Functional Convergence of U-Processes with Size-Dependent Kernels

Christian Döbler, Mikolaj Kasprzak and Giovanni Peccati

Unité de Recherche en Mathématiques, Université du Luxembourg

December 16, 2019

Abstract

We consider sequences of U-processes based on symmetric kernels of a fixed order, that possibly depend on the sample size. Our main contribution is the derivation of a set of analytic sufficient conditions, under which the aforementioned U-processes weakly converge to a linear combination of time-changed independent Brownian motions. In view of the underlying symmetric structure, the involved time-changes and weights remarkably depend only on the order of the U-statistic, and have consequently a universal nature. Checking these sufficient conditions requires calculations that have roughly the same complexity as those involved in the computation of fourth moments and cumulants. As such, when applied to the degenerate case, our findings are infinite-dimensional extensions of the central limit theorems (CLTs) proved in de Jong (1990) and Döbler and Peccati (2017). As important tools in our analysis, we exploit the multidimensional central limit theorems established in Döbler and Peccati (2019) together with upper bounds on absolute moments of degenerate U-statistics by Ibragimov and Sharakhmetov (2002), and also prove some novel multiplication formulae for degenerate symmetric U-statistics — allowing for different sample sizes — that are of independent interest. We provide applications to random geometric graphs and to a class of U-statistics of order two, whose Gaussian fluctuations have been recently studied by Robins *et al.* (2016), in connection with quadratic estimators in non-parametric models. In particular, our application to random graphs yields a class of new functional central limit theorems for subgraph counting statistics, extending previous findings in the literature. Finally, some connections with invariance principles in changepoint analysis are established.

Keywords: U-statistics, functional limit theorems, contractions, product formulae, Hoeffding decompositions, random geometric graphs, asymptotic properties of estimators, Stein's method

AMS 2010 Classification: 60F17, 60D05, 62G20,

1 Introduction

1.1 Overview

Consider a sequence $\{X_i : i = 1, 2, ...\}$ of independent and identically distributed (i.i.d.) random variables with values in some space $(E, \mathcal{E})$. The aim of this paper is to prove a class of Gaussian functional central limit theorems (FCLTs) involving general sequences of **U-processes** with symmetric kernels, that is, of càdlàg processes on the time interval $[0, 1]$, obtained by progressively revealing the argument of a symmetric U-statistic of order $p \geq 1$ based on the sample $(X_1, ..., X_n)$, for $n \geq 1$.

The type of weak convergence we deal with is in the large sample limit $n \to \infty$, and holds in the sense of the Skorohod space $D[0, 1]$ of càdlàg mappings on $[0, 1]$, endowed with Skorohod's J_1 topology (see e.g. [Bil99, p. 123]). The specific difficulty tackled in our work — marking a difference with previous contributions (see e.g. [Neu77, Hal79, MT84, NP87, AG93]) — is that we allow the kernels of the considered U-statistics to explicitly depend on the sample size n, and we do not assume a priori any form of Hoeffding degeneracy.

Despite the generality of the above setup, the limit processes displayed in our results always have the form

$$Z(t) = \sum_{k=1}^{p} \alpha_{k,p} Z_{k,p}(t), \quad t \in [0, 1], \tag{1.1}$$

where each $\alpha_{k,p} \in [0,\infty)$ is a constant depending on the sequence of U-statistics under study, and $\{Z_{k,p}(t) : t \in [0,1],\ 1 \leq k \leq p\}$ denotes a class of independent centered Gaussian processes obtained as follows: first consider a sequence $\{B_k(t) : t \in [0,1],\ 1 \leq k \leq p\}$ of independent standard Brownian motions on $[0,1]$, and then set

$$Z_{k,p}(t) := t^{p-k} B_k(t^k), \quad t \in [0,1], \tag{1.2}$$

in such a way that

$$\Gamma_{k,p}(s,t) := \mathbb{E}[Z_{k,p}(t) Z_{k,p}(s)] = (s \wedge t)^p (s \vee t)^{p-k}, \quad s,t \in [0,1], \tag{1.3}$$

and consequently

$$\Gamma(s,t) := \mathbb{E}[Z(t)Z(s)] = \sum_{k=1}^{p} \alpha_{k,p}^2 \, (s \wedge t)^p (s \vee t)^{p-k}, \quad s,t \in [0,1]. \tag{1.4}$$

Note that, in particular, the processes $(Z(t))_{t\in[0,1]}$ appearing in (1.1) all have continuous paths. Such a rigid asymptotic structure originates from the fact that we exclusively focus on *symmetric* U-statistics and i.i.d. samples: these strong assumptions yield then the emergence of the 'universal' time-changes $t \mapsto t^k$ and time-dependent weights $t \mapsto t^{p-k}$ from purely combinatorial considerations. One should compare this situation with the reference [Bas94], where a Gaussian FCLT is proved for sequences of non-symmetric homogenous sums, displaying as possible weak limits arbitrarily time-changed Brownian motions. We will see in Section 3 that our results contain a (weaker) version of Donsker's Theorem for sums of i.i.d. random variables — see [Bil99, p. 90].

The sufficient conditions for weak convergence discussed above are stated in the forthcoming Theorem 3.1 and Theorem 3.4, and are expressed in terms of the **contraction kernels** canonically associated with a given U-statistic (see Section 2.2 for definitions). We will see in Section 5 that the conditions derived in our paper are a slight strengthening of the sufficient conditions for one-dimensional CLTs derived in [DP19, Section 3] — see Remark 3.6 below for a full discussion of this point. Some of the additional requirements with respect to [DP19] (in particular, Assumption (a) in Theorem 3.1 and Assumption (a') in Theorem 3.4) are necessary and sufficient for the pointwise convergence of the covariance functions of the degenerate U-processes associated with a generic U-statistic via its Hoeffding decompositon: such a technical assumption (that can in principle be relaxed at the cost of more technical statements — see Remark 3.6-(ii)) is unavoidable in the case of degenerate U-statistics, and is automatically verified in the applications developed in Section 4.

As discussed in Section 5 — and similarly to the main findings of [DP19] — when applied to Hoeffding degenerate U-processes (see Section 2), the conditions expressed in Theorems 3.1 and 3.4 are roughly equivalent to the requirement that the joint cumulants of order ≤ 4 associated with the finite-dimensional restrictions of the U-processes under consideration converge to those of an appropriate Gaussian limit, and that such a convergence takes place at a rate of the type $O(1/n^\alpha)$, where $\alpha > 0$. As such, our findings can be regarded as functional versions of the well-known **de Jong CLT** for degenerate U-statistics, first proved in [dJ90] and then substantially extended in [DP17] (to which we refer for an overview of the relevant literature). To the best of our knowledge, apart from the reference [Bas94] (that only deals with homogeneous sums), ours is the first functional version of de Jong's CLT proved in the literature.

In the last four decades, numerous FCLTs for U-processes have been derived by several groups of authors; yet — to the best of our knowledge — none of them have a nature that is directly comparable to our findings. Among the large set of contributions in this domain, we refer the reader to the following relevant sample. References [Neu77, Hal79, MT84] contain functional limit theorems for sequences of degenerate U-statistics with a kernel independent on the sample size n: in such a framework, consistently with the known one-dimensional results (see e.g. [DM83]), the limit process lives in a Wiener chaos of order > 1 and it is therefore non-Gaussian. The already mentioned paper [Bas94] proves Gaussian FCLTs (in a spirit close to [dJ90]) for sequences of homogeneous sums: in this case, there is no overlap with our work since symmetric homogeneous sums (that are in principle contemplated in our setting) are necessarily multiples of degenerate U-statistics with a kernel not depending on the sample size n, and their asymptotic behaviour is consequently non-Gaussian (by virtue of [Neu77, Hal79]). References [NP87, AG93] are influential general contributions to the theory of U-processes, containing uniform FCLTs

for sequences of U-processes indexed by function classes not depending on the sample size, both in the non-degenerate and degenerate case. The recent contribution [CK17] deals with suprema of U-processes indexed by non-degenerate symmetric function classes possibly depending on the sample size and not necessarily verifying a FCLT, and also contains a detailed review of further relevant literature. See also [GM07, Che18], as well as [Bor96, DlPG99] for general references.

As discussed above, our main results are expressed in terms of explicit analytical quantities (e.g. norms of contraction kernels), and they are therefore particularly well-adapted to applications. As a demonstration of this fact, in Section 4 we deduce two new classes of FCLTs, respectively related to subgraph counting in geometric random graphs (retrieving novel functional versions of one-dimensional CLTs from [JJ86, Pen04, BG92, DP19])), and to quadratic U-statistics emerging e.g. in the non-parametric estimation of quadratic functionals of compactly supported densities (see e.g. [BR88, LM00, RLTvdV16]). In Section 3.2, we also illustrate some connections with invariance principles related to changepoint analysis, see e.g. [CH88, CH97, Gom04].

We eventually mention the challenging problem of deriving explicit rates of convergence for the FCLTs derived in this paper. While some promising partial results seem to be obtainable by adapting the infinite-dimensional 'generator approach' to Stein's method developed in [Bar90, BJ09, Kas17], we prefer to consider this point as a separate issue, and leave it open for subsequent research.

1.2 Notation and tightness criteria

From now on, every random object considered in the paper is assumed to be defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{E}$ denoting expectation with respect to $\mathbb{P}$. Given a collection of stochastic processes $\{X, X_n : n \geq 1\}$ with values in $D[0,1]$, we write $X_n \Longrightarrow X$ to indicate that X_n **weakly converges** to X, meaning that $\mathbb{E}[\varphi(X_n)] \to \mathbb{E}[\varphi(X)]$, as $n \to \infty$, for every bounded mapping $\varphi : D[0,1] \to \mathbb{R}$ which is continuous with respect to the Skorohod topology. Given two positive sequences $\{a_n, b_n\}$, we write $a_n \sim b_n$ whenever $a_n/b_n \to 1$, as $n \to \infty$. We will also use the notation $a_n \lesssim b_n$ to indicate that there exists an absolute finite constant C such that $a_n \leq C\, b_n$ for every n.

In several places of the present paper, tightness in the space $D[0,1]$ is established by using the following criterion. The argument in the proof reproduces the strategy adopted in [NN18, Lemma 3.1], and is reported for the sake of completeness.

Lemma 1.1. *Let $X_n = \{X_n(t) : t \in [0,1]\}$, $n \in \mathbb{N}$, be a sequence of stochastic processes with paths a.s. in $D[0,1]$. Suppose that there is a stochastic process $X = \{X(t) : t \in [0,1]\}$ whose paths are a.s. continuous and such that the finite-dimensional distributions of X_n, $n \in \mathbb{N}$, converge to those of X, as $n \to \infty$. Then, we have $X_n \Longrightarrow X$, if there are constants $C > 0$, $\beta > 0$ and $\alpha > 0$ such that, for all $n \in \mathbb{N}$ sufficiently large and for all $0 \leq s < t \leq 1$,*

$$\mathbb{E}\big|X_n(t) - X_n(s)\big|^{\beta} \leq C\left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n}\right)^{1+\alpha}. \tag{1.5}$$

In particular, in this case the sequence $(X_n)_{n\in\mathbb{N}}$ is tight in $D[0,1]$.

Proof. We are going to use the following well-known criterion from [Bil99]: let $X_n = \{X_n(t) : t \in [0,1]\}$, $n \in \mathbb{N}$, be a sequence of stochastic processes with paths a.s. in $D[0,1]$ such that there is a stochastic process $X = \{X(t) : t \in [0,1]\}$ whose paths are a.s. continuous and such that the finite-dimensional distributions of X_n, $n \in \mathbb{N}$, converge to those of X, as $n \to \infty$. Then, the sequence $(X_n)_{n\in\mathbb{N}}$, converges in distribution with respect to the Skorohod topology to X, if there are finite and strictly positive constants C_1, α and γ such that, for all $n \in \mathbb{N}$ sufficiently large and for all $0 \leq r \leq s \leq t \leq 1$,

$$\mathbb{E}\Big[\big|X_n(t) - X_n(s)\big|^{\gamma}\big|X_n(s) - X_n(r)\big|^{\gamma}\Big] \leq C_1(t-r)^{1+\alpha}. \tag{1.6}$$

Note that (1.6) is in fact a more specialized instance of formula (13.14) in [Bil99]. Now assume (1.5) and

observe that

$$\begin{aligned}
&\mathbb{E}\Big[\big|X_n(t)-X_n(s)\big|^{\beta/2}\big|X_n(s)-X_n(r)\big|^{\beta/2}\Big] \\
&\leq \sqrt{\mathbb{E}\Big[\big|X_n(t)-X_n(s)\big|^{\beta}\Big]}\sqrt{\mathbb{E}\Big[\big|X_n(s)-X_n(r)\big|^{\beta}\Big]} \\
&\leq \sqrt{C\bigg(\frac{\lfloor nt\rfloor-\lfloor ns\rfloor}{n}\bigg)^{1+\alpha}}\sqrt{C\bigg(\frac{\lfloor ns\rfloor-\lfloor nr\rfloor}{n}\bigg)^{1+\alpha}} \\
&\leq 3^{1+\alpha}C(t-r)^{1+\alpha}\,, \qquad (1.7)
\end{aligned}$$

where the last inequality follows from an argument used in the proof of [NN18, Lemma 3.1]. Hence, we conclude that (1.6) holds true with $\gamma=\beta/2$ and with $C_1=3^{1+\alpha}C$. □

1.3 Acknowledgments

We are grateful to Yannick Baraud, Andrew D. Barbour, Ivan Nourdin and Omar El-Dakkak for useful discussions. The research developed in this paper is supported by the FNR grant **FoRGES (R-AGR-3376-10)** at Luxembourg University.

1.4 Plan

Section 2 contains some general information about U-statistics and U-processes and several useful estimates for contraction operators. The main results of the paper, which give sufficient conditions for functional convergence of U-processes, are presented in Section 3.1; some connections to changepoint analysis are described in Section 3.2, whereas multidimensional extensions are discussed in Section 3.3. Section 4 deals with applications of our main results to subgraph counting in random geometric graphs and U-statistics of order 2 with a dominant diagonal component. Finally, Section 5 contains some further ancillary results, as well as the proofs of the main results.

2 General setup

2.1 Symmetric U-statistcs and U-processes

As before, we assume that $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables taking values in a measurable space $(E,\mathcal{E})$ (that we fix for the rest of this paper), and denote by μ their common distribution. For a fixed $p\in\mathbb{N}$, let

$$\psi:\big(E^p,\mathcal{E}^{\otimes p}\big)\to\big(\mathbb{R},\mathcal{B}(\mathbb{R})\big)$$

be a symmetric and measurable kernel of order p. By "symmetric" we mean that, for all $x=(x_1,\ldots,x_p)\in E^p$ and each $\sigma\in\mathbb{S}_p$ (the symmetric group acting on $\{1,\ldots,p\}$), one has that

$$\psi(x_1,\ldots,x_p)=\psi(x_{\sigma(1)},\ldots,x_{\sigma(p)})\,.$$

In general, the kernel $\psi=\psi_n$ can (and will most of the time) depend on an additional parameter $n\in\mathbb{N}$ (the size of the sample in the argument of the associated U-statistic), but we will often suppress such a dependence, in order to simplify the notation. We use the symbol μ^p to denote the p-th power of μ (which is a measure on $(E^p,\mathcal{E}^{\otimes p})$).

In what follows, we will write $X:=\{X_i:i\in\mathbb{N}\}$ and, for p,ψ,X as above and $n\in\mathbb{N}$, we define

$$J_p^{(n)}(\psi):=J_{p,X}^{(n)}(\psi):=\sum_{J\in\mathcal{D}_p(n)}\psi(X_j,j\in J)=\sum_{1\leq i_1<\ldots<i_p\leq n}\psi(X_{i_1},\ldots,X_{i_p})\,. \qquad (2.1)$$

where $\mathcal{D}_p(n)$ indicates the subsets of size p of $\{1,...,n\}$. We say that the random variable $J_p^{(n)}(\psi)$ is the **U-statistic of order p, based on $X_1,\ldots,X_n$ and generated by the kernel ψ**. For $p=0$ and a constant $c\in\mathbb{R}$ we further let $J_0(c):=c$.

Let $p \geq 1$, and let $\psi \in L^1(\mu^p)$ be symmetric. The kernel ψ is called **(completely) degenerate** or **canonical** with respect to μ, if

$$\int_E \psi(x_1, x_2, \ldots, x_p) d\mu(x_1) = 0 \quad \text{for } \mu^{p-1}\text{-a.a. } (x_2, \ldots, x_p) \in E^{p-1},$$

or, equivalently, if

$$\mathbb{E}\big[\psi(X_1, \ldots, X_p) \,\big|\, X_1, \ldots, X_{p-1}\big] = 0, \quad \mathbb{P}\text{-a.s.}.$$

Now assume that $\psi \in L^1(\mu^p)$ is a symmetric but not necessarily degenerate kernel. In this case, the random variable $J_p^{(n)}(\psi)$ can be written as the sum of its expectation and a linear combination of symmetric U-statistics with degenerate kernels of respective orders $1, \ldots, p$. More precisely, one has the following **Hoeffding decomposition** of $J_p^{(n)}(\psi)$:

$$J_p(\psi) = \mathbb{E}\big[J_p(\psi)\big] + \sum_{k=1}^{p} \binom{n-k}{p-k} J_k(\psi_k) = \sum_{k=0}^{p} \binom{n-k}{p-k} J_k(\psi_k), \tag{2.2}$$

where

$$\psi_k(x_1, \ldots, x_k) = \sum_{l=0}^{k} (-1)^{k-l} \sum_{1 \leq i_1 < \ldots < i_l \leq k} g_l(x_{i_1}, \ldots, x_{i_l}) \tag{2.3}$$

and the symmetric functions $g_l : E^l \to \mathbb{R}$ are defined by

$$g_l(y_1, \ldots, y_l) := \mathbb{E}\big[\psi(y_1, \ldots, y_l, X_1, \ldots, X_{p-l})\big], \tag{2.4}$$

in such a way that, for $1 \leq k \leq p$, ψ_k is symmetric and degenerate of order k. In particular, one has $g_0 \equiv \psi_0 \equiv \mathbb{E}\big[\psi(X_1, \ldots, X_p)\big]$ and $g_p = \psi$. See e.g. [Ser80, Vit92] for general references on Hoeffding decompositions.

Similarly to (2.1), we can naturally define a U**-process**

$$U = \{U(t) : t \in [0,1]\} \tag{2.5}$$

as follows. For $t \in [0,1]$, write $U(t) := U_n(t) := J_p^{(\lfloor nt \rfloor)}(\psi)$. Then, for every t one has the Hoeffding decomposition

$$U(t) := \mathbb{E}[U(t)] + \sum_{k=1}^{p} \binom{\lfloor nt \rfloor - k}{p-k} J_k^{(\lfloor nt \rfloor)}(\psi_k). \tag{2.6}$$

Whenever ψ is a symmetric element of $L^2(\mu^p)$, we will also make use of the notation

$$W(t) := W_n(t) = \frac{U(t) - \mathbb{E}[U(t)]}{\sigma_n}, \quad t \in [0,1], \tag{2.7}$$

where

$$\sigma_n^2 := \mathrm{Var}(U_n(1)) = \mathrm{Var}(J_p^{(n)}(\psi)) = \sum_{k=1}^{p} \binom{n-k}{p-k}^2 \binom{n}{k} \|\psi_k\|^2_{L^2(\mu^k)} \tag{2.8}$$

$$= \binom{n}{p} \sum_{k=1}^{p} \binom{p}{k} \binom{n-p}{p-k} \mathrm{Var}\big(g_k(X_1, \ldots, X_k)\big). \tag{2.9}$$

Setting

$$\varphi^{(k)} := \varphi^{(n,k)} := \frac{\binom{n-k}{p-k} \psi_k}{\sigma_n}, \quad 1 \leq k \leq p,$$

one has that each $\varphi^{(k)}$ is a degenerate kernel and, using the notation $V_k(t) := J_k^{(\lfloor nt \rfloor)}(\varphi^{(k)})$, one infers the following useful representation of W

$$W(t) = \sum_{k=1}^{p} \frac{\binom{\lfloor nt \rfloor - k}{p-k}}{\binom{n-k}{p-k}} V_k(t), \quad t \in [0,1].$$

It is immediately verified that, for each $n \in \mathbb{N}$, both $W_n := \{W_n(t) : t \in [0,1]\}$ and $U_n := \{U_n(t) : t \in [0,1]\}$ are random elements with values in $D[0,1]$. As anticipated, the aim of this paper is to deduce verifiable analytical conditions, under which the sequence $\{W_n : n \in \mathbb{N}\}$ converges in distribution to some continuous Gaussian process $Z = \{Z(t) : t \in [0,1]\}$ of the form (1.1).

2.2 Contractions

We will now define "contraction kernels" obtained from pairs of square-integrable mappings. These are one of the principal analytical tools exploited in our paper. Given integers $p, q \geq 1$, $0 \leq l \leq r \leq p \wedge q$ and two symmetric kernels $\psi \in L^2(\mu^p)$ and $\varphi \in L^2(\mu^q)$, we define the **contraction kernel** $\psi \star_r^l \varphi$ on $E^{p+q-r-l}$ by the relation

$$\begin{aligned}
&(\psi \star_r^l \varphi)(y_1, \dots, y_{r-l}, t_1, \dots, t_{p-r}, s_1, \dots, s_{q-r}) \\
&:= \int_{E^l} \Big(\psi\big(x_1, \dots, x_l, y_1, \dots, y_{r-l}, t_1, \dots, t_{p-r}\big) \\
&\qquad\qquad \cdot \varphi\big(x_1, \dots, x_l, y_1, \dots, y_{r-l}, s_1, \dots, s_{q-r}\big) \Big) d\mu^l(x_1, \dots, x_l) \qquad (2.10) \\
&= \mathbb{E}\Big[\psi\big(X_1, \dots, X_l, y_1, \dots, y_{r-l}, t_1, \dots, t_{p-r}\big) \\
&\qquad\qquad \cdot \varphi\big(X_1, \dots, X_l, y_1, \dots, y_{r-l}, s_1, \dots, s_{q-r}\big) \Big], \qquad (2.11)
\end{aligned}$$

for every $(y_1, \dots, y_{r-l}, t_1, \dots, t_{p-r}, s_1, \dots, s_{q-r})$ belonging to the set $A_0 \subset E^{p+q-r-l}$ such that the right-hand side of the previous equation is well-defined and finite, and we conventionally set it equal to zero otherwise. Given ψ, φ, r, l as above, we write that the kernel $\psi \star_r^l \varphi$ is **well-defined** if $\mu^{p+q-r-l}(A_0^c) = 0$ (where A_0 is the set defined above). Note that, in general, it is not clear that $\psi \star_r^l \varphi$ is well-defined in the sense specified above, or that the obtained contraction is square-integrable.

If $l = 0$, then (2.10) has to be understood as follows:

$$\begin{aligned}
&(\psi \star_r^0 \varphi)(y_1, \dots, y_r, t_1, \dots, t_{p-r}, s_1, \dots, s_{q-r}) \\
&= \psi(y_1, \dots, y_r, t_1, \dots, t_{p-r}) \varphi(y_1, \dots, y_r, s_1, \dots, s_{q-r}) \,.
\end{aligned}$$

In particular, if $l = r = 0$, then $\psi \star_r^l \varphi$ boils down to the **tensor product**

$$\psi \otimes \varphi : E^{p+q} \to \mathbb{R}$$

of ψ and φ, given by

$$(\psi \otimes \varphi)(x_1, \dots, x_{p+q}) = \psi(x_1, \dots, x_p) \cdot \varphi(x_{p+1}, \dots, x_{p+q}) \,.$$

We observe that $\psi \star_p^0 \psi = \psi^2$ is square-integrable if and only if $\psi \in L^4(\mu^p)$. As a consequence, $\psi \star_r^l \varphi$ might not be in $L^2(\mu^{p+q-r-l})$ even though $\psi \in L^2(\mu^p)$ and $\varphi \in L^2(\mu^q)$. Moreover, if $l = r = p$, then $\psi \star_p^p \psi = \|\psi\|^2_{L^2(\mu^p)}$ is constant.

The next result, taken from [DP19], lists the properties of contraction kernels that are useful for the present work.

Lemma 2.1 ([DP19], Lemma 2.4)**.** *Let $p, q \geq 1$ be integers and fix two symmetric kernels $\psi \in L^2(\mu^p)$ and $\varphi \in L^2(\mu^q)$.*

(i) *For all $0 \leq l \leq r \leq p \wedge q$ the function $\psi \star_r^l \varphi$ given by (2.10) is well-defined, in the sense specified at the beginning of the present subsection.*

(ii) *For all $0 \leq l \leq r \leq p \wedge q$ one has that*

$$\|\psi \star_r^l \varphi\|^2_{L^2(\mu^{p+q-r-l})} \leq \|\psi \star_p^{p-r+l} \psi\|_{L^2(\mu^{r-l})} \cdot \|\varphi \star_q^{q-r+l} \varphi\|_{L^2(\mu^{r-l})},$$

where both sides of the inequality might assume the value $+\infty$.

(iii) *For all* $0 \le l \le r \le p \wedge q$ *one has that*

$$\|\psi \star_r^l \varphi\|^2_{L^2(\mu^{p+q-r-l})} \le \|\psi \star_p^{p-r} \psi\|_{L^2(\mu^r)} \cdot \|\varphi \star_q^{q-r} \varphi\|_{L^2(\mu^r)},$$

where both sides of the inequality might assume the value $+\infty$.

(iv) *If* $\psi \in L^4(\mu^p)$ *and* $\varphi \in L^4(\mu^q)$, *then, for all* $0 \le l \le r \le p \wedge q$, *one has* $\psi \star_r^l \varphi \in L^2(\mu^{p+q-r-l})$ *and*

$$\|\psi \star_r^l \varphi\|_{L^2(\mu^{p+q-r-l})} \le \|\psi\|_{L^4(\mu^p)} \|\varphi\|_{L^4(\mu^q)}.$$

(v) *For all* $0 \le r \le p \wedge q$ *the function* $\psi \star_r^r \varphi$ *is in* $L^2(\mu^{p+q})$ *and*

$$\|\psi \star_r^r \varphi\|_{L^2(\mu^{p+q-2r})} \le \|\psi\|_{L^2(\mu^p)} \|\varphi\|_{L^2(\mu^q)}.$$

(vi) *If, for all* $0 \le l \le p-1$, $\psi \star_p^l \psi \in L^2(\mu^{p-l})$ *and, for all* $0 \le k \le q-1$, $\varphi \star_q^k \varphi \in L^2(\mu^{q-k})$, *then, for all* $0 \le l \le r \le p \wedge q$, *one has* $\psi \star_r^l \varphi \in L^2(\mu^{p+q-r-l})$ *and*

$$\begin{aligned} \|\psi \star_r^l \varphi\|^2_{L^2(\mu^{p+q-r-l})} &= \langle \psi \star_{p-l}^{p-r} \psi, \varphi \star_{q-l}^{q-r} \varphi \rangle_{L^2(\mu^{r+l})} \\ &\le \|\psi \star_r^l \psi\|_{L^2(\mu^{2p-r-l})} \cdot \|\varphi \star_r^l \varphi\|_{L^2(\mu^{2q-r-l})} < \infty . \end{aligned}$$

In what follows, for $p \in \mathbb{N}$ and a function $f : E^p \to \mathbb{R}$ we write $\tilde{f}$ for its **symmetrization**, i.e.

$$\tilde{f}(x_1, \dots, x_p) := \frac{1}{p!} \sum_{\sigma \in \mathbb{S}_p} f\big(x_{\sigma(1)}, \dots, x_{\sigma(p)}\big), \quad (x_1, \dots, x_p) \in E^p,$$

where $\mathbb{S}_p$ denotes the symmetric group acting on $\{1, \dots, p\}$. Note that, if $f \in L^q(\mu^p)$, then $\|\tilde{f}\|_{L^q(\mu^p)} \le \|f\|_{L^q(\mu^p)}$, by the triangle inequality.

3 Weak convergence of U-processes with symmetric kernels

3.1 Main results

For the rest of this section, we let $p \ge 1$ be an integer. Moreover, for positive integers $1 \le r, i, k \le p$ and $0 \le l \le p$ such that $0 \le l \le r \le i \wedge k$, we let $Q(i, k, r, l)$ be the set of quadruples (j, m, a, b) of nonnegative integers such that the following hold:

(1) $j \le i$ and $m \le k$.

(2) $b \le a \le r$.

(3) $b \le l$.

(4) $a - b \le r - l$.

(5) $j + m - a - b \le i + k - r - l \le i + k - 1$

(6) $a \le j \wedge m$.

(7) If $j = m = p$, then $b = l$ and $a = r \ge 1$.

The next statement is the main result of the paper.

Theorem 3.1 (Functional convergence of general U-statistics, I). *Let the assumptions and notation of Section 2 prevail, define the sequence*

$$W_n = \{W_n(t) : t \in [0,1]\}, \quad n \in \mathbb{N},$$

according to (2.7), and assume that the following three conditions (expressed by means of the notation (2.4)) are satisfied:

(a) *for all* $1 \leq k \leq p$, *the real limit*

$$b_k^2 := \lim_{n\to\infty} \frac{n^{2p-k}}{\sigma_n^2}\big(\|g_k\|^2_{L^2(\mu^k)} - (\mathbb{E}[\psi(X_1,\ldots,X_p)])^2\big) = \lim_{n\to\infty} \frac{n^{2p-k}}{\sigma_n^2}\operatorname{Var}\big(g_k(X_1,\ldots,X_k)\big)$$

exists;

(b) *for all* $1 \leq v \leq u \leq p$ *and all pairs* (l,r) *and quadruples* (j,m,a,b) *of integers such that* $1 \leq r \leq v$, $0 \leq l \leq r \wedge (u+v-r-1)$ *and* $(j,m,a,b) \in Q(v,u,r,l)$ *one has that*

$$\lim_{n\to\infty} \frac{n^{2p-\frac{u+v+r-l}{2}}}{\sigma_n^2} \|g_j \star_a^b g_m\|_{L^2(\mu^{j+m-a-b})} = 0\,;$$

(c) *there exists some* $\epsilon > 0$, *such that, for all* $1 \leq r \leq p$, *all* $0 \leq l \leq r-1$ *and all quadruples* $(j,m,a,b) \in Q(r,r,r,l)$, *the sequence*

$$\frac{n^{2p-r-\frac{r-l}{2}+\epsilon}}{\sigma_n^2} \|g_j \star_a^b g_m\|_{L^2(\mu^{j+m-a-b})}$$

is bounded.

Then, as $n \to \infty$, *one has that* $W_n \Longrightarrow Z$, *where* $Z = \{Z(t) : t \in [0,1]\}$ *is the centered Gaussian process defined in* (1.1), *for*

$$\alpha_{k,p}^2 = \frac{b_k^2}{k!(p-k)!^2}\,, \quad 1 \leq k \leq p.$$

Remark 3.2. The contractions appearing at Point (c) of Theorem 3.1 also appear at Point (b) of the same statement. In particular the requirement at Point (c) can be rephrased by saying that, for all $(j,m,a,b) \in Q(r,r,r,l)$, the sequence

$$\frac{n^{2p-r-\frac{r-l}{2}}}{\sigma_n^2} \|g_j \star_a^b g_m\|_{L^2(\mu^{j+m-a-b})}, \quad n \geq 1,$$

converges to zero as $O(1/n^\epsilon)$, for some $\epsilon > 0$. A similar remark applies to Point (b') and Point (c') of Theorem 3.4.

Remark 3.3. In the case $p = 2$, after removing redundant terms, verifying conditions (b) and (c) of Theorem 3.1 boils down to checking that the following quantities converge to zero, as $n \to \infty$:

1. $\frac{n^2}{\sigma_n^2}\|g_2\|_{L^2(\mu^2)}\,|\mathbb{E}g_2(X_1,X_2)|$
2. $\frac{n^2}{\sigma_n^2}\|g_1 \star_1^0 g_2\|_{L^2(\mu^2)}$
3. $\frac{n^{5/2}}{\sigma_n^2}\|g_1 \star_1^1 g_2\|_{L^2(\mu)}$
4. $\frac{n^{3/2}}{\sigma_n^2}\|g_2 \star_1^0 g_2\|_{L^2(\mu^3)}$
5. $\frac{n^2}{\sigma_n^2}\|g_2 \star_1^1 g_2\|_{L^2(\mu^2)}$
6. $\frac{n^{3/2}}{\sigma_n^2}\|g_2 \star_0^0 g_1\|_{L^2(\mu^3)} = \frac{n^{3/2}}{\sigma_n^2}\|g_1\|_{L^2(\mu)}\|g_2\|_{L^2(\mu^2)}$

and that the following sequences are bounded for some $\epsilon > 0$:

i) $n \mapsto \frac{n^{5/2+\epsilon}}{\sigma_n^2}\,(\mathbb{E}g_2(X_1,X_2))^2$

ii) $n \mapsto \frac{n^{5/2+\epsilon}}{\sigma_n^2}\|g_1\|_{L^2(\mu)}\,|\mathbb{E}g_2(X_1,X_2)|$

iii) $n \mapsto \frac{n^{5/2+\epsilon}}{\sigma_n^2}\|g_1\|^2_{L^4(\mu)}$

iv) $n \mapsto \frac{n^{1+\epsilon}}{\sigma_n^2}\|g_2\|^2_{L^4(\mu^2)}$.

Recall also that in this case we have that $g_2 = \psi$.

When the Hoeffding decomposition of a given U-statistic is directly provided, it is more convenient to work with the kernels $\{\psi_k\}$ defined in formula (2.3), rather than with the class $\{g_k\}$. The next statement allows one to obtain the same conclusion as in Theorem 3.1 by uniquely checking conditions related to the family $\{\psi_k\}$ defined in (2.3).

Theorem 3.4 (Functional convergence of general U-statistics, II). *The conclusion of Theorem 3.1 continues to hold if the following three conditions* (a'), (b') *and* (c') *replace conditions* (a), (b) *and* (c)*:*

(a') *for all* $1 \leq k \leq p$*, the real limit* $b_k^2 := \lim_{n\to\infty} \frac{n^{2p-k}}{\sigma_n^2}\|\psi_k\|^2_{L^2(\mu^k)}$ *exists;*

(b') *for all* $1 \leq v \leq u \leq p$ *and all pairs* (l, r) *of integers such that* $1 \leq r \leq v$, $0 \leq l \leq r \wedge (u + v - r - 1)$,

$$\lim_{n\to\infty} \frac{n^{2p-\frac{u+v+r-l}{2}}}{\sigma_n^2}\|\psi_v \star_r^l \psi_u\|_{L^2(\mu^{v+u-r-l})} = 0\,;$$

(c') *there exists some* $\epsilon > 0$*, such that, for all* $1 \leq r \leq p$ *and* $0 \leq l \leq r - 1$*, the sequence*

$$\frac{n^{2p-r-\frac{r-l}{2}+\epsilon}}{\sigma_n^2}\|\psi_r \star_r^l \psi_r\|_{L^2(\mu^{r-l})}$$

is bounded.

Remark 3.5. In case $p = 2$, verifying conditions (b') and (c') of Theorem 3.4 boils down to checking that the following quantities converge to zero as $n \to \infty$:

1. $\frac{n^2}{\sigma_n^2}\|\psi_1 \star_1^0 \psi_2\|_{L^2(\mu^2)}$
2. $\frac{n^{5/2}}{\sigma_n^2}\|\psi_1 \star_1^1 \psi_2\|_{L^2(\mu)}$
3. $\frac{n^{3/2}}{\sigma_n^2}\|\psi_2 \star_1^0 \psi_2\|_{L^2(\mu^3)}$
4. $\frac{n^2}{\sigma_n^2}\|\psi_2 \star_1^1 \psi_2\|_{L^2(\mu^2)}$

and the following sequences are bounded for some $\epsilon > 0$:

i) $n \mapsto \frac{n^{5/2+\epsilon}}{\sigma_n^2}\|\psi_1 \star_1^0 \psi_1\|_{L^2(\mu)}$

ii) $n \mapsto \frac{n^{1+\epsilon}}{\sigma_n^2}\|\psi_2 \star_2^0 \psi_2\|_{L^2(\mu^2)}$

iii) $n \mapsto \frac{n^{3/2+\epsilon}}{\sigma_n^2}\|\psi_2 \star_2^1 \psi_2\|_{L^2(\mu)}$.

Remark 3.6. (i) By inspection of pur proofs, one sees that Conditons (a) and (b) in Theorem 3.1 (resp. Conditions (a') and (b') in Theorem 3.4) are sufficient conditions for the convergence of the finite-dimensional distributions of W_n towards those of Z, whereas Conditons (c) and (c') therein imply tightness. Condition (b') in Theorem 3.4 implicitly appears in [DP19, Section 4 and Section 5], as an analytical sufficient condition ensuring that (in the notation of the present paper) $W_n(1)$ converges in distribution to a one-dimensional standard Gaussian random variable. On the other hand, Condition (b) in Theorem 3.1 is a substantial improvement of the sufficient conditions for one-dimensional asymptotic normality that can be deduced from [DP19, Theorem 5.2]. The difference between the conditions emerging from [DP19, Theorem 5.2] and those deduced in the present paper is explained by the fact that our findings use instead Lemma 5.7 below, which is a refined version of [DP19, Lemma 5.1].

(ii) It will become clear from the discussion to follow that Condition (a) in Theorem 3.1 and Condition (a') in Theorem 3.4 are equivalent for the same values of b_k^2, that is: for every $k = 1, ..., p$, the limit $\lim_{n\to\infty} \frac{n^{2p-k}}{\sigma_n^2} \operatorname{Var}\big(g_k(X_1, \ldots, X_k)\big)$ exists and is finite if and only if the same holds for $\lim_{n\to\infty} \frac{n^{2p-k}}{\sigma_n^2} \|\psi_k\|^2_{L^2(\mu^k)}$, and in this case the two limits coincide.

(iii) The proofs of Theorems 3.1 and 3.4 will show that Conditions (b) and (c) in Theorem 3.1 imply Conditons (b') and (c') in Theorem 3.4, whereas the opposite implication does not hold in general.

(iv) (*Relaxing Conditions* (a) *and* (a')) Suppose that all the assumptions of Theorem 3.4 are verified, *except* for Condition (a'). In such a situation, we can define $b_k^2(n) := \frac{n^{2p-k}}{\sigma_n^2} \|\psi_k\|^2_{L^2(\mu^k)}$, $k = 1, ..., p$, and observe that, for every k, the mapping $n \mapsto b_k^2(n)$ is bounded (by virtue of (2.8)). For every $n \geq 1$, we set moreover Z_n to be the Gaussian process obtained from (1.1), by replacing the coefficient $\alpha_{k,p}^2$ with

$$\alpha_{k,p}^2(n) := \frac{b_k^2(n)}{k!(p-k)!^2}, \quad 1 \leq k \leq p.$$

Then, a standard compactness argument combined with Theorem 3.1 yields the following conclusion: if ρ is any distance metrizing weak convergence on $D[0,1]$ (see e.g. [Dud02, Section 11.3]), one has that

$$\rho(W_n, Z_n) \longrightarrow 0, \quad n \to \infty, \tag{3.1}$$

where $\rho(W_n, Z_n)$ is shorthand for the distance between the distributions of W_n and Z_n, as random elements with values in $D[0,1]$. By virtue of Points (ii)-(iii) of the present remark, the exact same conclusion holds if one supposes that all assumptions of Theorem 3.1 are verified, except for Condition (a). In view of the content of Point (i) above, it follows that Theorem 3.1 and Theorem 3.4 contain and substantially extend the one-dimensional qualitative CLT stated in [DP19, Section 5]. One should notice that the techniques developed in [DP19] also allow one to deduce explicit rates of convergence, and that such a feature does not extend to our infinite-dimensional results.

The following corollary deals with the (simpler) situation of a degenerate kernel.

Corollary 3.7 (Degenerate kernels). *Let the assumptions and notation of Section 2 prevail, define the sequence W_n, $n \in \mathbb{N}$, according to (2.7), and suppose in addition that the kernel ψ is degenerate. Assume that, for all pairs (l, r) of integers such that $1 \leq r \leq p$ and $0 \leq l \leq r \wedge (2p - r - 1)$,*

$$\lim_{n\to\infty} n^{\frac{l-r}{2}} \frac{\|\psi \star_r^l \psi\|_{L^2(\mu^{2p-r-l})}}{\|\psi\|^2_{L^2(\mu^p)}} = 0 \tag{3.2}$$

and that, for all $0 \leq l \leq p-1$ and for some $\epsilon > 0$, the sequence

$$n \mapsto n^{\frac{l-p}{2}+\epsilon} \frac{\|\psi \star_p^l \psi\|_{L^2(\mu^{p-l})}}{\|\psi\|^2_{L^2(\mu^p)}} \tag{3.3}$$

is bounded. Then, as $n \to \infty$,

$$W_n \Longrightarrow \{B(t^p) : t \in [0,1]\},$$

where $B := \{B(t) : t \in [0,1]\}$ denotes a standard Brownian motion.

Remark 3.8. (i) Corollary 3.7 immediately implies (an L^4-version of) Donsker's theorem [Bil99, p. 90]. Indeed, for $p = 1$ and a kernel ψ not depending on n, conditions (3.2) and (3.3) are always satisfied since

$$\lim_{n\to\infty} n^{-1/2} \frac{\|\psi \star_1^0 \psi\|_{L^2(\mu)}}{\|\psi\|^2_{L^2(\mu)}} = \lim_{n\to\infty} n^{-1/2} \frac{\|\psi\|^2_{L^4(\mu)}}{\|\psi\|^2_{L^2(\mu)}} = 0$$

and, for any $\epsilon \in \big(0, \frac{1}{2}\big]$, the sequence

$$n \mapsto n^{-1/2+\epsilon} \frac{\|\psi \star_1^0 \psi\|_{L^2(\mu)}}{\|\psi\|^2_{L^2(\mu)}} = n^{-1/2+\epsilon} \frac{\|\psi\|^2_{L^4(\mu)}}{\|\psi\|^2_{L^2(\mu)}}$$

is bounded.

(ii) If $p \geq 2$ and the degenerate kernel ψ does not depend on n, then condition (3.2) is not satisfied for $l = r = 1$, since

$$\lim_{n\to\infty} n^{\frac{l-r}{2}} \frac{\|\psi \star_r^l \psi\|_{L^2(\mu^{2p-r-l})}}{\|\psi\|^2_{L^2(\mu^p)}} = \frac{\|\psi \star_1^1 \psi\|_{L^2(\mu^{2p-2})}}{\|\psi\|^2_{L^2(\mu^p)}} > 0.$$

This phenomenon is consistent with the known non-Gaussian fluctuations of degenerate U-processes of orders $p \geq 2$ having a kernel ψ independent of the sample size — see [Neu77, Hal79, DM83, MT84].

3.2 Connection to changepoint analysis

The techniques developed in the present paper can be used to characterize the weak convergence of families of processes that are more general than the ones defined in (2.7) and, in particular, to deal with limit theorems related to **changepoint analysis** (see e.g. [CH88, Fer94, GH95, CH97, Fer01, Gom04, HR14, RW19]).

In order to illustrate such a connection, we will show how to suitably adapt our results in order to generalise an influential invariance principle for order 2 symmetric U-statistics, originally proved by Csörgö and Horvath in [CH88]. As explained e.g. in [CH97, Gom04, RW19] such an invariance principle has been the starting point of a fruitful line of research, focussing on changepoint testing procedures based on generalisations of Wilcoxon-Mann-Whitney statistics. Further possible extensions of the results of this section, involving in particular antisymmetric kernels [CH88, GH95, CH97, Fer01] and rescaled U-processes [CH88, Fer94, RW19], are outside the scope of the present paper and will be investigated elsewhere.

As before, we start by fixing a sequence of i.i.d. random variables $X_1, X_2, ...$ with values in $(E, \mathcal{E})$, and with common distribution μ. We also consider a sequence of kernels $\{\psi^{(n)} : n \geq 1\}$, such that each mapping $\psi^{(n)} : E^2 \to \mathbb{R}$ is symmetric and square-integrable with respect to μ^2. For applications, $\psi^{(n)}$ is typically chosen in such a way that the quantity $\psi^{(n)}(x, y)$ is small whenever x, y are close, e.g. $\psi^{(n)}(x, y) = \|x - y\|^\beta$, $\beta > 0$ (assuming E is a normed space), but such a property has no impact on the convergence results discussed below. Here, to simplify the notation we assume from the start that each $\psi^{(n)}$ is centered, that is, $\mathbb{E}[\psi^{(n)}(X_1, X_2)] = 0$ for every n. We are interested in the family of U-processes $\{Y_n : n \geq 1\}$ given by

$$Y_n(t) := \sum_{1\leq i\leq \lfloor nt\rfloor < j\leq n} \psi^{(n)}(X_i, X_j), \quad t \in [0, 1].$$

Defining $\psi_u^{(n)}$, $u = 1, 2$, according to (2.3) and writing $\gamma_1(n) := \|\psi_1^{(n)}\|_{L^2(\mu)}$ and $\gamma_2(n) := \|\psi_2^{(n)}\|_{L^2(\mu^2)}$, one deduces immediately that, for $0 \leq s \leq t \leq 1$

$$\begin{aligned} &\mathrm{Cov}(Y_n(s), Y_n(t)) \\ &= \gamma_2(n)^2 \lfloor ns\rfloor(n - \lfloor nt\rfloor + 1) + \gamma_1(n)^2 \Big\{ \lfloor ns\rfloor(n - \lfloor nt\rfloor + 1)(n - \lfloor ns\rfloor + 1) \\ &\qquad + \lfloor ns\rfloor(n - \lfloor nt\rfloor + 1)(\lfloor nt\rfloor - \lfloor ns\rfloor + 1) + \lfloor ns\rfloor\lfloor nt\rfloor(n - \lfloor nt\rfloor + 1) \Big\}. \end{aligned} \tag{3.4}$$

We also set $\gamma_n^2 := \mathrm{Var}(Y_n(1/2))$ [1], and $\widetilde{Y}_n := Y_n/\gamma_n$. The next statement corresponds to one of the main findings in [CH88].

Theorem 3.9 (Csörgö and Horvath [CH88]). *Under the above assumptions, assume that $\psi^{(n)} \equiv \psi$ does not depend on n, and write $\gamma_1 \equiv \gamma_1(n)$, $n \geq 1$. Then, as $n \to \infty$, one has that $\gamma_n \sim \gamma_1 n^3/4$ and moreover $\widetilde{Y}_n \Longrightarrow 2A$, where $A = \{A(t) : t \in [0, 1]\}$ is defined as*

$$A(t) := (1 - t)B(t) + t(B(1) - B(t)), \quad t \in [0, 1],$$

with B a standard Brownian motion.

[1] The choice of $t = 1/2$ is arbitrary; one could set $\gamma_n^2 := \mathrm{Var}(Y_n(a))$ any number $a \in (0, 1)$ without changing the substance of the subsequent results.

The following statement (containing Theorem 3.9 as a special case) shows that, by allowing the kernels $\psi^{(n)}$ to explicitly depend on n, one can obtain a larger class of functional limit theorems. We recall that a centered Gaussian process $b = \{b(t) : t \in [0,1]\}$ is called a (standard) **Brownian bridge** if $\mathbb{E}[b(t)b(s)] = s \wedge t - st$, for all $s, t \in [0,1]$.

Theorem 3.10. *Let the above assumptions and notation prevail, and assume moreover that:* (I) *the kernels $\psi_1^{(n)}, \psi_2^{(n)}$ verify the asymptotic relations expressed in Conditions* (b') *and* (c') *of Theorem 3.4 for $p = 2$, and* (II) *as $n \to \infty$,*

$$n^{4-i}\,\frac{\gamma_n(i)^2}{\gamma_n^2} \longrightarrow c_i^2 \in [0,\infty), \quad i = 1, 2.$$

Then, $\gamma_n^2 \sim (c_1^2 + c_2^2)^{-1}(n^3\gamma_n(1)^2 + n^2\gamma_n(2)^2)$, and moreover $\widetilde{Y}_n \Longrightarrow c_1\,A + c_2\,b$, where b is a Brownian bridge independent of A.

Theorem 3.10 is proved in Section 5.5. An alternate class of FCLTs displaying Brownian bridges as limits can be found in [GH95]. In the forthcoming Corollary 4.4, we will present a direct application of Theorem 3.10 to edge counting in random geometric graphs.

3.3 Extension to vectors of U-processes

In this subsection we state multivariate extensions of Theorems 3.1 and 3.4. We first introduce the setup and some notation: Fix a dimension $d \geq 1$ and, for $1 \leq i \leq d$, let p_i be a positive integer and suppose that $\psi(i) = \psi^{(n)}(i) \in L^2(\mu^{p_i})$ is a symmetric kernel (that may again depend on the sample size n). Define the corresponding kernels $g_k(i)$ and $\psi_s(i)$ (which may also depend on n) for all $0 \leq k \leq p_i$ and $1 \leq s \leq p_i$ in the obvious way. W.l.o.g. we may assume that $1 \leq p_1 \leq p_2 \leq \ldots \leq p_d$. Moreover, for $1 \leq i \leq d$, $t \in [0,1]$ and $n \geq p_d$, let $U_i(t) := J_{p_i}^{(\lfloor nt \rfloor)}(\psi(i))$,

$$W_i(t) := W_i^{(n)}(t) := \frac{U_i(t) - \mathbb{E}[U_i(t)]}{\sqrt{\mathrm{Var}(U_i(1))}}$$

and $W_i := (W_i(t))_{t\in[0,1]}$. Then, with obvious notation we have that

$$W := W^{(n)} := (W_1, \ldots, W_d)^T \in D\big([0,1]; \mathbb{R}^d\big)\,.$$

In this section, the vector-valued Gaussian limiting processes $Z = (Z_1, \ldots, Z_d)^T$ will have zero mean, and a covariance structure that is given by

$$\mathrm{Cov}\big(Z_i(s), Z_j(t)\big) = \sum_{k=1}^{p_i \wedge p_j} \alpha_k(i,j)(s \wedge t)^k s^{p_i - k} t^{p_j - k}\,, \tag{3.5}$$

where $1 \leq i, j \leq d$, $s, t \in [0,1]$ and the $\alpha_k(i,j)$, $1 \leq k \leq p_i \wedge p_j$, are real numbers such that $\alpha_k(i,i) \geq 0$. Specializing (3.5) to the case $i = j$, $i = 1, ..., d$, one sees immediately that each process Z_i has the form (1.1), for $p = p_i$ and $\alpha_{k,p}^2 = \alpha_k(i,i)$, thus implying in particular that Z takes a.s. values in $C([0,1]; \mathbb{R}^d)$. For $1 \leq i \leq d$, we further write

$$\sigma_n^2(i) := \mathrm{Var}(U_i(1))\,.$$

The next two statements are multidimensional counterparts to Theorem 3.1 and Theorem 3.4.

Theorem 3.11 (Functional convergence of vectors of general U-statistics, I). *Let the assumptions and notation of this subsection prevail, and assume that the following three conditions are satisfied:*

(a) *For all $1 \leq i \leq j \leq d$ and for all $1 \leq k \leq p_i \wedge p_j$, the real limit*

$$b_k(i,j) := \lim_{n\to\infty} \frac{n^{p_i+p_j-k}}{\sigma_n(i)\sigma_n(j)} \Big(\langle g_k(i), g_k(j)\rangle_{L^2(\mu^k)} - \mathbb{E}\big[\psi(i)(X_1, \ldots, X_{p_i})\big] \cdot \mathbb{E}\big[\psi(j)(X_1, \ldots, X_{p_j})\big] \Big)$$

exists;

(b) *for all* $1 \leq i \leq k \leq d$, *all* $1 \leq v \leq p_i$, *all* $1 \leq u \leq p_k$ *and all pairs* (l, r) *and quadruples* (j, m, a, b) *of integers such that* $1 \leq r \leq v \wedge u$, $0 \leq l \leq r \wedge (u + v - r - 1)$ *and* $(j, m, a, b) \in Q(v, u, r, l)$ *one has that*

$$\lim_{n\to\infty} \frac{n^{p_i+p_k-\frac{u+v+r-l}{2}}}{\sigma_n(i)\sigma_n(k)} \|g_j(i) \star_a^b g_m(k)\|_{L^2(\mu^{j+m-a-b})} = 0\,;$$

(c) *there exists some* $\epsilon > 0$, *such that, for all* $1 \leq i \leq d$, *all* $1 \leq r \leq p_i$, *all* $0 \leq l \leq r-1$ *and all quadruples* $(j, m, a, b) \in Q(r, r, r, l)$, *the sequence*

$$\frac{n^{2p_i-r-\frac{r-l}{2}+\epsilon}}{\sigma_n(i)^2} \|g_j(i) \star_a^b g_m(i)\|_{L^2(\mu^{j+m-a-b})}$$

is bounded.

Then, as $n \to \infty$, *one has that* $W^{(n)} \Longrightarrow Z$, *where* $Z = \{Z(t) : t \in [0, 1]\}$ *is the centered, vector-valued Gaussian process defined by the covariance structure* (3.5) *and where, for* $1 \leq i, j \leq d$, *we have*

$$\alpha_k(i, j) = \frac{b_k(i, j)}{k!(p_i - k)!(p_j - k)!}\,, \quad 1 \leq k \leq p_i \wedge p_j.$$

Theorem 3.12 (Functional convergence of vectors of general U-statistics, II). *The conclusion of Theorem 3.11 continues to hold if the following three conditions* (a'), (b') *and* (c') *replace conditions* (a), (b) *and* (c)*:*

(a') *For all* $1 \leq i \leq j \leq d$ *and for all* $1 \leq k \leq p_i \wedge p_j$, *the real limit*

$$b_k(i, j) := \lim_{n\to\infty} \frac{n^{p_i+p_j-k}}{\sigma_n(i)\sigma_n(j)} \langle \psi_k(i), \psi_k(j) \rangle_{L^2(\mu^k)}$$

exists;

(b') *for all* $1 \leq i \leq k \leq d$, *all* $1 \leq v \leq p_i$, *all* $1 \leq u \leq p_k$ *and all pairs* (l, r) *of integers such that* $1 \leq r \leq v \wedge u$, $0 \leq l \leq r \wedge (u + v - r - 1)$,

$$\lim_{n\to\infty} \frac{n^{p_i+p_k-\frac{u+v+r-l}{2}}}{\sigma_n(i)\sigma_n(k)} \|\psi_v(i) \star_r^l \psi_u(k)\|_{L^2(\mu^{v+u-r-l})} = 0\,;$$

(c') *there exists some* $\epsilon > 0$, *such that, for all* $1 \leq i \leq d$, *all* $1 \leq r \leq p_i$ *and all* $0 \leq l \leq r-1$, *the sequence*

$$\frac{n^{2p_i-r-\frac{r-l}{2}+\epsilon}}{\sigma_n^2(i)} \|\psi_r(i) \star_r^l \psi_r(i)\|_{L^2(\mu^{r-l})}$$

is bounded.

4 Applications

4.1 Subgraph counting in random geometric graphs

Random geometric graphs are graphs whose vertices are random points scattered on some Euclidean domain, and whose edges are determined by some explicit geometric rule; in view of their wide applicability (for instance, to the modelling of telecommunication networks), these objects represent a very popular alternative to the combinatorial Erdös-Rényi random graphs. We refer to the texts [Pen04, PR16] for an introduction to this topic, and for an overview of related applications. Our aim is to use our main findings (Theorem 3.1 and Theorem 3.4) in order to establish a new collection of FCLTs for arbitrary subgraph counting statistics associated with generic sequences of random graphs. These FCLTs – whose statements appear in Theorem 4.2 below – hold in full generality and with minimal restrictions with respect to the already existing one-dimensional CLTs; as such, they substantially extend the one-dimensional CLTs proved in [Pen04, Section 3.5 and Section 3.4], as well as in [JJ86, BG92, DP19].

We fix a dimension $d \geq 1$ as well as a bounded and Lebesgue almost everywhere continuous probability density function f on $\mathbb{R}^d$. Let $\mu(dx) := f(x)dx$ be the corresponding probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and suppose that $X_1, X_2, \ldots$ are i.i.d. with distribution μ. Let $X := \{X_j : j \in \mathbb{N}\}$. We denote by $\{t_n : n \in \mathbb{N}\}$ a sequence of radii in $(0, \infty)$ such that $\lim_{n\to\infty} t_n = 0$. For each $n \in \mathbb{N}$, we denote by $G(X; t_n)$ the **random geometric graph** obtained as follows. The vertices of $G(X; t_n)$ are given by the set $V_n := \{X_1, \ldots, X_n\}$, which $\mathbb{P}$-a.s. has cardinality n, and two vertices X_i, X_j are connected if and only if $0 < \|X_i - X_j\|_2 < t_n$. Furthermore, let $p \geq 2$ be a fixed integer and suppose that Γ is a fixed connected graph on p vertices. For each n we denote by $G_n(\Gamma)$ the number of induced subgraphs of $G(X; t_n)$ which are isomorphic to Γ. Recall that an induced subgraph of $G(X; t_n)$ consists of a non-empty subset $V_n' \subseteq V_n$ with an edge set precisely given by the set of edges of $G(X; t_n)$ whose endpoints are both in V_n'. We will also have to assume that Γ is **feasible** for every $n \geq p$. This means that the probability that the restriction of $G(X; t_n)$ to $X_1, \ldots, X_p$ is isomorphic to Γ is strictly positive for $n \geq p$. Note that feasibility depends on the common distribution μ of the points. The quantity $G_n(\Gamma)$ is a symmetric U-statistic of $X_1, \ldots, X_n$, since

$$G_n(\Gamma) = \sum_{1 \leq i_1 < \ldots < i_p \leq n} \psi_{\Gamma, t_n}(X_{i_1}, \ldots, X_{i_p}),$$

where $\psi_{\Gamma,t_n}(x_1, \ldots, x_p)$ equals 1 if the graph with vertices $x_1, \ldots, x_p$ and edge set $\{\{x_i, x_j\} : 0 < \|x_i - x_j\|_2 < t_n\}$ is isomorphic to Γ, and equals 0 otherwise. We denote the corresponding normalized U-process by $\{W_n(t) : t \in [0, 1]\}$, i.e.

$$W_n(t) = \frac{U_n(t) - \mathbb{E}[U_n(t)]}{\mathrm{Var}\big(G_n(\Gamma)\big)^{1/2}},$$

where

$$U_n(s) := G_{\lfloor nt \rfloor}(\Gamma) = \sum_{1 \leq i_1 < \ldots < i_p \leq \lfloor nt \rfloor} \psi_{\Gamma, t_n}(X_{i_1}, \ldots, X_{i_p}), \quad 0 \leq t \leq 1.$$

For obtaining asymptotic normality one typically distinguishes between three different asymptotic regimes:

(R1) $nt_n^d \to 0$ and $n^p t_n^{d(p-1)} \to \infty$ as $n \to \infty$ (*sparse regime*)

(R2) $nt_n^d \to \infty$ as $n \to \infty$ (*dense regime*)

(R3) $nt_n^d \to \rho \in (0, \infty)$ as $n \to \infty$ (*thermodynamic regime*)

Note that we could rephrase the regimes **(R1)** and **(R2)** as follows:

(R1) $\left(\frac{1}{n}\right)^{\frac{p}{p-1}} \ll t_n^d \ll \frac{1}{n}$

(R2) $\frac{1}{n} \ll t_n^d$,

where, for positive sequence a_n and b_n we write $a_n \ll b_n$, $n \in \mathbb{N}$, if and only if $\lim_{n\to\infty} a_n/b_n = 0$. It turns out that, under regime **(R2)** one also has to take into account whether the common distribution μ of the X_j is the uniform distribution $\mathcal{U}(M)$ on some Borel subset $M \subseteq \mathbb{R}^d$, $0 < \lambda^d(M) < \infty$ with density $f(x) = \lambda^d(M)^{-1} \mathbf{1}_M(x)$, or not. To deal with this peculiarity, we will therefore distinguish between the following four cases:

(C1) $nt_n^d \to 0$ and $n^p t_n^{d(p-1)} \to \infty$ as $n \to \infty$.

(C2) $nt_n^d \to \infty$ as $n \to \infty$ and $\mu = \mathcal{U}(M)$ for some Borel subset $M \subseteq \mathbb{R}^d$ s.t. $0 < \lambda^d(M) < \infty$.

(C3) $nt_n^d \to \infty$ as $n \to \infty$, and μ is not a uniform distribution.

(C4) $nt_n^d \to \rho \in (0, \infty)$ as $n \to \infty$.

The following important variance estimates will be needed. Except for the case **(C2)**, which needs a special consideration, these have already been derived in the book [Pen04]. Since it does not make the argument much longer, we provide the whole proof.

Proposition 4.1. *Under all regimes* **(R1)**, **(R2)** *and* **(R3)** *it holds that* $\mathbb{E}[G_n(\Gamma)] \sim c n^p t_n^{d(p-1)}$ *for a constant* $c \in (0,\infty)$. *Moreover, there exist constants* $c_1, c_2, c_3, c_4 \in (0,\infty)$ *such that, as* $n \to \infty$,

(C1) $\mathrm{Var}(G_n(\Gamma)) \sim c_1 \cdot n^p t_n^{d(p-1)}$,

(C2) $\mathrm{Var}(G_n(\Gamma)) \geq c_2 \cdot n^{2p-2} t_n^{d(2p-3)}$ *for all* $n \in \mathbb{N}$,

(C3) $\mathrm{Var}(G_n(\Gamma)) \sim c_3 \cdot n^{2p-1} t_n^{d(2p-2)}$,

(C4) $\mathrm{Var}(G_n(\Gamma)) \sim c_4 \cdot n$.

Proof. For notational convenience, for $k = 0, 1, \dots, p$ we simply denote by g_k the function

$$(x_1, \dots, x_k) \mapsto \mathbb{E}\big[\psi_{\Gamma,t_n}(x_1, \dots, x_k, X_1, \dots, X_{p-k})\big],$$

corresponding to the kernel ψ_{Γ,t_n}, i.e. we suppress the dependence on n and on the graph Γ. We will make use of formula (2.9). First note that, for $1 \leq k \leq p$,

$$\mathrm{Var}\big(g_k(X_1, \dots, X_k)\big) = \|g_k\|^2_{L^2(\mu^k)} - \big(\mathbb{E}[\psi(X_1, \dots, X_p)]\big)^2 .$$

Hence, by (2.9), we have

$$\begin{aligned}\mathrm{Var}(G_n(\Gamma)) &= \binom{n}{p} \sum_{k=1}^{p} \binom{p}{k} \binom{n-p}{p-k} \Big(\|g_k\|^2_{L^2(\mu^k)} - \big(\mathbb{E}[\psi(X_1, \dots, X_p)]\big)^2 \Big) \\ &\sim \sum_{k=1}^{p} \frac{n^{2p-k}}{k!((p-k)!)^2} \Big(\|g_k\|^2_{L^2(\mu^k)} - \big(\mathbb{E}[\psi(X_1, \dots, X_p)]\big)^2 \Big) . \end{aligned} \tag{4.1}$$

For $k = 1, \dots, p$, we have that (e.g. by dominated convergence)

$$\begin{aligned}\|g_k\|^2_{L^2(\mu^k)} &= \int_{(\mathbb{R}^d)^k} g_k(x_1, \dots, x_k)^2 d\mu^k(x_1, \dots, x_k) \\ &= \int_{(\mathbb{R}^d)^k} \Big(\int_{(\mathbb{R}^d)^{p-k}} \psi_{\Gamma,t_n}(x_1, \dots, x_k, y_1, \dots, y_{p-k}) d\mu^{p-k}(y_1, \dots, y_{p-k}) \Big)^2 d\mu^k(x_1, \dots, x_k) \\ &= \int_{(\mathbb{R}^d)^k} \int_{(\mathbb{R}^d)^{p-k} \times (\mathbb{R}^d)^{p-k}} \psi_{\Gamma,t_n}(x_1, \dots, x_k, y_1, \dots, y_{p-k}) \psi_{\Gamma,t_n}(x_1, \dots, x_k, u_1, \dots, u_{p-k}) \\ &\qquad \prod_{j=1}^{k} f(x_j) dx_j \prod_{i=1}^{p-k} f(y_j) dy_j \prod_{l=1}^{p-k} f(u_l) du_l \\ &= \int_{\mathbb{R}^d} f(x_1) \int_{(\mathbb{R}^d)^{k-1}} \int_{(\mathbb{R}^d)^{p-k} \times (\mathbb{R}^d)^{p-k}} \psi_{\Gamma,1}(0, w_2 \dots, w_k, z_1, \dots, z_{p-k}) \psi_{\Gamma,1}(0, w_2, \dots, w_k, v_1, \dots, v_{p-k}) \\ &\qquad (t_n^d)^{2p-k-1} \prod_{j=2}^{k} f(x_1 + t_n w_j) dw_j \prod_{i=1}^{p-k} f(x_1 + t_n z_i) dz_i \prod_{l=1}^{p-k} f(x_1 + t_n u_l) du_l \, dx_1 \\ &\sim d_k (t_n^d)^{2p-k-1} , \end{aligned} \tag{4.2}$$

where, for $1 \leq k \leq p$,

$$\begin{aligned} d_k := \int_{\mathbb{R}^d} f(x_1)^{2p-k} dx_1 \int_{(\mathbb{R}^d)^{k-1}} \int_{(\mathbb{R}^d)^{p-k} \times (\mathbb{R}^d)^{p-k}} \psi_{\Gamma,1}(0, w_2 \dots, w_k, z_1, \dots, z_{p-k}) \times \\ \times \psi_{\Gamma,1}(0, w_2, \dots, w_k, v_1, \dots, v_{p-k}) \prod_{j=2}^{k} dw_j \prod_{i=1}^{p-k} dz_i \prod_{l=1}^{p-k} du_l . \end{aligned} \tag{4.3}$$

Also, from [Pen04, Proposition 3.1] we know that

$$\mathbb{E}\big[G_n(\Gamma)\big] \sim \frac{n^p}{p!} t_n^{d(p-1)} \nu ,$$

where

$$\nu = \nu(p,\Gamma) := \int_{\mathbb{R}^d} f(x)^p dx \int_{(\mathbb{R}^d)^{p-1}} \psi_{\Gamma,1}(0, y_2, \dots, y_p) dy_2 \dots dy_p > 0\,. \tag{4.4}$$

This implies that

$$\big(\mathbb{E}[\psi(X_1,\dots,X_p)]\big)^2 \sim \nu^2 (t_n^d)^{2p-2}\,. \tag{4.5}$$

Since $0 < t_n < 1$, for $2 \le k \le p$, this yields that

$$\begin{aligned}\mathrm{Var}\big(g_k(X_1,\dots,X_k)\big) &= \|g_k\|^2_{L^2(\mu^k)} - \big(\mathbb{E}[\psi(X_1,\dots,X_p)]\big)^2 \\ &\sim d_k (t_n^d)^{2p-k-1}\,.\end{aligned}$$

In order to discuss the case $k = 1$ we have to carefully compare d_1 to ν. Note that we have

$$d_1 = \int_{\mathbb{R}^d} f(x)^{2p-1} dx \bigg(\int_{(\mathbb{R}^d)^{p-1}} \psi_{\Gamma,1}(0, y_2, \dots, y_p) dy_2 \dots dy_p \bigg)^2$$

so that $d_1 > \nu^2$ if and only if

$$\int_{\mathbb{R}^d} f(x)^{2p-1} dx > \bigg(\int_{\mathbb{R}^d} f(x)^p dx \bigg)^2 .$$

By Jensen's inequality we have

$$\begin{aligned}\bigg(\int_{\mathbb{R}^d} f(x)^p dx \bigg)^2 &= \bigg(\int_{\mathbb{R}^d} f(x)^{p-1} d\mu(x) \bigg)^2 \le \int_{\mathbb{R}^d} f(x)^{2p-2} d\mu(x) \\ &= \int_{\mathbb{R}^d} f(x)^{2p-1} dx\end{aligned}$$

with equality, if and only if, $f(x)^{p-1}$ is μ-a.s. constant, i.e., if and only if μ is a uniform distribution on some Borel subset $M \subseteq \mathbb{R}^d$ s.t. $0 < \lambda^d(M) < \infty$. Thus, if μ is not a uniform distribution we obtain that

$$\mathrm{Var}\big(g_1(X_1)\big) \sim (d_1 - \nu^2)(t_n^d)^{2p-2}\,,$$

whereas, if μ is a uniform distribution we can only conclude that

$$\mathrm{Var}\big(g_1(X_1)\big) \lesssim (t_n^d)^{2p-2}$$

but, in general, we cannot give any lower asymptotic bound on $\mathrm{Var}\big(g_1(X_1)\big)$. Note that, for $1 \le k \le p-1$ we have

$$\frac{n^{2p-k}(t_n^d)^{2p-k-1}}{n^{2p-k-1}(t_n^d)^{2p-k-2}} = n t_n^d \overset{n\to\infty}{\longrightarrow} \begin{cases} 0\,, & \text{under } \mathbf{(R1)} \\ \infty\,, & \text{under } \mathbf{(R2)} \\ \rho\,, & \text{under } \mathbf{(R3)}\,. \end{cases} \tag{4.6}$$

This implies that there are positive constants c_1, c_3 and c_4 such that

$$\mathrm{Var}(G_n(\Gamma)) \sim \begin{cases} c_1 n^p (t_n^d)^{p-1}\,, & \text{in case } \mathbf{(C1)} \\ c_3 n^{2p-1}(t_n^d)^{2p-2}\,, & \text{in case } \mathbf{(C3)} \\ c_4 n\,, & \text{in case } \mathbf{(C4)}\,, \end{cases} \tag{4.7}$$

whereas, in case **(C2)** we can conclude (as claimed) that there is a positive constant c_2 such that

$$\mathrm{Var}(G_n(\Gamma)) \ge \binom{p}{2}\binom{n-p}{p-2}\Big(\|g_2\|^2_{L^2(\mu^2)} - \big(\mathbb{E}[\psi(X_1,\dots,X_p)]\big)^2 \Big) \sim c_2 n^{2p-2}(t_n^d)^{2p-3}\,. \tag{4.8}$$

□

The next collection of FCLTs, extending the one-dimensional CLT proved in [Pen04, DP19], is the main result of the section. Note that, in view of the large number of parameters, in the forthcoming Theorem 4.2 we choose to express the distribution of the limit process Z directly in terms of its covariance function (1.3), rather than using the representation (1.1). We will also need the following definition: For fixed $\rho \in (0,\infty)$, introduce the positive definite function $\Psi : [0,1] \times [0,1] \to \mathbb{R} : (s,t) \mapsto \Psi(s,t)$ given by

$$\Psi(s,t) := \left(\sum_{l=1}^{p} \frac{\rho^{2p-l-1}(d_l - \delta_{l,1}\nu^2)}{l!(p-l)!(p-l)!} \right)^{-1} \sum_{k=1}^{p} \frac{(s\wedge t)^p (s\vee t)^{p-k}}{k!(p-k)!(p-k)!} \rho^{2p-k-1}(d_k - \delta_{k,1}\nu^2) \tag{4.9}$$

where d_k, $1 \le k \le p$, and ν have been defined in (4.3) and (4.4), respectively.

Theorem 4.2 (FCLTs for subgraph counting in random geometric graphs). *Let the above assumptions and notation prevail. Then, in the cases* **(C3)** *and* **(C4)**, *the sequence* $W_n = \{W_n(t)) : t \in [0,1]\}$, $n \in \mathbb{N}$ *is such that* $W_n \Longrightarrow Z$, $n \to \infty$, *for some continuous Gaussian process* Z. *In the case* **(C1)**, *this convergence holds if, additionally, there is a* $\delta > 0$ *such that* $\left(\frac{1}{n}\right)^{\frac{p}{p-1}-\delta} \ll t_n^d \ll \frac{1}{n}$ *and in the case* **(C2)**, *this convergence holds if, additionally, there is a* $\delta > 0$ *such that* $\frac{1}{n} \ll t_n^d \ll n^{-1/2-\delta}$ *and if the limits* b_1^2 *and* b_2^2 *do exist, where the parameters* b_1^2 *and* b_2^2 *have been defined in Theorem 3.1). The covariance function* $\Gamma : [0,1] \times [0,1] \to \mathbb{R}$ *of* Z *is given by*

$$\Gamma(s,t) = \begin{cases} (s\wedge t)^p\,, & \text{under } \mathbf{(C1)} \\ \frac{(s\wedge t)^p (s\vee t)^{p-1}}{1+\lambda^{-1}} + \frac{(s\wedge t)^p (s\vee t)^{p-2}}{1+\lambda}\,, & \text{under } \mathbf{(C2)} \\ (s\wedge t)^p (s\vee t)^{p-1}\,, & \text{under } \mathbf{(C3)} \\ \Psi(s,t)\,, & \text{under } \mathbf{(C4)} \end{cases},$$

where $\Psi(s,t)$ *is defined in* (4.9) *and* $\lambda \in [0,+\infty]$ *is given in* (4.12) *below. In particular, in the case* **(C1)** *one has that* $W_n \Longrightarrow \{B(t^p) : t \in [0,1]\}$, *where* B *denotes a standard Brownian motion.*

Remark 4.3. Note that, interestingly, in the case **(C4)** the covariance function Ψ of the limiting process depends not only on ρ but also on the difference $d_1 - \nu^2$. In particular, the analytic properties of Ψ depend on whether μ is a uniform distribution or not.

Proof of Theorem 4.2. Let us first disregard the case **(C2)**. Then, by Proposition 4.1, (4.2) and (4.5) we have

$$\frac{n^{2p-k}}{\sigma_n^2}\left(\|g_k\|^2_{L^2(\mu^k)} - (\mathbb{E}[\psi(X_1,\dots,X_p)])^2\right) \sim \frac{n^{2p-k}\left(d_k (t_n^d)^{2p-k-1} - \nu^2 (t_n^d)^{2p-2}\right)}{\sum_{l=1}^{p} \frac{n^{2p-l}}{l!(p-l)!(p-l)!}\left(d_l (t_n^d)^{2p-l-1} - \nu^2 (t_n^d)^{2p-2}\right)}$$

By relation (4.6) this implies that

$$b_k^2 = \lim_{n\to\infty} \frac{n^{2p-k}}{\sigma_n^2}\left(\|g_k\|^2_{L^2(\mu^k)} - (\mathbb{E}[\psi(X_1,\dots,X_p)])^2\right) = \begin{cases} p!\delta_{k,p}\,, & \text{in case } \mathbf{(C1)} \\ (p-1)!(p-1)!\delta_{k,1}\,, & \text{in case } \mathbf{(C3)} \\ \frac{\rho^{2p-k-1}(d_k - \delta_{k,1}\nu^2)}{\sum_{l=1}^{p} \frac{\rho^{2p-l-1}(d_l-\delta_{l,1}\nu^2)}{l!(p-l)!(p-l)!}}\,, & \text{in case } \mathbf{(C4)} \end{cases} \tag{4.10}$$

In the case **(C2)**, in general we only know that

$$\sigma_n^2 \sim \frac{n^{2p-1}}{(p-1)!(p-1)!}\operatorname{Var}(g_1(X_1)) + d_2 \frac{n^{2p-2}(t_n^d)^{2p-3}}{2(p-2)!(p-2)!}\,,$$

where

$$\operatorname{Var}(g_1(X_1)) \lesssim (t_n^d)^{2p-2}\,.$$

Thus, using (4.6) we can conclude that

$$b_k^2 = \lim_{n\to\infty} \frac{n^{2p-k}}{\sigma_n^2}\left(\|g_k\|^2_{L^2(\mu^k)} - (\mathbb{E}[\psi(X_1,\dots,X_p)])^2\right) = 0\,, \quad 3 \le k \le p\,. \tag{4.11}$$

Hence, if both limits b_1^2 and b_2^2 do exist, depending on the precise order of $\operatorname{Var}(g_1(X_1))$, the three scenarios

(a) $b_1^2 = (p-1)!(p-1)!$ and $b_2^2 = 0$

(b) $b_1^2 = 0$ and $b_2^2 = 2(p-2)!(p-2)!$ and

(c) $b_1^2 = (p-1)!(p-1)!\big(1+\lambda^{-1}\big)^{-1}$ and $b_2^2 = 2(p-2)!(p-2)!\big(\lambda+1\big)^{-1}$ for some $\lambda \in (0,\infty)$

are possible, where

$$\lambda = \lim_{n\to\infty} \frac{n^{2p-1}}{(p-1)!(p-1)!}\operatorname{Var}(g_1(X_1)) \cdot \left(d_2 \frac{n^{2p-2}(t_n^d)^{2p-3}}{2(p-2)!(p-2)!}\right)^{-1}. \tag{4.12}$$

Note that (c) contains (a) and (b) if we allow for $\lambda = 0$ and $\lambda = +\infty$ by adopting the conventions that $a/0 := \infty$ and $a/\infty := 0$ for $a \in (0,\infty)$.

In particular, (4.10) and (4.11) make sure that condition (a) of Theorem 3.1 is always satisfied in this example and that the covariance function Γ of the potential limiting process given by

$$\Gamma(s,t) = \sum_{k=1}^{p} \frac{(s\wedge t)^p (s\vee t)^{p-k}}{k!(p-k)!(p-k)!} b_k^2$$

coincides with the one in the statement. Now fix integers $1 \le v \le u \le p$ and l, r such that $1 \le r \le v$ and $0 \le l \le r \wedge (u+v-r-1)$. The computations on pages 4196-4197 of [LRP13] show that for all

$$\begin{aligned}(j,m,a,b) \in P := \Big(\{(j,m,a,b) : 1 \le b \le a \le j \le m \text{ and } b < m\} \cup \{(j,m,a,b) : j = m = a \text{ and } b = 0\}\Big) \\ \cap\, Q(v,u,r,l)\end{aligned}$$

one has

$$\begin{aligned}\|g_j \star_a^b g_m\|^2_{L^2(\mu^{j+m-a-b})} &= O\big(t_n^{d(4p-(j+m+a-b)-1)}\big) \\ &= O\big(t_n^{d(4p-(u+v+r-l)-1)}\big),\end{aligned} \tag{4.13}$$

where the second relation follows from $0 < t_n < 1$ and the inequality $j+m+a-b \le v+u+r-l$ (we observe that the authors of [LRP13] actually deal with the rescaled measure $n\cdot\mu$, which is why they obtain another power of n as a prefactor). Now suppose that $(j,m,a,b) \in Q(v,u,r,l) \cap P$. We are going to repeatedly use (4.13) and Proposition 4.1 for the following estimates: In case **(C1)** we have

$$\begin{aligned}\frac{n^{4p+l-r-v-u}}{\sigma_n^4}\|g_j \star_a^b g_m\|^2_{L^2(\mu^{v+u-i-k})} &\lesssim \frac{n^{4p+l-r-v-u} t_n^{d(4p-u-v-r+l-1)}}{n^{2p} t_n^{d(2p-2)}} \\ &= n^{2p-(v+u+r-l)} t_n^{d(2p-(j+m+a-b)+1)} \\ &\le (nt_n^d)^{2p-(v+u+r-l)} t_n^d \\ &\le \Big(n^p t_n^{d(p-1)}\Big)^{-1},\end{aligned}$$

where we have used that $v+u+r-l \le 3p$ for the second inequality and, hence, under the assumptions of the theorem it follows that

$$\frac{n^{4p+l-r-v-u}}{\sigma_n^4}\|g_j \star_a^b g_m\|^2_{L^2(\mu^{v+u-i-k})} \lesssim n^{-\delta(p-1)}.$$

In case **(C2)** we obtain

$$\begin{aligned}\frac{n^{4p+l-r-v-u}}{\sigma_n^4}\|g_j \star_a^b g_m\|^2_{L^2(\mu^{v+u-i-k})} &\lesssim \frac{n^{4p+l-r-v-u} t_n^{d(4p-u-v-r+l-1)}}{n^{4p-4} t_n^{d(4p-6)}} \\ &= n^{4-(v+u+r-l)} t_n^{d(5-(j+m+a-b))} \\ &\le n^{4-(v+u+r-l)} t_n^{d(5-(v+u+r-l))} \\ &\le n t_n^{2d},\end{aligned}$$

where we have used that $(v+u+r-l) \geq 3$. Hence, under the assumptions of the theorem it follows that

$$\frac{n^{4p+l-r-v-u}}{\sigma_n^4} \, \|g_j \star_a^b g_m\|^2_{L^2(\mu^{v+u-i-k})} \lesssim n^{-2\delta} \, .$$

In case **(C3)** we similarly obtain

$$\begin{aligned}
\frac{n^{4p+l-r-v-u}}{\sigma_n^4} \, \|g_j \star_a^b g_m\|^2_{L^2(\mu^{v+u-i-k})} &\lesssim \frac{n^{4p+l-r-v-u} t_n^{d(4p-u-v-r+l-1)}}{n^{4p-2} t_n^{d(4p-4)}} \\
&= n^{2-(v+u+r-l)} t_n^{d(3-(j+m+a-b))} \\
&\leq n^{2-(v+u+r-l)} t_n^{d(3-(v+u+r-l))} \\
&\leq n^{-1} \, ,
\end{aligned}$$

where we have again used that $(v+u+r-l) \geq 3$. Finally, in case **(C4)** we have

$$\begin{aligned}
\frac{n^{4p+l-r-v-u}}{\sigma_n^4} \, \|g_j \star_a^b g_m\|^2_{L^2(\mu^{v+u-i-k})} &\lesssim \frac{n^{4p+l-r-v-u} t_n^{d(4p-u-v-r+l-1)}}{n} \\
&\leq \big(n t_n^d\big)^{4p+l-r-v-u} t_n^d \sim \rho^{4p+l-r-v-u} t_n^d \\
&= O(n^{-1}) \, .
\end{aligned}$$

Eventually, we have to deal with the quadruples $(j,m,a,b) \in Q(v,u,r,l) \setminus P$. In order to do this, we first remark that we have the asymptotic relations

$$\|g_m\|^2_{L^2(\mu^m)} \lesssim (t_n^d)^{2p-m-1} \, , \quad 1 \leq m \leq p \text{ and} \tag{4.14}$$

$$\mu^p(\psi_{\Gamma,t_n}) := \int_{(\mathbb{R}^d)^p} \psi_{\Gamma,t_n} d\mu^p \lesssim (t_n^d)^{p-1} \, . \tag{4.15}$$

Relation (4.14) follows from the computation

$$\begin{aligned}
\|g_m\|^2_{L^2(\mu^m)} &= \int_{(\mathbb{R}^d)^m} g_m^2(x_1,\dots,x_m) \prod_{j=1}^m f(x_j) dx_j \\
&= \int_{(\mathbb{R}^d)^m} \prod_{j=1}^m f(x_j) dx_j \int_{(\mathbb{R}^d)^{2p-2m}} \psi_{\Gamma,t_n}(x_1,\dots,x_p) \\
&\qquad \psi_{\Gamma,t_n}(x_1,\dots,x_m,z_{m+1},\dots,z_p) \prod_{l=m+1}^p f(x_l) f(z_l) dx_l dz_l \\
&= \int_{(\mathbb{R}^d)^m} \prod_{j=1}^m f(x_j) dx_j \int_{(\mathbb{R}^d)^{2p-2m}} \psi_{\Gamma,1}\big(0, t_n^{-1}(x_1-x_2)\dots, t_n^{-1}(x_1-x_p)\big) \\
&\qquad \psi_{\Gamma,1}\big(0, t_n^{-1}(x_1-x_2),\dots,t_n^{-1}(x_1-x_m), t_n^{-1}(x_1-z_{m+1}),\dots,t_n^{-1}(x_1-z_p)\big) \\
&\qquad \prod_{l=m+1}^p f(x_l) f(z_l) dx_l dz_l \\
&= (t_n^d)^{2p-m-1} \int_{\mathbb{R}^d} f(x_1) dx_1 \int_{(\mathbb{R}^d)^{m-1}} \prod_{j=2}^m f(x_1+t_n y_j) dy_j \int_{\mathbb{R}^{2p-2m}} \prod_{l=m+1}^p f(x_1+t_n u_l) f(x_1+t_n v_l) du_l dv_l \\
&\qquad \psi_{\Gamma,1}(0,y_2,\dots,y_m,u_{m+1},\dots,u_p) \psi_{\Gamma,1}(0,y_2,\dots,y_m,v_{m+1},\dots,v_p) \\
&\sim (t_n^d)^{2p-m-1} \int_{\mathbb{R}^d} f(x_1)^{2p-m} dx_1 \int_{(\mathbb{R}^d)^{m-1}} \prod_{j=2}^m dy_j \int_{\mathbb{R}^{2p-2m}} \prod_{l=m+1}^p du_l dv_l \\
&\qquad \psi_{\Gamma,1}(0,y_2,\dots,y_m,u_{m+1},\dots,u_p) \psi_{\Gamma,1}(0,y_2,\dots,y_m,v_{m+1},\dots,v_p) \\
&\lesssim (t_n^d)^{2p-m-1} \, ,
\end{aligned}$$

where made use of the translation invariance and scaling property of the kernel ψ_{Γ,t_n} as well as of the a.e.-continuity of f. The derivation of (4.15) is similar but easier and is for this reason omitted.

First, if $a = b = 0$ and $j, m \geq 1$, then we have

$$\begin{aligned}\|g_j \star_0^0 g_m\|^2_{L^2(\mu^{j+m})} &= \|g_j\|^2_{L^2(\mu^j)} \cdot \|g_m\|^2_{L^2(\mu^m)} \\ &\lesssim (t_n^d)^{2p-j-1}(t_n^d)^{2p-m-1} = (t_n^d)^{4p-(j+m)-2}\,.\end{aligned}$$

Now note that by the definition of the set $Q(v,u,r,l)$ we further have that

$$j + m = j + m - a - b \leq u + v - r - l$$

which implies that

$$\begin{aligned}\|g_j \star_0^0 g_m\|_{L^2(\mu^{j+m})} &\lesssim (t_n^d)^{4p-(u+v-r-l)-2} = (t_n^d)^{4p-(u+v+r-l)+2r-2} \\ &\leq (t_n^d)^{4p-(u+v+r-l)}\,,\end{aligned}$$

since $r \geq 1$. If $a = b = j = m = 0$, then we have

$$\begin{aligned}\|g_0 \star_0^0 g_0\|^2_{L^2(\mu^0)} &= \mu^p(\psi_{\Gamma,t_n})^4 \lesssim (t_n^d)^{4p-4} \\ &\leq (t_n^d)^{4p-(u+v+r-l)-1}\,,\end{aligned}$$

which provides a bound of the same order as (4.13). If $a = b = j = 0$ and $m \geq 1$, then using $m = j + m - a - b \leq u + v - r - l$ and $r \geq 1$,

$$\begin{aligned}\|g_0 \star_0^0 g_m\|^2_{L^2(\mu^m)} &= \mu^p(\psi_{\Gamma,t_n})^2\|g_m\|^2_{L^2(\mu^m)} \\ &\lesssim (t_n^d)^{2p-2}(t_n^d)^{2p-m-1} = (t_n^d)^{4p-m-3} \\ &\leq (t_n^d)^{4p-(u+v-r-l)-3} = (t_n^d)^{4p-(u+v+r-l)-3+2r} \\ &\leq (t_n^d)^{4p-(u+v+r-l)-1}\,,\end{aligned}$$

which again yields a bound of the same order as (4.13). The only remaining possibility is that $1 \leq a = b = m = j \leq p-1$. In this case, we first claim that

$$2j + 1 \leq u + v + r - l\,.$$

Indeed, if $j < u$, then $2j < u+v \leq u+v+r-l$ since $j \leq v$. On the other hand, if $j = u$, then $j = v$ and we must also have $r = j$ and $l \leq r-1 = j-1$ since $j = a \leq r \leq v = j$ and $0 \leq l \leq u+v-r-1 = j-1 = r-1$. Hence, $u + v + r - l \geq 2j + r - l \geq 2j + 1$. Thus, we obtain that

$$\begin{aligned}\|g_j \star_j^j g_j\|^2_{L^2(\mu^0)} &= \|g_j\|^4_{L^2(\mu^j)} \\ &\lesssim (t_n^d)^{2p-j-1}(t_n^d)^{2p-j-1} = (t_n^d)^{4p-(2j+1)-1} \\ &\leq (t_n^d)^{4p-(u+v+r-l)-1}\end{aligned}$$

which is the same bound as in (4.13). Since all these bounds are at most the same as the bound in (4.13) we conclude that the above estimates continue to hold for all $(j,m,a,b) \in Q(v,u,r,l) \setminus P$. Since the estimates just proven are independent of the variables v, u, c and r this implies that conditions (b) and (c) of Theorem 3.1 are satisfied in the asserted cases. □

As announced, the following statement is a consequence of Theorem 3.10 and provides a changepoint counterpart to the previous theorem, in the special case of edge counting.

Corollary 4.4. *Under the above assumptions and notation, suppose that the sequence $\{t_n\}$ verifies condition* **(C1)** *for $p = 2$, and write $\sigma_n^2 := \mathrm{Var}(G_n(\mathbf{edge}))$. Then, if there is a $\delta > 0$ such that $\left(\frac{1}{n}\right)^{2-\delta} \ll t_n^d \ll \frac{1}{n}$, the process $T_n := \{T_n(t) : t \in [0,1]\}$ defined by*

$$\begin{aligned}T_n(t) &:= \frac{1}{\sigma_n}\sum_{1\leq i\leq \lfloor nt\rfloor < j\leq n}\left(\mathbf{1}_{\{0<\|X_i-X_j\|<t_n\}} - \eta_n\right), \\ &= \frac{1}{\sigma_n}\sum_{1\leq i\leq \lfloor nt\rfloor < j\leq n}\mathbf{1}_{\{0<\|X_i-X_j\|<t_n\}} - \frac{\eta_n}{\sigma_n}\lfloor nt\rfloor(n - \lfloor nt\rfloor + 1),\end{aligned}$$

where $\eta_n := \mathbb{P}\left[0 < \|X_1 - X_2\| \leq t_n\right]$, is such that $T_n \Longrightarrow \sqrt{2}\,b$, where b is a standard Brownian bridge.

The proof of Corollary 4.4 (whose details are left to the reader) follows from the fact that, under the regime **(C1)** and in the notation of Theorem 3.10, one has that $\gamma_n^2 \sim \sigma_n^2/2$, and also $c_1 = 0$ and $c_2 = 2$. Writing $k := \lfloor nt \rfloor$ for a fixed t, the sum

$$S(n,t) := \sum_{1 \leq i \leq \lfloor nt \rfloor < j \leq n} \mathbf{1}_{\{0 < \|X_i - X_j\| \leq t_n\}}$$

counts the number of edges in $G(X; t_n)$ such that one endpoint belongs to the set $\{X_1, ..., X_k\}$ and the other belongs to $\{X_{k+1}, ..., X_n\}$; a small value of $S(n,t)$ implies that most distances between the elements of the two blocks of variables are larger than t_n. For testing procedures related to changepoint analysis (see [Fer01, CH97]), one is typically interested in understanding the asymptotic distribution of such quantities as

$$M_n := \max_{t \in [0,1]} (-T_n(t)), \quad \text{or} \quad A_n := \operatorname*{argmax}_{t \in [0,1]} (-T_n(t)),$$

where $\operatorname{argmax}_{t \in [0,1]} g(t)$ stands conventionally for the smallest maximizer of a function g admitting a maximum[2]. Corollary 4.4 immediately implies that M_n and A_n converge in distribution to $m := \sqrt{2} \max b_t$ and $a := \operatorname{argmax} b_t$, respectively. It is a well-known fact (see e.g. [Fer95, CH97]) that $m/\sqrt{2}$ is distributed according to the Kolmogorov-Smirnov law, whereas a is uniformly distributed on $[0,1]$. More general limit theorems (involving in particular an independent process A, as in Theorem 3.10) could be obtained by considering an adequately renormalized version of T_n under the remaining regimes **(C2)**—**(C4)**.

4.2 U-statistics of order 2 with a dominant diagonal component

4.2.1 General statements

In the paper [RLTvdV16], a remarkable collection of one-dimensional CLTs was proved for sequences of U-statistics of order 2 displaying size-dependent kernels, as well as dominant non-linear Hoeffding components. The Gaussian fluctuations of the U-statistics considered in [RLTvdV16] emerge asymptotically from the fact that the corresponding kernels tend to concentrate around a diagonal, a phenomenon leading to Gaussianity if one assumes some additional Lyapounov-type condition. The scope of the applications developed in [RLTvdV16] covers e.g. the estimation of quadratic functionals of densities and regression functions, as well as the estimation of mean responses with missing data (see Section 4.2.2 below, as well as [RLTvdV16, Section 3], and [BR88, Lau96, Lau97, LM00]).

Our aim in this section is to use our Theorem 3.1 in order to prove a functional version of the forthcoming Theorem 4.5, corresponding to a special (but fundamental) case of [RLTvdV16, Theorem 2.1]. Two explicit examples related to kernels based on wavelets and on Fourier bases, respectively, are studied in full detail in Section 4.2.2.

In order to state the announced results, we adopt a notation similar to [RLTvdV16] and consider a sequence of i.i.d. random variables $\{X_i : i \geq 1\}$ with values in the measurable space $(E, \mathcal{E})$ and with common distribution μ. We also consider a sequence $\{K_n : n \geq 1\} \subset L^2(\mu^2)$ of symmetric kernels

$$K_n : E \times E \to \mathbb{R} : (x,y) \mapsto K_n(x,y).$$

For every n, we define the constant σ_n and the processes $U_n = \{U_n(t) : t \in [0,1]\}$ and $W_n = \{W_n(t) : t \in [0,1]\}$ according to (2.5)—(2.9), in the special case $p = 2$ and $\psi = K_n$, that is:

$$W_n(t) = \frac{\sum_{1 \leq 1 < i < j \leq \lfloor nt \rfloor} K_n(X_i, X_j) - \lfloor nt \rfloor (\lfloor nt \rfloor - 1)\, \mathbb{E}\, K_n(X_1, X_2)}{\sigma_n}, \ t \in [0,1]. \tag{4.16}$$

We write $k_n := \mathbb{E}\, K_n^2(X_1, X_2) = \|K_n\|^2_{L^2(\mu^2)}$, and assume that

$$\frac{k_n}{n} \xrightarrow{n \to \infty} \infty, \tag{4.17}$$

[2] The domain of the argmax operator can be extended to $D[0,1]$ — see e.g. [Fer01, p. 491].

and moreover

$$\sup_n \sup_{\|f\|_{L^2(\mu)}=1} \int_E \left(\int_E f(v) K_n(x,v) \mu(dv) \right)^2 \mu(dx) = \sup_n \|K_n\|_{op}^2 < \infty, \tag{4.18}$$

where $\| \bullet \|_{op}$ denotes the operator norm of the Hilbert-Schmidt operator $f \mapsto \int_E K_n(\cdot, y) f(y) \mu(dy)$, and

$$\|K_n\|_\infty \lesssim k_n. \tag{4.19}$$

Assumptions (4.17) and (4.18) are easily checkable conditions implying that the linear part of the Hoeffding decomposition of $W_n(1)$ vanishes in $L^2(\mathbb{P})$ as $n \to \infty$. Assumption (4.19) can be relaxed (see e.g. formula (10) and Lemma 2.1 in [RLTvdV16]), but we decided to avoid such a level of generality — which is not needed for the examples developed below — in order to keep our paper within bounds.

Theorem 4.5 (**Theorem 2.1 and Lemma 2.1 in [RLTvdV16]**). *Assume that there exists a sequence of finite measurable partitions $\mathcal{P}_n := \{\mathcal{X}_{n,m} : m = 1, ..., M_n\}$, $n \geq 1$, of the set E, such that*

$$\frac{1}{k_n} \sum_m \int_{\mathcal{X}_{n,m}} \int_{\mathcal{X}_{n,m}} K_n^2 \, d\mu d\mu \xrightarrow{n\to\infty} 1, \tag{4.20}$$

$$\frac{1}{k_n} \max_m \int_{\mathcal{X}_{n,m}} \int_{\mathcal{X}_{n,m}} K_n^2 \, d\mu d\mu \xrightarrow{n\to\infty} 0, \tag{4.21}$$

$$\max_m \mu(\mathcal{X}_{n,m}) \frac{k_n}{n} \xrightarrow{n\to\infty} 0, \tag{4.22}$$

$$\liminf_{n\to\infty} n \min_m \mu(\mathcal{X}_{n,m}) > 0. \tag{4.23}$$

Then, as $n \to \infty$, one has that $\sigma_n^2 \sim \frac{n^2 k_n}{2}$ and

$$W_n(1) \xrightarrow{\text{Law}} Z,$$

where Z is a standard Gaussian random variable.

The main abstract result of the present section is the following functional version of the previous statement.

Theorem 4.6. *Let the setting and assumptions of Theorem 4.5 prevail. If, in addition, one has that*

$$\sup_n n^{1/2+\epsilon_1} \max_m \mu(\mathcal{X}_{n,m}) < \infty, \text{ for some } \epsilon_1 > 0, \tag{4.24}$$

$$\sup_n \max_m \mu(\mathcal{X}_{n,m}) \frac{k_n}{n^{1-\epsilon_2}} < \infty \text{ **or** } \liminf_{n\to\infty} n^{1-\epsilon_2} \min_m \mu(\mathcal{X}_{n,m}) > 0, \text{ for some } \epsilon_2 > 0, \tag{4.25}$$

$$\sup_n \frac{n^{1+\alpha_1}}{k_n} < \infty, \text{ for some } \alpha_1 > 0, \tag{4.26}$$

$$\sup_n \frac{n^{\alpha_2}}{k_n} \int_E \int_E \left| K_n(x,y) \mathbf{1}_{\{(x,y) \in \bigcap (\mathcal{X}_{n,m} \times \mathcal{X}_{n,m})^c\}} \right|^2 \mu(dx)\mu(dy) < \infty, \text{ for some } \alpha_2 > 0, \tag{4.27}$$

then

$$W_n \Longrightarrow \left\{ B(t^2) : t \in [0,1] \right\}, \quad n \to \infty,$$

where B is a standard Brownian motion issued from zero.

As we will point out in Remark 4.7 below each of the four assumptions (4.24), (4.25), (4.26) and (4.27) plays a substantially different role in the proof.

Proof of Theorem 4.6. Using the notation introduced in this section, for every n we define

$$Q_n := \bigcup_{m=1}^{M_n} \mathcal{X}_{m,n} \times \mathcal{X}_{m,n} \subset E^2,$$

and write $Q_n^c = \bigcap_{m=1}^{M_n} (\mathcal{X}_{m,n} \times \mathcal{X}_{m,n})^c$ in order to denote the complement of Q_n in E^2. We set $A_n := K_n \mathbf{1}_{Q_n}$ and $B_n := K_n - A_n = K_n \mathbf{1}_{Q_n^c}$, and define $T_n := \{T_n(t) : t \in [0,1]\}$ and $R_n := \{R_n(t) : t \in [0,1]\}$ as

$$\begin{aligned} T_n(t) &:= \frac{\sum_{1 \leq 1 < i < j \leq \lfloor nt \rfloor} A_n(X_i, X_j) - \lfloor nt \rfloor (\lfloor nt \rfloor - 1) \, \mathbb{E}\, A_n(X_1, X_2)}{\sigma_n}, \\ R_n(t) &:= \frac{\sum_{1 \leq 1 < i < j \leq \lfloor nt \rfloor} B_n(X_i, X_j) - \lfloor nt \rfloor (\lfloor nt \rfloor - 1) \, \mathbb{E}\, B_n(X_1, X_2)}{\sigma_n}, \end{aligned}$$

in such a way that $W_n = T_n + R_n$. Our first remark is that, for every $f \in L^2(\mu)$ with unit norm, one has that

$$\begin{aligned} &\int_E \left(\int_E f(v) A_n(x,v) \mu(dv) \right)^2 \mu(dx) = \sum_{m=1}^{M_n} \int_{\mathcal{X}_{m,n}} \left(\int_{\mathcal{X}_{m,n}} f(v) K_n(x,v) \mu(dv) \right)^2 \mu(dx) \\ &\leq \sum_{m=1}^{M_n} \int_E \left(\int_{\mathcal{X}_{m,n}} f(v) K_n(x,v) \mu(dv) \right)^2 \mu(dx) \leq \|K_n\|_{op} \sum_{m=1}^{M_n} \int_{\mathcal{X}_{m,n}} f(x)^2 \mu(dx) = \|K_n\|_{op}, \end{aligned}$$

from which we infer that

$$\sup_n \|A_n\|_{op}, \ \sup_n \|B_n\|_{op} < \infty, \tag{4.28}$$

where we have applied (4.18) and the triangle inequality in order to deal with B_n. Using the identity (2.9) (in the case $\psi = B_n$) together with (4.28), with the relations $\sigma_n^2 \sim n^2 k_n / 2$ and $n/k_n \to 0$ and with (4.20), shows immediately that, as $n \to \infty$, $\mathbb{E}[R_n(t)^2] = o(\sigma_n^2)$ for every $t \in [0,1]$. This in turn implies that $\sigma_n^2 \sim \mathrm{Var}(T_n(1))$. We will now study R_n and T_n separately, and prove that

(i) the sequence $\{R_n : n \geq 1\}$ is tight in $D[0,1]$, so that $R_n \Longrightarrow 0$ (zero function of $D[0,1]$);

(ii) the sequence $\{T_n : n \geq 1\}$ verifies the assumptions of Theorem 3.1 in the case $p = 2$, with $\alpha_{1,2} = 0$ and $\alpha_{2,2} = 1$, and therefore T_n weakly converges to $B(t^2)$ in $D[0,1]$.

[Proof of (i)] We first define the functions g_0, g_1, g_2 according to (2.4), in the case $p = 2$ and $\psi = B_n$, so that the Hoeffding decomposition of the U-statistic $R_n(t)$, $t \in [0,1]$, is

$$R_n(t) = \frac{\lfloor nt \rfloor - 1}{\sigma_n} \sum_{j=1}^{\lfloor nt \rfloor} [g_1(X_j) - g_0] + \frac{1}{\sigma_n} \sum_{1 \leq i < j \leq \lfloor nt \rfloor} [g_2(X_i, X_j) - g_1(X_i) - g_1(X_j) + g_0] := R_n'(t) + R_n''(t).$$

Now fix $0 \leq s < t \leq 1$. Then, writing $g_1 - g_0 := \psi_1$, as before, and using as always the symbol c in order to denote an absolute finite constant whose exact value might change from line to line,

$$\begin{aligned} &\mathbb{E}\left| R_n'(t) - R_n'(s) \right|^2 \\ &\leq \frac{c}{\sigma_n^2} \left\{ \mathbb{E} \left| \lfloor ns \rfloor \sum_{\lfloor ns \rfloor < j \leq \lfloor nt \rfloor} \psi_1(X_j) \right|^2 + \mathbb{E} \left| (\lfloor nt \rfloor - \lfloor ns \rfloor) \sum_{1 \leq j \leq \lfloor nt \rfloor} \psi_1(X_j) \right|^2 \right\} := \frac{c}{\sigma_n^2} (Y_1 + Y_2). \end{aligned}$$

We can assume without loss of generality that $\alpha_1 \in (0,1]$; we have that

$$\frac{Y_1}{\sigma_n^2} \leq \frac{c}{k_n} \mathbb{E} \left[\sum_{\lfloor ns \rfloor < j \leq \lfloor nt \rfloor} \psi_1(X_j)^2 \right] \leq \frac{c}{k_n} \left(\lfloor nt \rfloor - \lfloor ns \rfloor \right)^{1+\alpha_1} \mathbb{E} |\psi_1(X_1)|^2 \leq c \left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \right)^{1+\alpha_1} \tag{4.29}$$

where we have used (4.18) and (4.26) to deduce the last inequality. Analogously, one shows that

$$\frac{Y_2}{\sigma_n^2} \leq c \left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \right)^2 \frac{n}{k_n} \mathbb{E} |\psi_1(X_j)|^2 \leq c \left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \right)^{1+\alpha_1}, \tag{4.30}$$

where the last inequality follows again from (4.18) and from (4.17), as well as from the fact that $\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \in [0,1]$.

In order to deal with R'', we adopt as before the notation $\psi_2(X_i,X_j) := g_2(X_i,X_j) - g_1(X_i) - g_1(X_j) + g_0$, and observe that, for every $a \geq 1$, $\mathbb{E}[|\psi_2(X_i,X_j)|^a] \leq c \int_E \int_E |B_n|^a \, d\mu^2$, for some absolute constant c depending solely on a. For every n and every $0 \leq s < t \leq 1$, we define the set of integers

$$H_n(s,t) \quad := \quad \{(i,j) : 0 \leq i < j \leq \lfloor nt \rfloor\} \backslash \{(i,j) : 0 \leq i < j \leq \lfloor ns \rfloor\}$$

Clearly, $|H_n(s,t)| = \binom{\lfloor nt \rfloor}{2} - \binom{\lfloor ns \rfloor}{2} = \frac{1}{2}(\lfloor nt \rfloor + \lfloor ns \rfloor - 1)(\lfloor nt \rfloor - \lfloor ns \rfloor)$. For fixed $0 \leq s < t \leq 1$, one has that

$$\mathbb{E}\left|R_n''(t) - R_n''(s)\right|^2 = \frac{1}{\sigma_n^2}\,\mathbb{E}\left|\sum_{(i,j)\in H_n(s,t)} \psi_2(X_i,X_j)\right|^2.$$

In order to bound such a quantity, we use orthogonality of the summands in the above sum, fix an $\alpha_2 > 0$ such that condition (4.27) is satisfied, and note that

$$\begin{aligned}
\mathbb{E}\left|R_n''(t) - R_n''(s)\right|^2 &= \frac{1}{\sigma_n^2} \sum_{(i_1,i_2)\in H_n(s,t)} \mathbb{E}\left[\psi_2(X_{i_1},X_{i_2})^2\right] \\
&\leq \frac{c}{\sigma_n^2}|H_n(s,t)| \int_E \int_E B_n^2 \, d\mu^2 \\
&\leq \frac{c\int_E \int_E B_n^2 \, d\mu^2}{k_n} \cdot \frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \\
&\leq \frac{cn^{\alpha_2}\int_E \int_E B_n^2 \, d\mu^2}{k_n} \cdot \left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n}\right)^{1+\alpha_2}. \qquad (4.31)
\end{aligned}$$

We have therefore shown (in (4.29), (4.30) and (4.31)) that $\{R_n\}$ satisfies the tightness criterion of Lemma 1.1, for $\alpha = \min(\alpha_1, \alpha_2)$ and $\beta = 2$, and the proof of Point (i) is concluded.

[Proof of (ii)] In this part of the proof, we denote by g_0, g_1, g_2 the functions obtained from (2.4) by selecting $p = 2$ and $\psi = A_n$. Note that each of the three kernels g_i implicitly depends on n and that, by virtue of (4.28), one has

$$\sup_n \left(\mathbb{E}[g_1(X_1)^2] + |g_0|\right) < \infty. \qquad (4.32)$$

Since $g_2 = A_n$, $\sigma_n^2 \sim k_n n^2/2$ and (4.20) is in order, we see immediately that the constants b_1 and b_2 appearing at Point (a) of Theorem 3.1 are such that $b_1 = 0$ and $b_2^2 = 2$, yielding $\alpha_{1,2} = 0$ and $\alpha_{2,2} = 1$. In order to conclude our proof, we have now to check that the quantities appearing at Points 1.-6. of Remark 3.3 all converge to zero as $n \to \infty$ and that the quantities in points i)-iv) of the same remark are bounded for some $\epsilon > 0$. This is immediately done for the quantities at Points 1., i) and ii), by virtue of (4.32). To deal with the quantity at Point iii), we note that, for some $\epsilon > 0$,

$$\begin{aligned}
\frac{n^{5/2+\epsilon}}{\sigma_n^2}\|g_1\|^2_{L^4(\mu)} &\sim \frac{n^{1/2+\epsilon}}{k_n}\left[\int_E \left(\int_E K_n(x,y)\,\mathbf{1}\left[(x,y) \in \bigcup_m \mathcal{X}_{n,m} \times \mathcal{X}_{n,m}\right]\mu(dy)\right)^4 \mu(dx)\right]^{1/2} \\
&= \frac{n^{1/2+\epsilon}}{k_n}\left[\sum_m \int_{\mathcal{X}_{n,m}} \left(\int_{\mathcal{X}_{n,m}} K_n(x,y)\mu(dy)\right)^4 \mu(dx)\right]^{1/2} \\
&\leq \frac{n^{1/2+\epsilon}}{k_n}\left[\sum_m \int_{\mathcal{X}_{n,m}} \left(\int_{\mathcal{X}_{n,m}} K_n(x,y)\mu(dy)\right)^2 \left(\int_{\mathcal{X}_{n,m}} k_n\mu(dy)\right)^2 \mu(dx)\right]^{1/2}, \quad \text{by (4.20)} \\
&\leq n^{1/2+\epsilon} \max_m \mu(\mathcal{X}_{n,m}) \left[\sum_m \int_{\mathcal{X}_{n,m}} \left(\int_{\mathcal{X}_{n,m}} K_n(x,y)\mu(dy)\right)^2 \mu(dx)\right]^{1/2} \\
&\lesssim n^{1/2+\epsilon} \max_m \mu(\mathcal{X}_{n,m}), \quad \text{by (4.28)},
\end{aligned}$$

which is bounded by (4.24). On the other hand,

$$\frac{n^2}{\sigma_n^2}\|g_1 \star_1^0 g_2\|_{L^2} = \frac{n^2}{\sigma_n^2}\left(\int_E\int_E (g_1(x)g_2(x,y))^2\,\mu(dx)\mu(dy)\right)^{1/2}$$
$$\sim \frac{1}{k_n}\left[\int_E\int_E\left(\int_E K_n(x,z)\mathbf{1}\left[(x,z)\in\bigcup_m \mathcal{X}_{n,m}\times\mathcal{X}_{n,m}\right]\mu(dz)\right.\right.$$
$$\left.\left.\cdot\ K_n(x,y)\mathbf{1}\left[(x,y)\in\bigcup_m \mathcal{X}_{n,m}\times\mathcal{X}_{n,m}\right]\right)^2\mu(dx)\mu(dy)\right]^{1/2}$$
$$= \frac{1}{k_n}\left[\sum_m\int_{\mathcal{X}_{n,m}}\int_{\mathcal{X}_{n,m}}\left(\int_{\mathcal{X}_{n,m}} K_n(x,z)\mu(dz)\right)^2 K_n^2(x,y)\mu(dx)\mu(dy)\right]^{1/2}$$
$$\overset{(4.19)}{\leq} \frac{1}{k_n}\left[\sum_m\int_{\mathcal{X}_{n,m}}\int_{\mathcal{X}_{n,m}}\left(\int_{\mathcal{X}_{n,m}} K_n(x,z)\mu(dz)\right)^2 k_n^2\,\mu(dx)\mu(dy)\right]^{1/2}$$
$$\leq \sqrt{\max_m \mu(\mathcal{X}_{n,m})}\left[\sum_m\int_{\mathcal{X}_{n,m}}\left(\int_{\mathcal{X}_{n,m}} K_n(x,z)\mu(dz)\right)^2\mu(dx)\right]^{1/2} \xrightarrow{n\to\infty} 0,$$

by (4.24) and (4.28), showing that the quantity at Point 2. vanishes. We can deal at once with the quantities at Point 3. and 5. by means of the following considerations. For a fixed n, denote by $\{\lambda_j : j\geq 1\}$ and $\{e_j : j\geq 1\}$, respectively, the sequence of eigenvalues (taken in decreasing order) and eigenfunctions of the Hilbert-Schmidt operator on $L^2(\mu)$ given by $f\mapsto\int_E A_n(\cdot,y)f(y)\mu(dy)$. Then, such eigenfunctions form an orthonormal system in $L^2(\mu)$, and one has that $A_n = g_2 = \sum_i \lambda_i e_i\otimes e_i$, with convergence in $L^2(\mu^2)$. Such a relation yields that $\|A_n\|_{op} = \lambda_1$, $g_2\star_1^1 g_2 = \sum_i \lambda_i^2 e_i\otimes e_i$, $g_1 = \sum_i \lambda_i\mu_i e_i$ (where $\mu_i := \int_E e_i d\mu$) and $g_1\star_1^1 g_2 = \sum_i \lambda_i^2\mu_i e_i$. Since $|\mu_i|\leq 1$ (by Cauchy-Schwarz), we infer that

$$\|g_1\star_1^1 g_2\|_{L^2(\mu)},\ \|g_2\star_1^1 g_2\|_{L^2(\mu^2)} \leq \sqrt{\sum_i \lambda_i^4} \leq \|A_n\|_{op}\|A_n\|_{L^2(\mu^2)},$$

and the desired convergence to zero follows from (4.17), (4.20) and (4.28). The vanishing of the quantity at Point 4. follows from

$$\frac{n^{3/2}}{\sigma_n^2}\|g_2\star_1^0 g_2\|_{L^2} \sim \frac{1}{n^{1/2}k_n}\left(\sum_m\int_{\mathcal{X}_{n,m}}\int_{\mathcal{X}_{n,m}}\int_{\mathcal{X}_{n,m}} K_n^2(x,y)K_n^2(x,z)\mu(dx)\mu(dy)\mu(dz)\right)^{1/2}$$
$$\leq \frac{1}{n^{1/2}k_n}\left(\sum_m\int_{\mathcal{X}_{n,m}}\int_{\mathcal{X}_{n,m}}\left(\int_{\mathcal{X}_{n,m}} K_n^4(x,y)\mu(dx)\right)^{1/2}\left(\int_{\mathcal{X}_{n,m}} K_n^4(x,z)\mu(dx)\right)^{1/2}\mu(dy)\mu(dz)\right)^{1/2}$$
$$= \frac{1}{n^{1/2}k_n}\left\{\sum_m\left(\int_{\mathcal{X}_{n,m}}\left(\int_{\mathcal{X}_{n,m}} K_n^4(x,z)\mu(dx)\right)^{1/2}\mu(dz)\right)^2\right\}^{1/2}$$
$$\leq \left\{\frac{1}{k_n^2}\max_m\frac{\mu(\mathcal{X}_{m,n})}{n}\sum_m\int_{\mathcal{X}_{n,m}}\int_{\mathcal{X}_{n,m}} K_n^4(x,z)\mu(dx)\mu(dz)\right\}^{1/2} \xrightarrow{n\to\infty} 0$$

where we have applied (4.19) and (4.22). One has also that, for some $\epsilon>0$,

$$\frac{n^{1+\epsilon}}{\sigma_n^2}\|g_2\|_{L^4(\mu)}^2 \sim \frac{1}{n^{1-\epsilon}k_n}\left\{\sum_m\int_{\mathcal{X}_{n,m}}\int_{\mathcal{X}_{n,m}} K_n^4(x,y)\mu(dx)\mu(dy)\right\}^{1/2}$$
$$= n^{\epsilon}\left\{\frac{1}{k_n^2}\max_m\frac{\mu(\mathcal{X}_{n,m})}{n}\sum_m\int_{\mathcal{X}_{n,m}}\int_{\mathcal{X}_{n,m}} K_n^4(x,y)\mu(dx)\mu(dy)\right\}^{1/2}\frac{1}{n^{1/2}\max_m\mu(\mathcal{X}_{n,m})^{1/2}}$$

is bounded, by (4.19), (4.20), (4.22), (4.23) and (4.25) – this yields the boundedness of the sequence at Point iv). Finally, the convergence to zero of the quantity at Point 6. is a direct consequence of (4.17) and (4.20).

□

Remark 4.7. By inspection of the previous proof, one sees that Assumptions (4.26) and (4.27) imply that the off-diagonal part of the U-process W_n is tight in the space $D[0,1]$. On the other hand, Assumptions (4.24) and (4.25) are needed in order to ensure that the (dominating) diagonal component of W_n meets the requirements of Theorem 3.1. Note that Assumption (4.24) is such that (a) it does not appear in [RLTvdV16], and (b) it would be needed if one wanted to prove a one-dimensional CLT for $W_n(1)$ by using the techniques developed in [DP19]. This slight discrepancy between the assumptions of [DP19] and [RLTvdV16] is explained by the fact that the sufficient conditions discovered in [DP19] would imply not only a CLT for $W_n(1)$, but also that $\mathbb{E}[W_n(1)^4] \to 3$, and consequently need to be stronger.

4.2.2 Two examples

As an application of Theorem 4.6, we will consider two families of kernels satisfying the set of sufficient conditions for functional convergence pointed out in the previous section. As explained in [RLTvdV16, Section 3] both types of U-statistics can be used in the non-parametric estimation of quadratic functional of densities – see also [BR88, LM00]

(I) (*Wavelet-based kernels*) Following [RLTvdV16, Section 4.1], we consider expansions of functions $f \in L_2(\mathbb{R}^d)$ on an orthonormal basis of compactly supported, bounded wavelets of the form

$$f(x) = \sum_{j\in\mathbb{Z}^d} \sum_{v\in\{0,1\}^d} \langle f, \psi^v_{0,j}\rangle \, \psi^v_{0,j}(x) + \sum_{i=0}^{\infty} \sum_{j\in\mathbb{Z}^d} \sum_{v\in\{0,1\}^d\setminus\{0\}} \langle f, \psi^v_{i,j}\rangle \, \psi^v_{i,j}(x).$$

The functions $\psi^v_{i,j}$ are orthogonal for different indices (i,j,v) and given by scaled and translated versions of the 2^d base functions $\psi^v_{0,0}$:

$$\psi^v_{i,j}(x) = 2^{id/2}\psi^v_{0,0}(2^i x - j).$$

We concentrate on functions f with support in $E = [0,1]^d$. As noted in [RLTvdV16, Section 4.1], for each resolution level i and vector v, only the order 2^{id} elements $\psi^v_{i,j}$ are nonzero in E. We denote the corresponding set of indices j by J_i. We then truncate the expansion at the level of resolution $i = I$ and look at the kernel

$$K_n(x,y) = \sum_{j\in J_0} \sum_{v\in\{0,1\}^d} \psi^v_{0,j}(x)\psi^v_{0,j}(y) + \sum_{i=0}^{I} \sum_{j\in J_i} \sum_{v\in\{0,1\}^d\setminus\{0\}} \psi^v_{i,j}(x)\psi^v_{i,j}(y).$$

(II) (*Kernels based on Fourier expansions*) Any function $f \in L_2[-\pi,\pi]$ can be represented through the Fourier series $f = \sum_{j\in\mathbb{Z}} f_j e_j$ for $e_j(x) = e^{ijx/\sqrt{2\pi}}$ and $f_j = \int_{-\pi}^{\pi} f e_j d\lambda$, where λ is the Lebesgue measure. We can write $f_k = \sum_{|j|\le k} f_j e_j$ to obtain an orthogonal projection of f onto a $2k+1$-dimensional space. Assuming that k depends on n, we can also write down the corresponding kernel as:

$$K_n(x,y) = \sum_{|j|\le k} e_j(x)e_j(y) = \frac{\sin\left(\left(k+\frac{1}{2}\right)(x-y)\right)}{2\pi\sin\left(\frac{1}{2}(x-y)\right)}$$

and note that $K_n(x,y) = D_k(x-y)$, where D_k is the well-known Dirichlet kernel.

Theorem 4.8. *Let the above assumption and notation prevail.*

1. *Let μ be any probability measure on $[0,1]^d$ with a Lebesgue density that is bounded and bounded away from zero. The sequence of wavelet-based kernels $\{K_n : n \ge 1\}$ defined at Point* (I) *above satisfies the assumptions of Theorem 4.5, with respect to μ, as soon as $n \ll k_n \ll n^2$, for $k_n = 2^{Id}$. Moreover, a sufficient condition for such a sequence to satisfy the assumptions of Theorem 4.6 is $n^{1+\gamma_1} \lesssim k_n \lesssim n^{2-\gamma_2}$, for some $\gamma_1, \gamma_2 > 0$.*

2. *Let μ be any measure on $\mathbb{R}$ with a bounded Lebesgue density and $k_n = 2k+1$. The sequence of Fourier-based kernels $\{K_n : n \geq 1\}$ defined at Point (II) above satisfies the assumptions of Theorem 4.5 as soon as $n \ll k_n \ll n^2$. In addition, a sufficient condition for such a sequence to meet the assumptions of Theorem 4.6 is $n^{1+\eta_1} \lesssim k_n \lesssim n^{2-\eta_2}$, for some $\eta_1, \eta_2 > 0$.*

Proof.

1. For $n \ll k_n \ll n^2$ and K_n defined in point (I) above, the assumptions of Theorem 4.5 are verified in [RLTvdV16, Proposition 4.1]. The authors note that, by assumption, each function $\psi_{I,j}^v$ is supported within a set of the form $2^{-I}(C+j)$ for a given cube C that depends on the type of the wavelet, for any v. They take $\mathcal{X}_{n,m}$ to be blocks (cubes) of l_n^d adjacent cubes $2^I(C+j)$, giving $M_n = O(k_n/l_n^d)$ sets $\mathcal{X}_{n,m}$. In order for the assumptions (4.18)-(4.23) to be satisfied, the authors require that

   a) $M_n \to \infty$,
   b) $M_n \lesssim n$,
   c) $\frac{M_n}{k_n} \to 0$,
   d) $M_n^{-1} k_n / n \to 0$.

   Now assume $n^{1+\gamma_1} \lesssim k_n \lesssim n^{2-\gamma_2}$ for some $\gamma_1, \gamma_2 > 0$. Condition (4.26) is then automatically satisfied. As noted in the proof of [RLTvdV16, Proposition 4.1], $\mu(\mathcal{X}_{m,n})$ is of order $\frac{1}{M_n}$. Now, it is also noted in the proof of [RLTvdV16, Proposition 4.1] that, if $K_n(x_1, x_2) \neq 0$ then there exists some j such that $x_1, x_2 \in 2^{-I}(C+j)$. Moreover, the set of (x_1, x_2) in the complement of $\bigcup_m \mathcal{X}_{n,m} \times \mathcal{X}_{n,m}$ where $K_n(x_1, x_2) \neq 0$ is contained in the union U of all cubes $2^{-I}(C+j)$ that intersect the boundary of some $\mathcal{X}_{n,m}$. It is also noted that the number of such cubes is of order $M_n^{1/d} k_n^{1-1/d}$ and that $\mu\left(2^{-I}(C+j)\right) \lesssim \frac{1}{k_n}$. Therefore, using $\|K_n\|_\infty \lesssim k_n$, we note that, for any $\alpha_2 > 0$,

$$\begin{aligned}
&\frac{n^{\alpha_2}}{k_n} \int\int \left| K_n(x,y) \mathbf{1}\left[(x,y) \in \bigcap (\mathcal{X}_{n,m} \times \mathcal{X}_{n,m})^c\right] \right|^2 \mu(dx)\mu(dy) \\
\lesssim &\frac{n^{\alpha_2} k_n^2}{k_n} M_n^{1/d} k_n^{1-1/d} \left(\frac{1}{k_n}\right)^2 \\
= &n^{\alpha_2} \left(\frac{M_n}{k_n}\right)^{1/d}.
\end{aligned}$$

   Condition (4.27) requires this quantity to be bounded for some $\alpha_2 > 0$. It will indeed be bounded for $\alpha_2 \leq \frac{1}{2d}$ if we choose $M_n = k_n^{1/2}$. Moreover, for $M_n = k_n^{1/2}$, $\frac{M_n}{k_n} \to 0$. Also, $\frac{k_n}{M_n n} = \frac{k_n^{1/2}}{n} \to 0$, as $k_n \ll n^2$. Under the same assumption, $M_n \to \infty$ and $n^{1/2+\gamma_1/2} \lesssim M_n \lesssim n^{1-\gamma_2/2} \lesssim n$ and so conditions (4.24) and (4.25) are also satisfied (as $\mu(\mathcal{X}_{m,n})$ is of order $\frac{1}{M_n}$). Therefore all the conditions a), b), c), d) from above, as well as conditions (4.24)-(4.27), are satisfied. This finishes the proof.

2. For $n \ll k_n \ll n^2$, the assumptions (4.18)-(4.23) for kernel K_n of Point (II) are verified in [RLTvdV16, Proposition 4.2]. The authors take a partition $(-\pi, \pi] = \bigcup_m \mathcal{X}_{n,m}$ in $M_n = \frac{2\pi}{\delta}$ intervals of length δ for $\delta \to 0$ for $\frac{1}{\sqrt{k_n}} \ll \delta \ll \frac{n}{k_n}$ and introduce an $\epsilon > 0$ such that $\frac{1}{\sqrt{k_n}} \ll \epsilon \ll \delta$.

   Now, assume that $n^{1+\eta_1} \lesssim k_n \lesssim n^{2-\eta_2}$, for some $\eta_1, \eta_2 > 0$. This makes condition (4.26) readily satisfied. In order for condition (4.24) to be satisfied, we require that $n^{1/2+\epsilon_1}\delta$ is bounded for some $\epsilon_1 > 0$. Moreover, condition (4.25) will be satisfied if $\delta \lesssim \frac{n^{1-\epsilon_2}}{k_n}$ for some $\epsilon_2 > 0$.

   The authors of [RLTvdV16] note that the complement of $\bigcup_m \mathcal{X}_{n,m} \times \mathcal{X}_{n,m}$ is contained in $\{(x_1, x_2) : |x_1 - x_2| > \epsilon\}$ except for the set of $2(M_n - 1)$ triangles indicated in [RLTvdV16, Figure 3]. Now, by the argument of the proof of [RLTvdV16, Proposition 4.2], for any $\alpha_2 > 0$,

$$\frac{n^{\alpha_2}}{k_n} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathbf{1}[|x-y| > \epsilon] \left|K_n(x,y)\right|^2 \mu(dx)\mu(dy) \lesssim \epsilon n^{\alpha_2} + \frac{n^{\alpha_2}}{\epsilon^2 k_n},$$

which is bounded, if $\frac{n^{\alpha_2/2}}{k_n^{1/2}} \lesssim \epsilon \lesssim n^{-\alpha_2}$. Each of the remaining triangles in the complement of $\bigcup_m \mathcal{X}_{n,m} \times \mathcal{X}_{n,m}$ has sides of length of order ϵ. Hence, for a typical triangle Δ and an interval I of length of the order ϵ,

$$\frac{n^{\alpha_2}}{k_n} \int\int_{\Delta} |K_n(x,y)|^2 \, dxdy \underset{u=x-y}{\overset{v=y}{\lesssim}} \frac{n^{\alpha_2}}{k_n} \int_I \int_0^{\epsilon} |D_n(u)|^2 dudv \lesssim \frac{n^{\alpha_2}}{k_n} \epsilon k_n = \epsilon n^{\alpha_2}. \tag{4.33}$$

There are $2(M_n - 1)$ such triangles. Therefore, condition (4.26) will be satisfied if, in addition to $\frac{n^{\alpha_2/2}}{k_n^{1/2}} \lesssim \epsilon \lesssim n^{-\alpha_2}$, $M_n \epsilon n^{\alpha_2} = \frac{2\pi\epsilon n^{\alpha_2}}{\delta}$ is bounded, i.e. $\epsilon n^{\alpha_2} \lesssim \delta$.

Summing up, conditions (4.18) - (4.27) are satisfied if, for some $\delta, \epsilon, \epsilon_1, \epsilon_2, \alpha_2 > 0$,

$$\frac{n^{3\alpha_2/2}}{k_n^{1/2}} \lesssim \epsilon n^{\alpha_2} \lesssim \delta \lesssim \min\left(\frac{1}{n^{1/2+\epsilon_1}}, \frac{n^{1-\epsilon_2}}{k_n}\right).$$

Such choices of $\delta, \epsilon, \epsilon_1, \epsilon_2, \alpha_2 > 0$ exist. Indeed, let $\alpha_2 = \frac{\min(\eta_1,\eta_2)}{6}$, $\epsilon_1 = \frac{1}{2}\left(\eta_1 - \frac{\min(\eta_1,\eta_2)}{2}\right)$, $\epsilon_2 = \frac{1}{2}\left(\eta_2 - \frac{\min(\eta_1,\eta_2)}{2}\right)$. Then, under the assumption $n^{1+\eta_1} \lesssim k_n \lesssim n^{2-\eta_2}$, we have that $\frac{n^{3\alpha_2/2}}{k_n^{1/2}} \lesssim \min\left(\frac{1}{n^{1/2+\epsilon_1}}, \frac{n^{1-\epsilon_2}}{k_n}\right)$ and it suffices to choose $\delta = \frac{n^{3\alpha_2/2}}{k_n^{1/2}}$ and $\epsilon = \frac{n^{\alpha_2/2}}{k_n^{1/2}}$. This finishes the proof.

□

5 Technical results and proofs of main statements

Unless otherwise specified, for the rest of the section we adopt the same conventions and notation put forward in Section 2.

5.1 A new product formula

We start by proving a new product formula for symmetric U-statistics with arguments of possibly different sizes. In order to state it, we need to recall the Hoeffding decomposition of not necessarily symmetric kernel functions: Let $f \in L^1(\mu^p)$. Then, f can be decomposed as follows: For all $(x_1, \ldots, x_p) \in E^p$ one has

$$f(x_1, \ldots, x_p) = \sum_{J \subseteq [p]} f_J\big((x_i)_{i \in J}\big), \tag{5.1}$$

where we follow the convention that in $(x_i)_{i \in J}$ the coordinates i appear in increasing order, i.e. if $J = \{i_1, \ldots, i_k\}$ with $k = |J|$ and $1 \le i_1 < \ldots < i_k \le p$, then $(x_i)_{i \in J} = (x_{i_1}, \ldots, x_{i_k})$. The kernels f_J, $J \subseteq [p]$, are given by

$$f_J\big((x_i)_{i \in J}\big) = \sum_{K \subseteq J} (-1)^{|J|-|K|} \int_{E^{p-|K|}} f(x_1, \ldots, x_p) d\mu^{p-|K|}\big((x_i)_{i \in [p] \setminus K}\big) \tag{5.2}$$

and they are *canonical* with respect to μ in the sense that for each $\emptyset \neq J \subseteq p$ with $|J| = k$, each $j \in J$ and all $(x_i)_{i \in J \setminus \{j\}} \in E^{|J|-1}$ one has that

$$\int_E f_J\big(x_{i_1}, \ldots, x_{i_{l-1}}, y, x_{i_{l+1}}, \ldots, x_{i_k}\big) d\mu(y) = 0\,, \tag{5.3}$$

where we again suppose that $J = \{i_1, \ldots, i_k\}$, $1 \le i_1 < \ldots < i_k \le p$ and where $i_l = j$. For a detailed discussion and proofs of these facts we refer the reader to [Maj13, Chapter 9].

Note that, if the kernel f is symmetric as in Section 2, then we can define the (symmetric) functions g_k, $0 \le k \le p$ by

$$g_k(y_1, \ldots, y_k) = \int_{E^{p-k}} f\big(y_1, \ldots, y_k, x_1, \ldots, x_{p-k}\big) d\mu^{p-k}(x_1, \ldots, x_{p-k})$$

as before and we obtain that, for every subset $J \subseteq [p]$ with $1 \le k := |J| \le p$,

$$f_J(x_1,\dots,x_k) = \sum_{l=0}^{s} (-1)^{k-l} \sum_{1\le i_1<\dots<i_l\le k} g_l\big(x_{i_1},\dots,x_{i_l}\big) = f_k(x_1,\dots,x_k)\,,$$

where the symmetric and degenerate kernel f_k has been defined in (2.3).

For the statement of our product formula we have to fix some more notation: Let us fix two positive integers p and q. Then, for nonnegative integers l, n, m, r such that $n \le m$, $r \le p \wedge q$, $l \ge p+q-2r$ and sets $L \subseteq [m]$ with $|L| = l$, we denote by $\Pi_{r,n,m}(L)$ the collection of all triples

$$(A,B,C) \in \mathcal{D}_{2r+l-p-q}(n) \times \mathcal{D}_{p-r}(n) \times \mathcal{D}_{q-r}(m)$$

such that L is the disjoint union of A, B and C.

Proposition 5.1 (Product formula). *Let $p, q \ge 1$ be positive integers and assume that $\psi \in L^2(\mu^p)$ and $\varphi \in L^2(\mu^q)$ are degenerate, symmetric kernels of orders p and q respectively. Moreover, let $n \ge p$ and $m \ge q$ be positive integers with $m \ge n$. Then, whenever $n \ge p+q$ we have the Hoeffding decomposition:*

$$J_p^{(n)}(\psi) J_q^{(m)}(\varphi) = \sum_{\substack{M\subseteq[m]:\\ |M|\le p+q}} U_M = \sum_{k=|p-q|}^{p+q} \sum_{s=0}^{q\wedge k\wedge(m-n)} \sum_{\substack{M\subseteq[m]:\\ |M|=k,\\ |M\cap\{n+1,\dots,m\}|=s}} U_M\,,$$

where, for a set $M \subseteq [m]$ with $0 \le k := |M| \le p+q$ and $0 \le s := |M \cap \{n+1,\dots,m\}| \le q \wedge k \wedge (m-n)$, the Hoeffding component U_M is given by

$$U_M = \sum_{r=\lceil \frac{p+q-k}{2} \rceil}^{p\wedge(q-s)\wedge(p+q-k)} \binom{n-k+s}{p+q-r-k} \cdot \sum_{(A,B,C)\in\Pi_{r,n,m}(M)} \big(\psi \star_r^{p+q-r-k} \varphi\big)_M\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big)\,. \tag{5.4}$$

Moreover, for such an M, we further have the bound

$$\sqrt{\mathrm{Var}(U_M)} \le \sum_{r=\lceil \frac{p+q-k}{2} \rceil}^{p\wedge(q-s)\wedge(p+q-k)} \binom{n-k+s}{p+q-r-k} \binom{k-s}{2r+k-p-q,p-r,q-r-s} \big\|\psi \star_r^{p+q-r-k} \varphi\big\|_{L^2(\mu^k)}\,. \tag{5.5}$$

Remark 5.2. The above product formula is an extension of the one proved in [DP19, Proposition 2.6], for symmetric and degenerate U-statistics based on the same range $X_1,\dots,X_n$ of data. Indeed, suppose that $n = m$. Then, if $|M| = k$ and (necessarily) $s = 0$ it is not hard to verify that

$$\begin{aligned}&\sum_{(A,B,C)\in\Pi_{r,n,m}(M)} \big(\psi \star_r^{p+q-r-k} \varphi\big)_M\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big)\\ &= \frac{k!}{(2r+k-p-q)!(p-r)!(q-r)!} \big(\psi \widetilde{\star_r^{p+q-r-k}} \varphi\big)_k\big(X_i, i\in M\big)\end{aligned}$$

and the product formula reduces to the one in [DP19, Proposition 2.6]. The main difference in general is that, in the situation of Proposition 5.1 and for $n \ne m$, the product is no longer (in general) a finite sum of degenerate and symmetric U-statistics. However, its Hoeffding decomposition (in the sense of not necessarily symmetric statistics — see e.g. [KR82, DP17]) is still completely explicit and hence suitable for providing useful bounds.

Proof of Proposition 5.1. Write

$$W := J_p^{(n)}(\psi) = \sum_{J\in\mathcal{D}_p(n)} W_J \quad \text{and} \quad V := J_q^{(m)}(\varphi) = \sum_{K\in\mathcal{D}_q(m)} V_K$$

for the respective Hoeffding decompositions of W and V, i.e.

$$W_J = \psi(X_j, j \in J)\,, \quad J \in \mathcal{D}_p(n) \quad \text{and} \quad V_K = \varphi(X_i, i \in K)\,, \quad K \in \mathcal{D}_q(m)\,.$$

From Theorem 2.6 in [DP17] we know that the Hoeffding decomposition of VW is given by

$$VW = \sum_{\substack{M \subseteq [m]: \\ |M| \le p+q}} U_M\,,$$

where, for $M \subseteq [m]$ with $|M| \le p+q$ we have

$$U_M = \sum_{L \subseteq M} (-1)^{|M|-|L|} \sum_{\substack{J \in \mathcal{D}_p(n), K \in \mathcal{D}_q(m): \\ J \Delta K \subseteq L, \\ M \subseteq J \cup K}} \mathbb{E}\big[W_J V_K \,\big|\, \mathcal{F}_L\big]\,, \tag{5.6}$$

where $\mathcal{F}_L := \sigma(X_j, j \in L)$. Note that $U_M = 0$ whenever $|M| < |p-q|$ because $|J \Delta K| \ge |p-q|$ for all $J \in \mathcal{D}_p(n)$ and $K \in \mathcal{D}_q(m)$. Moreover, $U_M = 0$ if $|M \cap \{n+1, \dots, m\}| > q$ since $M \cap \{n+1, \dots, m\} \subseteq K$ and $|K| = q$. Hence, we have

$$VW = \sum_{\substack{M \subseteq [m]: \\ |M| \le p+q\,, \\ |M \cap \{n+1,\dots,m\}| \le q}} U_M = \sum_{k=|p-q|}^{p+q} \sum_{s=0}^{q \wedge k} \sum_{\substack{M \subseteq [m]: \\ |M| = k\,, \\ |M \cap \{n+1,\dots,m\}| = s}} U_M\,. \tag{5.7}$$

Since $K \cap \{n+1, \dots, m\} \subseteq J \Delta K \subseteq L \subseteq M$ and $M \cap \{n+1, \dots, m\} \subseteq K \cap \{n+1, \dots, m\}$ this also implies that we can restrict our attention to sets M and L that satisfy

$$M \cap \{n+1, \dots, m\} = L \cap \{n+1, \dots, m\} = K \cap \{n+1, \dots, m\}\,.$$

Writing $k := |M|$, $r := |J \cap K|$, $l := |L|$ and $s := |M \cap \{n+1, \dots, m\}|$ it follows that $r \le p \wedge (q-s)$ and, since

$$|J \cap K| = |J| + |K| - |J \cup K| = p + q - |J \cup K|\,, \tag{5.8}$$

it follows from $M \subseteq J \cup K$ that $r \le p+q-k$. Moreover, since $|J \cup K| = |J \cap K| + |J \Delta K|$ and $J \Delta K \subseteq L \subseteq M$, it follows again from (5.8) that $2r \ge p+q-l \ge p+q-k$. In particular, we have $l \ge |p-q| \vee (p+q-2r) = p+q-2r$. Moreover, note that

$$\begin{aligned}
|J \cap K \cap L| &= |L| - |L \cap (J \Delta K)| = |L| - |J \Delta K| = l - (p+q-2r) = 2r + l - p - q\,, \\
|(J \cap K) \setminus L| &= |(J \cap K)| - |J \cap K \cap L| = r - (2r + l - p - q) = p + q - r - l \quad \text{and} \\
|L \setminus (J \cap K)| &= |L| - |J \cap K \cap L| = l - (2r + l - p - q) = p + q - 2r\,.
\end{aligned}$$

Note that we have

$$\begin{aligned}
\mathbb{E}\big[W_J V_K \,\big|\, (X_i)_{i \in L}\big] &= \mathbb{E}\big[\psi(X_j, j \in J)\varphi(X_k, k \in K) \,\big|\, X_i, i \in L\big] \\
&= \big(\psi \star_r^{p+q-r-l} \varphi\big)\big((X_i)_{i \in L \cap J \cap K}, (X_j)_{j \in J \setminus K}, (X_k)_{k \in K \setminus J}\big)\,.
\end{aligned} \tag{5.9}$$

Let us now fix M and L. Then, for each $(A, B, C) \in \Pi_{r,n,m}(L)$, there are precisely

$$\binom{n-k+s}{p+q-r-k}$$

pairs $(J, K) \in \mathcal{D}_p(n) \times \mathcal{D}_q(m)$ such that $M \subseteq J \cup K$, $J \cap K \cap L = A$, $J \setminus K = B$ and $K \setminus J = C$. Indeed, given these restrictions it only remains to choose the elements of $(J \cap K) \setminus L \subseteq [n]$ in such a way that

$$M \setminus L \subseteq (J \cap K) \setminus L\,.$$

The claim now follows from the facts that $|M \cap [n]| = k - s$,

$$\big((J \cap K) \setminus L\big) \setminus (M \setminus L) = (J \cap K) \setminus M$$

and

$$\big|(J\cap K)\setminus L\big| - \big|M\setminus L\big| = p+q-r-l-(k-l) = p+q-r-k\,.$$

Thus we have proved that

$$\begin{aligned}
U_M &= \sum_{l=|p-q|}^{k}\sum_{\substack{L\subseteq M:\\ |L|=l}}(-1)^{|M|-|L|}\sum_{r=\lceil\frac{p+q-l}{2}\rceil}^{p\wedge(q-s)\wedge(p+q-k)}\binom{n-k+s}{p+q-r-k}\\
&\qquad\cdot\sum_{(A,B,C)\in\Pi_{r,n,m}(L)}\Big(\psi\star_r^{p+q-r-l}\varphi\Big)\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big)\\
&= \sum_{r=\lceil\frac{p+q-k}{2}\rceil}^{p\wedge(q-s)\wedge(p+q-k)}\binom{n-k+s}{p+q-r-k}\sum_{l=p+q-2r}^{k}\sum_{\substack{L\subseteq M:\\ |L|=l}}(-1)^{|M|-|L|}\\
&\qquad\sum_{(A,B,C)\in\Pi_{r,n,m}(L)}\Big(\psi\star_r^{p+q-r-l}\varphi\Big)\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big)\,,
\end{aligned}\tag{5.10}$$

Now, suppose that $(A,B,C)\in\Pi_{r,n,m}(M)$ such that, in particular, $|A| = 2r+k-p-q$. Moreover, suppose that

$$T_L(A,B,C) := \mathbb{E}\Big[\big(\psi\star_r^{p+q-r-k}\varphi\big)\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big)\,\Big|\,\mathcal{F}_L\Big]\neq 0\,.$$

Then, it is easy to see that $B\cup C\subseteq L$, that

$$T_L(A,B,C) = \big(\psi\star_r^{p+q-r-l}\varphi\big)\big((X_i)_{i\in A\cap L},(X_i)_{i\in B},(X_i)_{i\in C}\big)$$

and that $(A\cap L,B,C)\in\Pi_{r,n,m}(L)$. Moreover, for each given $(A,B,C)\in\Pi_{r,n,m}(L)$, there is a unique $(\hat A,B,C)\in\Pi_{r,n,m}(M)$ such that $(\hat A\cap L,B,C) = (A,B,C)$, namely one has to take $\hat A = A\cup(M\setminus L)$. From these observations we infer that

$$\sum_{(A,B,C)\in\Pi_{r,n,m}(L)}\Big(\psi\star_r^{p+q-r-l}\varphi\Big)\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big) = \sum_{(\hat A,B,C)\in\Pi_{r,n,m}(M)}T_L(\hat A,B,C)\,.\tag{5.11}$$

Now, recall that by the Hoeffding decomposition for non-symmetric kernels, for each $(A,B,C)\in\Pi_{r,n,m}(M)$ we have that

$$\begin{aligned}
&\big(\psi\star_r^{p+q-r-k}\varphi\big)_M\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big)\\
&= \sum_{L\subseteq M}(-1)^{|M|-|L|}\mathbb{E}\Big[\big(\psi\star_r^{p+q-r-k}\varphi\big)\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big)\,\Big|\,\mathcal{F}_L\Big]\\
&= \sum_{L\subseteq M}(-1)^{|M|-|L|}T_L(A,B,C)\,.
\end{aligned}\tag{5.12}$$

Thus, from (5.10), (5.11) and (5.12) we can conclude that

$$\begin{aligned}
U_M &= \sum_{r=\lceil\frac{p+q-k}{2}\rceil}^{p\wedge(q-s)\wedge(p+q-k)}\binom{n-k+s}{p+q-r-k}\sum_{l=p+q-2r}^{k}\sum_{\substack{L\subseteq M:\\ |L|=l}}(-1)^{|M|-|L|}\sum_{(A,B,C)\in\Pi_{r,n,m}(M)}T_L(A,B,C)\\
&= \sum_{r=\lceil\frac{p+q-k}{2}\rceil}^{p\wedge(q-s)\wedge(p+q-k)}\binom{n-k+s}{p+q-r-k}\sum_{(A,B,C)\in\Pi_{r,n,m}(M)}\sum_{L\subseteq M}(-1)^{|M|-|L|}T_L(A,B,C)\\
&= \sum_{r=\lceil\frac{p+q-k}{2}\rceil}^{p\wedge(q-s)\wedge(p+q-k)}\binom{n-k+s}{p+q-r-k}\sum_{(A,B,C)\in\Pi_{r,n,m}(M)}\big(\psi\star_r^{p+q-r-k}\varphi\big)_M\big((X_i)_{i\in A},(X_i)_{i\in B},(X_i)_{i\in C}\big)\,,
\end{aligned}\tag{5.13}$$

as claimed. The bound (5.5) then follows immediately from

$$\left\| \left(\psi \star_r^{p+q-r-k} \varphi\right)_M \right\|_{L^2(\mu^k)} \leq \left\| \psi \star_r^{p+q-r-k} \varphi \right\|_{L^2(\mu^k)}$$

and from the fact that

$$|\Pi_{r,n,m}(M)| = \binom{k-s}{2r+k-p-q, p-r, q-r-s}.$$

□

In the next subsection, we focus on convergence of finite-dimensional distributions (f.d.d.) for processes of the form (2.7). Our approach makes important use of the general (quantitative) CLTs from [DP19].

5.2 F.d.d. convergence

5.2.1 A general qualitative multivariate CLT

Fix a positive integer d and, for $1 \leq i \leq d$ and $n \in \mathbb{N}$, let $p_i \leq n_i = n_{i,n} \leq n$ be positive integers. We will always assume that the sequences $\{n_{i,n} : n \in \mathbb{N}\}$ diverge to ∞ as $n \to \infty$, for each $i = 1, \ldots, d$. Moreover, let $\psi^{(i)} = \psi^{(i,n)} \in L^4(\mu^{p_i})$ be degenerate kernels . Define

$$\varphi^{(i)} = \varphi^{(i,n)} := \frac{\psi^{(i)}}{\sqrt{\binom{n_i}{p_i}}}$$

as well as

$$\sigma_n(i)^2 := \mathrm{Var}\left(J_{p_i}^{(n_i)}(\varphi^{(i)})\right) = \|\psi^{(i)}\|^2_{L^2(\mu^{p_i})}.$$

For $i = 1, \ldots, d$ write $Y(i) := J_{p_i}^{(n_i)}(\varphi^{(i)})$ as well as

$$Y = Y_n := (Y_1, \ldots, Y_d)^T.$$

Then, Y is a centered random vector with components in $L^4(\mathbb{P})$. We will write $\mathbf{V} = \mathbf{V}_n = \{v_{i,k} : 1 \leq i, k \leq d\}$ for its covariance matrix. Throughout the section, we denote by $Z = Z_n = (Z_1, \ldots, Z_d)^T \sim N_d(0, \mathbf{V})$ a centered Gaussian vector with the same covariance matrix as Y. Note that, due to degeneracy, we have $v_{i,k} = 0$ unless $p_i = p_k$. The following finite-dimensional CLT is one of our crucial tools.

Proposition 5.3. *With the above notation and definitions, assume that* $\mathbf{C} := \lim_{n\to\infty} \mathbf{V}_n \in \mathbb{R}^{d\times d}$ *exists. Then,* Y_n *converges in distribution to* $N_d(0, \mathbf{C})$*, provided Conditions* (i)-(iii) *below hold for all* $1 \leq i \leq k \leq d$*:*

(i) $\lim_{n\to\infty} n^{a/2-r} \|\psi^{(i,n)} \star_r^{a-r} \psi^{(k,n)}\|_{L^2(\mu^{p_i+p_k-a})} = 0$ *for all pairs* (a, r) *of integers such that* $1 \leq a \leq \min(p_i + p_k - 1, 2(p_i \wedge p_k))$ *and* $\lceil \frac{a}{2} \rceil \leq r \leq a \wedge p_i \wedge p_k$*,*

(ii) $\lim_{n\to\infty} n^{a/2-r} \|\psi^{(i,n)}\|_{L^2(\mu^{p_i})} \|\psi^{(i,n)} \star_r^{a-r} \psi^{(i,n)}\|_{L^2(\mu^{2p_i-a})} = 0$ *for all for all pairs* (a, r) *of integers such that* $1 \leq a \leq 2p_i - 1$ *and* $\lceil \frac{a}{2} \rceil \leq r \leq a \wedge p_i$*, and*

(iii) $\lim_{n\to\infty} \dfrac{\|\psi^{(i,n)}\|^3_{L^2(\mu^{p_i})}}{\sqrt{n}} = 0.$

Proof. For $1 \leq i, k \leq d$, we use the notation

$$Y(i)Y(k) = \sum_{M\subseteq[n]:|M|\leq p_i+p_k} U_M(i,k)$$

to indicate the Hoeffding decomposition of $Y(i)Y(k)$. The following bound is taken from Lemma 4.1 in [DP19]: for $h \in C^3(\mathbb{R}^d)$ whose partial derivatives up to order three are all bounded, there exist constants

$\tilde{M}_2(h), M_3(h) \in (0,\infty)$ such that

$$\begin{aligned}\big|\mathbb{E}[h(Y)] - \mathbb{E}[h(Z)]\big| &\le \frac{1}{4p_1}\tilde{M}_2(h)\sum_{i,k=1}^{d}(p_i+p_k)\Big(\sum_{\substack{M\subseteq[n]:\\ |M|\le p_i+p_k-1}}\mathrm{Var}\big(U_M(i,k)\big)\Big)^{1/2}\\ &+\frac{2M_3(h)\sqrt{d}}{9p_1}\sum_{i=1}^{d}p_i\sigma_n(i)\Big(\sum_{\substack{M\subseteq[n]:\\ |M|\le 2p_i-1}}\mathrm{Var}\big(U_M(i,i)\big)\Big)^{1/2}\\ &+\frac{\sqrt{2d}M_3(h)}{9p_1\sqrt{n}}\sum_{i=1}^{d}p_i^{3/2}\sigma_n(i)^3\sqrt{\kappa_{p_i}}\,,\end{aligned}$$

and each finite constant κ_{p_i} only depends on p_i, $1\le i\le d$. We now apply Proposition 5.1 in order to bound

$$\Big(\sum_{\substack{M\subseteq[n]:\\ |M|\le p_i+p_k-1}}\mathrm{Var}\big(U_M(i,k)\big)\Big)^{1/2}$$

for $1\le i\le k\le d$. We will, for notational convenience, assume that $n_i\le n_k$. Moreover, for integers $p,q\ge 0$ we will write $M(p,q):=\min(2(p\wedge q),p+q-1)$. From (5.5) we know that for $M\subseteq[n_k]$ such that $|M|=p_i+p_k-a$ for some $a\in\{1,\dots,M(p_i,p_k)\}$
and $|M\cap\{n_i+1,\dots,n_k\}|=s\in\{0,1,\dots,p_k\wedge(p_i+p_k-a)\wedge(n_k-n_i)\}$ we have

$$\begin{aligned}\sqrt{\mathrm{Var}\big(U_M(i,k)\big)} &\le \sum_{r=\lceil\frac{a}{2}\rceil}^{a\wedge p_i\wedge(p_k-s)}\binom{n_i-p_i-p_k+a+s}{a-r}\binom{p_i+p_k-a}{p_i-r,p_k-r,2r-a}\|\varphi^{(i)}\star_r^{a-r}\varphi^{(k)}\|_{L^2(\mu^{p_i+p_k-a})}\\ &=: b_{i,k}(a,s)\,.\end{aligned}\tag{5.14}$$

Then, we have

$$\begin{aligned}\sum_{\substack{M\subseteq[n]:\\ |M|\le p_i+p_k-1}}\mathrm{Var}\big(U_M(i,k)\big) &= \sum_{a=1}^{M(p_i,p_k)}\sum_{s=0}^{p_k\wedge(p_i+p_k-a)\wedge(n_k-n_i)}\sum_{\substack{M\subseteq[n_k]:\\ |M|=p_i+p_k-a\,,\\ |M\cap\{n_i+1,\dots,n_k\}|=s}}\mathrm{Var}\big(U_M(i,k)\big)\\ &\le \sum_{a=1}^{M(p_i,p_k)}\sum_{s=0}^{p_k\wedge(p_i+p_k-a)\wedge(n_k-n_i)}\sum_{\substack{M\subseteq[n_k]:\\ |M|=p_i+p_k-a\,,\\ |M\cap\{n_i+1,\dots,n_k\}|=s}}b_{i,k}^2(a,s)\\ &= \sum_{a=1}^{M(p_i,p_k)}\sum_{s=0}^{p_k\wedge(p_i+p_k-a)\wedge(n_k-n_i)}\binom{n_k-n_i}{s}\binom{n_i}{p_i+p_k-a-s}b_{i,k}^2(a,s)\\ &\le \sum_{a=1}^{M(p_i,p_k)}\binom{n_k}{p_i+p_k-a}\sum_{s=0}^{p_k\wedge(p_i+p_k-a)\wedge(n_k-n_i)}b_{i,k}^2(a,s)\,.\end{aligned}$$

Now, writing

$$K(p_i,p_k,a,r):=\binom{p_i+p_k-a}{p_i-r,p_k-r,2r-a}$$

and using the inequality (5.14), we obtain

$$\begin{aligned}
&\Big(\sum_{\substack{M\subseteq[n]:\\ |M|\le p_i+p_k-1}} \mathrm{Var}\big(U_M(i,k)\big)\Big)^{1/2}\\
&\le \sum_{a=1}^{M(p_i,p_k)} \sqrt{\binom{n_k}{p_i+p_k-a}} \sum_{s=0}^{p_k\wedge(p_i+p_k-a)\wedge(n_k-n_i)} \sum_{r=\lceil \frac{a}{2}\rceil}^{a\wedge p_i\wedge(p_k-s)} \binom{n_i-p_i-p_k+a+s}{a-r}\\
&\qquad\qquad \cdot K(p_i,p_k,a,r)\|\varphi^{(i)}\star_r^{a-r}\varphi^{(k)}\|_{L^2(\mu^{p_i+p_k-a})}\\
&\le \frac{p_k+1}{\sqrt{\binom{n_i}{p_i}}\sqrt{\binom{n_k}{p_k}}} \sum_{a=1}^{M(p_i,p_k)} n_k^{(p_i+p_k-a)/2} \sum_{r=\lceil \frac{a}{2}\rceil}^{a\wedge p_i\wedge(p_k-s)} n_i^{a-r} K(p_i,p_k,a,r)\|\psi^{(i)}\star_r^{a-r}\psi^{(k)}\|_{L^2(\mu^{p_i+p_k-a})}\\
&\le C(p_i,p_k) \sum_{a=1}^{M(p_i,p_k)} n_k^{(p_i-a)/2} \sum_{r=\lceil \frac{a}{2}\rceil}^{a\wedge p_i\wedge p_k} n_i^{a-r-p_i/2}\|\psi^{(i)}\star_r^{a-r}\psi^{(k)}\|_{L^2(\mu^{p_i+p_k-a})},
\end{aligned}$$

where the finite constant $C(p_i,p_k)$ only depends on p_i and p_k. Since we also have $n_i, n_k \le n$ we can further bound

$$\begin{aligned}
&\Big(\sum_{\substack{M\subseteq[n]:\\ |M|\le p_i+p_k-1}} \mathrm{Var}\big(U_M(i,k)\big)\Big)^{1/2}\\
&\le C(p_i,p_k) \sum_{a=1}^{M(p_i,p_k)} \sum_{r=\lceil \frac{a}{2}\rceil}^{a\wedge p_i\wedge p_k} n^{a/2-r}\|\psi^{(i)}\star_r^{a-r}\psi^{(k)}\|_{L^2(\mu^{p_i+p_k-a})},
\end{aligned}$$

and the desired conclusion follows immediately. □

Remark 5.4. Note that the conditions (i)-(iii) in Proposition 5.3 are the same as those we would obtain in the case $n_{i,n} = n$ for $i = 1,\dots,d$. In particular they make sure that the vector

$$V := \big(J_{p_1}^{(n)}(\varphi^{(1)}),\dots,J_{p_d}^{(n)}(\varphi^{(d)})\big)^T$$

converges in distribution to $N_d(0,\Gamma)$, whenever $\Gamma := \lim_{n\to\infty}\mathbb{E}\big[VV^T\big]$ exists.

Remark 5.5. Let the integers $\{n_{i,n}\}$ and kernels $\{\varphi^{(i,n)}\}$ be defined as above. For each $i = 1,...,d$ and $n \ge 1$ consider sets of pairs of integers of the type

$$A(i,n) \subset \{(k_1,...,k_{p_i}) : 1 \le k_1 < ... < k_{p_i} \le n\},$$

and assume that, as $n\to\infty$ and for every $i = 1,...,d$, $|A(i,n)| \sim \binom{n_i}{p_i}$. For every $i = 1,...,d$ now set

$$H(i,n) := \sum_{(k_1,...,k_{p_i})\in A(i,n)} \varphi^{(i,n)}(X_{k_1},...,X_{k_{p_i}}),$$

and write $\mathbf{K}_n$, $n \ge 1$, to denote the covariance matrix of the vector $H_n = (H(1,n),....,H(d,n))$. Then, the proof of Proposition 5.3 can be straightforwardly adapted to show that, if $\mathbf{K}_n$ converges to a positive definite matrix $\mathbf{K}_\star$ and Conditions (i)–(iii) in Proposition 5.3 are veirified, then H_n converges in distribution to $Z \sim N_d(0,\mathbf{K}_\star)$. Such a conclusion plays a role in the proof of Theorem 3.10.

5.2.2 F.d.d. convergence for general symmetric U-processes

Let $\psi : E^p \to \mathbb{R}$ be a symmetric kernel of order p which is not necessarily degenerate and which might explicitly depend on the sample size n. Fix time points $0 \le t_1 < \dots < t_m \le 1$. Then, for each $j = 1,\dots,m$, the random variable $F_j := J_p^{(\lfloor nt_j\rfloor)}(\psi)$ has the Hoeffding decomposition

$$F_j = \mathbb{E}[F_j] + \sum_{k=1}^{p}\binom{\lfloor nt_j\rfloor - k}{p-k} J_k^{(\lfloor nt_j\rfloor)}(\psi_k),$$

where the symmetric and degenerate kernels $\psi_k : E^k \to \mathbb{R}$ of order k are given by

$$\psi_k(x_1, \dots, x_k) = \sum_{l=0}^{k} (-1)^{k-l} \sum_{1 \le i_1 < \dots < i_l \le k} g_l(x_{i_1}, \dots, x_{i_l})$$

and the symmetric functions $g_l : E^l \to \mathbb{R}$ are defined by

$$g_l(y_1, \dots, y_l) := \mathbb{E}\big[\psi(y_1, \dots, y_l, X_1, \dots, X_{p-l})\big] .$$

Without loss of generality, we can thus assume that $t_1 > 0$ and also that $0 < \|\psi\|_{L^2(\mu^p)} < +\infty$ which implies that

$$0 < \sigma_n^2 := \mathrm{Var}(J_p^{(n)}(\psi)) = \sum_{k=1}^{p} \binom{n-k}{p-k}^2 \binom{n}{k} \|\psi_k\|^2_{L^2(\mu^k)} < +\infty$$

for all $n \ge p$. We will further write

$$W_j := \frac{F_j - \mathbb{E}[F_j]}{\sigma_n}$$

for $j = 1, \dots, m$. Our goal is to use Proposition 5.3 in order to find conditions ensuring that the vector $W := W_n := (W_1, \dots, W_m)^T$ converges to some multivariate normal distribution $N_m(0, \mathbf{D})$, which requires in particular that the limit $\mathbf{D} := \lim_{n\to\infty} \mathbb{E}[WW^T] \in \mathbb{R}^{m\times m}$ exists. Let us write $d := mp$ and for $i = 1, \dots, d$ let $i = k_i p + s_i$, where $k_i \in \{0, 1, \dots, m-1\}$ and $s_i \in \{1, \dots, p\}$ as well as $n_i := \lfloor n t_{k_i+1} \rfloor$. Moreover, similarly as in [DP19, Section 5], we define

$$\varphi^{(i)} := \varphi^{(n,i)} := \frac{\binom{n_i - s_i}{p-s_i} \psi_{s_i}}{\sigma_n} \quad \text{and}$$
$$\psi^{(i)} := \psi^{(n,i)} := \sqrt{\binom{n_i}{s_i}} \varphi^{(i)} .$$

With this notation at hand, we define the random vector $Y := (Y_1, \dots, Y_d)^T$, where $Y_i := J_{s_i}^{(n_i)}(\varphi^{(i)})$, $1 \le i \le d$. In this way, our notation is fitted to the framework of Subsection 5.2.1. We are going to reformulate the conditions from Proposition 5.3. First note that $\mathbb{E}[Y_i Y_j] = 0$ whenever $s_i \ne s_j$, due to the degeneracy of the involved kernels. On the other hand, if $s_i = s_j = s$, then

$$\mathbb{E}[Y_i Y_j] = \frac{\binom{n_i \wedge n_j}{s} \binom{n_i - s}{p-s} \binom{n_j - s}{p-s}}{\sigma_n^2} \|\psi_s\|^2_{L^2(\mu^s)} .$$

Since $n_i = \lfloor n t_i \rfloor$ for $1 \le i \le d$, the covariance matrix of Y thus converges to some limit $\Gamma \in \mathbb{R}^{d\times d}$ if and only if the real limit

$$\lim_{n\to\infty} \frac{n^{2p-s}}{\sigma_n^2} \|\psi_s\|^2_{L^2(\mu^s)} \tag{5.15}$$

exists for $s = 1, \dots, p$. Moreover, for $1 \le i, k \le d$ we have

$$n^{a/2-r} \|\psi^{(i,n)} \star_r^{a-r} \psi^{(k,n)}\|_{L^2(\mu^{s_i+s_k-a})} = n^{a/2-r} \frac{\sqrt{\binom{n_i}{s_i}} \sqrt{\binom{n_k}{s_k}} \binom{n_i - s_i}{p-s_i} \binom{n_k - s_k}{p-s_k}}{\sigma_n^2} \|\psi_{s_i} \star_r^{a-r} \psi_{s_k}\|_{L^2(\mu^{s_i+s_k-a})} .$$

Thus, $\lim_{n\to\infty} n^{a/2-r} \|\psi^{(i,n)} \star_r^{a-r} \psi^{(k,n)}\|_{L^2(\mu^{s_i+s_k-a})} = 0$ if and only if

$$\lim_{n\to\infty} \frac{n^{2p + \frac{a-s_i-s_k}{2} - r}}{\sigma_n^2} \|\psi_{s_i} \star_r^{a-r} \psi_{s_k}\|_{L^2(\mu^{s_i+s_k-a})} = 0 .$$

Further, for $i = 1, \dots, d$, we have

$$\frac{\|\psi^{(i,n)}\|^3_{L^2(\mu^{s_i})}}{\sqrt{n}} = \frac{\binom{n_i}{s_i}^{3/2} \binom{n_i - s_i}{p-s_i}^{3/2}}{\sqrt{n}\sigma_n^3} \|\psi_{s_i}\|^3_{L^2(\mu^{s_i})} .$$

Hence, $\lim_{n\to\infty} \frac{\|\psi^{(i,n)}\|^3_{L^2(\mu^{s_i})}}{\sqrt{n}} = 0$ if and only if

$$\lim_{n\to\infty} \frac{n^{3p-\frac{3s_i+1}{2}}}{\sigma_n^3}\|\psi_{s_i}\|^3_{L^2(\mu^{s_i})} = \left(\lim_{n\to\infty} n^{-1/3}\frac{n^{2p-s_i}}{\sigma_n^2}\|\psi_{s_i}\|^2_{L^2(\mu^{s_i})}\right)^{3/2} = 0\,,$$

which is implied by (5.15). Taking into account that, for $j = 1,\dots,m$,

$$1 = \mathrm{Var}\big(J_p^{(n)}(\psi/\sigma_n)\big) \geq \mathrm{Var}(W_j) = \sum_{i=(j-1)p+1}^{jp} \|\psi^{(i,n)}\|^2_{L^2(\mu^{s_i})}$$

and that $(W_1,\dots,W_m)^T$ is obtained from $(Y_1,\dots,Y_d)^T$ by applying a linear functional, from Proposition 5.3 we thus deduce the following result. Note that we also apply the reindexing $l := a - r$.

Theorem 5.6. *With the above notation and definitions, the vector $(W_1,\dots,W_m)^T$ converges, as $n\to\infty$, to a multivariate normal distribution, whenever the following conditions hold for all $1 \leq v \leq u \leq p$:*

(a) *The real limit $\lim_{n\to\infty} \frac{n^{2p-v}}{\sigma_n^2}\|\psi_v\|^2_{L^2(\mu^v)}$ does exist and*

(b) $\lim_{n\to\infty} \frac{n^{2p-\frac{u+v+r-l}{2}}}{\sigma_n^2}\|\psi_v \star_r^l \psi_u\|_{L^2(\mu^{v+u-r-l})} = 0$ *for all pairs (l,r) of integers such that $1 \leq r \leq v$ and $0 \leq l \leq r \wedge (u+v-r-1)$.*

Due to the complicated expressions of the kernels ψ_s, the following result, which is a rectified version of Lemma 5.1 of [DP19], is often useful for bounding the contraction norms appearing in Theorem 5.6. Recall the definition of the set $Q(i,k,r,l)$ defined before Theorem 3.1.

Lemma 5.7. *For positive integers $1 \leq r,i,k \leq p$ and $0 \leq l \leq p$ such that $0 \leq l \leq r \leq i \wedge k$ there exists a constant $K(i,k,r,l) \in (0,\infty)$ which only depends on i,k,r and l such that*

$$\|\psi_i \star_r^l \psi_k\|_{L^2(\mu^{i+k-r-l})} \leq K(i,k,r,l) \max_{(j,m,a,b)\in Q(i,k,r,l)} \|g_j \star_a^b g_m\|_{L^2(\mu^{j+m-a-b})}\,.$$

Proof of Lemma 5.7. Recall that we have

$$\begin{aligned}
&\big(\psi_i \star_r^l \psi_k\big)(x_1,\dots,x_{k+i-2r},y_{l+1},\dots,y_r)\\
&= \int_{E^l} \psi_i(y_1,\dots,y_r,x_1,\dots,x_{i-r})\\
&\qquad\qquad \cdot \psi_k(y_1,\dots,y_r,x_{i-r+1},\dots,x_{k+i-2r})d\mu^{\otimes l}(y_1,\dots,y_l)\,. \qquad (5.16)
\end{aligned}$$

Recalling also the expression (2.3) of the kernels ψ_i and ψ_k, respectively, and taking into account that μ is a probability measure as well as that, for $k \geq s$, the functions g_k from (2.4) satsify

$$\int_{E^{k-s}} g_k(x_1,\dots,x_k)d\mu^{\otimes k-s}(x_{s+1},\dots,x_k) = g_s(x_1,\dots,x_s) \qquad (5.17)$$

by virtue of Fubini's theorem, we see that $\big(\psi_i \star_r^l \psi_k\big)(x_1,\dots,x_{k+i-2r},y_{l+1},\dots,y_r)$ is a linear combination, with coefficients only depending on i,k,r and l but not on n, of expressions of the form

$$\begin{aligned}
&G_{(a,b,j,m)}(x_{i_1},\dots,x_{i_{j-b-d}},y_{q_1},\dots,y_{q_c},x_{k_1},\dots,x_{k_{m-b-e}})\\
&:= \int_{E^t} g_j(u_1,\dots,u_b,y_{m_1},\dots,y_{m_d},x_{i_1},\dots,x_{i_{j-b-d}})\\
&\qquad\qquad \cdot g_m(u_1,\dots,u_b,y_{n_1},\dots,y_{n_e},x_{k_1},\dots,x_{k_{m-b-e}})d\mu^{\otimes b}(u_1,\dots,u_b)\\
&= \big(g_j \star_a^b g_m\big)(x_{i_1},\dots,x_{i_{j-b-d}},y_{q_1},\dots,y_{q_c},x_{k_1},\dots,x_{k_{m-b-e}})\,,
\end{aligned}$$

where $0 \leq j \leq i$, $0 \leq m \leq k$, $0 \leq b \leq l$, $0 \leq b \leq a \leq r$, $1 \leq i_1 < \ldots < i_{j-b-d} \leq i-r$, $i-r+1 \leq k_1 < \ldots < k_{l-b-e} \leq k+i-2r$ such that, in particular, the sets $\{i_1,\dots,i_{j-t-a}\}$ and $\{k_1,\dots,k_{m-t-b}\}$ are disjoint. Furthermore, we have $l+1 \leq m_1 < \ldots < m_d \leq r$, $l+1 \leq n_1 < \ldots < n_e \leq r$,

$l+1 \le q_1 < \ldots < q_c \le r$ such that $\{q_1,\ldots,q_c\} = \{m_1,\ldots,m_d\} \cup \{n_1,\ldots,n_e\}$, $d \le j-b$, $e \le m-b$ and $a := b + |\{m_1,\ldots,m_d\} \cap \{n_1,\ldots,n_e\}| \le b+c$. Note that $c \le r-l$ and, hence, also $a-b \le c \le r-l$ as well as $a \le (b+d) \wedge (b+e) \le j \wedge m$. Moreover, the number $j+m-a-b$ of arguments of the function $g_j \star_a^b g_m$ is at most as large as the number $i+k-r-l$ of arguments of the function $\psi_i \star_r^l \psi_k$. Finally, if $j=m=p$, then $i=k=p$ and $g_j = g_m = \psi$. This also implies that $b=l$ and $a=r$. Hence, we conclude that $(j,m,a,b) \in Q(i,k,r,l)$.

Now, using the fact that μ is a probability measure, we obtain that

$$\begin{aligned}\int_{E^{k+i-r-l}} & G^2_{(a,b,j,m)}(x_{i_1},\ldots,x_{i_{j-b-d}},y_{q_1},\ldots,y_{q_c},x_{k_1},\ldots,x_{k_{m-b-e}}) \\ & \qquad d\mu^{\otimes i+k-l-r}(x_1,\ldots,x_{k+i-2r},y_{l+1},\ldots,y_r) \\ &= \int_{E^{j+m-a-b}} \big(g_j \star_a^b g_m\big)^2 d\mu^{\otimes j+m-a-b} = \|g_j \star_a^b g_m\|^2_{L^2(\mu^{\otimes j+m-a-b})}\,. \end{aligned} \tag{5.18}$$

Since $\psi_i \star_r^l \psi_k$ is a finite linear combination with coefficients depending uniquely on i,k,r and l of the $G_{(a,b,j,m)}$, the claim thus follows from (5.18) and Minkowski's inequality. □

The next result is a direct consequence of Theorem 5.6, of Lemma 5.7 and of the fact that, for all $1 \le v \le p$, we have

$$\|\psi_v\|^2_{L^2(\mu^v)} = \sum_{l=1}^{v} \binom{v}{l} (-1)^{v-l} \|\hat{g}_l\|^2_{L^2(\mu^l)}\,,$$

where $\hat{g}_l := g_l - \mathbb{E}[\psi(X_1,\ldots,X_p)]\,, 1 \le l \le p$ (see e.g. [Vit92, Theorem 4.3]).

Corollary 5.8. *With the above notation and definitions, the vector $(W_1,\ldots,W_m)^T$ converges, as $n \to \infty$, to a multivariate normal distribution if the following conditions hold for all $1 \le v \le u \le p$:*

(a) *the real limit $\lim_{n\to\infty} \frac{n^{2p-v}}{\sigma_n^2}\big(\|g_v\|^2_{L^2(\mu^v)} - (\mathbb{E}[\psi(X_1,\ldots,X_p)])^2\big)$ exists, and*

(b) $\lim\limits_{n\to\infty} \dfrac{n^{2p-\frac{u+v+r-l}{2}}}{\sigma_n^2} \|g_j \star_i^k g_m\|_{L^2(\mu^{j+m-i-k})} = 0$ *for all pairs (l,r) and quadruples (j,m,i,k) of integers such that $1 \le r \le v$, $0 \le l \le r \wedge (u+v-r-1)$ and $(j,m,i,k) \in Q(v,u,r,l)$.*

Remark 5.9. (i) The conditions given in Theorem 5.6 and in Corollary 5.8 do not depend on the finite sequence $0 < t_1 < \ldots < t_m \le 1$ used to define the vector $(W_1,\ldots,W_m)^T$. Hence, both statements yield sufficient conditions for f.d.d. convergence of the sequence $(W_n(t))_{t\in[0,1]}$ of processes.

(ii) Note that the respective conditions (b) of Theorem 5.6 and of Corollary 5.8 are the same as those we would get to obtain the asymptotic normality of the single U-statistic $J_p^{(n)}(\psi/\sigma_n)$ (see [DP19, Section 5]).

(iii) It is easy to see from the following computation that the respective conditions (a) in Theorem 5.6 and in Corollary 5.8 imply that the covariance function of W_n converges pointwise to an explicit limit. Such a condition was not necessary in the univariate case dealt with in [DP19, Section 5] since there the U-statistic could simply be normalized to have variance one. For $s,t \in [0,1]$ we have

$$\begin{aligned}\mathrm{Cov}\big(W_n(s),W_n(t)\big) &= \mathbb{E}\big[W_n(s)W_n(t)\big] = \sigma_n^{-2} \sum_{k,l=1}^{p} \binom{\lfloor ns\rfloor - k}{p-k}\binom{\lfloor nt\rfloor - l}{p-l} \mathbb{E}\big[J_k^{(\lfloor ns\rfloor)}(\psi_k) J_l^{(\lfloor nt\rfloor)}(\psi_l)\big] \\ &= \sigma_n^{-2} \sum_{k=1}^{p} \binom{\lfloor ns\rfloor - k}{p-k}\binom{\lfloor nt\rfloor - k}{p-k}\binom{\lfloor ns\rfloor \wedge \lfloor nt\rfloor}{k} \|\psi_k\|^2_{L^2(\mu^k)} \\ &\sim \sigma_n^{-2} \sum_{k=1}^{p} \frac{\lfloor ns\rfloor^{p-k}}{(p-k)!}\frac{\lfloor nt\rfloor^{p-k}}{(p-k)!}\frac{\big(\lfloor ns\rfloor \wedge \lfloor nt\rfloor\big)^k}{k!} \|\psi_k\|^2_{L^2(\mu^k)} \\ &\sim \sum_{k=1}^{p} \frac{(s\wedge t)^p (s\vee t)^{p-k}}{k!(p-k)!(p-k)!}\frac{n^{2p-k}}{\sigma_n^2} \|\psi_k\|^2_{L^2(\mu^k)}\,. \end{aligned}$$

Note further that, using (2.8), we can conclude that for fixed $s, t \in [0,1]$, the sequence $\mathrm{Cov}\big(W_n(s), W_n(t)\big)$, $n \in \mathbb{N}$, is always bounded.

5.3 Criteria for tightness

We are going to establish tightness using Lemma 1.1 and, as a result, obtain the following theorem:

Theorem 5.10 (**Tightness of general U-processes**). *Let $p \in \mathbb{N}$ and suppose that $\psi = \psi(n) \in L^4(\mu^p)$, $n \geq p$, is a sequence of symmetric kernels. For $t \in [0,1]$ let $U(t) := U_n(t) := J_p^{(\lfloor nt \rfloor)}(\psi)$ and define $W(t) = W_n(t)$ by (2.7), where $\sigma_n^2 := \mathrm{Var}(U_n(1)) = \mathrm{Var}(J_p^{(n)}(\psi))$. Suppose that there is an a.s. continuous Gaussian process $Z = (Z(t))_{t\in[0,1]}$ such that the finite-dimensional distributions of W_n, $n \in \mathbb{N}$, converge to those of Z. Consider the following conditions:*

(i) *There is an $\epsilon > 0$ such that for all $1 \leq r \leq p$ and all $0 \leq l \leq r-1$, the sequence*

$$\frac{n^{2p-r-\frac{r-l}{2}+\epsilon}}{\sigma_n^2} \big\| \psi_r \star_r^l \psi_r \big\|_{L^2(\mu^{r-l})}$$

is bounded.

(ii) *There is an $\epsilon > 0$ such that for all $1 \leq r \leq p$, all $0 \leq l \leq r-1$ and for all quadruples $(j,m,a,b) \in Q(r,r,r,l)$ the sequence*

$$\frac{n^{2p-r-\frac{r-l}{2}+\epsilon}}{\sigma_n^2} \| g_j \star_a^b g_m \|_{L^2(\mu^{j+m-a-b})}$$

is bounded.

Then, one has that (ii) $\to$ (i), *and also that* (i) *is sufficient in order for the sequence W_n, $n \in \mathbb{N}$, to be tight in $D[0,1]$.*

The proof of Theorem 5.10 is detailed in the forthcoming Subsections 5.3.1 and 5.3.2. There, we are however not going to establish (1.6) directly, but will show that there is a finite constant $C_1 > 0$ such that, under the assumptions of Theorem 5.10, for all $n \in \mathbb{N}$ and all $0 \leq s \leq t \leq 1$ we have the inequality

$$\mathbb{E}\big|W_n(t) - W_n(s)\big|^4 \leq C_1 \left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \right)^{1+\epsilon}, \tag{5.19}$$

where ϵ is the same as in the statement of Theorem 5.10. This is sufficient by Lemma 1.1.

5.3.1 Proof of Theorem 5.10, I: degenerate kernels

Throughout the present and subsequent section, we can assume without loss of generality that $\epsilon \in (0,1]$ Let us first assume that W_n is a U-process of order p based on a degenerate kernel φ, i.e. for $0 \leq t \leq 1$ we have

$$W(t) := W_n(t) := J_p^{(\lfloor nt \rfloor)}(\varphi)\,.$$

For $0 \leq s \leq t \leq 1$ let $I(n,s,t) := \{\lfloor ns \rfloor + 1, \ldots \lfloor nt \rfloor\}$ and for $J \in \mathcal{D}_p(\lfloor nt \rfloor)$ write

$$V_J := \mathbf{1}_{\{J \cap I(n,s,t) \neq \emptyset\}} \varphi(X_i, i \in J)\,.$$

Then we have that

$$W_n(t) - W_n(s) = \sum_{J \in \mathcal{D}_p(\lfloor nt \rfloor)} V_J$$

is a degenerate (non-symmetric) U-statistic of order p, based on $X_1, \dots, X_{\lfloor nt \rfloor}$. In [IS02, Theorem 2], the following bounds are given:

$$
\begin{aligned}
&A \max_{k=0,\dots,p} \max_{1\le j_1<\dots<j_k\le p} \sum_{1\le i_{j_1}<\dots<i_{j_k}\le \lfloor nt\rfloor} \mathbb{E}\Bigg[\Bigg(\sum_{\substack{i_s:1\le s\le p,\\ s\notin\{j_1,\dots,j_k\},\\ 1\le i_1<\dots<i_p\le\lfloor nt\rfloor}} \mathbb{E}\big[V^2_{\{i_1,\dots,i_p\}} \,\big|\, X_{i_{j_1}},\dots,X_{i_{j_1}}\big]\Bigg)^2\Bigg] \\
&\le \mathbb{E}\big|W_n(t)-W_n(s)\big|^4 \qquad (5.20)\\
&\le B \max_{k=0,\dots,p} \max_{1\le j_1<\dots<j_k\le p} \sum_{1\le i_{j_1}<\dots<i_{j_k}\le \lfloor nt\rfloor} \mathbb{E}\Bigg[\Bigg(\sum_{\substack{i_s:1\le s\le p,\\ s\notin\{j_1,\dots,j_k\},\\ 1\le i_1<\dots<i_p\le\lfloor nt\rfloor}} \mathbb{E}\big[V^2_{\{i_1,\dots,i_p\}} \,\big|\, X_{i_{j_1}},\dots,X_{i_{j_1}}\big]\Bigg)^2\Bigg],
\end{aligned}
$$

where A and B are finite constants which only depend on p. We will now make use of the upper bound in (5.20). Note that for each $k = 0,\dots,p$, each labelling $1 \le j_1 < \dots < j_k \le p$ and all choices $1 \le i_{j_1} < \dots < i_{j_k} \le \lfloor nt \rfloor$ we have that

$$
\begin{aligned}
&\mathbb{E}\Bigg[\Bigg(\sum_{\substack{i_s:1\le s\le p,\\ s\notin\{j_1,\dots,j_k\},\\ 1\le i_1<\dots<i_p\le\lfloor nt\rfloor}} \mathbb{E}\big[V^2_{\{i_1,\dots,i_p\}} \,\big|\, X_{i_{j_1}},\dots,X_{i_{j_k}}\big]\Bigg)^2\Bigg] \\
&\le \mathbb{E}\Bigg[\Bigg(\sum_{\substack{L\in\mathcal{D}_{p-k}(\lfloor nt\rfloor):\\ L\cap\{i_{j_1},\dots,i_{j_k}\}=\emptyset}} \mathbb{E}\big[V^2_{\{i_{j_1},\dots,i_{j_k}\}\cup L} \,\big|\, X_{i_{j_1}},\dots,X_{i_{j_k}}\big]\Bigg)^2\Bigg] \qquad (5.21)
\end{aligned}
$$

and, that for each set $L \in \mathcal{D}_{p-k}(\lfloor nt \rfloor)$ such that $L \cap \{i_{j_1},\dots,i_{j_k}\} = \emptyset$ and $\big(L \cup \{i_{j_1},\dots,i_{j_k}\}\big) \cap I(n,s,t) \ne \emptyset$ we have

$$
\mathbb{E}\big[V^2_{\{i_{j_1},\dots,i_{j_k}\}\cup L} \,\big|\, X_{i_{j_1}},\dots,X_{i_{j_1}}\big] = \big(\varphi \star_p^{p-k} \varphi\big)\big(X_{i_{j_1}},\dots,X_{i_{j_k}}\big).
$$

Thus, we can further bound

$$
\begin{aligned}
&\mathbb{E}\big|W_n(t)-W_n(s)\big|^4 \le B \max_{k=0,\dots,p} \sum_{K\in\mathcal{D}_k(\lfloor nt\rfloor)} \mathbb{E}\Bigg[\Bigg(\sum_{\substack{L\in\mathcal{D}_{p-k}(\lfloor nt\rfloor):\\ L\cap K=\emptyset,\\ (L\cup K)\cap I(n,s,t)\ne\emptyset}} \mathbb{E}\big[V^2_{K\cup L} \,\big|\, \mathcal{F}_K\big]\Bigg)^2\Bigg] \qquad (5.22)\\
&= B \max_{k=0,\dots,p}\Bigg(\sum_{\substack{K\in\mathcal{D}_k(\lfloor nt\rfloor):\\ K\cap I(n,s,t)=\emptyset}} \mathbb{E}\Bigg[\Bigg(\binom{\lfloor nt\rfloor-k}{p-k}-\binom{\lfloor ns\rfloor-k}{p-k}\Bigg)^2 \big(\varphi\star_p^{p-k}\varphi\big)^2(X_i, i\in K)\Bigg] \\
&\qquad + \sum_{\substack{K\in\mathcal{D}_k(\lfloor nt\rfloor):\\ K\cap I(n,s,t)\ne\emptyset}} \mathbb{E}\Bigg[\binom{\lfloor nt\rfloor-k}{p-k}^2 \big(\varphi\star_p^{p-k}\varphi\big)^2(X_i, i\in K)\Bigg]\Bigg) \\
&\le B \max_{k=0,\dots,p-1}\binom{\lfloor ns\rfloor}{k}\Bigg(\binom{\lfloor nt\rfloor-k}{p-k}-\binom{\lfloor ns\rfloor-k}{p-k}\Bigg)^2 \mathbb{E}\Big[\big(\varphi\star_p^{p-k}\varphi\big)^2(X_i, i\in K)\Big] \\
&\quad + B \max_{k=1,\dots,p}\Bigg(\binom{\lfloor nt\rfloor}{k}-\binom{\lfloor ns\rfloor}{k}\Bigg)\binom{\lfloor nt\rfloor-k}{p-k}^2 \mathbb{E}\Big[\big(\varphi\star_p^{p-k}\varphi\big)^2(X_i, i\in K)\Big] \\
&= B \max_{k=0,\dots,p-1}\binom{n-k}{p-k}^2\binom{\lfloor ns\rfloor}{k}\binom{n-k}{p-k}^{-2}\Bigg(\binom{\lfloor nt\rfloor-k}{p-k}-\binom{\lfloor ns\rfloor-k}{p-k}\Bigg)^2 \big\|\varphi\star_p^{p-k}\varphi\big\|^2_{L^2(\mu^k)} \\
&\quad + B \max_{k=1,\dots,p}\binom{n}{k}\binom{n}{k}^{-1}\Bigg(\binom{\lfloor nt\rfloor}{k}-\binom{\lfloor ns\rfloor}{k}\Bigg)\binom{\lfloor nt\rfloor-k}{p-k}^2 \big\|\varphi\star_p^{p-k}\varphi\big\|^2_{L^2(\mu^k)} \\
&\lesssim \max_{k=0,\dots,p-1}\Bigg(\frac{\lfloor nt\rfloor-\lfloor ns\rfloor}{n}\Bigg)^2 n^{2p-k}\big\|\varphi\star_p^{p-k}\varphi\big\|^2_{L^2(\mu^k)}
\end{aligned}
$$

$$
\begin{aligned}
&+ \max_{k=1,\dots,p} \frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} n^{2p-k} \left\| \varphi \star_p^{p-k} \varphi \right\|^2_{L^2(\mu^k)} \\
&\lesssim \left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \right)^2 n^{2p} \|\varphi\|^4_{L^2(\mu^p)} + \frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \max_{k=1,\dots,p} n^{2p-k} \left\| \varphi \star_p^{p-k} \varphi \right\|^2_{L^2(\mu^k)} \\
&\lesssim \left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \right)^{1+\epsilon} \left(n^{2p} \|\varphi\|^4_{L^2(\mu^p)} + \max_{k=1,\dots,p} n^{2p-k+\epsilon} \left\| \varphi \star_p^{p-k} \varphi \right\|^2_{L^2(\mu^k)} \right), \qquad (5.23)
\end{aligned}
$$

for each $\epsilon \in (0,1]$. Now, if, for some $\epsilon \in (0,1]$, there is a $C_1 = C_1(\epsilon) \in (0,\infty)$ such that for all $n \in \mathbb{N}$ we have

$$
n^{2p} \|\varphi\|^4_{L^2(\mu^p)} + \max_{k=1,\dots,p} n^{2p-k+\epsilon} \left\| \varphi \star_p^{p-k} \varphi \right\|^2_{L^2(\mu^k)} \leq C_1 \qquad (5.24)
$$

we conclude from (5.23) that (5.19) is satisfied. This concludes the argument in the case of degenerate kernels.

Remark 5.11. Incidentally, one can show that inequality (5.22) also holds in the opposite direction when the constant B appearing there is replaced with a small enough positive constant C, which only depends on p. Our way of bounding $\mathbb{E}|W_n(t) - W_n(s)|^4$ is therefore optimal with respect to the order in n.

5.3.2 Proof of Theorem 5.10, II: general kernels

For $t \in [0,1]$ recall the definition (2.7) of $W_n(t)$ and (2.8) of σ_n^2. Then, defining for $1 \leq r \leq p$,

$$
\varphi^{(r)} := \varphi^{(n,r)} := \frac{\binom{n-r}{p-r} \psi_r}{\sigma_n},
$$

the $\varphi^{(r)}$ are degenerate kernels and, with the notation $V_r(t) := J_r^{(\lfloor nt \rfloor)}(\varphi^{(r)})$, we have

$$
W(t) = \sum_{r=1}^{p} \frac{\binom{\lfloor nt \rfloor - r}{p-r}}{\binom{n-r}{p-r}} V_r(t).
$$

Since

$$
\lim_{n\to\infty} \frac{\binom{\lfloor nt \rfloor - r}{p-r}}{\binom{n-r}{p-r}} = t^r
$$

for each $r = 1, \dots, p$, by an application of Prohorov's theorem, it follows that the sequence $(W_n(t))_{t\in[0,1]}$, $n \in \mathbb{N}$, is tight whenever $(V_r(t))_{t\in[0,1]}$ is tight for every $r = 1, \dots, p$.

Sufficient conditions for this to hold were given in the previous subsection. Indeed, $(V_r(t))_{t\in[0,1]}$ is tight if for some $\epsilon \in (0,1]$ the sequence

$$
n^{2r} \|\varphi^{(r)}\|^4_{L^2(\mu^r)} + \max_{k=1,\dots,r} n^{2r-k+\epsilon} \left\| \varphi^{(r)} \star_r^{r-k} \varphi^{(r)} \right\|^2_{L^2(\mu^r)} \qquad (5.25)
$$

is bounded from above by a constant. Now, first note that from (2.8) we see that there is a finite constant L_r such that

$$
n^{2r} \|\varphi^{(r)}\|^4_{L^2(\mu^r)} = n^{2r} \binom{n-r}{p-r}^4 \frac{\|\psi_r\|^4_{L^2(\mu^r)}}{\sigma_n^4} \leq L_r \qquad (5.26)
$$

for all $n \in \mathbb{N}$. Moreover, for $k = 1, \dots, r$ we have that

$$
\begin{aligned}
n^{2r-k+\epsilon} \left\| \varphi^{(r)} \star_r^{r-k} \varphi^{(r)} \right\|^2_{L^2(\mu^r)} &= n^{2r-k+\epsilon} \binom{n-r}{p-r}^4 \frac{\left\| \psi_r \star_r^{r-k} \psi_r \right\|^2_{L^2(\mu^r)}}{\sigma_n^4} \\
&\leq D_r n^{4p-2r-k+\epsilon} \frac{\left\| \psi_r \star_r^{r-k} \psi_r \right\|^2_{L^2(\mu^r)}}{\sigma_n^4},
\end{aligned}
$$

where D_r is a finite constant only depending on r. The proof is concluded by letting $l := r - k$ and by applying Lemma 5.7. □

5.4 Proofs of Theorem 3.1, Theorem 3.4 and Corollary 3.7

Proof of Theorem 3.1. It is well-known that f.d.d. convergence and tightness together imply weak convergence in $D[0,1]$ (see e.g. [Bil99, Section 13]). F.d.d. convergence and Gaussianity of the limit process Z now follow from Corollary 5.8. Moreover, the needed formula for the covariance Γ of Z follows from Remark 5.9 (iii) as

$$\begin{aligned}\lim_{n\to\infty}\mathrm{Cov}\big(W_n(s),W_n(t)\big) &= \lim_{n\to\infty}\sigma_n^{-2}\sum_{k=1}^{p}\frac{(s\wedge t)^p n^{2p-k}(s\vee t)^{p-k}}{k!(p-k)!(p-k)!}\|\psi_k\|^2_{L^2(\mu^k)}\\ &= \sum_{k=1}^{p}\frac{(s\wedge t)^p(s\vee t)^{p-k}}{k!(p-k)!(p-k)!}\sum_{l=1}^{k}\binom{k}{l}(-1)^{k-l}\lim_{n\to\infty}\frac{n^{2p-k}}{\sigma_n^2}\left(\|g_l\|^2_{L^2(\mu^l)}-\left(\int_E\psi d\mu\right)^2\right)\\ &= \sum_{k=1}^{p}\frac{(s\wedge t)^p(s\vee t)^{p-k}}{k!(p-k)!(p-k)!}b_k^2\,.\end{aligned}$$

Since this is the same covariance function as that of the process given in (1.1) (with $\alpha^2_{k,p}$ as given in the statement) we conclude that the limiting process Z may be chosen to have a.s. continuous paths. Now, tightness is implied by Theorem 5.10-(ii). □

Proof of Theorem 3.4. The proof is similar to that of Theorem 3.1 and follows from Theorems 5.6 and Theorem 5.10-(i). □

Proof of Corollary 3.7. We directly use Theorem 3.1. In this case, we have $g_k = 0$ for all $0 \le k \le p-1$ and $g_p = \psi$. Moreover,

$$\sigma_n^2 = \binom{n}{p}\|\psi\|^2_{L^2(\mu^p)} \sim \frac{n^p}{p!}\|\psi\|^2_{L^2(\mu^p)}\,.$$

Hence, we have $b_k^2 = 0$ for all $1 \le k \le p-1$ and $b_p^2 = p!$. From this, we obtain that $\Gamma(s,t) = (s\wedge t)^p$, implying the result. □

Outline of the proof of Theorem 3.12. Finite-dimensional convergence can again be proved by means of Proposition 5.3 with the dimension d appearing there given by $m\sum_{j=1}^d p_j$, where m is the number of points $0 \le t_1 < t_2 < \ldots < t_m \le 1$ under consideration. The computation in the proof of Theorem 3.1 can be easily generalized to yield the claimed limiting covariance structure. Hence, as already pointed out, the vector-valued limiting process Z is an element of $C([0,1];\mathbb{R}^d)$. Next, computations very similar to those leading to Theorem 5.6 yield Conditions (a') and (b'). Finally, Condition (c') is obtained by observing that the laws of the family $\{W^{(n)} : n \ge 1\}$ are tight whenever those of $\{W^{(n)}(i) : n \ge 1\}$ are tight for each $1 \le i \le d$; this last point is established by means of Theorem 5.10. □

Proof of Theorem 3.11. From the expresssion (2.3) it is a simple combinatorial exercise to deduce that

$$\begin{aligned}\langle\psi_k(i),\psi_k(j)\rangle_{L^2(\mu^k)} &= \sum_{q=0}^{k}(-1)^{k-q}\binom{k}{q}\langle g_q(i),g_q(j)\rangle_{L^2(\mu^q)}\\ &= \sum_{q=1}^{k}(-1)^{k-q}\binom{k}{q}\langle \hat g_q(i),\hat g_q(j)\rangle_{L^2(\mu^q)}\,,\end{aligned}$$

where $\hat g_q(i) = g_q(i) - \mathbb{E}[\psi(i)(X_1,\ldots,X_{p_i})]$. Thus, Condition (a) is yielded by Condition (a') of Theorem 3.12. Finally, Conditions (b) and (c) are derived from Conditions (b') and (c') of Theorem 3.12 by an application of Lemma 5.7. □

5.5 Proof of Theorem 3.10

For every n, define $\pi_n : [n] \to [n]$ to be the bijection given by $\pi_n(i) := n - i + 1$, $i = 1, ..., n$, and also set $\beta(n,t) := n - \lfloor nt\rfloor + 1$. We define, for $t \in [0,1]$,

$$U_n(t) := \sum_{1\le i<j\le\lfloor nt\rfloor}\psi^{(n)}(X_i,X_j),$$

and

$$I_n(t) := \sum_{(i,j):\pi_n(j)<\pi_n(i)<\beta(n,t)} \psi^{(n)}(X_i, X_j).$$

Then, one has that $Y_n(t) = U_n(1) - U_n(t) - I_n(t)$, in such a way that the tightness of $\{\widetilde{Y}_n\}$ in $D[0,1]$ follows from a direct application of Theorem 5.10 first to U_n/γ_n and then to I_n/γ_n. The asymptotic Gaussianity of the finite-dimensional distributions of $Y^{(n)}$ now follows from Remark 5.5, and one can check that the covariance function of $Y^{(n)}$ converges to that of $c_1 A + c_2 b$ by a direct computation.

□

References

[AG93] Miguel A. Arcones and Evarist Giné. Limit theorems for U-processes. *Ann. Probab.*, 21(3):1494–1542, 1993.

[Bar90] A. D. Barbour. Stein's Method for Diffusion Approximations. *Probab. Theory and Relat. Fields*, 84:297–322, 1990.

[Bas94] A. Basalykas. Functional limit theorems for random multilinear forms. *Stoch. Proc. Appl.*, 53:175–191, 1994.

[BG92] R. N. Bhattacharya and J. K. Ghosh. A class of U-statistics and asymptotic normality of the number of k-clusters. *J. Multivariate Anal.*, 43(2):300–330, 1992.

[Bil99] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[BJ09] A. Barbour and S. Janson. A Functional Combinatorial Central Limit Theorem. *Electron. J. Probab.*, 14:2352–2370, 2009.

[Bor96] V. Yu. Borovskikh. *U-Statistics in Banach Spaces*. V.S.P. Intl Science, 1996.

[BR88] P. J. Bickel and Y. Ritov. Estimating integrated squared density serivatives: sharp best order of convergence estimates. *Sankhya Ser. A*, 50(3):381–393, 1988.

[CH88] M. Csörgö and L. Horvath. Invariance principles for changepoint problems. *J. Multivariate Anal.*, 27:151–168, 1988.

[CH97] M. Csörgö and L. Horvath. *Limit theorems in changepoint analysis*. Wiley, 1997.

[Che18] Xiaohui Chen. Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.*, 46(2):642–678, 2018.

[CK17] X. Chen and K. Kato. Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. *Probab. Theory Relat. Fields, to appear*, 2017.

[dJ90] P. de Jong. A central limit theorem for generalized multilinear forms. *J. Multivariate Anal.*, 34(2):275–289, 1990.

[DlPG99] V. H. De la Pena and E. Giné. *Decoupling: From Dependence to Independence. Springer, 1999.* Springer-Verlag, 1999.

[DM83] E. B. Dynkin and A. Mandelbaum. Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.*, 11(3):739–745, 1983.

[DP17] C. Döbler and G. Peccati. Quantitative de Jong Theorems in any dimension. *Electron. J. Probab.*, 22(2):1–35, 2017.

[DP19] C. Döbler and G. Peccati. Quantitative CLTs for symmetric U-statistics using contractions. *Electron. J. Probab.*, 24:1–43, 2019.

[Dud02] R. M. Dudley. *Real analysis and probability*. Cambridge University Press, 2002.

[Fer94] D. Ferger. An extension of the Csörgő-Horváth functional limit theorem and Its applications to changepoint problems. *J. Multivariate Anal.*, 51(2):338–351, 1994.

[Fer95] D. Ferger. The joint distribution of the running maximum and its location of D-valued Markov processes. *Journal of Applied Probability*, 32(3):842–845, 1995.

[Fer01] D. Ferger. Analysis of change-point estimators under the null hypothesis. *Bernoulli*, 7(3):487–506, 2001.

[GH95] E. Gombay and L. Horvath. An application of U-statistics to change-point analysis. *Acta Sci. Math. (Szeged)*, 60:345–357, 1995.

[GM07] E. Giné and D. Mason. On local U-statistic processes and the estimation of densities of functons of several sample variables. *Ann. Statist.*, 35(3):1105–1145, 2007.

[Gom04] E. Gombay. U-Statistics in Sequential Tests and Change Detection. *Sequential Analysis: Design Methods and Applications*, 23(2):257–274, 2004.

[Hal79] P. Hall. On the invariance principle for U-statistics. *Stoch. Proc. Appl.*, 9:163–174, 1979.

[HR14] L. Horvath and G. Rice. Extensions of some classical methods in change point analysis. *TEST*, 23(2):219–255, 2014.

[IS02] R. Ibragimov and Sh. Sharakhmetov. Bounds on moments of symmetric statistics. *Studia Sci. Math. Hungar.*, 39(3-4):251–275, 2002.

[JJ86] S. R. Jammalamadaka and S. Janson. Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances. *Ann. Probab.*, 14(4):1347–1358, 1986.

[Kas17] M.J. Kasprzak. Stein's method for multivariate Brownian approximations of sums under dependence . arXiv:1708.02521, 2017.

[KR82] S. Karlin and Y. Rinott. Applications of Anova Type Decompositions for Comparisons of Conditional Variance Statistics Including Jackknife Estimates. *Ann. Statist.*, 10(2):485–501, 1982.

[Lau96] B. Laurent. Efficient estimation of integral functionals of a density. *Ann. Stat.*, 24(2):659–681, 1996.

[Lau97] B. Laurent. Estimation of integral functionals of a density and its derivatives. *Bernoulli*, 3(2):181–211, 1997.

[LM00] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 2000.

[LRP13] R. Lachièze-Rey and G. Peccati. Fine Gaussian fluctuations on the Poisson space II: rescaled kernels, marked processes and geometric U-statistics. *Stochastic Process. Appl.*, 123(12):4186–4218, 2013.

[Maj13] P. Major. *On the estimation of multiple random integrals and U-statistics*, volume 2079 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2013.

[MT84] A. Mandelbaum and M. S. Taqqu. Invariance Principle for Symmetric Statistics. *Ann. Statist.*, 12(2):483–496, 1984.

[Neu77] G. Neuhaus. Functional limit theorems for U-statistics in the degenerate case. *J. Multivariate Anal.*, 7(3):424–439, 1977.

[NN18] I. Nourdin and D. Nualart. The functional Breuer-Major theorem. *Probab. Theory Relat. Fields, to appear*, 2018.

[NP87] D. Nolan and D. Pollard. Functional limit theorems for U-processes. *Ann. Statist.*, 15:780–799, 1987.

[Pen04] M. Penrose. *Geometric Random Graphs*. Oxford, 2004.

[PR16] G. Peccati and M. Reitzner, editors. *Stochastic analysis for Poisson point processes*. Springer-Verlag, 2016.

[RLTvdV16] J. Robins, L. Li, E. Tchetgen, and A. van der Vaart. Asymptotic Normality of Quandratic Estimators. *Stoch. Processes Their Appl.*, 126:3733–3759, 2016.

[RW19] A. Račkauskas and M. Wendler. Convergence of U-Processes in Hölder Spaces with Application to Robust Detection of a Changed Segment. *Arxiv Preprint*, 2019.

[Ser80] R. J. Serfling. *Approximation Theorems in Mathematical Statistics*. Wiley, 1980.

[Vit92] R. A. Vitale. Covariances of symmetric statistics. *J. Multivariate Anal.*, 41(1):14–26, 1992.