Limit theorems for von Mises statistics of a measure preserving transformation

Manfred Denker · Mikhail Gordin

Abstract For a measure preserving transformation T of a probability space $(X, \mathcal{F}, \mu)$ and some $d \geq 1$ we investigate almost sure and distributional convergence of random variables of the form

$$x \to \frac{1}{C_n} \sum_{0 \leq i_1, \ldots, i_d < n} f(T^{i_1}x, \ldots, T^{i_d}x), \ n = 1, 2, \ldots,$$

where $C_1, C_2, \ldots$ are normalizing constants and the kernel f belongs to an appropriate subspace in some $L_p(X^d, \mathcal{F}^{\otimes d}, \mu^d)$. We establish a form of the individual ergodic theorem for such sequences. Using a filtration compatible with T and the martingale approximation, we prove a central limit theorem in the non-degenerate case; for a class of canonical (totally degenerate) kernels and $d = 2$, we also show that the convergence holds in distribution towards a quadratic form $\sum_{m=1}^{\infty} \lambda_m \eta_m^2$ in independent standard Gaussian variables $\eta_1, \eta_2, \ldots$.

Keywords Von Mises statistic · Measure preserving transformation · Projective tensor product · Ergodic theorem · Hoeffding decomposition · Central limit theorem · Martingale approximation

M. Denker (✉)
Mathematics Department, Pennsylvania State University, State College, PA 16802, USA
e-mail: denker@math.psu.edu

M. Gordin
St. Petersburg Division, V. A. Steklov Institute of Mathematics,
27 Fontanka emb., Saint Petersburg 191023, Russia

M. Gordin
Faculty of Mathematics and Mechanics, Saint Petersburg State University,
28 Universitetski av., Petrodvorets, Saint Petersburg 198504, Russia
e-mail: gordin@pdmi.ras.ru

1 Introduction

1.1 Objectives and contents

The present paper aims to extend the theory of von Mises statistics for independent, identically distributed random variables to the realm of strictly stationary processes. Every stationary process will be investigated together with a respective measure preserving transformation of the main probability space. Such a transformation is the only structure used in the present article to establish a Strong Law of Large Numbers (SLLN) for von Mises statistics. The Central Limit Theorem (CLT) and other weak convergence results are treated in the framework of a filtration compatible with the transformation. A stationary processes generating such a filtration will appear only in applications. It turns out that a considerable part of the limit theory can be developed on this basis. One of the objectives of the paper is to show that such a relatively modest additional structure creates a suitable setting to apply some form of the martingale approximation; indeed, the latter is our main tool when proving the CLT-type results. Below, we will explain another objective of the present work and its results; the latter are collected in four statements.

Let T be a measure preserving transformation of a probability space $(X, \mathcal{F}, \mu)$. For every $d \geq 1$ and every suitable (see the next paragraph for the elaboration) measurable function $f : X^d \to \mathbb{R}$, called a *kernel*, we investigate, after normalizing appropriately, the asymptotic behavior of random variables

$$x \mapsto \sum_{0 \leq i_1 < n, \ldots, 0 \leq i_d < n} f(T^{i_1} x, \ldots, T^{i_d} x), \; n = 1, 2, \ldots, \tag{1}$$

as n tends to ∞. Every function of the form (1), normalized by some constant or not, will be called a *von Mises statistic* (or a *V-statistic*) for the transformation T and the kernel f. Notice that the same class of statistics is determined by *symmetric* kernels, so we will assume that f is symmetric whenever it is needed.

At first glance the summands in (1) can be defined in two steps. Firstly, the functions $(x_1, \ldots, x_d) \mapsto f(T^{i_1} x_1, \ldots, T^{i_d} x_d)$ can be obtained using the dynamics coordinate-wise; secondly, they should be restricted to the main diagonal of X^d. The second step, however, requires some care. Analysis and clarification of the concept of restriction became another important objective of this work. This is a crucial point determining substantially the approach in the present paper. If $f : X^d \to \mathbb{R}$ is a measurable function on the Cartesian power $(X^d, \mathcal{F}^{\otimes d}, \mu^d)$, it is viewed, as usual, not as an individual function, but rather as an equivalence class of individual functions any two of which agree on some set of measure 1. Such an equivalence class, in general, does not have a well-defined restriction to a subset of measure zero, like the main diagonal is in the case of the atomless space $(X, \mathcal{F}, \mu)$. However, some equivalence classes may contain individual functions with well-defined restrictions (for example, continuous

functions, assuming that X is the unit interval with the Lebesgue measure μ). A simple but important observation made in this article is that suitable nice functions on product probability spaces can be described in purely measure-theoretical terms. The key concept here is the *projective tensor product* of Banach spaces. First we show that, under appropriate assumptions, the elements of a respective abstract Banach space can be viewed as functions from some $L_p(\mu^d)$. In particular, every such a function determines an equivalence class discussed above. Analogously to the situation with continuous functions, nice representatives (non-unique) can be found within every such equivalence class; in view of specific properties of projective tensor products, they can be represented by absolutely convergent series of the products of functions in separate variables. Furthermore, such 'special representatives' can be restricted to the main diagonal in a correct way. Notice, that the main diagonal is considered here as a probability measure space whose measure is the image of μ under the map $x \mapsto \underbrace{(x, \ldots, x)}_{d \text{ times}}$; correctness means here that possible uncertainty in the choice of the restricted function concerns only sets of measure 0 on the diagonal. We emphasize that this procedure of 'naive restriction' applies to 'special representatives' of equivalence classes only. Different choice of a representative within the same equivalence class may lead to misunderstandings which can be observed in the literature. In the present paper, however, another approach to the restriction problem is developed. Using general properties of projective tensor products, a *restriction operator* is defined. We will see that this operator agrees with the 'naive restriction' in the case of the sums of product functions and their proper limits. On the other hand, for every equivalence class of measurable functions discussed above, the restriction operator can (or can not) be applied to the *entire equivalence class* and sends it, if applicable, to an equivalence class of functions on the diagonal; thus, no special choice of a representative within the class is needed. Moreover, we show in Proposition 2 that the correct restriction can be obtained as the result of a natural procedure combining approximation and regularization (compare with the Steklov smoothing operators and Theorem 8.4 in [29]). Finally, we obtain, along with the *correctness of the restriction*, its *continuous dependence on the kernel*; this continuity is critical for our approach. The above discussion introduces the following result which summarizes Lemma 1 and a particular case of Proposition 1 in Sect. 2 where also some information on projective tensor products can be found. We denote by $L_p(\mu^d)$ the space $L_p(X^d, \mathcal{F}^{\otimes d}, \mu^d)$ and by $|\cdot|_p$ the norm in any space L_p.

Statement A *Let $p \in [1, \infty)$ and $dr = p$. Then the projective tensor product $L_{p,\pi}(\mu^d)$ of d copies of $L_p(\mu)$ is contractively embedded into $L_p(\mu^d)$ as a dense subspace. The embedding is given by a linear map sending an elementary tensor $f_1 \otimes \cdots \otimes f_d$ to the function $(x_1, \ldots, x_d) \mapsto f_1(x_1) \cdots f_d(x_d)$. Moreover, the linear map D_d defined on elementary tensors $f_1 \otimes \cdots \otimes f_d$ by the relation*

$$D_d(f_1 \otimes \cdots \otimes f_d)(x) = f_1(x) \cdots f_d(x), \quad x \in X,$$

is a norm 1 linear map of $L_{p,\pi}(\mu^d) \subset L_p(\mu^d)$ to $L_r(\mu)$.

We shall see that the map D_d is compatible with the dynamics defined by T in the sense that for every $x \in X$ and $n_1, \ldots, n_d \in \mathbb{Z}_+$

$$D_d\big((f_1 \circ T^{n_1}) \otimes \cdots \otimes (f_d \circ T^{n_d})\big)(x) = (f_1 \circ T^{n_1})(x) \cdots (f_d \circ T^{n_d})(x).$$

For $\mathbf{k} = (k_1, \ldots, k_d)$ and $\mathbf{n} = (n_1, \ldots, n_d)$ we use the notation $\mathbf{k} < \mathbf{n}$ ($\mathbf{k} \leq \mathbf{n}$) if $k_l < n_l$ (respectively, $k_l \leq n_l$) for every $l = 1, \ldots, d$; we set for a function $f : X^d \to \mathbb{R}$

$$\big(V^{\mathbf{k}} f\big)(x_1, \ldots, x_n) = f(T^{k_1} x_1, \ldots, T^{k_d} x_d), \quad x_1, \ldots, x_d \in X.$$

The operators $V^{\mathbf{k}}$ act on every space $L_p(\mu^d)$ and also on every $L_{p, \pi}(\mu^d)$ ($1 \leq p \leq \infty$). So do the (pre-)adjoint operators $V^{*\mathbf{k}}$ (details are contained in Sect. 2).

Statement A leads to the following version of the multivariate ergodic theorem (Corollary 2 in Sect. 3).

Statement B *Let $p = dr$, $1 \leq r, p < \infty$. Then for $f \in L_{p, \pi}(\mu^d)$ we have*

$$\frac{1}{n_1 n_2 \ldots n_d} \sum_{\mathbf{0} \leq \mathbf{k} < \mathbf{n}} D_d V^{\mathbf{k}} f \to D_d E_{inv, \pi}^{\otimes d} f \tag{2}$$

almost surely and in the norm of $L_r(\mu)$ as $n_1, \ldots, n_d \to \infty$.

Here E_{inv} is the conditional expectation operator with respect to the σ-algebra of T-invariant sets, and $E_{inv, \pi}^{\otimes d}$ is the d-th projective tensor power of E_{inv}.

The distributional limit theorems rely on the Hoeffding decomposition. For every $m \in \{1, \ldots, d\}$ let $L_p^{sym}(\mu^m)$ be the subspace of symmetric elements of $L_p(\mu^m)$, $\mathcal{S}_d^m$ be the collection of all subsets of $\{1, \ldots, d\}$ of cardinality m and, for every $S \in \mathcal{S}_d^m$, let π_S be the projection map from X^d onto X^m which only keeps coordinates with indices in S. The *symmetric Hoeffding decomposition* asserts the existence of operators $R_m : L_p^{sym}(\mu^d) \to L_p^{sym}(\mu^m)$ such that every $f \in L_p^{sym}(\mu^d)$ can be represented in a unique way in the form

$$f = \sum_{m=0}^{d} \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S$$

(see Sect. 4 for details). The same or analogous notation will be applied to the spaces $L_{p, \pi}(\mu^d)$.

In the following Statement C (Theorem 2 in Sect. 7) we assume that T is an *exact* transformation in the sense that $\bigcap_{n \geq 1} T^{-n} \mathcal{F} = \mathcal{N}$, where $\mathcal{N}$ is the trivial sub-σ-field of $\mathcal{F}$. Let E denote the expectation operator. Using the Hoeffding decomposition and applying to every of its components the multiparameter martingale-coboundary representation [33], we prove

Statement C *Let T be an exact transformation and $f \in L_2^{sym}(\mu^d)$ be a real-valued kernel. Assume that for every $m = 1, \ldots, d$, $R_m f \in L_{2m, \pi}^{sym}(\mu^m)$ and the series*

$$\sum_{0 \leq \mathbf{k} < \infty} V^{*\mathbf{k}} R_m f \left(\stackrel{def}{=} \lim_{\mathbf{n} \to \infty} \sum_{0 \leq \mathbf{k} < \mathbf{n}} V^{*\mathbf{k}} R_m f \right) \tag{3}$$

converges in $L_{2m, \pi}(\mu^m)$ (here $\mathbf{k} = (k_1, \ldots, k_m)$, $\mathbf{n} = (n_1, \ldots, n_m)$). Then

$$V_n^{(d)} f \stackrel{def}{=} \frac{1}{n^{d-1/2}} \sum_{0 \leq k_1, \ldots, k_d \leq n-1} D_d V^{(k_1, \ldots, k_d)}(f - R_0 f)$$

converges in distribution to a centered Gaussian random variable with variance $d^2\sigma^2(f) \geq 0$, where

$$\sigma^2(f) = \left| \sum_{k=0}^{\infty} V^{*k} R_1 f \right|_2^2 - \left| \sum_{k=1}^{\infty} V^{*k} R_1 f \right|_2^2 \geq 0.$$

The convergence of the second moments

$$E(V_n^{(d)} f)^2 \underset{n \to \infty}{\to} d^2\sigma^2(f)$$

holds as well.

This CLT is complemented by Theorem 3 in Sect. 7 which asserts, under weaker assumptions, only the convergence of the first absolute moments (besides the convergence to the Gaussian distribution). Last, in Theorem 4 of Sect. 8, we prove the following distributional result when $d = 2$ and f is a symmetric *canonical* kernel (that is $R_0 f = 0$ and $R_1 f = 0$).

Statement D *Let $d = 2$. For every canonical f satisfying the assumptions in Statement C there exists an absolutely summable real sequence $(\lambda_m)_{m \in \mathbb{N}}$ such that the random variables*

$$\frac{1}{n} \sum_{0 \leq k_1, k_2 \leq n-1} D V^{(k_1, k_2)} f$$

converge in distribution, as $n \to \infty$, to

$$\xi = \sum_{m=1}^{\infty} \lambda_m \eta_m^2$$

where $(\eta_m)_{m \geq 1}$ is a sequence of independent standard Gaussian random variables. Moreover,

$$E\left(\frac{1}{n}\sum_{0\le i_1,\,i_2\le n-1}D_2V^{(i_1,\,i_2)}f\right)\underset{n\to\infty}{\to}\sum_{m=1}^{\infty}\lambda_m.$$

The main limit theorems are presented with proofs in Sects. 3, 7 and 8. Section 2 contains necessary preliminary material; in particular, the restriction operator is introduced there. The Hoeffding decomposition and filtrations are discussed, respectively, in Sects. 4 and 5. Section 6 contains the main part of the preparatory work for the rest of the paper. It is here that the martingale decomposition undergoes the projective tensor multiplication, leading from the classical Burkholder martingale inequality to upper bounds for certain multiparameter sums. These bounds allow (Sect. 7) to neglect the influence of higher degree summands in the Hoeffding decomposition to the asymptotic behavior when proving the CLT in the non-degenerate case. They are also applied in Sect. 8 in the proof of Statement D to show that the contribution of "partial coboundaries" vanishes in the limit; this reduces the proof to the particular case of a kernel with maximal possible martingale difference properties. Some examples (in fact, mostly general results treating entire classes of stationary processes and kernels) are collected in Sect. 9.

The above stated results, along with their modification for the case of invertible transformations (see Remark 8) and the examples in Sect. 9, clearly show that a substantial part of the limit theory for V-statistics of stationary processes can be developed, basing exclusively on projective tensor products and martingale approximations. The latter is presented only in its original primitive form (moreover, only the adapted case is considered). Using more recent developments could substantially relax many assumptions in the paper. Many other limit results can be established similarly or at the expense of small additional efforts. However, we believe that this presentation is more suitable for introducing the subject.

Remark 1 For a given function f defined on $(X^d, \mathcal{F}^{\otimes d}, \mu^d)$, a natural question arises to decide whether $f \in L_{p,\pi}(\mu^d)$ and to bound its norm. For $d = 2$ and some $p \in (1,\infty]$, $p' \in [1,\infty)$, $p^{-1} + (p')^{-1} = 1$, an equivalent question is whether the integral operator from $L_{p'}$ to L_p with the kernel f is *nuclear* [49]. There is an extensive literature on the topic, especially on nuclear (or *trace class*: see [47] and also [49] where Exercise 2.12 shows the difference between the complex and the real cases) operators in Hilbert spaces. Criteria for integral operators to be nuclear can be traced back to classical papers of Fredholm and Carleman (see monographs [28,29] and references therein; in [29] also nuclear operators in Banach spaces are considered). A special class consists of positive semidefinite kernels. For example, the well-known Mercer's theorem implies $f \in L_{2,\pi}(\mu^2)$ for such kernels under the additional assumption that X is a compact space and f is continuous.

To the best of our knowledge, for $d \ge 3$, much fewer literature exists on this topic. The main tool here is the expansion of f into a functional series whose summands are products of sufficiently regular functions in separate variables $x_1, \ldots, x_d$ (see Proposition 6 and Sect. 9 for some examples).

Remark 2 The U-statistics [that is, for symmetric kernels f, the off-diagonal modification of sums (1)] are mentioned but not treated in the present paper. Under some

strengthening our assumptions (the series in (3), (29) and (32) should converge uncon-
ditionally; for example, this will be the case if we are in the position to check the
assumptions of Proposition 6) the conclusions of Theorems 2, 3 and 4 can be refor-
mulated for U-statistics. Notice that both advantages of U-statistics compared to
V-statistics in the i.i.d. case (to be unbiased estimates of the mean value of the kernel
with i.i.d. arguments; to require weaker assumptions imposed on the kernel) in general
are no longer valid in the dependent case.

1.2 Some history and earlier results

The theory of U- and V-statistics for i.i.d. variables is well developed (see [2, 15, 17,
35, 36, 40] and references therein). Degenerate von Mises statistics for independent
variables have first been treated by von Mises in [52] and Filippova in [27]. Neuhaus
[46] proved a functional form of the weak convergence for degenerate kernels of
degree 2. Although he dealt with the U-statistics only, the method applies as well to
von Mises statistics with properly modified limit distributions. In [23] the functional
form of Filippova's result is obtained with the distributional limit presented by multiple
stochastic integrals with respect to the Kiefer–Müller process. Many fine results on
U-statistics (maximal inequalities, large deviations, functional CLT) are included or
surveyed in [17] and [45].

For non-independent random variables some progress has been made for weakly
dependent and associated processes (see [18, 19] and references therein). More gener-
ally, the Strong Law of Large Numbers (SLLN) for von Mises statistics of an ergodic
stationary real-valued processes $\xi = (\xi_n)_{n \geq 0}$ with one-dimensional distribution ν has
been treated in [1], where it is shown, among other important results and interesting
examples, that almost surely we have

$$n^{-d} \sum_{0 \leq i_1 < n, \ldots, 0 \leq i_d < n} F(\xi_{i_1}, \ldots, \xi_{i_d}) \underset{n \to \infty}{\to} \int_{X^d} F(x_1, \ldots, x_d) \nu(dx_1) \cdots \nu(dx_d), \quad (4)$$

the assumptions ranging from continuity of the kernel F to the weak Bernoulli property
of ξ. One of the results in [1] on von Mises statistics is a SLLN under the assumption
that the kernel is bounded by a product of functions in separate variables. In case
of functionals of mixing processes a form of the SLLN has been proven in [10]
which is not contained in [1]. In almost all other papers the CLT (sometimes together
with its functional form) has been considered. Yoshihara [53] was the first to give a
probabilistic treatment of the CLT question when the process is *absolutely regular*.
Other *mixing conditions* are investigated in [4,5,7–9,20,39,50,51,54]. Functionals of
absolutely regular processes have been studied in [21]. In [22] these results were used
to construct a new type of asymptotically distribution free confidence intervals for
the correlation dimension (see [34]). Later many limit results have been considerably
improved in [10] and [11] by establishing a functional form of the central limit theorem.
In the weakly dependent case we mention the works of Babbel [4,5] and Amanov [3]
where various types of mixing conditions are considered, including strong mixing.

The above list is incomplete, more information is contained in the surveys [18] and [19].

Notice that in a recent paper [41], independently of our research, for a certain class of canonical symmetric kernels of degree 2 (in 9.2.1 we call them *martingale kernels*) a limit distribution of V-statistics is derived which has the same form as in the i.i.d. case. This conclusion agrees with ours in Statement D above; the result in [41] is a rather particular case of our Statement D (see 9.2.1 for more details). The paper [41] and the subsequent papers [42,43] also develop impressive statistical applications of this and other limit results; some new, compared to [41], limit theorems in [42,43] are developed by means of methods different from those used in the present paper; the corresponding assumptions about the process include some decay of the Kantorovich distance between the conditional and the unconditional distributions of the process given its past; also some form of the Lipschitz condition is imposed on the kernel. The spectral decomposition of the kernel or, alternatively, its approximation by Lipschitz continuous wavelets are used there to derive the results.[1]

2 Preliminaries

2.1 Multiparameter actions

Let T be a measure preserving transformation of a probability space $(X, \mathcal{F}, \mu)$ (which is assumed to be standard, that is a Lebesgue space in the sense of Rokhlin [48]). For every $p \in [1, \infty]$ we set $L_p(\mu) = L_p(X, \mathcal{F}, \mu)$, choosing $\mathbb{C}$ as the field of scalars and denoting by $|\cdot|_p$ the norm of $L_p(\mu)$. Define an isometry $V : L_p(\mu) \to L_p(\mu)$ by the relation $Vf = f \circ T$. For every $p \in [1, \infty)$ let $V^* : L_{p'}(\mu) \to L_{p'}(\mu)$ be the adjoint operator of $V : L_p(\mu) \to L_p(\mu)$ where $p^{-1} + p'^{-1} = 1$. The preadjoint operator (acting in $L_1(X, \mathcal{F}, \mu)$) of the operator $V : L_\infty(\mu) \to L_\infty(\mu)$ will be loosely called the adjoint of V and denoted by V^*, too, whenever this does not lead to a misunderstanding. Analogous notations and agreements will be applied to other measure spaces, their transformations and related operators.

For every $i = 1, \ldots, d$ let $(X_i, \mathcal{F}_i, \mu_i, T_i)$ be a probability space with a measure preserving transformation T_i; let V_i, V_i^* be the corresponding operators. We assume that these spaces are copies of $(X, \mathcal{F}, \mu)$. The direct product $\prod_{1 \le i \le d}(X_i, \mathcal{F}_i, \mu_i)$ will be denoted by $(X^d, \mathcal{F}^{\otimes d}, \mu^d)$. Unlike the spaces, the transformations $T_1, \ldots, T_d$ can be different; however, from Sect. 6 on we assume that they are copies of the same transformation T. The notation $L_p(\mu^d)$ should be understood correspondingly. Let $\mathbb{Z}_+ = \{0, 1, \ldots\}$. For every $\mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{Z}_+^d$ we set $T^{\mathbf{n}}(x_1, \ldots, x_d) = (T_1^{n_1} x_1, \ldots, T_d^{n_d} x_d)$. Define a representation of the semigroup $\mathbb{Z}_+^d$ by isometries in $L_p(\mu^d)$ via

$$V^{\mathbf{n}} f = f \circ T^{\mathbf{n}}, \ f \in L_p(\mu^d).$$

---

[1] Though all this creates a favorable environment for employing our Proposition 6, we do not investigate this possible application in the present paper.

We do not assume that the transformation T is invertible. The CLT proved below will hold for the class of essentially noninvertible T (known as *exact transformations*). The family of adjoint operators $(V^{\mathbf{n}*})_{\mathbf{n}\in\mathbb{Z}_+^d}$ is also a representation of $\mathbb{Z}_+^d$ (by *coisometries* in this case). Note that these two representations do not commute with each other in the noninvertible case (otherwise they clearly commute). However, if $\mathbf{e}_1, \ldots, \mathbf{e}_d$ denote the standard basis of $\mathbb{Z}_+^d$, the operators $V^{\mathbf{e}_i}$ and $V^{*\mathbf{e}_j}$ commute for $i \neq j$ because they act on different coordinates in X^d. This will be used in the proof of Lemma 5.

2.2 Tensor products and products of functions

We discuss here conditions on kernels under which V-statistics are well-defined. Recall the concept of the projective tensor product of Banach spaces [16,49]. The main field is assumed to be $\mathbb{C}$ or $\mathbb{R}$.

Let $B_1, \ldots, B_d$ be Banach spaces with norms $|\cdots|_{B_1}, \ldots, |\cdot|_{B_d}$ and let $B_1 \otimes \ldots \otimes B_d$ be their algebraic tensor product. Elements of $B_1 \otimes \cdots \otimes B_d$, representable in the form $f_1 \otimes \cdots \otimes f_d$, are called *elementary tensors*. The *projective tensor product* of $d \geq 2$ Banach spaces denoted by $B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_d$ is, by definition, the completion of the algebraic tensor product with respect to the *projective norm* defined as the supremum of all *cross norms* on $B_1 \otimes \cdots \otimes B_d$. Recall that a norm on $B_1 \otimes \cdots \otimes B_d$ is said to be a cross norm whenever it equals $\prod_{i=1}^d |f_i|_{B_i}$ for every elementary tensor $f_1 \otimes \cdots \otimes f_d$.

Recall that for every $i = 1, \ldots, d$ $(X_i, \mathcal{F}_i, \mu_i)$ is a copy of $(X, \mathcal{F}, \mu)$. For $p_1, \ldots, p_d \in [1, \infty]$ we denote by $|\cdot|_{p_1, \ldots, p_d, \pi}$ the norm of the space

$$L_{p_1}(X_1, \mathcal{F}_1, \mu_1) \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi L_{p_d}(X_d, \mathcal{F}_d, \mu_d).$$

If $p_1 = \cdots = p_d = p \in [1, \infty]$, the above projective tensor product and its norm will be denoted by $L_{p,\pi}(\mu^d)$ and $|\cdot|_{p,d,\pi}$, respectively. We show in the following lemma that $L_{p,\pi}(\mu^d)$ can be thought of as a subspace of $L_p(\mu^d)$; hence, its elements can be viewed as functions on X^d. Some useful properties of these functions are established in 2.3.

Lemma 1 *For every $p \in [1, \infty]$ there exists a unique linear map*

$$J_d : L_{p,\pi}(\mu^d) \to L_p(\mu^d)$$

of norm 1 *which sends every elementary tensor $f_1 \otimes \cdots \otimes f_d$ to the function $(x_1, \ldots, x_d) \mapsto f_1(x_1) \cdots f_d(x_d)$. Moreover, J_d maps $L_{p,\pi}(\mu^d)$ into $L_p(\mu^d)$ injectively. For $p \in [1, \infty) J_d(L_{p,\pi}(\mu^d))$ is dense in $L_p(\mu^d)$.*

Proof The case $d = 1$ is trivial, so we assume $d \geq 2$. For every $p \in [1, \infty]$, let us define a linear map J_d of norm 1,

$$J_d : L_{p,\pi}(\mu^d) \to L_p(\mu^d).$$

When we need to specify p we shall use the notation $J_{d,p}$. First, sending every elementary tensor $f_1 \otimes \cdots \otimes f_d$ to the function $(x_1, \ldots, x_d) \mapsto f_1(x_1) \cdots f_d(x_d)$, we define a d-linear map of norm 1 from $L_p(X_1, \mathcal{F}_1, \mu_1) \times \cdots \times L_p(X_d, \mathcal{F}_d, \mu_d)$ to $L_p(\mu^d)$. Then, by a general property of the projective tensor product (see [49], Theorem 2.9, for $d = 2$; use induction and associativity for $d > 2$) this map extends to $L_{p,\pi}(\mu^d)$ uniquely with norm 1. Denote this resulting map by J_d. Its image is dense in $L_{p,\pi}(\mu^d)$ for $p < \infty$ since so is the image under J_d of the algebraic tensor product.

We prove now that $J_d (= J_{d,p})$ is injective. For $p = 1$ it is so because $J_{d,1}$ is an isometric isomorphism between its domain and its range ([49], Exercise 2.8). Let now for some $p > 1$ $I_{1,p} : L_p(\mu) \to L_1(\mu)$ and $I_{d,p} : L_p(\mu^d) \to L_1(\mu^d)$ be the inclusion operators (of norm 1 each). By the *metric mapping property* ([16], 12.1) of the projective tensor norm, the inclusion $I_{1,p}$ gives rise to the norm 1 mapping $A_d : L_{p,\pi}(\mu^d) \to L_{1,\pi}(\mu^d)$ (notice that $L_{1,\pi}(\mu^d)$ and $L_1(\mu^d)$ are identified by $J_{d,1}$). Since the spaces L_p have the *approximation property*, the operator A_d is injective as a projective tensor product of injective operators $I_{1,p}$ (see Corollary 4 (1), subsection 5.8, in [16]; then use induction). Starting with algebraic tensor products and passing, in view of boundedness of all operators involved, to the completions with respect to corresponding norms, we obtain that the mappings $J_{d,1} A_d : L_{p,\pi}(\mu^d) \to L_1(\mu^d)$ and $I_{d,p} J_{d,p} : L_{p,\pi}(\mu^d) \to L_1(\mu^d)$ agree. Since A_d and $J_{d,1}$ are injective, so is $J_{d,p}$. $\qquad\qquad\square$

In view of the properties of J_d we shall, when possible, omit the symbol J_d and consider $L_{p,\pi}(\mu^d)$ as a subspace of $L_p(\mu^d)$. Set for $\mathbf{n} = (n_1, \ldots, n_d)$

$$V_\pi^{\mathbf{n}} = V_1^{n_1} \otimes_\pi \cdots \otimes_\pi V_d^{n_d}, \quad V_\pi^{*\mathbf{n}} = V_1^{*n_1} \otimes_\pi \cdots \otimes_\pi V_d^{*n_d}. \tag{5}$$

The operators $(V_\pi^{\mathbf{n}}, V_\pi^{*\mathbf{n}})_{\mathbf{n} \in \mathbb{Z}_+^d}$ have properties very similar to those of $(V^{\mathbf{n}}, V^{*\mathbf{n}})_{\mathbf{n} \in \mathbb{Z}_+^d}$; in particular, they have norm 1 with respect to the projective tensor norm. The relations $J_{d,p} V_\pi^{\mathbf{n}} = V^{\mathbf{n}} J_{d,p}, J_{d,p} V_\pi^{*\mathbf{n}} = V^{*\mathbf{n}} J_{d,p}, \mathbf{n} \in \mathbb{Z}_+^d$, are obvious for elementary tensors and immediately extend to the general case. It follows from these relations that the space $L_{p,\pi}(\mu^d)$ is preserved by the operators $(V^{\mathbf{n}}, V^{*\mathbf{n}})_{\mathbf{n} \in \mathbb{Z}_+^d}$. From now on we shall use the notation $(V^{\mathbf{n}}, V^{*\mathbf{n}})_{\mathbf{n} \in \mathbb{Z}_+^d}$ also to denote the restrictions of these families to the space $L_{p,\pi}(\mu^d) \subset L_p(\mu^d)$.

Remark 3 The space $L_{2,\pi}(\mu^2)$ can be identified with the space of nuclear (or *trace class*) operators from $L_2(\mu)^*$ to $L_2(\mu)$ ([49]). The operator J_2 in Lemma 1 transforms such (integral) operators to their kernels which form a subspace of $L_2(\mu^2)$.

2.3 Restriction to the diagonal

In the following Proposition 1, for every $p_1, \ldots, p_d \in [1, \infty]$ with $p_1^{-1} + \cdots + p_d^{-1} = 1$ and for every $f \in L_{p_1}(\mu) \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi L_{p_d}(\mu)$, we define a function $D_d f \in L_1(\mu)$. In the case of $1 \leq p_1 = \cdots = p_d = p \leq \infty$ the embedding J_d (Lemma 1) allows us to consider the space $L_p(\mu) \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi L_p(\mu)$ as a subspace of the $L_p(\mu^d)$ and interpret

its elements as functions defined on X^d. Then $D_d f$ plays the role of the restriction of f to the principal diagonal $\{(x_1, \ldots, x_d) : x_1 = \cdots = x_d)\} \subset X^d$. In this particular case the term 'restriction' can be justified by an approximation procedure described in Proposition 2 below.

Proposition 1 *Let $p_1, \ldots, p_d \in [1, \infty], r \in [1, \infty]$ satisfy*

$$\sum_{i=1}^{d} \frac{1}{p_i} = \frac{1}{r}.$$

Then

(1) *the map $\mathcal{D}$, sending every d-tuple $(f_1, \ldots, f_d) \in L_{p_1}(\mu) \times \cdots \times L_{p_d}(\mu)$ to the function*

$$x \mapsto f_1(x) \cdots f_d(x),$$

is a norm 1 d-linear map from $L_{p_1}(\mu) \times \cdots \times L_{p_d}(\mu)$ to $L_r(\mu)$;
(2) *there exists a unique linear map (of norm 1)*

$$D_d : L_{p_1}(\mu) \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi L_{p_d}(\mu) \to L_r(\mu)$$

such that for every $d-$tuple $(f_1, \ldots, f_d) \in L_{p_1}(\mu) \times \cdots \times L_{p_d}(\mu)$

$$D_d(f_1 \otimes \ldots \otimes f_d) = \mathcal{D}(f_1, \ldots, f_d).$$

Proof The first assertion is a consequence of the multiple Hölder inequality (Exercise 6.11.2 in [24]). The second one follows from the linearization property of the projective tensor products with respect to polylinear maps. For the case of bilinear maps see Theorem 2.9 in [49]; for $d > 2$ use induction and associativity. □

If $p_1 = \cdots = p_d = p$, the space $L_{p,\pi}(\mu^d) = L_p(\mu) \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi L_p(\mu)$ is embedded into $L_p(\mu^d)$ by the operator J_d (Lemma 1); we omit J_d and treat an $f \in L_{p,\pi}(\mu^d)$ as a function. For every finite measurable partition $\mathcal{A} = \{A_1 \ldots, A_m\}$ let us denote by $\mathcal{F}_\mathcal{A}$ the σ-field of all possible unions of atoms of $\mathcal{A}$ and by $E(\cdot \,|\, \mathcal{A})$ the corresponding conditional expectation. Let $(\mathcal{A}_n)_{n \geq 1}$ be a refining sequence of finite measurable partitions $\mathcal{A}_n = \{A_{1,n}, \ldots, A_{m_n, n}\}$ such that $\mathcal{F}$ is the smallest σ-field containing all $\mathcal{F}_{\mathcal{A}_n}, n \geq 1$. Let I_A denote the indicator of the set A.

Proposition 2 *Let $d \geq 1, p_1 = \cdots = p_d = p \in [d, \infty)$ and $r = p/d$. Define the sequence $(D_{d,n})_{n \geq 1}$ of operators $D_{d,n} : L_{p,\pi}(\mu^d) \to L_r(\mu)$ by*

$$D_{d,n} f = \sum_{i=1}^{m_n} \frac{I_{A_{i,n}}}{\mu(A_{i,n})^d} \int_{A_{i,n}^d} f(x_1, \ldots, x_d) \mu(dx_1) \cdots \mu(dx_d).$$

Then $D_{d,n} \underset{n \to \infty}{\to} D_d$ in the strong operator topology.

Proof First let us verify (again using the Hölder inequality) that $D_{d,n}$ as a map from $L_{p,\pi}(\mu^d)$ to $L_r(\mu)$ does not increase the norms of elementary tensors. From the relation

$$D_{d,n}(f_1 \otimes \cdots \otimes f_d) = D_d(E(f_1|\mathcal{A}_n) \otimes \cdots \otimes E(f_d|\mathcal{A}_n))$$
$$= E(f_1|\mathcal{A}_n) \cdots E(f_d|\mathcal{A}_n) \qquad (6)$$

it follows that

$$|D_{d,n}(f_1 \otimes \cdots \otimes f_d)|_r = |E(f_1|\mathcal{A}_n) \cdots E(f_d|\mathcal{A}_n)|_r$$
$$\leq |E(f_1|\mathcal{A}_n)|_p \cdots |E(f_d|\mathcal{A}_n)|_p \leq |f_1|_p \cdots |f_d|_p.$$

By the properties of the projective norm, this implies that the norm of every $D_{d,n} :$ $L_{p,\pi}(\mu^d) \to L_r(\mu)$ is also bounded by 1.

Now, using (6), standard properties of conditional expectations and the Hölder inequality, we obtain

$$|D_{d,n}(f_1 \otimes \cdots \otimes f_d) - f_1 \cdots f_d|_r \leq \sum_{k=1}^{d} |E(f_k|\mathcal{A}_n) - f_k|_p \prod_{m=1,\, m\neq k}^{d} |f_m|_p.$$

From the martingale convergence theorem for the space L_p we conclude that every sequence $(D_{d,n}(f_1 \otimes \cdots \otimes f_d))_{n\geq 1}$ converges in the norm of $L_r(\mu)$ to the function $f_1(\cdot) \cdots f_d(\cdot)$. The analogous conclusion holds for finite linear combinations of elementary tensors. Since the norms of the operators $D_{d,n}$ are uniformly bounded, the proposition follows. $\square$

The following corollary will be used in the proof of Proposition 3.

Corollary 1 *The restriction operator D_d preserves positivity of real valued functions.*

Thus, the function $D_d f \in L_r(\mu)$ is a well-defined substitute for the naive restriction of f to the principal diagonal. For example, for $\mathbf{n} = (n_1, \ldots, n_d)$ the function $D_d V^{\mathbf{n}} f$ can be viewed as a substitute for the function $x \mapsto f(T^{n_1}x, \ldots, T^{n_d}x)$.

3 Strong law of large numbers

3.1 A multivariate ergodic theorem

If T is an ergodic transformation of a probability space, a von Mises statistic may be considered as an estimate for the multiple integral of the kernel with respect to the invariant measure. Consistency is one of the desirable statistical properties of a sequence of estimates; this raises the question of an appropriate ergodic theorem. Proposition 3, the main result of this subsection, states such a theorem in a general setting. It asserts, in the ergodic case, the convergence of multiparameter sums (7) to the average of the kernel with respect to the product measure. This reminds of a

Wiener-type ergodic theorem ([24], Theorem 8.6.9) specialized to the case of d one-parameter coordinatewise actions on the product of d probability spaces. However, not only the assumptions, but also the conclusions in these results are different: unlike the Wiener theorem, our result asserts the convergence for almost all initial points with respect to a probability measure which is in general *neither absolutely continuous* with respect to the product measure (being supported on the main diagonal) *nor invariant* under the multiparameter action.

We do not assume here symmetry of the kernel and perform summation over rectangular coordinate domains (which is common in the multiparameter ergodic theorems, see [24], Chapter 8) rather than over coordinate cubes involved in the definition of V-statistics. In this subsection we consider several possibly different μ–preserving transformations $T_1, \ldots, T_d$ of the space $(X, \mathcal{F}, \mu)$, using the notation $T^{(n_1,...,n_d)}(x_1, \ldots, x_d) = (T_1^{n_1} x_1, \ldots, T_d^{n_d} x_d)$ and $V^{(n_1,...,n_d)} f = f \circ T^{(n_1,...,n_d)}$.

Transformations considered in this subsection in general are not ergodic, so we need some notations to include the non-ergodic case. Recall that $A \in \mathcal{F}$ is said to be $T-$*invariant* if $T^{-1} A = A$. For every $l \in \{1, \ldots, d\}$ let $\mathcal{F}_{inv, l}$ denote the σ-field of all T_l-invariant measurable sets in $(X, \mathcal{F}, \mu)$, and let $E_{inv, l}$ be the corresponding conditional expectation considered as an operator in $L_{p_l}(X, \mathcal{F}, \mu)$.

Proposition 3 *Let $p = rd$ for some integer $d \geq 1$ and a real number $r \in [1, \infty)$. Let $T_1, \ldots, T_d$ be measure preserving transformations of a probability space $(X, \mathcal{F}, \mu)$ and $f \in L_{p,\pi}(\mu^d)$. Then, with $\mathbf{n} = (n_1, \ldots, n_d)$, we have*

$$\frac{1}{n_1 \cdots n_d} \sum_{\mathbf{0} \leq \mathbf{k} < \mathbf{n}} D_d V^{\mathbf{k}} f \underset{n_1, ..., n_d \to \infty}{\to} D_d(E_{inv, 1} \otimes_\pi \cdots \otimes_\pi E_{inv, d}) f \qquad (7)$$

with probability 1 and in $L_r(\mu)$.

Remark 4 The main point of Proposition 3 is the convergence with probability 1 in the case $d \geq 2$. As to the convergence in L_r, it is not hard to prove, for every $d \geq 2$ and $p_1, \ldots, p_d, r \in (1, \infty)$, satisfying $\sum_{i=1}^d p_i^{-1} = r^{-1}$, the following multiple statistical ergodic theorem:

$$\frac{1}{n_1 \cdots n_d} \sum_{\mathbf{0} \leq \mathbf{k} < \mathbf{n}} V^{\mathbf{k}} \underset{n_1, ..., n_d \to \infty}{\to} (E_{inv, 1} \otimes_\pi \cdots \otimes_\pi E_{inv, d}),$$

asserting the strong convergence in the space $L_{p_1}(X_1, \mathcal{F}_1, \mu_1) \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi L_{p_d} (X_d, \mathcal{F}_d, \mu_d)$. Applying the operator D_d to the both sides of this relation, we obtain the convergence in the L_r-norm. Choosing $p_1 = \cdots = p_d = rd = p$, the convergence in $L_r(\mu)$ in Proposition 3 follows for $d \geq 2$. The proof of Proposition 3 contains a second argument of this fact.

The next lemma will be used in the proof of Proposition 3.

Lemma 2 *Let d, p, r and the transformations $T_1, \ldots, T_d$ satisfy the conditions of Proposition 3. Let, moreover, $p > 1$. Then there exists a constant $C = C(r, d)$ such*

that for every $f \in L_{p,\pi}(\mu^d)$ we have the inequality

$$\left\| \sup_{\substack{1 \leq n_1 < \infty \\ \cdots \\ 1 \leq n_d < \infty}} \left| \frac{\sum_{k_1=0}^{n_1-1} \cdots \sum_{k_d=0}^{n_d-1} D_d V^{(k_1, \ldots, k_d)} f}{n_1 \cdots n_d} \right| \right\|_r \leq C|f|_{p,d,\pi}.$$

Proof For the proof we will use the bound in [24], Theorem 8.6.8. Note that this result is the lemma for $d = 1$.

Let now $d \geq 2$. According to one of the properties of the projective tensor norm ([49], Proposition 2.8), for every $f \in L_{p,\pi}(\mu^d)$ and $\epsilon > 0$ there exists a bounded family of functions $f_{i,l} \in L_p(\mu)$ ($1 \leq i < \infty, 1 \leq l \leq d$) such that $f = \sum_i f_{i,1} \otimes_\pi \cdots \otimes_\pi f_{i,d}$ and

$$\sum_i |f_{i,1}|_p \cdots |f_{i,d}|_p \leq |f|_{p,d,\pi} + \epsilon.$$

Then we have, using Corollary 1, that

$$\left\| \sup_{\substack{1 \leq n_1 < \infty \\ \cdots \\ 1 \leq n_d < \infty}} \left| (n_1 \cdots n_d)^{-1} \sum_{k_1=0}^{n_1-1} \cdots \sum_{k_d=0}^{n_d-1} D_d V^{(k_1, \ldots, k_d)} f \right| \right\|_r$$

$$\leq \sum_i \left\| D_d \left(\sup_{1 \leq n_1 < \infty} \frac{|\sum_{k_1=0}^{n_1-1} V_1^{k_1} f_{i,1}|}{n_1} \cdots \sup_{1 \leq n_d < \infty} \frac{|\sum_{k_d=0}^{n_d-1} V_d^{k_d} f_{i,d}|}{n_d} \right) \right\|_r$$

$$\leq \sum_i C|f_{i,1}|_p \cdots |f_{i,d}|_p \leq C(|f|_{p,d,\pi} + \epsilon).$$

In the above formulas $V_1, \ldots, V_d$ are the dynamical operators associated with the transformations $T_1, \ldots, T_d$. □

Proof of Proposition 3 For $d = 1$ the assertions of the proposition are the classical individual and statistical ergodic theorems. Let now $d \geq 2$, hence $p \geq 2$. In view of Lemma 2, the proof is straightforward. First we prove the assertions of the proposition for elementary tensors $f = f_1 \otimes \cdots \otimes f_d$ with $f_l \in L_p(\mu)$, $1 \leq l \leq d$. Then the corresponding normalized V-statistic can be written in the product form

$$\frac{\sum_{k_1=0}^{n_1-1} V_1^{k_1} f_1}{n_1} \cdots \frac{\sum_{k_d=0}^{n_d-1} V_d^{k_d} f_d}{n_d},$$

where by the individual ergodic theorem the l-th term in the product converges to $E_{inv,\,l}\,f_l$ with probability 1. Hence, the product tends with probability 1 to

$$(E_{inv,\,1}f_1)\cdots(E_{inv,\,d}f_d) = D_d E_{inv,\,1}\otimes_\pi \cdots \otimes_\pi E_{inv,\,d}f.$$

The same conclusion holds for finite sums of elementary tensors which are dense in the space $L_{p,\pi}(\mu^d)$. Let now $f \in L_{p,\pi}(\mu^d)$. Fix an $\epsilon > 0$. There exists an element $f_\epsilon \in L_{p,\pi}(\mu^d)$ with $|f - f_\epsilon|_{p,\,d,\,\pi} < \epsilon$ such that the a.s. assertion of the proposition holds for f_ϵ and with probability 1

$$0 \le \xi \overset{\text{def}}{=} \varlimsup_{\substack{n_1\to\infty \\ \cdots \\ n_d\to\infty}} \left| \frac{1}{n_1\cdots n_d} \sum_{0\le\mathbf{k}<\mathbf{n}} D_d V^{\mathbf{k}} f - D_d(E_{inv,\,1}\otimes_\pi \cdots \otimes_\pi E_{inv,\,d})f \right|$$

$$\le \varlimsup_{\substack{n_1\to\infty \\ \cdots \\ n_d\to\infty}} \left| \frac{1}{n_1\cdots n_d} \sum_{0\le\mathbf{k}<\mathbf{n}} D_d V^{\mathbf{k}}(f-f_\epsilon) \right| + \left| D_d(E_{inv,\,1}\otimes_\pi \cdots \otimes_\pi E_{inv,\,d})(f_\epsilon - f) \right|$$

$$+ \varlimsup_{\substack{n_1\to\infty \\ \cdots \\ n_d\to\infty}} \left| \frac{1}{n_1\cdots n_d} \sum_{0\le\mathbf{k}<\mathbf{n}} D_d V^{\mathbf{k}} f_\epsilon - D_d(E_{inv,\,1}\otimes_\pi \cdots \otimes_\pi E_{inv,\,d})f_\epsilon \right|$$

$$\overset{\text{def}}{=} \xi_{1,\,\epsilon} + \xi_{2,\,\epsilon} + \xi_{3,\,\epsilon}.$$

Since the operators D_d and $D_d(E_{inv,1}\otimes_\pi \cdots \otimes_\pi E_{inv,d})$ are of norm 1, we have $|\xi_{2,\,\epsilon}|_r \le \epsilon$, and, in view of the individual ergodic theorem and Lemma 2, $\xi_{3,\,\epsilon} = 0$ and $|\xi_{1,\,\epsilon}|_r \le C\epsilon$. This implies $\xi = 0$ which proves the convergence with probability 1. To establish the L_r-convergence, we observe that we have the convergence with probability 1 along with the domination by an L_r-function given by Lemma 2. Hence, we can apply Theorem 3.3.7 in [24]. □

3.2 Applications to the SLLN for von Mises statistics

We return here to the assumption that the transformations $T_1, \ldots, T_d$ are copies of the same transformation T. For simplicity we assume that T is ergodic. Symmetry of the kernel is not assumed.

Theorem 1 *Let $r = p/d$ for some integer $d \ge 2$ and a real number $p \ge d$. Let T be an ergodic measure preserving transformation of a probability space $(X, \mathcal{F}, \mu)$. Assume also that $f \in L_{p,\pi}(\mu^d)$. Then, as $n \to \infty$, the sequence*

$$\frac{1}{n^d} \sum_{0\le k_1,\ldots,k_d\le n-1} D_d V^{(k_1,\ldots,k_d)} f \tag{8}$$

converges with probability 1 and in $L_r(\mu)$ to the limit

$$\int_{X^d} (J_d f)(x_1, \ldots, x_d)\mu(dx_1)\cdots\mu(x_d).$$

Here $J_d : L_{p,\pi}(\mu^d) \to L_p(\mu^d)$ is the operator introduced in Lemma 1.

Proof The theorem follows from Proposition 3. We only need to identify the limits. Since the limit expressions given in Proposition 3 and in the theorem are both continuous in the projective norm, it suffices to check that these expressions agree for elementary tensors $f_1 \otimes \cdots \otimes f_d$. It is straightforward to check that in the ergodic case both expressions reduce to $Ef_1 \cdots Ef_d$, where E denotes the integral with respect to μ. □

Corollary 2 *In the case $p = d$ Theorem 1 applies and gives the convergence with probability 1 and in $L_1(\mu)$.*

Remark 5 Examples show that it is possible to extend the class of kernels to which the conclusion in Corollary 2 applies to such kernels $f \in L_p(\mu^d)$ which can be "sandwiched" between decreasing and increasing sequences of some $L_{p,\pi}(\mu^d)$-kernels whose common $L_p(\mu^d)$−limit is f (notice that bounding by products plays some role in [1]). This indicates that probably more appropriate functional spaces can be found in order to treat the SLLN.

Corollary 3 *Let T be an ergodic measure preserving transformation of a probability space $(X, \mathcal{F}, \mu)$ and let $(e_k)_{k=0}^{\infty}$ be a sequence of functions in $L_d(\mu)$ such that $e_0 \equiv 1$ and for every $k \geq 1$ $|e_k|_d = 1$, $\int_X e_k(x)\mu(dx) = 0$. Let $f \in L_d(\mu^d)$ admit the representation*

$$f(x_1, \ldots, x_d) = \sum_{0 \leq \mathbf{k} < \infty} \lambda_{\mathbf{k}}(f) \, e_{k_1}(x_1) \cdots e_{k_d}(x_d)$$

for some family $(\lambda_{\mathbf{k}}(f))_{0 < \mathbf{k} < \infty}$ satisfying the condition

$$\sum_{0 \leq \mathbf{k} < \infty} |\lambda_{\mathbf{k}}(f)| < \infty.$$

Then Corollary 2 applies to f.

Proof The series representing f obviously converges in $L_{p,\pi}(\mu^d)$, and the corollary follows. □

4 The Hoeffding decomposition

In this section we recall well-known properties of the Hoeffding decomposition for kernels in the spaces L_p, omitting proofs (see [25] for the proofs in the symmetric case). It is not hard to see that the results and formulas related to this decomposition (both general and symmetric) apply also to the spaces $L_{p,\pi}$ and, in case $\mu_1 = \cdots = \mu_d = \mu$, to their symmetric subspaces.

4.1 The Hoeffding decomposition for general kernels

Let $(X_1, \mathcal{F}_1, \mu_1), \ldots, (X_d, \mathcal{F}_d, \mu_d)$ be probability spaces. We do not assume in this subsection that all $(X_l, \mathcal{F}_l, \mu_l)$, $l = 1, \ldots, d$, are copies of the same probability space. Let $\mathcal{S}_d$ ($\mathcal{S}_d^m$) be the set of all subsets (respectively, of all m-subsets) of $\{1, \ldots, d\}$. For every $S \subset \{1, \ldots, d\}$ we define

$$(X^S, \mathcal{F}^{\otimes S}, \mu^S) = \left(\prod_{l \in S} X_l, \bigotimes_{l \in S} \mathcal{F}_l, \prod_{l \in S} \mu_l \right), \quad L_p(\mu^S) = L_p(X^S, \mathcal{F}^{\otimes S}, \mu^S).$$

Denoting the conditional expectation with respect to a σ-field $\mathcal{G} \subset \mathcal{F}$ by $E^{\mathcal{G}}$ and the projection map from $X^{\{1,\ldots,d\}}$ onto $X^{\{l\}} = X_l$ ($l = 1, \ldots, d$) by π_l, we set for every $S \in \mathcal{S}_d$

$$\mathcal{F}^S = \bigvee_{l \in S} \pi_l^{-1}(\mathcal{F}_l), \quad E^S = E^{\mathcal{F}_S}, \quad \check{E}^l = E^{\{1,\ldots,d\}\setminus\{l\}}.$$

In other terms, applying $\check{E}^l$, one integrates out the l-th variable.

The identity operator I in $L_p(\mu^{\{1\ldots d\}})$ ($p \in [1, \infty]$) decomposes as

$$I = \prod_{l=1}^{d} \left(\check{E}^l + (I - \check{E}^l) \right) = \sum_{m=0}^{d} \sum_{S \in \mathcal{S}_d^m} Q_S,$$

where $Q_S = \prod_{l \notin S} \check{E}^l \prod_{l' \in S}(I - \check{E}^{l'})$. In general, the Hoeffding decomposition assigns to every $f \in L_p(\mu^{\{1,\ldots,d\}})$ the family $(R_S f)_{S \in \mathcal{S}_d}$ such that

(i) for every $S \in \mathcal{S}_d$ $R_S f \in L_p(\mu^S)$;
(ii) for every $S = \{l_1, \ldots, l_m\} \in \mathcal{S}_d^m$

$$(R_S f) \circ \pi_S = Q_S f,$$

where $\pi_S : X^d \mapsto X^S$ is defined by $\pi_S(x_1, \ldots, x_d) = (x_{l_1}, \ldots, x_{l_m})$;
(iii) every $R_S f$ is *canonical* (or, using an alternate terminology, *totally degenerate*) that is for every $l \in S$, $f \in L_p^{\{1,\ldots,d\}}$

$$\check{E}^l \left((R_S f) \circ \pi_S \right) = 0.$$

Kernels of the form $(R_S f) \circ \pi_S$ will be also called canonical. $R_S f$ (or $(R_S f) \circ \pi_S$) is said to have the *degree m* whenever it does not vanish identically and $S \in \mathcal{S}_d^m$. Every kernel $f \in L_p(\mu^{\{1,\ldots,d\}})$ can be represented in a unique way as a sum of canonical kernels (the Hoeffding decomposition) as follows

$$f = \sum_{m=0}^{d} \sum_{S \in \mathcal{S}_d^m} (R_S f) \circ \pi_S. \tag{9}$$

As said before, the Hoeffding decomposition also holds for $L_{p,\pi}(\mu^{\{1,\ldots,d\}}) \overset{\text{def}}{=\!=} L_p(\mu_1)\hat{\otimes}_\pi \cdots \hat{\otimes}_\pi L_p(\mu_d)$, and we shall use the above notation for the operators on these spaces as well.

The *degree* of a kernel f with decomposition (9) (or the decomposition (10) below) is, by definition, the smallest degree of non-vanishing summands in (9). A kernel f in (9) is called *degenerate* if the degree of $f - R_\emptyset f$ is greater than 1 and *non-degenerate* if it equals 1.

4.2 The Hoeffding decomposition of symmetric kernels

We assume in this subsection that all spaces $(X_l, \mathcal{F}_l, \mu_l)$, $l = 1, \ldots, d$, are copies of the same probability space $(X, \mathcal{F}, \mu)$. $L_p(\mu^d)$ and $L_{p,\pi}(\mu^d)$ denote, respectively, the usual L_p–spaces of the product of d identical probability spaces and the projective tensor product $\underbrace{L_p(\mu)\hat{\otimes}_\pi \cdots \hat{\otimes}_\pi L_p(\mu)}_{d \text{ times}}$ with the norms $|\cdot|_p$ and $|\cdot|_{p,d,\pi}$, respectively.

There is an isometric action of the symmetric group $\mathbf{S}_d$ by permutations of the multipliers on every of these spaces. The fixed points of these actions form closed subspaces called *symmetric*; their denotations will contain the superscript sym; their elements are called *symmetric functions*. The next property of the Hoeffding decomposition is specific for the symmetric case.

iv) whenever the function f belongs to $L_p^{sym}(\mu^d)$, the canonical function $R_S f$ does not depend on the choice of $S \in \mathcal{S}_d^m$ and is symmetric; thus, in this case there exist operators $R_m : L_p^{sym}(\mu^d) \to L_p^{sym}(\mu^m)$ such that for every $S = \{i_1, \ldots, i_m\} \in \mathcal{S}_d^m$

$$(R_m f) \circ \pi_S = Q_S f.$$

Furthermore, every $f \in L_p^{sym}(\mu^d)$ can be represented in a unique way in the form

$$f = \sum_{m=0}^{d} \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S. \tag{10}$$

Remark 6 We illustrate the difference between general and symmetric kernels for $d = 2$. For a general kernel $f \in L_p(\mu^2)$ we have

$$f(x_1, x_2) = f_\emptyset + f_{\{1\}}(x_1) + f_{\{2\}}(x_2) + f_{\{1,2\}}(x_1, x_2),$$

where

$$f_\emptyset = \int_{X^2} f(z_1, z_2)\mu(dz_1)\mu(dz_2),$$

$$f_{\{1\}}(x_1) = \int_X f(x_1, z_2)\mu(dz_2) - f_\emptyset, \quad f_{\{2\}}(x_2) = \int_X f(z_1, x_2)\mu(dz_1) - f_\emptyset,$$

$$f_{\{1,2\}}(x_1, x_2) = f(x_1, x_2) - f_{\{1\}}(x_1) - f_{\{2\}}(x_2) - f_\emptyset.$$

Notice, in order to illustrate the notion of canonical kernels, that we have for almost every $x_1, x_2 \in X$,

$$\int_X f_{\{1\}}(z)\mu(dz) = 0, \int_X f_{\{2\}}(z)\mu(dz) = 0,$$

$$\int_X f_{\{1,2\}}(z_1, x_2)\mu(dz_1) = \int_X f_{\{1,2\}}(x_1, z_2)\mu(dz_2) = 0.$$

For a kernel $f \in L_p^{sym}(\mu^2)$ the above relations reduce to

$$f(x_1, x_2) = f_0 + f_1(x_1) + f_1(x_2) + f_2(x_1, x_2),$$

where

$$f_0 = \int_{X^2} f(z_1, z_2)\mu(dz_1)\mu(dz_2),$$

$$f_1(x) = \int_X f(x, z)\mu(dz) - f_0 \left(= \int_X f(z, x)\mu(dz) - f_0 \right),$$

$$f_2(x_1, x_2) = f(x_1, x_2) - f_1(x_1) - f_1(x_2) - f_0.$$

Here $\int_X f_1(z)\mu(dz) = 0$, $f_2 \in L_p^{sym}(\mu^2)$ and for almost every $x \in X$ we have

$$\int_X f_2(z, x)\mu(dz) \left(= \int_X f_2(x, z)\mu(dz) \right) = 0.$$

5 Filtrations: exactness and Kolmogorov property

In the remaining part of the paper we deal with distributional convergence of von Mises statistics for a measure preserving transformation. Our tool here is a kind of martingale approximation. For $d = 1$ this approximation goes back to [30,32] and [44] (in the latter paper only Harris recurrent Markov chains were considered) and was developed for higher dimensional random arrays in [33].

The additional structure needed is a filtration compatible with the dynamics defined by a measure preserving transformation. From now on we restrict ourselves to a class of measure preserving transformations of probability spaces, which are *exact* [48]. Let T be a measure preserving transformation of a probability space $(X, \mathcal{F}, \mu)$. The transformation T defines a decreasing filtration $(T^{-k}\mathcal{F})_{k \geq 0}$. Exactness of T means that $\bigcap_{k \geq 0} T^{-k}\mathcal{F} = \mathcal{N}$, where $\mathcal{N}$ is the trivial σ-field of the space $(X, \mathcal{F}, \mu)$. As can easily be seen, every exact transformation is ergodic. The standard assumption of the ergodic theory is that $(X, \mathcal{F}, \mu)$ is a Lebesgue space in the sense of Rokhlin. Under this assumption it can be shown that, except for the case of the one point measure space, the Lebesgue space with an exact transformation is an atomless measure space,

hence, is isomorphic to the unit interval with the Lebesgue measure. As before, by V^* we denote the adjoint (for $p > 1$) and the preadjoint (for $p = 1$) of the operator V. As the operator V acts as an isometry in all L_p spaces, preserves constants and positivity, the operator V^* also acts on all these spaces as a contraction which preserves constants and positivity. The operator V^* is a particular case of a Markov transition operator.

For every $k \geq 0$ we have the relations $V^{*k} V^k = I$ and $V^k V^{*k} = E_k$, where I is the identity operator and $E_k = E^{T^{-k}\mathcal{F}}$, the corresponding conditional expectation. Let E denote the expectation operator. We can easily conclude (for example, from known facts about the convergence of reversed martingales) that the exactness of T is equivalent to the the strong convergence $V^{*n} \underset{n\to\infty}{\to} E$ in every space $L_p(\mu)$ with $1 \leq p < \infty$. In the sequel the strong convergence of the series

$$\sum_{k\geq 0} V^{*k} f \tag{11}$$

and other similar conditions will be imposed on f. Set

$$L_p^0(\mu) = \{f \in L_p(\mu), Ef = 0\}.$$

Assuming T is exact, for every $1 \leq p < \infty$ the series (11) converges in the norm of $L_p(\mu)$ if and only if f can be represented in the form $f = (I - V^*)g$ with some $g \in L_p(\mu)$ (such g is unique up to an additive constant which can be fixed by the condition $g \in L_p^0(\mu)$). Observe that, in view of exactness, such f's form a dense subspace in $L_p^0(\mu)$.

Remark 7 In the rest of the paper we will mainly restrict ourselves to exact transformations. This is just done to simplify the statements of the results and make the notation more convenient. We could easily extend these results to ergodic transformations T and to kernels $f \in L_p(\mu^d)$ satisfying the additional condition $E(f \mid \mathcal{F}_1 \otimes \cdots \otimes T_l^{-n}\mathcal{F}_l \otimes \cdots \otimes \mathcal{F}_d) \underset{n\to\infty}{\to} \check{E}^l f, l = 1, \ldots, d$. Here T_l is the copy of T acting on the l-th coordinate in X^d, $\check{E}^l$ was defined in Sect. 4.1.

Remark 8 The results of the next sections are primarily concerned with exact (hence, non-invertible) transformations; however, they can be converted into some results on invertible transformations furnished with an additional structure. Indeed, assume that an invertible measure preserving T acts on $(X, \mathcal{F}, \mu)$ and we are given a σ-field $\mathcal{F}_0 \subset \mathcal{F}$ such that $T^{-1}\mathcal{F}_0 \supseteq \mathcal{F}_0$. Then a theory, totally parallel to that we develop in the following sections for the exact case, applies to kernels measurable with respect to $\mathcal{F}_0^{\otimes d}$. The restriction of T^{-1} to $\mathcal{F}_0$ corresponds to a non-invertible transformation. We leave details of this correspondence to the reader; it will be used when considering applications in Sect. 9. Just notice that the counterpart of exactness for an invertible T is the property $\bigcap_{k\geq 0} T^k \mathcal{F}_0 = \mathcal{N}$. If, moreover, $\bigvee_{k\geq 0} T^{-k}\mathcal{F}_0 = \mathcal{F}$, the transformation T is called *Kolmogorov*. Similarly to the exactness property in Remark 7, the Kolmogorov property can be relaxed to the requirement that T is ergodic and f satisfies an analogue of the additional condition there.

6 Growth rates for multiparameter sums

It follows from Lemma 1 for $p \in [1, \infty)$ that the space $L_{p,\pi}^{sym}(\mu^m)$ can be identified, using the injective map J_m, with a (non-closed) dense subspace of $L_p^{sym}(\mu^m)$. As we warned the reader above, the symbol J_m will be omitted and the relation $L_{p,\pi}^{sym}(\mu^m) \subset L_p^{sym}(\mu^m)$ will be assumed instead of $J_m(L_{p,\pi}^{sym}(\mu^m)) \subset L_p^{sym}(\mu^m)$. In particular, it makes sense to speak of canonical elements of $L_{p,\pi}^{sym}(\mu^m)$.

A noninvertible measure preserving transformation T of a probability space $(X, \mathcal{F}, \mu)$ has a natural decreasing filtration given by $(T^{-n}\mathcal{F})_{n \geq 0}$. We shall use the following consequence of the Burkholder inequality.

Lemma 3 *For every* $p \in [2, \infty)$ *there exists a constant* $C(p)$ *such that for every stationary sequence* $(\xi_n)_{n \in \mathbb{Z}}$ *of martingale differences in* $L_p(\mu)$ *we have*

$$\left| \sum_{k=1}^{n-1} \xi_k \right|_p \leq C(p)\sqrt{n} \, |\xi_0|_p .$$

Proof Let $p \in [2, \infty)$. Using the Burkholder inequality (Theorem 9 in [13]) for the original sequence and then applying the triangle inequality for the space $L_{p/2}$ to the sequence $(\xi_n^2)_{n \in \mathbb{Z}}$, we obtain

$$\frac{1}{\sqrt{n}} \left| \sum_{k=1}^{n-1} \xi_k \right|_p \leq C(p) \left| \left(\frac{1}{n} \sum_{k=1}^{n-1} \xi_k^2 \right)^{1/2} \right|_p \leq C(p) |\xi_0^2|_{p/2}^{1/2} = C(p) |\xi_0|_p .$$

$\square$

For every m, $0 \leq m \leq d$, let $\mathcal{S}_m$ ($\mathcal{S}_m^s$) be the set of all subsets (respectively, of all subsets of cardinality $s \in \{0, \ldots, m\}$) of the set $\{1, \ldots, m\}$. For every $S \in \mathcal{S}_m$ define a subsemigroup $\mathbb{Z}_+^{m,S} \subseteq \mathbb{Z}_+^m$ by

$$\mathbb{Z}_+^{m,S} = \{(n_1, \ldots, n_m) \in \mathbb{Z}_+^m : n_k = 0 \text{ for all } k \notin S\}.$$

In this section we write $\mathbf{k}$ and $\mathbf{n}$ for $(k_1, \ldots, k_m)$ and $(n_1, \ldots, n_m)$, respectively; the notation $\mathbf{k} < \mathbf{k}'$ ($\mathbf{k} \leq \mathbf{k}'$) means that $k_1 < k_1', \ldots, k_m < k_m'$ (respectively, $k_1 \leq k_1', \ldots, k_m \leq k_m'$).

Lemma 4 *Let* $m \in \{1, \ldots, d\}$ *and let* $\mathbf{e}_1, \ldots, \mathbf{e}_m$ *denote the standard basis of* $\mathbb{Z}_+^m$. *Then, for every real* $p \in [2, \infty)$ *and every integer* $s \in \{1, \ldots, m\}$, *there exists a constant* $C(p, s) > 0$ *with the following property: For every* $S \in \mathcal{S}_m^s$ *and* $f \in L_{p,\pi}(\mu^m)$, *satisfying*

$$V^{*\mathbf{e}_l} f = 0, \, l \in S, \tag{12}$$

the relation

$$\left| \sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^{m,S} \\ 0 \le k_l \le n_l - 1, \, l \in S}} V^{\mathbf{k}} f \right|_{p, m, \pi} \le C(p, s) \left(\prod_{l \in S} \sqrt{n_l} \right) |f|_{p, m, \pi} \tag{13}$$

holds for every family $(n_l)_{l \in S}$ *of natural numbers. Moreover, if* $p \ge m$ *and* $r = p/m$, *we also obtain*

$$\left| \sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^{m,S} \\ 0 \le k_l \le n_l - 1, \, l \in S}} D_m V^{\mathbf{k}} f \right|_r \le C(p, s) \left(\prod_{l \in S} \sqrt{n_l} \right) |f|_{p, m, \pi}$$

for every $(n_l)_{l \in S}$.

Proof Let s and S be as in the statement of the lemma. Since the norm of the map $D_m \colon L_{p, \pi}(\mu^m) \to L_r(\mu)$ is 1, it suffices to prove (13). Let $\mathbf{0}_m$ denote the neutral element of $\mathbb{Z}_+^m$. Set

$$M_{p, \mathbf{0}_m, \pi}^S = \{f \in L_{p, \pi}(\mu^m) : V^{*\mathbf{e}_l} f = 0 \quad \text{for every } l \in S\}.$$

Observe that the subspace $M_{p, \mathbf{0}_m, \pi}^S \subset L_{p, \pi}(\mu^m)$ itself can be represented as the projective tensor product of s copies of the subspace $M_{p, 0} \overset{\text{def}}{=} \{f \in L_p(\mu) : Vf = 0\}$ and $m - s$ copies of the space $L_p(\mu)$. Notice that the relations (12) are equivalent to the following description of the corresponding subspace in terms of projections:

$$(I - V^{\mathbf{e}_l} V^{*\mathbf{e}_l}) f = f \quad \text{for every } l \in S.$$

The subspace $M_{p, \mathbf{0}_m, \pi}^S$ can also be described as the range of the projection

$$\prod_{l \in S} (I - V^{\mathbf{e}_l} V^{*\mathbf{e}_l}).$$

We need now the following consequence of Proposition 2.4 in [49]. In general, for some Banach spaces A_l and their closed subspaces $B_l \subset A_l$, $l = 1, \dots, m$, we only have a canonical linear map $i : B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_m \to A_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi A_m$ of norm 1. However, if every B_l is a complemented subspace in the corresponding A_l (that is the range of a bounded projection $\varphi_l : A_l \to B_l$) then this map is a topological linear isomorphism onto its range (the latter is closed in $A_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi A_m$). Moreover, if every φ_l is a projection of norm 1 then this map is an isometry.

Thus, if bounded projections $(\varphi_l)_{l=1,\dots,m}$ exist, we can consider $B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_m$ as a closed subspace of $A_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi A_m$, the map $\varphi_1 \otimes_\pi \cdots \otimes_\pi \varphi_m$ being a bounded

projection of $A_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi A_m$ onto its subspace $B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_m$. The latter subspace can be described by

$$B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_m = \left\{ f \in A_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi A_m : (\varphi_1 \otimes_\pi \cdots \otimes_\pi \varphi_m) f = f \right\}$$

or, equivalently, by

$$B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_m = \big\{ f \in A_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi A_m :$$
$$\big((I - \varphi_1) \otimes_\pi \cdots \otimes_\pi I\big) f = 0; \ldots; \big(I \otimes_\pi I \otimes_\pi \cdots \otimes_\pi (I - \varphi_m)\big) f = 0 \big\}.$$

Moreover, the projective tensor norm on the space $B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_m$ and the norm induced by its embedding into $A_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi A_m$ are equivalent.

We will apply this assertion to the case when $A_l = L_p(\mu)$ for every $l \in \{1, \ldots, m\}$, $B_l = M_{p,0}$, $\varphi_l = I - VV^*$ for $l \in S$, and $B_l = L_p(\mu)$, $\varphi_l = I$ for $l \notin S$. Since VV^* is a conditional expectation, it is clear that φ_l is bounded for every l (in fact its norm does not exceed $2^{1-(2/p)}$). With this notation we have that $M^S_{p, \mathbf{0}_m, \pi}$ and $B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_m$ are isomorphic as topological vector spaces. Observe that we have here a vector space which is equipped with two possibly norms: the norm inherited from $L_{p,\pi}(\mu^m)$ and the projective tensor product norm, respectively. According to one of the properties of the projective tensor norm ([49], Proposition 2.8), for every $f \in M^S_{p, \mathbf{0}_m, \pi}$ and $\epsilon > 0$ there exists a bounded family of functions $f_{i,l} \in B_l (1 \le i < \infty, 1 \le l \le m)$ such that

$$f = \sum_i f_{i,1} \otimes \cdots \otimes f_{i,m} \quad \text{and} \quad \sum_i |f_{i,1}|_p \cdots |f_{i,m}|_p \le C'(p,s) |f|_{p,m,\pi} + \epsilon.$$

The constant $C'(p, s)$ appears here because we put into the right hand side the inherited norm $|f|_{p,m,\pi}$ of f rather than its norm in $B_1 \hat{\otimes}_\pi \cdots \hat{\otimes}_\pi B_m$. For $l = 1, \ldots, m$ and every i let $F_{i,l} = \sum_{0 \le k \le n_l - 1} V^k f_{i,l}$ if $l \in S$, and $F_{i,l} = f_{i,l}$ if $l \notin S$. Then, applying Lemma 3 to the sums $\sum_{k=0}^{n_l-1} V^k f_{i,l}$ for $l \in S$ (in this case the summands form a stationary sequence of reversed martingale differences), it follows that

$$\left| \sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^{m,S} \\ 0 \le k_l \le n_l - 1, l \in S}} V^{\mathbf{k}} f \right|_{p,m,\pi}$$

$$\le \sum_i \left| \sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^{m,S} \\ 0 \le k_l \le n_l - 1, l \in S}} V^{\mathbf{k}} \big(f_{i,1} \otimes \cdots \otimes f_{i,m} \big) \right|_{p,m,\pi} = \sum_i \big| F_{i,1} \otimes \cdots \otimes F_{i,m} \big|_{p,m,\pi}$$

$$= \sum_i \prod_{l \in \{1,\ldots,m\}} |F_{i,l}|_p = \sum_i \prod_{l \in S} \left| \sum_{k=0}^{n_l-1} V^k f_{i,l} \right|_p \prod_{l \notin S} |f_{i,l}|_p$$

$$\leq C^s(p) \left(\prod_{l \in S} \sqrt{n_l}\right) \sum_i \prod_{l \in \{1,\dots,m\}} |f_{i,1}|_p \cdots |f_{i,m}|_p$$

$$\leq C^s(p) \left(\prod_{l \in S} \sqrt{n_l}\right) \left(C'(p,s)|f|_{p,m,\pi} + \epsilon\right).$$

Thus inequality (13) follows with $C(p,s) = C^s(p) C'(p,s)$.

Remark 9 Every f satisfying the assumptions of the above lemma is S-canonical in the following sense: since every operator $V^{*\mathbf{e}_l}$ preserves the integrals with respect to the l–th variable, it follows from (12) that, under the assumptions of Lemma 4, integrating f over the l–th variable returns 0 whenever $l \in S$. This implies the assertion.

The following lemma provides a condition under which the martingale-coboundary decomposition is valid.

Lemma 5 *Let $p \in [1,\infty]$ and $f \in L_{p,\pi}(\mu^m)$ be a canonical kernel such that the series in the right hand side of*

$$g = \sum_{\mathbf{0} \leq \mathbf{k} < \infty} V^{*\mathbf{k}} f \quad \left(\overset{\text{def}}{=} \lim_{\substack{n_1 \to \infty \\ \dots \\ n_m \to \infty}} \sum_{\mathbf{0} \leq \mathbf{k} < \mathbf{n}} V^{*\mathbf{k}} f\right) \tag{14}$$

converges in $L_{p,\pi}(\mu^m)$. Then f can be represented in the form

$$f = \sum_{S \in \mathcal{S}_m} A^S f, \tag{15}$$

where for every $S \in \mathcal{S}_m$

$$A^S f = \left(\prod_{l \notin S}(I - V^{\mathbf{e}_l} V^{*\mathbf{e}_l}) \prod_{l \in S}(V^{\mathbf{e}_l} - I)\right) h^S \tag{16}$$

and the function $h^S \in L_{p,\pi}(\mu^m)$ is defined by the equation

$$h^S = \left(\prod_{l \in S} V^{*\mathbf{e}_l}\right) g. \tag{17}$$

The functions g and $(h^S)_{S \in \mathcal{S}_m}$ are canonical; the summands of the form (16) in (15) are uniquely determined.

Proof The results and the proofs in [33], developed originally for the L_p–spaces, apply to the $L_{p,\pi}$-spaces without any changes. The requirement of *complete commutativity* imposed in [33] on the multiparameter dynamical system and the invariant measure is obviously fulfilled for a direct product with a coordinatewise action which we deal

with in the present paper. Hence, by Proposition 3 in [33], the convergence of the series (14) implies that the *Poisson equation* (see [33]) is solvable for f; therefore, we may apply Proposition 1 in [33] to f. Then we obtain the representation (15) with $A^S f$ defined by formulas (16), (17) and the assertion on the uniqueness of the summands of the form (16). Notice that the operator V^* preserves integrals of functions with respect to μ; as a consequence, every $V^{*\mathbf{n}}$ maps canonical functions to canonical ones. Being according to (14) a limit of canonical functions, g is canonical. In view of (17), all h^S are canonical, too.   □

Proposition 4 *Let $0 \le s \le m$ and f be a kernel satisfying the assumptions of Lemma 5 for some $p \in [2, \infty)$. Let $A^S f$ be defined by formulas (16) and (17). Then there exists a constant $C_{p,m,s} > 0$ such that for every $S \in \mathcal{S}_m^s$ and every $n_1, \dots, n_m$*

$$\left| \sum_{0 \le \mathbf{k} < \mathbf{n}} V^{\mathbf{k}} A^S f \right|_{p,m,\pi} \le C_{p,m,s} \left(\prod_{l \notin S} \sqrt{n_l} \right) |g|_{p,m,\pi}, \tag{18}$$

where g is defined in (14). Moreover, for $p \ge m$

$$\left| \sum_{0 \le \mathbf{k} < \mathbf{n}} D_m V^{\mathbf{k}} A^S f \right|_r \le C_{p,m,s} \left(\prod_{l \notin S} \sqrt{n_l} \right) |g|_{p,m,\pi} \tag{19}$$

holds with $r = p/m$.

Proof Setting $\overline{S} = \{1, \dots, m\} \setminus S$, we have

$$\sum_{0 \le \mathbf{k} < \mathbf{n}} V^{\mathbf{k}} A^S f = \sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^{m, \overline{S}} \\ 0 \le k_t \le n_t - 1, t \in \overline{S}}} V^{\mathbf{k}} \prod_{r \notin S} (I - V^{\mathbf{e}_r} V^{*\mathbf{e}_r})$$

$$\times \sum_{\substack{\mathbf{l} \in \mathbb{Z}_+^{m, S} \\ 0 \le l_u \le n_u - 1, u \in S}} V^{\mathbf{l}} \prod_{u \in S} (V^{\mathbf{e}_u} - I) h^S$$

$$= \sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^{m, \overline{S}} \\ 0 \le k_t \le n_t - 1, t \in \overline{S}}} V^{\mathbf{k}} \prod_{r \notin S} (I - V^{\mathbf{e}_r} V^{*\mathbf{e}_r}) \prod_{u \in S} (V^{n_u \mathbf{e}_u} - I) h^S. \tag{20}$$

Since for $l \notin S$

$$V^{*\mathbf{e}_l} \prod_{r \notin S} (I - V^{\mathbf{e}_r} V^{*\mathbf{e}_r}) \prod_{u \in S} (V^{n_u \mathbf{e}_u} - I) h^S = 0$$

and

$$\left| \prod_{r \notin S} (I - V^{\mathbf{e}_r} V^{*\mathbf{e}_r}) \prod_{u \in S} (V^{n_u \mathbf{e}_u} - I) h^S \right|_{p,m,\pi} \le 2^m |g|_{p,m,\pi},$$

the proposition follows with $C_{p,m,s} = 2^m C(p,s)$ from Lemma 4 and formula (20).

$\square$

Proposition 5 *Let $p \geq 2$ and $f \in L_{p,\pi}(\mu^m)$ be a canonical kernel such that the series on the right hand side of*

$$g = \sum_{0 \leq \mathbf{k} < \infty} V^{*\mathbf{k}} f$$

converges in $L_{p,\pi}(\mu^m)$. Then for every $n_1, \ldots, n_m$ the following inequality holds

$$\left| \sum_{0 \leq \mathbf{k} < \mathbf{n}} V^{\mathbf{k}} f \right|_{p,m,\pi} \leq C_{p,m} \sqrt{n_1 \cdots n_m} \, |g|_{p,m,\pi}, \tag{21}$$

where $C_{m,p}$ is a constant depending only on m and p. If, in addition, $p \in [m, \infty)$ then, with $r = p/m$, we also have that

$$\left| \sum_{0 \leq \mathbf{k} < \mathbf{n}} D_m V^{\mathbf{k}} f \right|_r \leq C_{p,m} \sqrt{n_1 \cdots n_m} \, |g|_{p,m,\pi}.$$

Proof Again, since the norm of the operator $D_m : L_{p,\pi}(\mu^m) \to L_r(\mu)$ is 1, we only need to prove (21). As $n_1 \geq 1, \ldots, n_m \geq 1$, we have for every $S \in \mathcal{S}_m \prod_{l \in S} \frac{1}{n_l} \leq 1$. Using this relation along with (15) and (18) we obtain (21) with $C_{p,m} = \sum_{s=0}^{m} \binom{m}{s} C_{p,m,s}$. $\square$

The following sufficient condition for convergence of the series in (14) will be used in Sect. 9 when considering applications. Expansion of a kernel into an absolutely convergent series whose summands are products of functions in separate variables is natural in the context of the limit theory of U- and V-statistics (see, for example, [9]). Projective tensor products call for using such series to representing arbitrary elements (see Proposition 2.8 in [49]). Neither uniqueness of the representation, nor linear independence of the 'basis' is assumed. Notice that we used such a decomposition in Corollary 3.

Proposition 6 *Let, for some $p \in [1, \infty]$, $(e_k)_{k=0}^{\infty}$ be a sequence of functions such that $e_0 \equiv 1$ and for every $k \geq 1$ $e_k \in L_p(\mu)$ with $\int_X e_k(x)\mu(dx) = 0$. Assume that for every $k \geq 1$*

$$C_{p,k} \stackrel{\text{def}}{=} \sum_{n \geq 0} |V^{*n} e_k|_p < \infty.$$

Suppose that $f \in L_p(\mu^m)$ admits a representation

$$f(x_1, \ldots, x_m) = \sum_{0 < \mathbf{k} < \infty} \lambda_{\mathbf{k}}(f) \, e_{k_1}(x_1) \cdots e_{k_m}(x_m) \tag{22}$$

where $(\lambda_{\mathbf{k}}(f))_{0<\mathbf{k}<\infty}$ is a family of constants satisfying

$$C_p(f) \overset{\text{def}}{=} \sum_{0<\mathbf{k}<\infty} |\lambda_{\mathbf{k}}(f)| C_{p,k_1} \cdots C_{p,k_m} < \infty. \tag{23}$$

Then f is a canonical kernel of degree m, $f \in L_{p,\pi}(\mu^m)$, the series in (14) converges in $L_{p,\pi}(\mu^m)$ and its sum g satisfies the inequality

$$|g|_{p,m,\pi} \leq C_p(f). \tag{24}$$

Proof For every $k \geq 1 C_{p,k} \geq |e_k|_p$. Hence,

$$\sum_{0<\mathbf{k}<\infty} |\lambda_{\mathbf{k}}(f)| |e_{k_1}|_p \cdots |e_{k_m}|_p \leq C_p(f) < \infty.$$

Then, according to [49], $|f|_{p,m,\pi} \leq C_p(f) < \infty$ and $f \in L_{p,\pi}(\mu^m)$; f is canonical because so is every term of the series in (22). Now we obtain

$$
\begin{aligned}
|g|_{p,m,\pi} &\leq \sum_{0\leq\mathbf{n}<\infty} \sum_{0<\mathbf{k}<\infty} |\lambda_{\mathbf{k}}(f)| |V^{*\mathbf{n}}(e_{k_1}\cdots e_{k_m})|_{p,m,\pi} \\
&= \sum_{0\leq\mathbf{n}<\infty} \sum_{0<\mathbf{k}<\infty} |\lambda_{\mathbf{k}}(f)| |V^{*n_1}e_{k_1}|_p \cdots |V^{*n_m}e_{k_m}|_p \\
&= \sum_{0<\mathbf{k}<\infty} |\lambda_{\mathbf{k}}(f)| C_{p,k_1} \cdots C_{p,k_m} = C_p(f) < \infty.
\end{aligned}
$$

$\square$

7 Central limit theorems in the non-degenerate case

$N(m, \sigma^2)$ will denote the Gaussian distribution in $\mathbb{R}$ with mean value $m \in \mathbb{R}$ and variance $\sigma^2 \geq 0$ including the case $\sigma^2 = 0$ of the Dirac measure at $m \in \mathbb{R}$. We first prove a central limit theorem together with the convergence of the second moments.

Theorem 2 *Let $f \in L_2^{sym}(\mu^d)$ be a real valued kernel with the symmetric Hoeffding decomposition*

$$f = \sum_{m=0}^{d} \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S.$$

Assume that for every $m = 1, \ldots, d R_m f \in L_{2m,\pi}^{sym}(\mu^m)$ and that the series

$$\sum_{\substack{\mathbf{k}\in\mathbb{Z}_+^m \\ 0\leq\mathbf{k}<\infty}} V^{*\mathbf{k}} R_m f \left(\overset{\text{def}}{=} \lim_{\substack{n_1\to\infty \\ \cdots \\ n_m\to\infty}} \sum_{\substack{\mathbf{k}\in\mathbb{Z}_+^m \\ 0\leq\mathbf{k}<\mathbf{n}}} V^{*\mathbf{k}} R_m f \right) \tag{25}$$

converges in $L_{2m,\pi}(\mu^m)$. *Then the sequence*

$$V_n^{(d)} f = \frac{1}{n^{d-1/2}} \sum_{\substack{0 \le k_1 \le n-1 \\ \cdots \\ 0 \le k_d \le n-1}} D_d V^{(k_1, \ldots, k_d)}(f - R_0 f)$$

converges in distribution to $N(0, d^2\sigma^2(f))$, *where*

$$\sigma^2(f) = \left| \sum_{k=0}^{\infty} V^{*k} R_1 f \right|_2^2 - \left| \sum_{k=1}^{\infty} V^{*k} R_1 f \right|_2^2 \ge 0.$$

The convergence of the second moments

$$E(V_n^{(d)} f)^2 \xrightarrow[n \to \infty]{} d^2\sigma^2(f)$$

holds as well.

Remark 10 According to the standard terminology, a kernel f is called *non-degenerate* if $R_1 f$ does not vanish identically, otherwise f is called *degenerate*. In the case of i.i.d. variables such non-degeneracy is equivalent to the non-degeneracy of the limit Gaussian distribution using normalization by the constants $n^{d-1/2}$. However, in the general stationary dependent case such a *statical non-degeneracy* may occur together with the degeneracy of the limit distribution. This phenomenon can be viewed as a *dynamical degeneracy*.

Proof Decompose $f - R_0 f$ in the following way:

$$f - R_0 f = \sum_{m=1}^{d} \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S = \sum_{m=1}^{d} f_m,$$

where

$$f_m = \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S, \qquad m = 1, \ldots, d.$$

In order to prove the theorem it suffices to establish that

(1) $V_n^{(d)} f_1$ converges in distribution to $N(0, d^2\sigma^2(f))$,

(2) $|V_n^{(d)} f_1|_2^2 \xrightarrow[n \to \infty]{} d^2\sigma^2(f)$,

(3) $|V_n^{(d)} \sum_{m=2}^{d} f_m|_2 \xrightarrow[n \to \infty]{} 0$.

In view of the equality

$$D_d \left(V^{(k_1, \ldots, k_d)} \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S \right) = \sum_{S = \{i_1, \ldots, i_m\} \in \mathcal{S}_d^m} D_m V^{(k_{i_1}, \ldots, k_{i_m})} R_m f$$

we obtain

$$
\begin{aligned}
V_n^{(d)} f_m &= V_n^{(d)} \left(\sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S \right) \\
&= \frac{1}{n^{d-1/2}} \sum_{\substack{0 \le k_1 \le n-1 \\ \cdots \\ 0 \le k_d \le n-1}} D_d \left(V^{(k_1, \ldots, k_d)} \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S \right) \\
&= \frac{1}{n^{d-1/2}} \sum_{\substack{0 \le k_1 \le n-1 \\ \cdots \\ 0 \le k_d \le n-1}} \sum_{S = \{i_1, \ldots, i_m\} \in \mathcal{S}_d^m} D_m V^{(k_{i_1}, \ldots, k_{i_m})} R_m f \\
&= \frac{\binom{d}{m}}{n^{m-1/2}} D_m \sum_{\substack{0 \le k_1 \le n-1 \\ \cdots \\ 0 \le k_m \le n-1}} V^{(k_1, \ldots, k_m)} R_m f
\end{aligned}
\tag{26}
$$

for every $m = 1, \ldots, d$. It follows from (26), Proposition 5 with $p = 2m$ and the assumptions of the theorem that the function f_m satisfies the inequality

$$
|V_n^{(d)} f_m|_2 \le C_m \binom{d}{m} n^{-(m-1)/2} |g_m|_{2m, m, \pi}
$$

where g_m denotes the sum of the series (25). This bound for $m \ge 2$ proves (3).

Consider now the sums involving f_1. We obtain from (26) that

$$
V_n^{(d)} f_1 = d \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} V^k R_1 f,
\tag{27}
$$

where $R_1 f$ has the representation $R_1 f = g_1 - V^* g_1$ with g_1 denoting the series (25) for $m = 1$. This representation can be rewritten as

$$
R_1 f = (I - V V^*) g_1 + (V - I) V^* g_1.
\tag{28}
$$

Here the first summand gives, under the action of the operators $(V^k)_{k \ge 0}$, an ergodic stationary sequence of reversed square integrable martingale differences $(V^k (I - V V^*) g_1)_{k \ge 0}$. By the Billingsley–Ibragimov CLT [6,37], the variables $1/\sqrt{n} \sum_{k=0}^{n-1} V^k (I - V V^*) g_1$ converge in distribution, along with the variance, to the required centered Gaussian law. The second summand in (28) only makes a uniformly L_2–bounded contributions to each of the sums $\sum_{k=0}^{n-1} V^k R_1 f$. Thus, the convergence to the Gaussian distribution in (1) is established.

The convergence of the second moments can be concluded as follows. In the situation of the Billingsley–Ibragimov CLT we have

$$\left| \frac{\sum_{k=0}^{n-1} V^k (I - VV^*) g_1}{\sqrt{n}} \right|_2^2 = |(I - VV^*) g_1|_2^2 = |g_1|_2^2 - |V^* g_1|_2^2 = \sigma^2(f).$$

This implies, in view of (27), (28) and the triangle inequality, that

$$||V_n^{(d)} f_1|_2 - d\sigma(f)| \leq \frac{2d|g_1|_2}{\sqrt{n}},$$

which proves (2) and, together with (3), the convergence of the second moments. □

Under somewhat weaker assumptions we have the following central limit theorem with the convergence of the first absolute moment.

Theorem 3 *Let $f \in L_1^{sym}(\mu^d)$ be a real valued kernel with the symmetric Hoeffding decomposition*

$$f = \sum_{m=0}^{d} \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S.$$

Assume that

(1) *for every $m = 1, \ldots, d$ $R_m f \in L_{m,\pi}^{sym}(\mu^m)$ and the series*

$$\sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^m \\ 0 \leq \mathbf{k} < \infty}} V^{*\mathbf{k}} R_m f \quad \left(\overset{def}{=} \lim_{\substack{n_1 \to \infty \\ \cdots \\ n_m \to \infty}} \sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^m \\ 0 \leq \mathbf{k} < \mathbf{n}}} V^{*\mathbf{k}} R_m f \right) \tag{29}$$

   converges in $L_{m,\pi}(\mu^m)$,
(2) *$R_1 f$ satisfies the relation*

$$\left| \sum_{k=0}^{n-1} V^k R_1 f \right|_1 = O(\sqrt{n}) \quad \text{as } n \to \infty. \tag{30}$$

Then there exists $\sigma^2(f) \geq 0$ such that the sequence

$$V_n^{(d)} f = \frac{1}{n^{d-1/2}} \sum_{\substack{0 \leq k_1 \leq n-1 \\ \cdots \\ 0 \leq k_d \leq n-1}} D_d V^{(k_1, \ldots, k_d)} (f - R_0 f), \quad n \geq 1,$$

converges in distribution to $N(0, d^2\sigma^2(f))$ as $n \to \infty$. The convergence of the first absolute moments

$$E| V_n^{(d)} f| \underset{n\to\infty}{\to} d\sqrt{\frac{2}{\pi}} \sigma(f) \tag{31}$$

holds as well.

Proof The proof is parallel to that of Theorem 2, so we will concentrate on the essential changes in the proof. Consider the Hoeffding decomposition of $f - R_0 f$

$$f - R_0 f = \sum_{m=1}^{d} \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S = \sum_{m=1}^{d} f_m$$

with

$$f_m = \sum_{S \in \mathcal{S}_d^m} (R_m f) \circ \pi_S, \quad m = 1, \ldots, d.$$

In order to prove the theorem it suffices to establish that

1) for some $\sigma(f) \geq 0$, $V_n^{(d)} f_1$ converges in distribution to $N(0, d^2\sigma^2(f))$,
2) $|V_n^{(d)} f_1|_1 \underset{n\to\infty}{\to} d\sqrt{\frac{2}{\pi}} \sigma(f)$,
3) $|V_n^{(d)} \sum_{m=2}^{d} f_m|_1 \underset{n\to\infty}{\to} 0$.

Analogously to the proof of Theorem 2, the functions $f_m, 1 \leq m \leq d$, can be shown to satisfy the inequality

$$|V_n^{(d)} f_m|_1 \leq C_m \binom{d}{m} n^{-(m-1)/2} | g_m|_{m, m, \pi},$$

where g_m denotes the sum of the series (29). For $m \geq 2$ the latter bound implies the convergence in $L_1(\mu)$ to zero, proving 3). Taking $m = 1$, we obtain

$$V_n^{(d)} f_1 = d\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} V^k R_1 f,$$

where $R_1 f$ has the representation $R_1 f = g_1 - V^* g_1$ with $g_1 \in L_1(\mu)$ denoting the sum of the series (29) with $m = 1$. As in the proof of Theorem 2, $R_1 f$ can be represented in the form

$$R_1 f = (I - V V^*) g_1 + (V - I) V^* g_1,$$

where the first summand defines an ergodic stationary sequence of reversed martingale difference $(V^k (I - V V^*) g_1)_{k \geq 0}$, while the second one only contributes a

uniformly L_1-bounded amount to each of the sums $\sum_{k=0}^{n-1} V^k R_1 f$. However, now we only have $(I - VV^*)g_1 \in L_1(\mu)$, while we need $(I - VV^*)g_1 \in L_2(\mu)$ to apply the Billingsley–Ibragimov CLT. The latter can be concluded, as suggested in [31], from (30) using another Burkholder inequality (Theorem 8 in [13]) and the ergodic theorem (see [12] for details). This proves the convergence in distribution. The convergence of the first moments can be concluded similarly to the corresponding part in the proof of Theorem 2. □

Remark 11 In the statement of Theorem 3 the requirement (30) can be substituted by the relation

$$
\left| \sum_{\substack{0 \le k_1 \le n-1 \\ \cdots \\ 0 \le k_d \le n-1}} D_d V^{(k_1, \ldots, k_d)}(f - R_0 f) \right|_1 = O(n^{d-1/2}) \quad \text{as } n \to \infty.
$$

8 A limit theorem for canonical kernels of degree 2

Apart from non-degenerate kernels of the previous section, a different type of von Mises statistics emerges from canonical symmetric kernels of degree $d \ge 2$. Limit distributions of V-statistics defined by such kernels are usually described in terms of series (or polynomials) in Gaussian variables, or in terms of multiple stochastic integrals. In the case of V-statistics of dependent variables some descriptions of the limits in terms of dependent Gaussian variables or non-orthogonal stochastic integrals are known [7,8,26]. A rather attractive way is to present the limit distribution, like in the i.i.d. case, in terms of *independent* Gaussian variables. This will be done below in the case $d = 2$ and is based on the *diagonalization* of the symmetric kernel. The point is that the diagonalization here is applied, instead of the original kernel, to a *martingale kernel* which emerges as a leading summand in the *martingale-coboundary representation* of the original kernel. Notice that the diagonalization of martingale kernels is also used in [41]; in the present work, however, martingale kernels are considered as a subclass to which the study of much more general kernels is reduced.

We assume that $f = f_2$ in terms of the Hoeffding decomposition for symmetric kernels (see Remark 6 in Sect. 4.2). Let θ denote the involution in $(X^2, \mathcal{F}^{\otimes 2}, \mu^2)$ interchanging the multipliers in the Cartesian product. We consider the spaces $L_{2,\pi}(\mu^2)$ and $L_{2,\pi}^{sym}(\mu^2)$ as embedded in $L_2(\mu^2)$.

Proposition 7 Let $f (= f_2) \in L_{2,\pi}^{sym}(\mu^2)$ be a canonical kernel of degree 2. If the limit

$$
g \stackrel{def}{=} \lim_{n_1, n_2 \to \infty} \sum_{\substack{0 \le i_1 \le n_1-1 \\ 0 \le i_2 \le n_2-1}} V^{*(i_1, i_2)} f \tag{32}
$$

exists in $L_{2,\pi}(\mu^2)$, then f admits a unique representation of the form [2]

$$f = g^{\emptyset} + (V^{(1,0)} - I)g^{\{1\}} + (V^{(0,1)} - I)g^{\{2\}} + (V^{(1,0)} - I)(V^{(0,1)} - I)g^{\{1,2\}}, \quad (33)$$

where

$$E(g^{\emptyset}|T^{-(1,0)}\mathcal{F}^{\otimes 2}) = 0, \qquad E(g^{\emptyset}|T^{-(0,1)}\mathcal{F}^{\otimes 2}) = 0,$$
$$E(g^{\{1\}}|T^{-(0,1)}\mathcal{F}^{\otimes 2}) = 0, \quad E(g^{\{2\}}|T^{-(1,0)}\mathcal{F}^{\otimes 2}) = 0 \qquad (34)$$

and $g^{\emptyset}$, $g^{\{1\}}$, $g^{\{2\}}$, $g^{\{1,2\}}$ are canonical. The functions $g^{\emptyset}$, $g^{\{1\}}$, $g^{\{2\}}$, $g^{\{1,2\}}$ in (33) *are uniquely determined by the above properties; moreover, $g^{\emptyset}$, $g^{\{1,2\}} \in L_{2,\pi}^{sym}(\mu^2)$, $g^{\{1\}}$, $g^{\{2\}} \in L_{2,\pi}(\mu^2)$, $g^{\{1\}} \circ \theta = g^{\{2\}}$ and $g^{\{2\}} \circ \theta = g^{\{1\}}$.*

Proof Up to the details related to symmetry the proposition follows from Example 2.1 in [33]. We propose, however, a partially independent proof based on the decomposition of f presented by Lemma 5 with $m = 2$. Set $\mathbf{e}_1 = (1,0)$, $\mathbf{e}_2 = (0,1)$. Vanishing of conditional expectations follows from the presence of the operators $I - V^{\mathbf{e}_l}V^{*\mathbf{e}_l}$ $(l = 1, 2)$ in corresponding summands in view of (16) (recall that $V^{\mathbf{e}_1}$, $V^{*\mathbf{e}_1}$ commute with $V^{\mathbf{e}_2}$, $V^{*\mathbf{e}_2}$). Further, the operators $I - V^{\mathbf{e}_l}V^{*\mathbf{e}_l}$ preserve canonicity since so do I, $V^{\mathbf{e}_l}$ and $V^{*\mathbf{e}_l}$. Hence, the functions $g^{\emptyset}$, $g^{\{1\}}$, $g^{\{2\}}$, $g^{\{1,2\}}$ are canonical because so are the functions h^S in Lemma 5; this lemma also implies the uniqueness of the summands in the representation (33). To establish the uniqueness claimed in the proposition we need to prove that canonical solutions to the equations

$$(V^{\mathbf{e}_1} - I)g^{\{1\}} = 0, (V^{\mathbf{e}_2} - I)g^{\{2\}} = 0, (V^{\mathbf{e}_1} - I)(V^{\mathbf{e}_2} - I)g^{\{1,2\}} = 0$$

vanish. Applying $V^{*\mathbf{e}_l}$ to the first equation, we obtain $(I - V^{*\mathbf{e}_l})g^{\{1\}} = 0$ or $g^{\{1\}} = V^{*\mathbf{e}_l}g^{\{1\}}$. Iterating the latter equation gives $g^{\{1\}} = V^{*n\mathbf{e}_1}g^{\{1\}}$ for every $n \geq 1$. For a canonical $g^{\{1\}}$ the right hand side of the last equation tends to 0 as $n \to \infty$; hence $g^{\{1\}} = 0$. Other equations can be treated similarly. The symmetry of $g^{\emptyset}$ and $g^{\{1,2\}}$ follows from the symmetry of f and the uniqueness. Then we apply θ to the decomposition of a symmetric f and use the uniqueness of summands in the decomposition (33) with symmetric $g^{\emptyset}$ and $g^{\{1,2\}}$. By uniqueness we obtain $g^{\{1\}} \circ \theta = g^{\{2\}}$ and $g^{\{2\}} \circ \theta = g^{\{1\}}$. □

Assume, in addition, that the kernel f is real-valued. The function $g^{\emptyset}$ is the real-valued kernel of a symmetric trace class integral operator in $L_2(\mu)$. Hence, it admits the eigenfunction decomposition

$$g^{\emptyset}(x_1, x_2) = \sum_{m=1}^{\infty} \lambda_m \varphi_m(x_1)\varphi_m(x_2) \qquad (35)$$

where $(\varphi_m)_{m \geq 1}$ is a normalized orthogonal sequence in $L_2(\mu)$ and $(\lambda_m)_{m \geq 1}$ is a real sequence (of not necessarily distinct numbers) for which $\sum_{m=1}^{\infty} |\lambda_m| < \infty$. We shall

---

[2] Upper indices here follow Lemma 5.

assume that $\lambda_m \neq 0$ for every $m \geq 1$. Moreover, since we consider $L_2(\mu)$ over $\mathbb{C}$, we assume that the functions $(\varphi_m)_{m \geq 1}$ are chosen real-valued (this is always possible since g is symmetric and real-valued).

Theorem 4 *Let f be a real-valued canonical kernel satisfying the assumptions of Proposition 7. Then, as $n \to \infty$, the sequence of random variables*

$$\frac{1}{n} \sum_{0 \leq i_1, i_2 \leq n-1} D_2 V^{(i_1, i_2)} f, \quad n \geq 1,$$

converges in distribution to

$$\xi \stackrel{def}{=} \sum_{m=1}^{\infty} \lambda_m \eta_m^2,$$

where $(\eta_m)_{m=1}^{\infty}$ is a sequence of independent standard Gaussian variables. Moreover,

$$E\left(\frac{1}{n} \sum_{0 \leq i_1, i_2 \leq n-1} D_2 V^{(i_1, i_2)} f\right) \xrightarrow[n \to \infty]{} \sum_{m=1}^{\infty} \lambda_m.$$

Proof Setting in (19) $m = 2, p = 2, r = 1$, we obtain with $s = 1$

$$\left| \sum_{0 \leq i_1, i_2 \leq n-1} D_2 V^{(i_1, i_2)} \left((V^{(1,0)} - I) g^{\{1\}} + (V^{(0,1)} - I) g^{\{2\}} \right) \right|_1 \leq 2 C_{2,2,1} \sqrt{n} \, |g|_{2,2,\pi}$$

and with $s = 2$

$$\left| \sum_{0 \leq i_1, i_2 \leq n-1} D_2 V^{(i_1, i_2)} \left((V^{(1,0)} - I)(V^{(0,1)} - I) g^{\{1,2\}} \right) \right|_1 \leq C_{2,2,2} \, |g|_{2,2,\pi}.$$

These two inequalities and decomposition (33) imply that

$$\left| \frac{1}{n} \sum_{0 \leq i_1, i_2 \leq n-1} D_2 V^{(i_1, i_2)} (f - g^{\emptyset}) \right|_1 \xrightarrow[n \to \infty]{} 0$$

which reduces the proof to the special case of the kernel $g^{\emptyset}$.

Let us show next that for every $m \geq 1$

$$E(\varphi_m | T^{-1} \mathcal{F}) = 0.$$

We have $\mu \times \mu$—almost surely

$$0 = E(g^{\emptyset} | T^{-(0,1)} \mathcal{F}^{\otimes 2})(x_1, x_2) = \sum_{l=1}^{\infty} \lambda_l \varphi_l(x_1) E(\varphi_l | T^{-1} \mathcal{F})(x_2)$$

which for every m implies, via multiplying by $\varphi_m(x_1)$ and integrating over x_1 with respect to μ, that $\lambda_m E(\varphi_m | T^{-1}\mathcal{F})(x_2) = 0$. Thus we have

$$E(\varphi_m | T^{-1}\mathcal{F}) = 0$$

μ—almost surely for every $m \geq 1$.

Define now a random variable ξ_N and a truncated kernel $g_N^{\emptyset}$ by setting

$$\xi_N = \sum_{m=1}^{N} \lambda_m \eta_m^2, \quad g_N^{\emptyset}(x_1, x_2) = \sum_{m=1}^{N} \lambda_m \varphi_m(x_1) \varphi_m(x_2).$$

Observe that for every N the assertions of the theorem on the convergence in distribution and the convergence of the first moments hold for for $g_N^{\emptyset}$ and ξ_N, when f is replaced by $g_N^{\emptyset}$. Indeed, the Billingsley–Ibragimov theorem applies to reversed $\mathbb{R}^N$-valued martingale differences (this is straightforward via the Cramer–Wold device). So, the random vectors

$$\left(\frac{1}{\sqrt{n}} \sum_{k=o}^{n-1} \varphi_1 \circ T^k, \ldots, \frac{1}{\sqrt{n}} \sum_{k=o}^{n-1} \varphi_N \circ T^k \right)$$

converge in distribution to $(\eta_1, \ldots, \eta_N)$ as $n \to \infty$. Hence, the random variables

$$\frac{1}{n} \sum_{0 \leq i_1, i_2 \leq n-1} D_2 V^{(i_1, i_2)} g_{\emptyset}^{(N)} = \sum_{m=1}^{N} \lambda_m \left(\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \varphi_m \circ T^k \right)^2$$

converge in distribution to $\sum_{m=1}^{N} \lambda_m \eta_m^2$ as $n \to \infty$. The convergence of the first moments follows here from the convergence of the second moments in the CLT for martingale differences. Observe now that

$$|\xi - \xi_N|_1 = \left| \sum_{m=N+1}^{\infty} \lambda_m \eta_m^2 \right|_1 \leq \sum_{m=N+1}^{\infty} |\lambda_m| \xrightarrow[N \to \infty]{} 0.$$

Hence, $(\xi_n)_{n\geq 1}$ converges to ξ in distribution along with the first moment. Combining this with the fact that

$$
\left| \frac{1}{n} \sum_{0 \leq i_1, i_2 \leq n-1} D_2 V^{(i_1, i_2)} g^\emptyset - \frac{1}{n} \sum_{0 \leq i_1, i_2 \leq n-1} D_2 V^{(i_1, i_2)} g_N^\emptyset \right|_1
$$

$$
\leq \left| \sum_{m=N+1}^{\infty} \lambda_m \left(\frac{1}{\sqrt{n}} \sum_{0 \leq i \leq n-1} \varphi_m \circ T^i \right) \otimes \left(\frac{1}{\sqrt{n}} \sum_{0 \leq i \leq n-1} \varphi_m \circ T^i \right) \right|_{2, 2, \pi}
$$

$$
\leq \sum_{m=N+1}^{\infty} |\lambda_m| \underset{N \to \infty}{\to} 0
$$

holds uniformly in n (we used here that the functions $(\varphi_m \circ T^i)_{1 \leq m, \, 1 \leq i}$ are orthonormal), the proof is completed. $\qquad\square$

9 Exemplary applications

In this section we show how the results of the present paper can be applied in situations familiar to specialists in limit theorems for dynamical systems or weakly dependent random variables. We develop only a few of all possible applications and we do not optimize our assumptions. Instead, we show how certain earlier known and some new results can be deduced from ours. Applications of Theorem 1 were given in Corollaries 2 and 3.

9.1 Doubling transformation

Let $X = \{z \in \mathbb{C} : |z| = 1\}$, μ be the probability Haar measure on X, $Tz = z^2$, $z \in X$. Clearly,

$$
(Vf)(x) = f(x^2), \quad (V^*f)(x) = 1/2 \sum_{\{u : u^2 = x\}} f(u).
$$

T is known to be exact [48]. If $f_1 \in L^2(\mu)$ and $\int_X f_1(x)\mu(dx) = 0$ then the series

$$
\sum_{k \geq 0} V^{*k} f_1
$$

converges in $L^2(\mu)$ under very mild conditions. For example, the condition

$$
\sum_{k \geq 0} w^{(2)}(f_1, 2^{-k}) < \infty
$$

is sufficient. Here $w^{(2)}(f_1, \cdot)$ is the modulus of continuity of f_1 in $L^2(\mu)$.

(a) **Translation-invariant kernels** Let now $f \in L^2(\mu^2)$ be of the form

$$f(x_1, x_2) = g(x_1 x_2^{-1}) \tag{36}$$

with some $g(x) = \sum_{k \in \mathbb{Z}} g_k x^k \in L^2(\mu)$. Assume that $f = f_2$ (that is f is canonical), real-valued and symmetric. This means that $g_0 = 0$, g_k are real and satisfy $g_{-k} = g_k$ for all $k \in \mathbb{Z}$. Assume, moreover, that $f_2 \in L_{2,\pi}^{sym}(\mu^2)$. In our setup this is equivalent to the relation

$$\sum_{k \in \mathbb{Z}} |g_k| < \infty. \tag{37}$$

The condition of the existence of the limit

$$\lim_{n \to \infty} \sum_{0 \le i_1, i_2 \le n-1} V^{*(i_1, i_2)} f_2$$

in $L_{2,\pi}(\mu^2)$ is satisfied if the series $\sum_{k=0}^{\infty} n V^{*n} g$ is norm convergent in the space of absolutely convergent trigonometric series, that is

$$\sum_{k \in \mathbb{Z}} \sum_{n \ge 0} n |g_{2^n k}| < \infty.$$

The latter condition holds, for example, if for some $C > 0$ and $\delta > 0$

$$|g_m| \le \frac{C}{|m|(\log |m|)^{1+\delta}}$$

for every $m \in \mathbb{Z}$, $m \ne 0$. This condition is a very mild strengthening of (37).

(b) **General kernels** Consider now (compare Proposition 6) a general kernel $f \in L_2(X^2, \mathcal{F}^{\otimes 2}, \mu^2)$ with Fourier expansion

$$f(x_1, x_2) = \sum_{k_1, k_2 \in \mathbb{Z}} f_{k_1, k_2} x_1^{k_1} x_2^{k_2}, \quad x_1, x_2 \in X.$$

Assume that the kernel f is real-valued and symmetric, that is $f_{-k_1, -k_2} = \overline{f}_{k_1, k_2}$ and $f_{k_2, k_1} = f_{k_1, k_2}$ for $k_1, k_2 \in \mathbb{Z}$. Following the notation of Remark 6, we have $f_0 = f_{0,0}$, $f_1(x) = \sum_{k \in \mathbb{Z} \setminus \{0\}} f_{k,0} x^k$, $f_2(x_1, x_2) = \sum_{k_1, k_2 \in \mathbb{Z} \setminus \{0\}} f_{k_1, k_2} x_1^{k_1} x_2^{k_2}$. The kernel f satisfies all conditions of Theorems 2 and 4 whenever

$$\sum_{n \ge 0} \left(\sum_{k \in \mathbb{Z} \setminus \{0\}} |f_{2^n k, 0}|_2^2 \right)^{1/2} < \infty \quad \text{and} \quad \sum_{n_1, n_2 \ge 0} \sum_{k_1, k_2 \in \mathbb{Z} \setminus \{0\}} |f_{2^{n_1} k_1, 2^{n_2} k_2}| < \infty.$$

Remark 12 In this subsection we gave applications of our results to the simplest example of a *differentiable expanding map*. This is based on the group structure of the example and its Fourier analysis. A more general approach can be developed on the basis of the *transfer operator* (V^* in our setting) restricted to some spaces of nice (smooth, Hölder or Sobolev) functions.

9.2 Stationary processes (martingale kernels, mixing conditions, Markov processes)

Let $\xi = (\xi_n)_{n \in \mathbb{Z}}$ be an ergodic stationary random process defined on the space $(X, \mathcal{F}, \mu)$ where an invertible measure preserving transformation T acts so that $\xi_{n+1} = \xi_n \circ T$, $n \in \mathbb{Z}$. We assume that all ξ_n take values in a probability space $(Y, \mathcal{G}, \nu)$, ν being the common distribution of $(\xi_n)_{n \in \mathbb{Z}}$. Let $(X^d, \mathcal{F}^{\otimes d}, \mu^d)$ be the d-th Cartesian power of $(X, \mathcal{F}, \mu)$ with the coordinatewise action of $(T^{\mathbf{n}})_{\mathbf{n} \in \mathbb{Z}^d}$ and the corresponding operators $(V^{\mathbf{n}})_{\mathbf{n} \in \mathbb{Z}^d}$; let, furthermore, $(\xi_n^{(i)})_{n \in \mathbb{Z}}$, $1 \leq i \leq d$, be independent copies of $(\xi_n)_{n \in \mathbb{Z}}$ defined on $(X^d, \mathcal{F}^{\otimes d}, \mu^d)$ so that $\xi_n^{(i)}(x_1, \ldots, x_d) = \xi_n(x_i)$, where $x_1, \ldots, x_d \in X$, $1 \leq i \leq d, n \in \mathbb{Z}$. Assume now that we are given some $F \in L_{p, \pi}(Y^d, \mathcal{G}^{\otimes d}, \nu^d)$ for some $d \in \mathbb{N}$ and $p \in [1, \infty)$. Then $f = F(\xi_0^{(1)}, \ldots, \xi_0^{(d)}) \in L_{p, \pi}(X^d, \mathcal{F}^{\otimes d}, \mu^d)$, $F(\xi_{n_1}^{(1)}, \ldots, \xi_{n_d}^{(d)}) = V^{\mathbf{n}} f$ and $F(\xi_{n_1}, \ldots, \xi_{n_d}) = D_d V^{\mathbf{n}} f$ for every $\mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{Z}^d$.

In the rest of the paper, instead of saying that an assertion of the previous part of the paper applies to a kernel f and a transformation T, we will usually say that this assertion applies to the kernel F (the process ξ will be omitted).

9.2.1 Martingale kernels

Let $d = 2$. Set $\mathcal{F}_0 = \sigma(\xi_0, \xi_{-1}, \ldots)$, the σ-field generated by $\xi_0, \xi_{-1}, \ldots$, and $\mathcal{F}_n = T^{-n}\mathcal{F}_0 = \sigma(\xi_n, \xi_{n-1}, \ldots)$. Assume that $f = F(\xi_0^{(1)}, \xi_0^{(2)})$ is a canonical kernel. Obviously, it is measurable with respect to $\mathcal{F}_0^{(1)} \otimes \mathcal{F}_0^{(2)} \left(\stackrel{\text{def}}{=} \sigma((\xi_0^{(1)}, \xi_{-1}^{(1)}, \ldots, \xi_0^{(2)}, \xi_{-1}^{(2)}, \ldots))\right)$.

The equivalent of (32) for invertible T is the existence of the limit

$$\lim_{n_1, n_2 \to \infty} \sum_{\substack{0 \leq i_1 \leq n_1 - 1 \\ 0 \leq i_2 \leq n_2 - 1}} V^{(i_1, i_2)} E\left(F(\xi_0^{(1)}, \xi_0^{(2)}) \mid \mathcal{F}_{-i_1}^{(1)} \otimes \mathcal{F}_{-i_2}^{(2)}\right)$$

in the space $L_{2, \pi}(X^2, \mathcal{F}^{\otimes 2}, \mu^2)$. To compare our Theorem 4 (in its invertible modification, see Remark 8) with the main limit theorem in [41] notice that it is assumed there that the kernel F is *symmetric* and satisfies $E\left(F(\xi_0^{(1)}, \xi_0^{(2)}) \mid \mathcal{F}_{-1}^{(1)} \otimes \mathcal{F}_0^{(2)}\right) = 0$. This implies that a non-vanishing summand may appear in the above sum only for $i_1 = i_2 = 0$, so we have nothing more to check in this case.

9.2.2 Processes satisfying mixing conditions

For $k \in \mathbb{Z}$ we set $\mathcal{F}_k = \sigma(\xi_l, l \leq k)$, $\mathcal{F}^k = \sigma(\xi_l, l \geq k)$, $\mathcal{F}(k) = \sigma(\xi_k)$; let E_k, E^k, $E(k)$ denote the corresponding conditional expectation operators, E being the unconditional expectation. For the system of σ-fields $(\mathcal{F}_k, \mathcal{F}^k)_{k \in \mathbb{Z}}$ and $n \in \mathbb{Z}_+$ define the well-known *mixing coefficients* by setting

$$\alpha(n) = \sup_{A \in \mathcal{F}_k, B \in \mathcal{F}^{k+n}} |\mu(A \cap B) - \mu(A)\mu(B)|,$$

$$\varphi(n) = \sup_{A \in \mathcal{F}_k, \, \mu(A) > 0, B \in \mathcal{F}^{k+n}} \mu^{-1}(A)|\mu(A \cap B) - \mu(A)\mu(B)|,$$

$$\psi(n) = \sup_{A \in \mathcal{F}_k, B \in \mathcal{F}^{k+n}, \, \mu(A)\mu(B) > 0} \mu^{-1}(A)\mu^{-1}(B)|\mu(A \cap B) - \mu(A)\mu(B)|.$$

For the norms of the operators $E^{k+n}E_k - E$, $E_k E^{k+n} - E$ which act from $L_q(\mu)$ to $L_p(\mu)(p, q \in [1, \infty])$ certain bounds in terms of the mixing coefficients are known [14,38]. Indeed, we have for $1 \leq p \leq q \leq \infty$

$$\max(|E^{k+n}E_k - E|_{q, p}, |E_k E^{k+n} - E|_{q, p}) \leq C(q, p)\alpha(n)^{p^{-1} - q^{-1}} \quad (38)$$

$$|E_k E^{k+n} - E|_{q, p} \leq 2\varphi(n)^{1 - q^{-1}}, \quad (39)$$

and for $1 \leq p, q \leq \infty$

$$\max(|E^{k+n}E_k - E|_{q, p}, |E_k E^{k+n} - E|_{q, p}) \leq \psi(n). \quad (40)$$

Notice that if at least one of the mixing coefficients tends to 0, the process ξ is Kolmogorov (see Remark 8) and consequently ergodic. Set

$$M_{q, p} = \sum_{n \geq 0} |E_k E^{k+n} - E|_{q, p}, \quad M'_{q, p} = \sum_{n \geq 0} |E^{k+n}E_k - E|_{q, p}$$

(in view of stationarity $M_{q, p}$ and $M'_{q, p}$ do not depend on k).

In the rest of 9.2.2 we show how Proposition 6 (more precisely, its analogue for an invertible T) can be used applying the results of the paper to V-statistics of a process ξ with suitable mixing properties.

Let $(\epsilon_k)_{k=0}^{\infty}$ be a sequence of functions satisfying

$$\epsilon_k \in L_q(Y, \mathcal{G}, \nu), \quad |\epsilon_k|_q = 1 \quad (k \geq 0),$$

$$\epsilon_0 \equiv 1, \quad \int_Y \epsilon_k(y)\nu(dx) = 0 \quad (k \geq 1). \quad (41)$$

Set $e_k = \epsilon_k \circ \xi_0, k \geq 0$, and fix some $p \in [1, q]$. Observe that for every $k \geq 1$

$$C_{p, k} = \sum_{n \geq 0} |E_{-n}e_k|_p = \sum_{n \geq 0} |(E_{-n}E^0 - E)e_k|_p \leq M_{q, p}|\epsilon_k|_q = M_{q, p}. \quad (42)$$

Assume that a function $F \in L_q(Y^m, \mathcal{G}^{\otimes m}, \nu^m)$ expands into the series

$$F(y_1, \ldots, y_m) = \sum_{0 < \mathbf{k} < \infty} \lambda_{\mathbf{k}}^F \, \epsilon_{k_1}(y_1) \cdots \epsilon_{k_m}(y_m) \tag{43}$$

with some family $(\lambda_{\mathbf{k}}^F)_{0 < \mathbf{k} < \infty}$. For the following expression of the type of (23) we have

$$C_p^F \stackrel{\text{def}}{=} \sum_{0 < \mathbf{k} < \infty} |\lambda_{\mathbf{k}}^F| \, C_{p, k_1} \cdots C_{p, k_m} \le (M_{q, p})^m \sum_{0 < \mathbf{k} < \infty} |\lambda_{\mathbf{k}}^F|, \tag{44}$$

so $C_p^F < \infty$ whenever

$$M_{q, p} < \infty$$

(this is a condition on the mixing rate of the process ξ) and expansion (43) of the function F satisfies the condition

$$\sum_{0 < \mathbf{k} < \infty} |\lambda_{\mathbf{k}}^F| < \infty. \tag{45}$$

Thus, *the invertible version of Proposition* 6 *applies to the kernel* $f : (x_1, \ldots, x_m) \mapsto$ $F(\xi_0(x_1), \ldots, \xi_0(x_m))$ *with some* $p \in [1, \infty]$ *and the system* $(e_k)_{k=0}^{\infty}$ *if, for a certain* $q \in [p, \infty]$, *the system* $(\epsilon_k)_{k=0}^{\infty}$ *satisfies the conditions* (41), $F \in L_q(Y^m, \mathcal{G}^{\otimes m}, \nu^m)$ *admits the representation* (43), *satisfying* (45), *and we have* $M_{q, p} < \infty$ *for the process* ξ. We now indicate conditions (stated in terms of α, φ and ψ) under which Theorems 2, 3 and 4 of the paper, in their invertible forms and numerated by 2′, 3′ and 4′, apply to an F. Theorem 3 needs more substantial changes in case of the mixing coefficient φ. Below $(\epsilon_k)_{k \ge 0}$ is a system satisfying (41) with some parameter q.

   (a) Let $q \in [2d, \infty]$. We will use (38), (39) and (40), substituting there, in place of the pair (q, p), the pair $(q, 2d)$; we will employ Proposition 6 and formulas (42), (44) with $p = 2d$. Theorem 2′ applies to an $F \in L_2^{sym}(\nu^d)$ if

(1) at least one of the series

$$\sum_{n \ge 0} \alpha(n)^{(2d)^{-1} - q^{-1}}, \sum_{n \ge 0} \varphi(n)^{1 - q^{-1}}, \sum_{n \ge 0} \psi(n) \tag{46}$$

   converges (for $q = 2d$ the convergence of the α-series means that $\alpha(n) = 0$ for $n \ge n_0$), and

(2) for every $m = 2, \ldots, d$ $R_m F$ belongs to $L_q^{sym}(\nu^m)$ and admits the representation

$$R_m F(y_1, \ldots, y_m) = \sum_{0 < \mathbf{k} < \infty} \lambda_{\mathbf{k}}^{R_m F} \, \epsilon_{k_1}(y_1) \cdots \epsilon_{k_m}(y_m) \tag{47}$$

   where the coefficients satisfy $\sum_{0 < \mathbf{k} < \infty} |\lambda_{\mathbf{k}}^{R_m F}| < \infty$.

Under condition 2) with $q = 2d$ Theorem $2'$ applies, in particular, if $\sum_{n \geq 0} \varphi(n)^{1-(2d)^{-1}} < \infty$.

(b) To simplify the statements involving φ assume that $d \geq 2$. Let $q \in [d, \infty]$. We will use (38), (39) and (40), substituting there, in place of the pair (q, p), the pair (q, d); we will employ Proposition 6 and formulas (42), (44) with $p = d$. Theorem $3'$ applies to an $F \in L_1^{sym}(\nu^d)$ if

(1) at least one of the series

$$\sum_{n \geq 0} \alpha(n)^{d^{-1}-q^{-1}} \sum_{n \geq 0} \varphi(n)^{1-q^{-1}}, \ \sum_{n \geq 0} \psi(n) \tag{48}$$

converges (if $q = d$ the convergence of the α-series means that $\alpha(n) = 0$ for $n \geq n_0$);

(2) $R_1 F$ satisfies the relation (30):

$$\left| \sum_{k=0}^{n-1} (R_1 F) \circ \xi_k \right|_1 = O(\sqrt{n});$$

(3) for every $m = 2, \ldots, d$ $R_m F$ belongs to $L_q^{sym}(\nu^m)$ and admits the representation

$$R_m F(y_1, \ldots, y_m) = \sum_{0 < \mathbf{k} < \infty} \lambda_{\mathbf{k}}^{R_m F} \epsilon_{k_1}(y_1) \cdots \epsilon_{k_m}(y_m) \tag{49}$$

where the coefficients satisfy $\sum_{0 < \mathbf{k} < \infty} |\lambda_{\mathbf{k}}^{R_m F}| < \infty$.

Under conditions 2) and 3) Theorem $3'$ applies, in particular, if $q = 2d$ and $\sum_{n \geq 0} \alpha(n)^{1/2d} < \infty$.

(c) Theorem 4 leads to a result on mixing processes in the following way. Let $F \in L_{2,\pi}^{sym}(Y^2, \mathcal{G}^{\otimes 2}, \nu^2)$ be a canonical function. Hence, it is the kernel of a *nuclear* (or *trace class*) symmetric integral operator in $L_2(\nu)$ vanishing on constant functions. The general theory says that in $L_2(\nu)$ there exists an orthogonal normalized sequence $\epsilon_0 \equiv 1, \epsilon_1, \ldots$ and a real sequence $\gamma_1, \gamma_2, \ldots$ such that

$$F(x_1, x_2) = \sum_{k=1}^{\infty} \gamma_k \epsilon_k(x_1) \epsilon_k(x_2), \tag{50}$$

where $\sum_{k=1}^{\infty} |\gamma_k| < \infty$ ($k = 0$ is omitted because F is canonical). Neglecting the assumption of canonicity and symmetry, such functions form exactly the space $L_{2,\pi}(\nu^2)$; the projective norm agrees for symmetric functions with the sum of moduli of the eigenvalues of the corresponding integral operators. Thus $f : (x_1, x_2) \mapsto F(\xi_0(x_1), \xi_0(x_2))$ is a function to apply Proposition 6 with $d = 2$ and $e_k = \epsilon_k \circ \xi_0$ ($k \geq 0$). Then for $k \geq 1$ $C_{2,k} \leq M_{2,2}$. The latter quantity is bounded above by any of the series $\sum_{n \geq 0} \varphi(n)^{1/2}, \sum_{n \geq 0} \psi(n)$. Thus, the invertible version of Theorem 4 applies whenever at least one of these series converges.

Remark 13 The last assertion under the assumption $\sum_{n \geq 0} \varphi(n)^{1/2} < \infty$ is, up to inessential details, Theorem 5 in [26]. In [9] the authors express their doubts on correctness in [26] to substituting a dependent process into the function (50). Our conclusion agrees with that of [26]. In our paper the correctness is a simple consequence of general properties of projective tensor products. However, an elementary reasoning shows that the series (50) absolutely converges in $L_1(X^2, \kappa)$ where κ is an arbitrary probability on X^2 with one-dimensional marginals μ.

9.2.3 Discrete time Markov processes

Let $\xi = (\xi_n)_{n \in \mathbb{Z}}$ be a stationary Markov process defined on the space $(X, \mathcal{F}, \mu)$ where an invertible measure preserving transformation T acts so that $\xi_{n+1} = \xi_n \circ T$, $n \in \mathbb{Z}$. We assume that all ξ_n take values in a probability space $(Y, \mathcal{G}, \nu)$, Y being the *state space* of ξ and ν its *stationary distribution*. We will use the notations $\mathcal{F}_k$, $\mathcal{F}^k$, $\mathcal{F}(k)$, E_k, E^k, $E(k)$ and E as introduced above.

Let Q be the *transition operator* of ξ acting on every space $L_p(\nu)$, $1 \leq p \leq \infty$, with norm 1 and satisfying $E_k f(\xi_{k+1}) = (Qf)(\xi_k)$ for every $f \in L_1(\nu)$ and $k \in \mathbb{Z}$. Assuming $\mathcal{F} = \sigma(\xi_l, l \in \mathbb{Z})$, the process ξ (that is the transformation T) is ergodic if and only if for the transition operator $Q : L_2(\nu) \to L_2(\nu)$ every solution to the equation $Qf = f$ is a constant. To stay within the assumptions of the present paper we assume a stronger relation $Q^n h \underset{n \to \infty}{\to} \int h(y)\nu(dy)(h \in L_1(\nu))$ which implies the Kolmogorov property of ξ.

Let $d \geq 1$ and $(\epsilon_k)_{k=0}^\infty$ be a sequence of functions satisfying (41) with $q = 2d$. Let I_ν denote the identity operator in every space $L_q(\nu)$. Assume that for some $C > 0$ and every $k \geq 1$ the equation $(I_\nu - Q)\phi_k = \epsilon_k$ is solvable and $|\phi_k|_{2d} \leq C$ (notice that the latter condition is fulfilled if the restriction $(I_\nu - Q)|_{L_{2d}^0}$ is invertible, L_{2d}^0 denoting the subspace of functions in L_{2d} with integral 0). Let $F \in L_2(Y^d, \mathcal{G}^{\otimes d}, \nu^d)$ satisfy assumption 2) of paragraph **a)** in 9.2.2 with $q = 2d$. Let, finally, the equation $(I_\nu - Q)g = R_1 F$ have a solution $g \in L_2(\nu)$. Then Theorem 2$'$ applies to $f = F(\xi_0^{(1)}, \ldots, \xi_0^{(d)})$.

References

1. Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T., Weiss, B.: Strong laws for L- and U-statistics. Trans. Am. Math. Soc. **348**(7), 2845–2866 (1992)
2. Arcones, M.A., Giné, E.: Limit theorems for U-processes. Ann. Probab. **21**(3), 1494–1542 (1993)

3. Amanov, A.K.: On the limit distribution of the Mises functional with degenerate kernel for dependent random variables. (Russian) Probabilistic models and mathematical statistics, Collect. Artic., Tashkent, 7–14 (1987)

4. Babbel, B.: Schwache Invarianzprinzipien für verallgemeinerte von Mises Funktionale und U-Statistiken von Funktionalen schwach abhängiger Prozesse im Mehrstichprobenfall. (Weak invariance principles for generalized von Mises functionals and U-statistics of functionals of weakly dependent processes in the multisample case) (German). Göttingen (FRG):Univ. Göttingen, Mathematisch-Naturwissenschaftlicher Fachbereich, Diss (1987)

5. Babbel, B.: Invariance principles for U-statistics and von Mises functionals. J. Stat. Plann. Inference **22**(3), 337–354 (1989)

6. Billingsley, P.: The Lindeberg–Levy theorem for martingales. Proc. Am. Math. Soc. **12**, 788–792 (1961)

7. Borisov, I.S., Bystrov, A.A.: Stochastic integrals and asymptotic analysis of canonical von Mises statistics based on dependent observations. High dimensional probability. IMS Lecture Notes Monograph Series, vol. 51, pp. 1–17. Inst. Math. Statist., Beachwood (2006)

8. Borisov, I.S., Bystrov, A.A.: Limit theorems for canonical von Mises statistics constructed from dependent observations. (Russian). Sibirsk. Mat. Zh. **47**(6), 1205–1217 (2006); transl. in. Siberian Math. J. **47**(6), 980–989 (2006)

9. Borisov, I.S., Volod'ko, N.V.: Orthogonal series and limit theorems for canonical U- and V-statistics of stationarily connected observations. (Russian) Mat. Tr. **11**(1), 25–48 (2008); Transl. in Siberian Adv. Math. **18**(4), 242–257 (2008)

10. Borovkova, S., Burton, R., Dehling, H.: Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. Trans. Am. Math. Soc. **353**(11), 4261–4318 (2001) (electronic)

11. Borovkova, S., Burton, R., Dehling, H.: From dimension estimation to asymptotics of dependent U-statistics. In: Limit Theorems in Probability and Statistics, vol. 1, pp. 201–234 (Balatonlelle, 1999). János Bolyai Mathamatical Society, Budapest (2002)

12. Bradley, R.C.: On some results of M. I. Gordin: a clarification of a misunderstanding. J. Theor. Probab. **1**(2), 115–119 (1988)

13. Burkholder, D.L.: Martingale transforms. Ann. Math. Stat. **37**, 1494–1504 (1966)

14. Davydov, YuA: Convergence of distributions generated by stationary stochastic processes (Russian). Prob. Theory Appl. **13**(4), 730–737 (1968)

15. Dehling, H., Denker, M., Philipp, W.: Invariance principles for von Mises and U-statistics. Z. Wahrsch. Verw. Gebiete **67**(2), 139–167 (1984)

16. Defant, A., Floret, K.: Tensor norms and operator ideals. North-Holland Mathematics Studies, vol. 176. North-Holland Publishing Co., Amsterdam (1993)

17. de la Peña, V.H., Giné, E.: Decoupling. From dependence to independence. Randomly stopped processes. U-statistics and processes. Martingales and beyond. Probab. Appl. (New York), Springer, New York (1999)

18. Dehling, H., Taqqu, M.S.: The limit behavior of empirical processes and symmetric statistics for stationary sequences. In: Proceedings of the 46th Session of the International Statistical Institute, vol. 4 (Tokyo, 1987). Bull. Inst. Intern. Stat. **52**(4), 217–234 (1987)

19. Dehling, H.: Limit theorems for dependent U-statistics. Dependence in probability and statistics. Lecture Notes in Statist, vol. 187, pp. 65–86 . Springer, New York (2006)

20. Dehling, H., Wendler, M.: Central limit theorem and a bootstrap for U-statistics of strongly mixing data. J. Mult. Anal. **101**, 126–137 (2010)

21. Denker, M., Keller, G.: On U-statistics and v. Mises' statistics for weakly dependent processes. Z. Wahrsch. Verw. Gebiete **64**(4), 505–522 (1983)

22. Denker, M., Keller, G.: Rigorous statistical procedures for data from dynamical systems. J. Stat. Phys. **44**(1–2), 67–93 (1986)

23. Denker, M., Grillenberger, C., Keller, G.: A note on invariance principles for v. Mises' statistics. Metrika **32**(3–4), 197–214 (1985)

24. Dunford, N., Schwartz, J.T.: Linear Operators. I. General Theory. Interscience Publishers Inc., New York (1958)

25. Dynkin, E.B., Mandelbaum, A.: Symmetric statistics, Poisson point processes, and multiple Wiener integrals. Ann. Stat. **11**(3), 739–746 (1983)

26. Eagleson, G.K.: Orthogonal expansions and U-statistics. Austral. J. Stat. **21**(3), 221–237 (1979)

27. Filippova, A.A.: Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. Theory Probab. Appl. **7**, 24–57 (1962)
28. Gohberg, I.C., Krein, M.G.: Introduction to the theory of linear nonselfadjoint operators. Translated from the Russian 1965 original. Translations of Mathematical Monographs, vol. 18. American Mathematical Society, Providence (1969)
29. Gohberg, I., Krupnik, N., Goldberg S.: Traces and determinants of linear operators. Oper. Theory Adv. Appl. (Birkhäuser) **116** (2000)
30. Gordin M.I.: On the central limit theorem for stationary processes. (Russian) Dokl. Akad. Nauk SSSR **188**, 739–741 (1969). Transl. in Soviet Math. Dokl. **10**, 1174–1176 (1969)
31. Gordin, M.I.: Central limit theorem for stationary processes without the assumption of finite variance. (Russian). Abstracts of Communications, T. 1: A-K. International Conference on Probability Theory and Mathematical Statistics, pp 173–174, June 25–30, Vilnius (1973)
32. Gordin, M.I., Lifshits, B.A.: Central limit theorem for stationary Markov processes. (Russian) Dokl. Akad. Nauk SSSR **239**, 766–767 (1978); Transl. in Soviet Math. Dokl. **19**(2), 392–394 (1978)
33. Gordin, M.I.: Martingale-coboundary representation for a class of random fields. (Russian). Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov (POMI), **364** (2009), Veroyatnost i Statistika. no. 14.2, pp. 88–108, 236; Transl.: J. Math. Sci. (NY) **163**(2), 363–374 (2009)
34. Grassberger, P., Procaccia, I.: Characterization of strange attractors. Phys. Rev. Lett. **50**, 346–349 (1983)
35. Halmos, P.: The theory of unbiased estimation. Ann. Math. Stat. **17**, 34–43 (1946)
36. Hoeffding, W.: A class of statistics with asymptotically normal distribution. Ann. Math. Stat. **19**, 293–325 (1948)
37. Ibragimov, I.A.: A central limit theorem for a class of dependent random variables (Russian). Prob.Theory Appl. **8**(1), 89–94 (1963)
38. Ibragimov, I.A., Linnik, YuV: Independent and Stationry Sequences of Random Variables. Wolters-Noordhof Publishing, Groningen (1971)
39. Khashimov Sh.A.: Asymptotic normality of a generalized von Mises functional for m-dependent variables. Theory of Random Processes and its Applications (Russian), pp. 141–148. "Naukova Dumka", Kiev (1990)
40. Koroljuk, V.S., Borovskich, Yu.V.: Theory of U-statistics. Math. Appl. **273** Translated from the 1989 Russian original. Kluwer Academic Publications Group, Dordrecht (1989)
41. Leucht, A., Neumann, M.H.: Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. Ann. Inst. Stat. Math. **65**, 349–386 (2013)
42. Leucht, A., Neumann, M.H.: Dependent wild bootstrap for degenerate U- and V-statistics. J. Multivar. Anal. **117**, 257–280 (2013)
43. Leucht, A.: Degenerate U- and V-statistics under weak dependence: asymptotic theory and bootstrap consistency. Bernoulli **18**(4), 552–585 (2012)
44. Maigret, N.: Théorème de limite centrale fonctionnel pour une chaîne de Markov récurrente au sens de Harris et positive. Ann. Inst. H. Poincaré Sect. B (N.S.) **14**(4), 425–440 (1978)
45. Major, P.: Tail behaviour of multiple random integrals and U-statistics. J. Probab. Surv. **2**, 448–505 (2005)
46. Neuhaus, G.: Functional limit theorems for U-statistics in the degenerate case. J. Multivar. Anal. **7**(3), 424–439 (1977)
47. Reed, M., Simon, B.: Methods of modern mathematical physics. Functional Analysis, vol. 1. Academic Press Inc., New York (1980)
48. Rohlin, V.A.: Exact endomorphisms of a Lebesgue space (Russian). Izv. Akad. Nauk SSSR Ser. Mat. **25**, 499–530 (1961)
49. Ryan, R.A.: Introduction to Tensor Products of Banach Spaces. Springer, London (2002)
50. Sharipov, OSh: The invariance principle for U-statistics and von Mises functionals of weakly dependent observations. (Russian). Teor. Veroyatnost. i Primenen. **47**(4), 814–817 (2002); Transl. in Theory Probab. Appl. **47**(4), 730–733 (2003)
51. Volod'ko, N.V.: Limit theorems for canonical von Mises statistics and U-statistics of m-dependent observations. (Russian). Teor. Veroyatn. Primen., **55**(2), 226–249 (2010). Transl. in. Theory Probab. Appl. **55**(2), 271–290 (2011)
52. von Mises, R.: On the asymptotic distribution of differentiable statistical functions. Ann. Math. Stat. **18**, 309–348 (1947)

53. Yoshihara, K.-I.: Limiting behavior of U-statistics for stationary, absolutely regular processes.linebreak
    Z. Wahrsch. Verw. Gebiete **35**(3), 237–252 (1976)
54. Yoshihara, K.-I.: Limiting behavior of U-statistics for strongly mixing sequences. Yokohama Math. J.
    39(2), 107–113 (1992)