds. Then (1.15) takes place.

2. Proofs

2.1. **Proof of Theorem 1.1.** The symmetry assumption guarantees the following decomposition

$$U_{m,n,h} = \frac{1}{m!\binom{n}{m}} \sum_{i_1,\dots,i_m=1}^{n} h\left(X_{i_1},\dots,X_{i_m}\right) - \frac{1}{m!\binom{n}{m}} \sum_{(i_\ell)_{\ell\in[\![1,m]\!]}\in J_n} h\left(X_{i_1},\dots,X_{i_m}\right), \tag{2.1}$$

where J_n denotes the set of elements $(i_\ell)_{\ell\in[\![1,m]\!]} \in [\![1,n]\!]^m$ for which there exist at least two distinct indexes ℓ and ℓ' for which $i_\ell = i_{\ell'}$. Since h is bounded and $\mathrm{Card}\left(J_n\right)/\binom{n}{m}$ goes to 0 as n goes to infinity, it suffices to prove that for almost every $\omega\in\Omega$,

$$\lim_{n\to\infty} \frac{1}{n^m} \sum_{i_1,\dots,i_m=1}^{n} h\left(X_{i_1}(\omega),\dots,X_{i_m}(\omega)\right) = \int_{S^m} h\left(x_1,\dots,x_m\right) d\mu_\omega\left(x_1\right)\dots d\mu_\omega\left(x_m\right), \tag{2.2}$$

where μ_ω is defined as in (1.4). Observe that for each $\omega\in\Omega$,

$$\frac{1}{n^m} \sum_{i_1,\dots,i_m=1}^{n} h\left(X_{i_1}(\omega),\dots,X_{i_m}(\omega)\right) = \int_{S^m} h\left(x_1,\dots,x_m\right) d\nu_{n,\omega}\left(x_1\right)\dots d\nu_{n,\omega}\left(x_m\right), \tag{2.3}$$

where

$$\nu_{n,\omega} = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}. \tag{2.4}$$

Separability of S guarantees the existence of a countable collection $(f_k)_{k\geqslant 1}$ of continuous and bounded functions from S to $\mathbb{R}$ such that a sequence $(\mu_n)_{n\geqslant 1}$ of probability measures converges weakly to a probability measure μ if and only if for each $k\geqslant 1$, $\int f_k d\mu_n \to \int f_k d\mu$. By the ergodic theorem, we know that for each $k\geqslant 1$, there exists a set Ω_k having probability one for which the convergence

$$\lim_{n\to\infty} \int f_k\left(x\right) d\nu_{n,\omega} = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} f_k\left(X_j\left(\omega\right)\right) = \mathbb{E}\left[f_k\left(X_0\right) \mid \mathcal{I}\right](\omega) = \int f_k\left(x\right) d\mu_\omega\left(x\right). \tag{2.5}$$

holds for each $\omega\in\Omega_k$. Therefore, for each ω belonging to the set of probability one $\Omega' := \bigcap_{k\geqslant 1}\Omega_k$, the sequence $(\nu_{n,\omega})_{n\geqslant 1}$ converges weakly to μ_ω.

Recall that Theorem 3.2 (page 21) of Billingsley (1968) shows that if $\mu_n \to \mu$ and $\mu'_n \to \mu'$ in distribution on metric spaces S_1 and S_2 respectively, then $\mu_n\times\mu'_n \to \mu\times\mu'$ in distribution on $S_1\times S_2$. Applying inductively this result and using assumptions (A.1.2) and (A.1.3) shows that for each $\omega\in\Omega'$, (1.7) holds. Indeed, we know that if S' is a separable metric space, $(m_n)_{n\geqslant 1}$ is a sequence of probability measures which converges weakly to m and $g\colon S'\to\mathbb{R}$ is bounded and $m\left(D\left(g\right)\right) = 0$, then $\int g dm_n \to \int g dm$. We use this for each fixed $\omega\in\Omega'$ with $M_n = \nu_{n,\omega}\times\dots\times\nu_{n,\omega}$, $S' = S^m$ and $g = h$.

2.2. **Proof of Corollary 1.2.** Let h_R be as in (2.13). Observe that assumption (A.2.3) guarantee that $I_m(S,h,\cdot)$ defined as in (1.5) belongs to $\mathbb{L}^p$. Moreover, the triangle inequality implies

$$\|U_{m,n,h} - I_m(S,h,\cdot)\|_p \leqslant \|U_{m,n,h} - U_{m,n,h_R}\|_p + \|U_{m,n,h_R} - I_m(S,h_R,\cdot)\|_p + \|I_m(S,h,\cdot) - I_m(S,h_R,\cdot)\|_p .$$

Using assumption (A.2.2) combined with the triangle inequality, one gets

$$\lim_{R\to\infty} \sup_{n\geqslant m} \|U_{m,n,h} - U_{m,n,h_R}\|_p = 0.$$

Moreover, assumption (A.2.3) combined with monotone convergence shows that

$$\lim_{R\to\infty} \|I_m(S,h,\cdot) - I_m(S,h_R,\cdot)\|_p = 0$$

hence it suffices to show that for each fixed $R>0$, $\|U_{m,n,h_R} - I_m(S,h_R,\cdot)\|_p \to 0$ as n goes to infinity. This follows from an application of Theorem 1.1 with h replaced by h_R (note that continuity of ϕ_R guarantees that $D(h_R) \subset D(h)$), which gives that $U_{m,n,h_R} \to I_m(S,h_R,\cdot)$ almost surely and the dominated convergence theorem allows to conclude.

2.3. **Convergence of weighted averages.** The proof of Theorems 1.3 and 1.4 rests on weighted versions of the ergodic theorem, which read as follows.

Lemma 2.1. *Let T be a measure preserving map on the probability space $(\Omega,\mathcal{F},\mathbb{P})$ and let $\mathcal{I}$ be the σ-algebra of T invariance sets. Let $p\geqslant 1$ and let $(f_j)_{j\geqslant 1}$ be a sequence of functions such that $\|f_j - f\|_p \to 0$. Then for each $m\geqslant 0$, the following convergence holds:*

$$\lim_{n\to\infty} \left\| \frac{1}{\binom{n}{m+1}} \sum_{j=m+1}^{n} \binom{j-1}{m} f_j\circ T^j - \mathbb{E}[f\mid\mathcal{I}] \right\|_p = 0. \tag{2.6}$$

Proof. First observe that since T is measure preserving,

$$\begin{aligned} &\left\| \frac{1}{\binom{n}{m+1}} \sum_{j=m+1}^{n} \binom{j-1}{m} f_j\circ T^j - \mathbb{E}[f\mid\mathcal{I}] \right\|_p \\ &\qquad \leqslant \frac{1}{\binom{n}{m+1}} \sum_{j=m+1}^{n} \binom{j-1}{m} \|f_j - f\|_p + \left\| \frac{1}{\binom{n}{m+1}} \sum_{j=m+1}^{n} \binom{j-1}{m} f\circ T^j - \mathbb{E}[f\mid\mathcal{I}] \right\|_p . \end{aligned} \tag{2.7}$$

The first term goes to zero as n goes to infinity from the elementary fact that $\sum_{i=1}^n c_i x_i / \left(\sum_{j=1}^n c_j\right) \to 0$ if $c_j>0$ and $\sum_{j=1}^n c_j \to \infty$. For the second term, we assume for without loss of generality that $\mathbb{E}[f\mid\mathcal{I}]=0$, otherwise, we replace f by $f-\mathbb{E}[f\mid\mathcal{I}]$. Let $S_j := \sum_{i=1}^j f\circ T^i$. Then

$$\sum_{j=m+1}^{n} \binom{j-1}{m} f\circ T^j = \sum_{j=m+1}^{n} \binom{j-1}{m} S_j - \sum_{j=m}^{n-1} \binom{j}{m} S_j \tag{2.8}$$

and it follows that

$$\left\| \frac{1}{\binom{n}{m+1}} \sum_{j=m+1}^{n} \binom{j-1}{m} f\circ T^j \right\|_p \leqslant \frac{\binom{n-1}{m}}{\binom{n}{m+1}} \|S_n\|_p + \frac{1}{\binom{n}{m+1}} \sum_{j=m}^{n-1} \left(\binom{j}{m} - \binom{j-1}{m} \right) \|S_j\|_p . \tag{2.9}$$

Since $\|S_n\|_p / n \to 0$, the first term of the right hand side of (2.9) goes to 0 as n goes to infinity. For the second term, one has for each $m \leqslant R \leqslant n-1$ that

$$\frac{1}{\binom{n}{m+1}} \sum_{j=m}^{n-1} \left(\binom{j}{m} - \binom{j-1}{m} \right) \|S_j\|_p \leqslant \frac{1}{\binom{n}{m+1}} \sum_{j=m}^{R} \left(\binom{j}{m} - \binom{j-1}{m} \right) \|S_j\|_p + \frac{1}{\binom{n}{m+1}} \sum_{j=R}^{n-1} j \left(\binom{j}{m} - \binom{j-1}{m} \right) \sup_{k \geqslant R} \frac{\|S_k\|_p}{k} \tag{2.10}$$

hence

$$\limsup_{n \to \infty} \frac{1}{\binom{n}{m+1}} \sum_{j=m}^{n-1} \left(\binom{j}{m} - \binom{j-1}{m} \right) \|S_j\|_p \leqslant \sup_{k \geqslant R} \frac{\|S_k\|_p}{k} \tag{2.11}$$

and we conclude using again that $\frac{\|S_k\|_p}{k} \to 0$ as k goes to infinity. □

2.4. **Proof of Theorem 1.3.** The proofs will lead us to consider truncated versions of the kernel h. Define for each fixed $R > 0$ the maps $\phi_R \colon \mathbb{R} \to \mathbb{R}$ by

$$\phi_R(t) := \begin{cases} -R & \text{if } t < -R, \\ t & \text{if } -R \leqslant t < R, \\ R & \text{if } t \geqslant R \end{cases} \tag{2.12}$$

and $h_R \colon S^m \to \mathbb{R}$ by

$$h_R(x_1, \ldots, x_m) := \phi_R(h(x_1, \ldots, x_m)), \quad x_1, \ldots, x_m \in S. \tag{2.13}$$

Then $|h_R|$ is bounded by R and since $D(h_R) \subset D(h)$, the equality $\mathbb{P}_{X_0} \times \mathbb{P}_{X_0}(D(h_R)) = 0$ holds. We claim that it suffices to prove that (1.11) holds for each R with h replaced by h_R. Indeed, by the triangle inequality,

$$\sup_{n \geqslant 2} \|U_{2,n,h} - U_{2,n,h_R}\|_p \leqslant \sup_{1 \leqslant i < j} \left\| h(X_i, X_j) \mathbf{1}_{|h(X_i, X_j)| > R} \right\|_p \tag{2.14}$$

and

$$\int_\Omega \left| \int_{S^2} h(x,y) \, d\mu_\omega(x) \, d\mu_\omega(y) - \int_{S^2} h_R(x,y) \, d\mu_\omega(x) \, d\mu_\omega(y) \right|^p d\mathbb{P}(\omega) \leqslant \int_\Omega \int_{S^2} |h(x,y)|^p \mathbf{1}_{|h(x,y)| > R} d\mu_\omega(x) \, d\mu_\omega(y) \, d\mathbb{P}(\omega), \tag{2.15}$$

hence assumptions (A.3.1) and (A.3.3) allows us to choose R making the previous quantities as small as we wish. Defining

$$d_{j,R} := \frac{1}{j} \sum_{i=1}^{j-1} h_R(X_{-i}, X_0), \tag{2.16}$$

we get that

$$U_{2,n,h_R} = \frac{1}{\binom{n}{2}} \sum_{j=2}^{n} j d_{j,R} \circ T^j. \tag{2.17}$$

We first show that there exists a set of probability one Ω' such that for each $\omega \in \Omega'$,

$$d_{j,R}(\omega) \to \int_{S^2} h_R(x,y) \, d\mu_\omega(x) \, d\delta_{X_0(\omega)} =: Y_R(\omega). \tag{2.18}$$

First, separability of S guarantees the existence of a countable collection $(f_k)_{k\geqslant 1}$ of continuous and bounded functions from S to $\mathbb{R}$ such that a sequence $(\mu_n)_{n\geqslant 1}$ of probability measures converges weakly to a probability measure μ if and only if for each $k\geqslant 1$, $\int f_k d\mu_n \to \inf f_k d\mu$.

Taking $\mu_{n,\omega} := n^{-1}\sum_{i=1}^n \delta_{X_{-i}(\omega)}$, the ergodic theorem furnishes for each $k\geqslant 1$ a set Ω_k having probability one for which the convergence

$$\lim_{n\to\infty}\int f_k(x)\,d\mu_{n,\omega} = \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n f_k(X_j(\omega)) = \mathbb{E}[f_k(X_0)\mid\mathcal{I}](\omega) = \int f_k(x)\,d\mu_\omega(x). \tag{2.19}$$

holds for each $\omega\in\Omega_k$. Consequently, for each $\omega\in\Omega' := \bigcap_{k\geqslant 1}\Omega_k$, one has $\mu_{n,\omega}\to\mu_\omega$ weakly in S and by Theorem 3.2 (page 21) of Billingsley (1968), we get that $\mu_{n,\omega}\times\delta_{X_0(\omega)}\to\mu_\omega\times\delta_{X_0(\omega)}$ weakly in S^2. Since h_R is bounded and for almost every $\omega\in\Omega$, $\mu_\omega\times\delta_{X_0(\omega)}(D(h_R)) = 0$, we get (2.18). Moreover, $|d_{j,R}|\leqslant R$ hence by dominated convergence,

$$\lim_{j\to\infty}\left\|d_{j,R} - \int_{S^2} h_R(x,y)\,d\mu_\omega(x)\,d\delta_{X_0(\omega)}\right\|_p = 0. \tag{2.20}$$

By (2.17) and the fact that T is measure preserving, we infer that

$$\|U_{2,n,h_R} - \mathbb{E}[Y_R\mid\mathcal{I}]\|_p \leqslant \frac{1}{\binom{n}{2}}\sum_{j=2}^n j\,\|d_{j,R}-Y_R\|_p + \left\|\frac{1}{\binom{n}{2}}\sum_{j=2}^n jY_R\circ T^j - \mathbb{E}[Y_R\mid\mathcal{I}]\right\|_p. \tag{2.21}$$

An application of (2.18) combined with the dominated convergence theorem shows that the first term of the right hand side of (2.21) goes to 0 as n goes to infinity. Then Lemma 2.1 with $m=1$ shows that $\|U_{2,n,h_R} - \mathbb{E}[Y_R\mid\mathcal{I}]\|_p\to 0$. It remains to check that

$$\mathbb{E}[Y_R\mid\mathcal{I}](\omega) = \int_{S^2} h(x,y)\,d\mu_\omega(x)\,d\mu_\omega(y). \tag{2.22}$$

Observe that by the ergodic theorem, $\mathbb{E}[Y_R\mid\mathcal{I}](\omega) = \lim_{n\to\infty} n^{-1}\sum_{k=1}^n Y_R(T^k\omega)$. Since $\mu_{T^k\omega} = \mu_\omega$, it follows that

$$\mathbb{E}[Y_R\mid\mathcal{I}](\omega) = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\int_{S^2} h_R(x,y)\,d\mu_\omega(x)\,d\delta_{X_k(\omega)}(y). \tag{2.23}$$

Using similar arguments as before gives that for almost every ω, $\mu_\omega\times\left(n^{-1}\sum_{k=1}^n\delta_{X_k(\omega)}\right)$ converges in distribution to $\mu_\omega\times\mu_\omega$. This ends the proof of Theorem 1.3.

2.5. **Proof of Theorem 1.4.** We start by proving Theorem 1.4 in the case where $h(x_1,\dots,x_m) = \prod_{\ell=1}^m \mathbf{1}_{x_\ell\in A_\ell}$. We show by induction over m that if $A_\ell, \ell\in[\![1,m]\!]$ are Borel subsets of $\mathbb{R}^d$ and $\left(X\circ T^i\right)_{i\in\mathbb{Z}}$ a stationary sequence with invariance σ-algebra $\mathcal{I}$, then

$$\lim_{n\to\infty}\left\|\frac{1}{\binom{n}{m}}\sum_{(i_\ell)_{\ell\in[\![1,m]\!]}\in \mathrm{Inc}_n^m}\prod_{\ell=1}^m \mathbf{1}_{X_{i_\ell}\in A_\ell} - \prod_{\ell=1}^m\mathbb{E}[\mathbf{1}_{X_0\in A_\ell}\mid\mathcal{I}]\right\|_p = 0. \tag{2.24}$$

The case $m=1$ is a direct consequence of the ergodic theorem. Let us show the case $m=2$. We start from

$$\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\mathbf{1}_{X_i\in A_1}\mathbf{1}_{X_j\in A_2} = \frac{1}{\binom{n}{2}}\sum_{j=2}^n j\left(\frac{1}{j}\sum_{k=1}^{j-1}\mathbf{1}_{X_{-k}\in A_1}\mathbf{1}_{X_0\in A_2}\right)\circ T^j, \tag{2.25}$$

where the change of index $k=j-i$ for a fixed j has been done. Let $f_j := \frac{1}{j}\sum_{k=1}^{j-1}\mathbf{1}_{X_{-k}\in A_1}\mathbf{1}_{X_0\in A_2}$. By the ergodic theorem, $f_j\to\mathbb{E}[\mathbf{1}_{X_0\in A_1}\mid\mathcal{I}]\,\mathbf{1}_{X_0\in A_2}$, which gives (2.24) for $m=2$.

Suppose now that (2.24) holds for each Borel subset $A_1,\dots,A_m$ of $\mathbb{R}^d$ and each strictly stationary sequence $\left(X_0\circ T^i\right)_{i\in\mathbb{Z}}$. Let $A_1,\dots,A_{m+1}$ be Borel subsets of $\mathbb{R}^d$ and let $\left(X_0\circ T^i\right)_{i\in\mathbb{Z}}$ be a strictly stationary sequence. We start from

$$\frac{1}{\binom{n}{m+1}}\sum_{(i_\ell)_{\ell\in[\![1,m+1]\!]}\in\mathrm{Inc}_n^{m+1}}\prod_{\ell\in[\![1,m+1]\!]}\mathbf{1}_{X_{i_\ell}\in A_\ell} = \frac{1}{\binom{n}{m+1}}\sum_{j=m+1}^{n}\left(\mathbf{1}_{X_0\in A_{m+1}}\sum_{(i_\ell)_{\ell\in[\![1,m]\!]}\in\mathrm{Inc}_{j-1}^{m}}\prod_{\ell\in[\![1,m]\!]}\mathbf{1}_{X_{i_\ell-j}\in A_\ell}\right)\circ T^j. \tag{2.26}$$

Define

$$f_j := \frac{1}{\binom{j-1}{m}}\mathbf{1}_{X_0\in A_{m+1}}\sum_{(i_\ell)_{\ell\in[\![1,m]\!]}\in\mathrm{Inc}_{j-1}^{m}}\prod_{\ell\in[\![1,m]\!]}\mathbf{1}_{X_{i_\ell-j}\in A_\ell}. \tag{2.27}$$

Doing the changes of index $k_1=j-i_m,\dots,k_m=j-i_1$, the previous expression can be rewritten as

$$f_j = \mathbf{1}_{X_0\in A_{m+1}}\frac{1}{\binom{j-1}{m}}\sum_{(k_\ell)_{\ell\in[\![1,m]\!]}\in\mathrm{Inc}_{j-1}^{m}}\prod_{\ell\in[\![1,m]\!]}\mathbf{1}_{X_{-k_\ell}\in A_{m-\ell+1}} \tag{2.28}$$

and using the induction assumption, we derive that

$$\lim_{j\to\infty}\left\|f_j-\mathbf{1}_{X_0\in A_{m+1}}\prod_{\ell\in[\![1,m]\!]}\mathbb{E}\left[\mathbf{1}_{X_0\in A_{m-\ell+1}}\mid\mathcal{I}\right]\right\|_p. \tag{2.29}$$

Then we conclude by (2.6).

We now show (1.15) in the general case. Fix a positive ε and define for a positive K

$$h^{(K)}\left(x_1,\dots,x_m\right)=h\left(x_1,\dots,x_m\right)\mathbf{1}_{|h(x_1,\dots,x_m)|\leqslant K}\prod_{\ell=1}^{m}\mathbf{1}_{|x_\ell|_d\leqslant K}, \tag{2.30}$$

where $|\cdot|_d$ denotes the Euclidean norm on $\mathbb{R}^d$. Observe that by the triangle inequality,

$$\begin{aligned}\left\|U_{m,n,h}-U_{m,n,h^{(K)}}\right\|_p &\leqslant \sup_{1\leqslant i_1<\dots<i_m}\left\|h\left(X_{i_1},\dots,X_{i_m}\right)\mathbf{1}_{|h(X_{i_1},\dots,X_{i_m})|>K}\right\|_p\\ &\quad+\sum_{\ell=1}^{m}\sup_{1\leqslant i_1<\dots<i_m}\left\|h\left(X_{i_1},\dots,X_{i_m}\right)\mathbf{1}_{|x_\ell|_d>K}\right\|_p,\end{aligned} \tag{2.31}$$

hence by assumption (A.4.1), we can find K' such that for each $K\geqslant K'$,

$$\sup_{n\geqslant m}\left\|U_{m,n,h}-U_{m,n,h^{(K)}}\right\|_p\leqslant\varepsilon. \tag{2.32}$$

Moreover, by assumption (A.4.4), we can choose K'' such that for each $K\geqslant K''$,

$$\int_\Omega I_m\left(\mathbb{R}^d,\left|h-h^{(K)}\right|^p,\omega\right)d\mathbb{P}\left(\omega\right)\leqslant\varepsilon^p. \tag{2.33}$$

Let $K_0=\max\left\{K',K''\right\}$. Observe that in assumptions (A.4.2) and (A.4.3) , we can assume without loss of generality that $q_0=q_1$. By standard results in measure theory, we know that we can find an integer J, constants $c_1,\dots,c_J$ and Borel subsets $A_{\ell,j}$, $\ell\in[\![1,m]\!],j\in[\![1,J]\!]$ such that

$$\int_{\left(\mathbb{R}^d\right)^m}\left|h^{(K_0)}\left(x_1,\dots,x_m\right)-\widetilde{h}^{(K_0)}\left(x_1,\dots,x_m\right)\right|^{p\frac{q_0}{q_0-1}}dx_1\dots dx_m<\left(M_1+M_2\right)^p\varepsilon^{p\frac{q_0}{q_0-1}}, \tag{2.34}$$

where

$$\widetilde{h}^{(K_0)}(x_1,\dots,x_m)=\sum_{j=1}^{J}c_j\prod_{\ell=1}^{m}\mathbf{1}_{x_\ell\in A_{\ell,j}}. \tag{2.35}$$

Notice that for each $1\leqslant i_1<\cdots<i_m$,

$$\begin{aligned}\left\|h^{(K_0)}(X_{i_1},\dots,X_{i_m})-\widetilde{h}^{(K_0)}(X_{i_1},\dots,X_{i_m})\right\|_p^p \\ =\int_{\left(\mathbb{R}^d\right)^m}\left|h^{(K_0)}(x_1,\dots,x_m)-\widetilde{h}^{(K_0)}(x_1,\dots,x_m)\right|^p f_{i_1,\dots,i_m}(x_1,\dots,x_m)\,dx_1\dots dx_m\end{aligned} \tag{2.36}$$

hence using Hölder's inequality, (2.34) and assumption (A.4.2), we derive that

$$\sup_{1\leqslant i_1<\cdots<i_m}\left\|h^{(K_0)}(X_{i_1},\dots,X_{i_m})-\widetilde{h}^{(K_0)}(X_{i_1},\dots,X_{i_m})\right\|_p\leqslant\varepsilon \tag{2.37}$$

and by the triangle inequality,

$$\sup_{N\geqslant m}\left\|U_{m,N,h^{(K_0)}}-U_{m,N,\widetilde{h}^{(K_0)}}\right\|_p\leqslant\varepsilon. \tag{2.38}$$

Moreover, using Hölder's inequality, we find that

$$\int_\Omega I_m\left(\mathbb{R}^d,\left|h^{(K_0)}-\widetilde{h}^{(K_0)}\right|^p,\omega\right)d\mathbb{P}(\omega)\leqslant\varepsilon^p. \tag{2.39}$$

As a consequence,

$$\begin{aligned}\left\|U_{m,n,h}-I_m\left(\mathbb{R}^d,h,\cdot\right)\right\|_p &\leqslant \sup_{N\geqslant m}\left\|U_{m,N,h}-U_{m,N,h^{(K_0)}}\right\|_p+\sup_{N\geqslant m}\left\|U_{m,N,h^{(K_0)}}-U_{m,N,\widetilde{h}^{(K_0)}}\right\|_p \\ &+\left\|U_{m,n,\widetilde{h}^{(K_0)}}-I_m\left(\mathbb{R}^d,\widetilde{h}^{(K_0)},\cdot\right)\right\|_p+\left\|I_m\left(\mathbb{R}^d,\widetilde{h}^{(K_0)},\cdot\right)-I_m\left(\mathbb{R}^d,h^{(K_0)},\cdot\right)\right\|_p \\ &+\left\|I_m\left(\mathbb{R}^d,\ h^{(K_0)},\cdot\right)-I_m\left(\mathbb{R}^d,h,\cdot\right)\right\|_p.\end{aligned}$$

By (2.24) and (2.35), we can find n_0 such that for each $n\geqslant n_0$, $\left\|U_{m,n,\widetilde{h}^{(K_0)}}-I_m\left(\mathbb{R}^d,\widetilde{h}^{(K_0)},\cdot\right)\right\|_p\leqslant\varepsilon$ hence we derive that for such n's, $\left\|U_{m,n,h}-I_m\left(\mathbb{R}^d,h,\cdot\right)\right\|_p\leqslant 4\varepsilon$. This ends the proof of Theorem 1.4.

2.6. **Proof of the results of Subsection 1.3.**

Proof of Corollary 1.5. This is an application of Theorem 1.1. Assumption (A.1.1) and (A.1.2) are clear. In order to check Assumption (A.1.3), we notice that

$$\begin{aligned}D(h)\subset\left\{(x_1,x_2,x_3)\in\mathbb{R}^3,2x_1=x_2+x_3\right\}\cup\left\{(x_1,x_2,x_3)\in\mathbb{R}^3,2x_2=x_1+x_3\right\} \\ \cup\left\{(x_1,x_2,x_3)\in\mathbb{R}^3,2x_3=x_1+x_2\right\}.\end{aligned} \tag{2.40}$$

It suffices to show that for almost every ω, $\int_{\mathbb{R}^3}\mathbf{1}_{2x_1=x_2+x_3}d\mu_\omega(x_1)\,d\mu_\omega(x_2)\,d\mu_\omega(x_3)=0$, since the treatment of the other terms is completely similar. Observe that $\int\mathbf{1}_{2x_1=x_2+x_3}d\mu_\omega(x_3)=\mathbb{E}\left[\mathbf{1}_{X_0=2x_1-x_2}\mid\mathcal{I}\right](\omega)$, and the expectation of $\mathbb{E}\left[\mathbf{1}_{X_0=2x_1-x_2}\mid\mathcal{I}\right]=0$ is zero hence this random variable equals 0 almost surely. □

Proof of Corollary 1.6. This is an application of Corollary 1.2. Assumption (A.2.1) is by definition satisfied. Let us check assumption (A.2.2). By definition of h, it suffices to check that the family $\left\{\left|f(X_{i_1},X_{i_2},X_{i_3},X_{i_4})\,f\left(X'_{i_1},X'_{i_2},X'_{i_5},X'_{i_6}\right)\right|^p,1\leqslant i_1<\cdots<i_6\right\}$ is uniformly integrable. Using the elementary fact that if $(Y_j)_{j\geqslant1}$ is independent of $\left(Y'_j\right)_{j\geqslant1}$ and both sequences are uniformly integrable, then

so is $\left(Y_j Y'_j\right)_{j\geqslant 1}$, it suffices to prove that $\{|f(X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4})|^p, 1 \leqslant i_1 < \cdots < i_4\}$ is uniformly integrable, since the argument for the other term is completely similar. By definition of f and the triangle inequality, $|f(X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4})| \leqslant 2\left(d(X_{i_1}, x_0) + d(X_{i_2}, x_0) + d(X_{i_3}, x_0) + d(X_{i_4}, x_0)\right)$ and uniform integrability follows from finiteness of $\mathbb{E}\left[d(X_0, x_0)^p\right]$. Let us check (A.2.3). Using the definition of h and f and the triangle inequality, it suffices to prove that $\int_{S^2\times S^2} d(x_1, x_0)^p\, d(y_2, x_0)^p\, d\mu_\omega(x_1, y_1)\, d\mu_\omega(x_2, y_2)$ and $\int_{S^2} d(x_1, x_0)^p\, d(y_1, x_0)^p\, d\mu_\omega(x_1, y_1)$ are finite. This follows from the fact that $\int u(x)\, v(y)\, d\mu_\omega(x, y) = \mathbb{E}\left[u(X_0)\, v(X'_0) \mid \mathcal{I}\right](\omega)$ and integrability of $d(X_0, x)^p\, d(X'_0, x)^p$. Continuity of h guarantees (A.2.4). □

Proof of Corollary 1.7. Condition (1.24) shows that the density of the vector $(X_{i_1}, \ldots, X_{i_m})$ is bounded over $\mathbb{R}^m$ and the bound is uniform with respect to $(i_1, \ldots, i_m)$. Therefore, Assumption (A.4.2) is satisfied. By (1.25), the sequence $\left(X_0 \circ T^i\right)_{i\in\mathbb{Z}}$ is ergodic (see Maruyama (1970)). (A.4.3) holds because f_ω is the density of X_0 and (A.4.4) follows from (1.26). □

Acknowledgement The author would like to thank the referees for many valuable comments that improved the note.

References

J. Aaronson, R. Burton, H. Dehling, D. Gilat, T. Hill, and B. Weiss. Strong laws for L- and U-statistics. *Trans. Amer. Math. Soc.*, 348(7):2845–2866, 1996. ISSN 0002-9947. URL https://doi.org/10.1090/S0002-9947-96-01681-9.

M. A. Arcones, *The law of large numbers for U-statistics under absolute regularity*, Electron. Comm. Probab. **3** (1998), 13–19. MR 1624866

P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.

S. Borovkova, R. Burton, and H. Dehling. From dimension estimation to asymptotics of dependent U-statistics. In *Limit theorems in probability and statistics, Vol. I (Balatonlelle, 1999)*, pages 201–234. János Bolyai Math. Soc., Budapest, 2002.

L. Breiman *Probability*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1968.

H. Dehling, D. Giraudo, and D. Volný. Some remarks on the ergodic theorem for U-statistics, 2023. C. R. Math. Acad. Sci. Paris 361 (2023), 1511–1519 URL https://10.5802/crmath.494

H. G. Dehling and O. Sh. Sharipov, *Marcinkiewicz-Zygmund strong laws for U-statistics of weakly dependent observations*, Statist. Probab. Lett. **79** (2009), no. 19, 2028–2036. MR 2571765

M. Denker and M. Gordin. Limit theorems for von Mises statistics of a measure preserving transformation. *Probab. Theory Related Fields*, 160(1-2):1–45, 2014. ISSN 0178-8051. URL https://doi.org/10.1007/s00440-013-0522-z.

D. Giraudo, *Limit theorems for U-statistics of Bernoulli data*, ALEA Lat. Am. J. Probab. Math. Stat. **18** (2021), no. 1, 793–828. MR 4243516

W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19: 293–325, 1948. ISSN 0003-4851. URL https://doi.org/10.1214/aoms/1177730196.

R. Lachièze-Rey and M. Reitzner. U-statistics in stochastic geometry. In *Stochastic analysis for Poisson point processes*, volume 7 of *Bocconi Springer Ser.*, pages 229–253. Bocconi Univ. Press, 2016.

R. Lyons. Distance covariance in metric spaces. *Ann. Probab.*, 41(5):3284–3305, 2013. ISSN 0091-1798. URL https://doi.org/10.1214/12-AOP803.

G. Maruyama Infinitely divisible processes. *Teor. Verojatnost. i Primenen.*, 15, 3–23 (1970) URL https:doi/10.1137/1115001

(†) Institut de Recherche Mathématique Avancée UMR 7501, Université de Strasbourg and CNRS 7 rue René Descartes 67000 Strasbourg, France