Bernoulli **20**(2), 2014, 846–877
DOI: 10.3150/13-BEJ508

Continuous mapping approach to the asymptotics of U- and V-statistics

ERIC BEUTNER[1] and HENRYK ZÄHLE[2]

[1]*Department of Quantitative Economics, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, The Netherlands. E-mail: e.beutner@maastrichtuniversity.nl*
[2]*Department of Mathematics, Saarland University, Postfach 151150, D-66041 Saarbrücken, Germany. E-mail: zaehle@math.uni-sb.de*

We derive a new representation for U- and V-statistics. Using this representation, the asymptotic distribution of U- and V-statistics can be derived by a direct application of the Continuous Mapping theorem. That novel approach not only encompasses most of the results on the asymptotic distribution known in literature, but also allows for the first time a unifying treatment of non-degenerate and degenerate U- and V-statistics. Moreover, it yields a new and powerful tool to derive the asymptotic distribution of very general U- and V-statistics based on long-memory sequences. This will be exemplified by several astonishing examples. In particular, we shall present examples where weak convergence of U- or V-statistics occurs at the rate a_n^3 and a_n^4, respectively, when a_n is the rate of weak convergence of the empirical process. We also introduce the notion of asymptotic (non-) degeneracy which often appears in the presence of long-memory sequences.

Keywords: Appell polynomials; central and non-central weak limit theorems; empirical process; Hoeffding decomposition; non-degenerate and degenerate U- and V-statistics; strong limit theorems; strongly dependent data; von Mises decomposition; weakly dependent data

1. Introduction

The study of the asymptotic distribution of U- and V- (von Mises-) statistics goes back to Halmos [16], Hoeffding [18] and von Mises [33]. Different approaches have been proposed to obtain the asymptotic distribution of these statistics. The most-used one is certainly based on the Hoeffding decomposition of a U-statistic; see, for instance, Dehling [9], Denker [12], Koroljuk and Borovskich [21], Lee [22], Serfling [28]. Recently, Beutner and Zähle [6] showed that the asymptotic distribution of U- and V-statistics can be obtained by using the concept of quasi-Hadamard differentiability introduced in Beutner and Zähle [5]. This concept led to new results for U- and V-statistics based on weakly dependent data and was shown in Beutner, Wu and Zähle [4] to be even suitable for a certain class of U- and V-statistics based on long-memory sequences. However, a general result that allows to deduce non-central limit theorems for general U- and V-statistics based on long-memory sequences is still missing. This is due to the fact that for long-memory sequences several parts of the Hoeffding decomposition may contribute to the limiting distribution; see Dehling and Taqqu [11] and our discussion before Corollary 4.3 in Section 4.

In this article, we derive a new representation of U- and V-statistics. Based on this representation, the asymptotic distribution of U- and V-statistics, subject to certain regularity conditions, can be inferred by a direct application of the Continuous Mapping theorem. It turns out that the continuous mapping approach does not only cover the majority of the results known in literature

1350-7265 © 2014 ISI/BS

and allows a unifying treatment of non-degenerate and degenerate U- and V-statistics (see also Section 2 for the definitions of non-degeneracy and degeneracy), but also supplements the existing theorems for U- and V-statistics based on long-memory sequences. We shall further see that the continuous mapping approach allows us to establish strong laws for U- and V-statistics. Using the continuous mapping approach it will also be seen that, once the new representation is established, the asymptotic distributions of several degenerate U- and V-statistics that are usually derived on a case by case basis, are direct consequences of more general results. Finally, we will demonstrate that, under certain conditions on the kernel, the continuous mapping approach is also suitable to derive the asymptotic distribution of two-sample U- and V-statistics; see Beutner and Zähle [7].

To explain our approach, we first of all recall that U- and V-statistics (of degree 2) are non-parametric estimators for the characteristic

$$V_g(F) := \iint g(x_1, x_2)\,\mathrm{d}F(x_1)\,\mathrm{d}F(x_2) \tag{1}$$

of a distribution function (df) F on the real line, where $g:\mathbb{R}^2 \to \mathbb{R}$ is some measurable function and it is assumed that the double integral in (1) exists. Given a sequence $(X_i)_{i\in\mathbb{N}}$ of random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ being identically distributed according to F, the V-statistic based on F_n is given by

$$V_g(F_n) = \iint g(x_1, x_2)\,\mathrm{d}F_n(x_1)\,\mathrm{d}F_n(x_2), \tag{2}$$

where F_n denotes some estimate of F based on $X_1, \dots, X_n$, and it is assumed that the integral in (2) exists for all $n \in \mathbb{N}$. The corresponding U-statistic is given by $U_{g,n} := \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1,\neq i}^n g(X_i, X_j)$. Assuming $\iint |g(x_1, x_2)|\,\mathrm{d}F_n(x_1)\,\mathrm{d}F(x_2) < \infty$, we obtain from (2) the decomposition

$$\begin{aligned} V_g(F_n) - V_g(F) = {} & \sum_{i=1}^{2}\left(\int g_{i,F}(x)\,\mathrm{d}F_n(x) - \int g_{i,F}(x)\,\mathrm{d}F(x)\right) \\ & + \iint g(x_1, x_2)\,\mathrm{d}(F_n - F)(x_1)\,\mathrm{d}(F_n - F)(x_2) \end{aligned} \tag{3}$$

with $g_{1,F}(\cdot) := \int g(\cdot, x_2)\,\mathrm{d}F(x_2)$ and $g_{2,F}(\cdot) := \int g(x_1, \cdot)\,\mathrm{d}F(x_1)$. This decomposition is sometimes called von Mises decomposition of $V_g(F_n) - V_g(F)$; see Koroljuk and Borovskich [21], page 40. If F_n is the empirical df $\hat{F}_n := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[X_i,\infty)}$ of $X_1, \dots, X_n$, then the first line and the second line on the right-hand side of (3) are the linear part and the degenerate part of the von Mises decomposition, respectively. In this case, the linear part of the von Mises decomposition coincides with the linear part of the Hoeffding decomposition of $U_{g,n} - V_g(F)$, and the degenerate part differs from the degenerate part of the Hoeffding decomposition of $U_{g,n} - V_g(F)$ only by

$$\frac{1}{n^2(n-1)} \sum_{i=1}^n \sum_{j=1,\neq i}^n g(X_i, X_j) - \frac{1}{n^2}\sum_{i=1}^n g(X_i, X_i). \tag{4}$$

While the linear part can usually be treated by a central limit theorem (applied to the random variables $Y_i := g_{1,F}(X_i) + g_{2,F}(X_i)$, $i \in \mathbb{N}$), it is exactly the degenerate part that causes the main difficulties in deriving the asymptotic distribution of U- and V-statistics.

Now let us suppose that we may apply a one-dimensional integration-by-parts formula to the two summands in the first line on the right-hand side of (3) and a two-dimensional integration-by-parts formula to the second line on the right-hand side of (3). Notice that this assumption in particular implies that $g_{1,F}$ and $g_{2,F}$ generate (possibly signed) measures on $\mathbb{R}$ and that g generates a (possibly signed) measure on $\mathbb{R}^2$. We then have the following representation (assuming that expressions like $\lim_{x\to\infty}(F_n - F)(x)g_{i,F}(x)$ are equal to zero $\mathbb{P}$-a.s.)

$$\begin{aligned} V_g(F_n) - V_g(F) = &-\sum_{i=1}^{2} \int \big[F_n(x-) - F(x-)\big]\, \mathrm{d}g_{i,F}(x) \\ &+ \iint (F_n - F)(x_1-)(F_n - F)(x_2-)\, \mathrm{d}g(x_1, x_2), \end{aligned} \tag{5}$$

where we refer to the sum of the first two lines on the right-hand side as the linear part and to the last line on the right-hand side as the degenerate part of the representation. Of course, they coincide with the linear part and the degenerate part of the von Mises decomposition (3). The representation (5) is the sum of the three mappings

$$\begin{aligned} &\Phi_{i,g} : \mathbf{V} \longrightarrow \mathbb{R}, \qquad \Phi_{i,g}(f) := -\int f(x-)\, \mathrm{d}g_{i,F}(x), \qquad i = 1, 2, \\ &\Phi_{3,g} : \mathbf{V} \longrightarrow \mathbb{R}, \qquad \Phi_{3,g}(f) := \iint f(x_1-) f(x_2-)\, \mathrm{d}g(x_1, x_2) \end{aligned} \tag{6}$$

applied to $F_n - F$, where $\mathbf{V}$ is some suitable space consisting of càdlàg functions on $\overline{\mathbb{R}}$. Of course, if g is symmetric then (5) can be represented using two mappings only. Now, on one hand, if the functions $g_{i,F}$, $i = 1, 2$, generate *finite* (possibly signed) measures on $\mathbb{R}$, and if g generates a *finite* (possibly signed) measure on $\mathbb{R}^2$, then the mappings $\Phi_{i,g}$, $i = 1, 2, 3$, are continuous if we endow $\mathbf{V}$ with the uniform sup-metric $d_\infty(f, h) := \|f - h\|_\infty$. On the other hand, if the (possibly signed) measure generated by $g_{i,F}$ is not finite but only σ-finite, then the map $\Phi_{i,g}$ is obviously not continuous w.r.t. the uniform sup-metric d_∞. However, if we assume, for example, $\int (1/\phi(x))|\,\mathrm{d}g_{i,F}|(x) < \infty$ for $i = 1, 2$, where $\phi : \mathbb{R} \to [1, \infty)$ is any continuous function, and $|\mathrm{d}g_{i,F}|$ denotes the total variation measure generated by $g_{i,F}$, then we still have $\Phi_{i,g}(f_n) \to \Phi_{i,g}(f)$ for $i = 1, 2$ when the sequence (f_n) converges to f in the weighted sup-metric $d_\phi(f, g) := \|(f - h)\phi\|_\infty$. If in addition $\iint (\phi(x_1)\phi(x_2))^{-1} |\mathrm{d}g|(x_1, x_2) < \infty$, then we also have $\Phi_{3,g}(f_n) \to \Phi_{3,g}(f)$ when the sequence (f_n) converges to f in the weighted sup-metric d_ϕ, and $|\mathrm{d}g|$ denotes the total variation measure generated by g. That is, under appropriate conditions, we have

$$a_n\big(V_g(F_n) - V_g(F)\big) = \sum_{i=1}^{2} \Phi_{i,g}\big(a_n(F_n - F)\big) + \Phi_{3,g}\big(\sqrt{a_n}(F_n - F)\big) \tag{7}$$

with *continuous* mappings $\Phi_{i,g}$, $i = 1, 2, 3$, and a_n a strictly positive real number. Therefore, once the representation (5) has been established, one only needs weak convergence of the process $a_n(F_n - F)$ w.r.t. the weighted sup-metric d_ϕ to make use of the Continuous Mapping theorem. The latter is not problematic. For instance, weak convergence of empirical processes w.r.t. weighted sup-metrics has been established under various conditions; see, for instance, Beutner, Wu and Zähle [4], Chen and Fan [8], Shao and Yu [29], Shorack and Wellner [30], Wu [34,36], Yukich [37]. One of the advantages of this approach lies in the fact that weak convergence of $a_n(F_n - F)$ implies that $\sqrt{a_n}(F_n - F)$ converges in probability to zero, and hence that $\Phi_{3,g}(\sqrt{a_n}(F_n - F))$ converges in probability to zero. Thus, with the continuous mapping approach we can easily deal with the degenerate part of a non-degenerate V-statistic. For a degenerate V-statistic the linear part vanishes, that is, $\Phi_{1,g} \equiv \Phi_{2,g} \equiv 0$, and so in this case (7) multiplied by a_n reads as $a_n^2(V_g(F_n) - V_g(F)) = \Phi_{3,g}(a_n(F_n - F))$. That is, with the continuous mapping approach we can also easily deal with the degenerate part of a degenerate V-statistic. Moreover, the continuous mapping approach can also provide a simple way to derive the asymptotic distribution of a V-statistic when both terms of the von Mises decomposition contribute to the asymptotic distribution, or when the scaling sequence has to be chosen as the cube (a_n^3) or the fourth power (a_n^4) of the scaling sequence (a_n) of $\hat{F}_n - F$. This will be illustrated in Section 4 in the context of long-memory data; see Examples 4.7, 4.8 and 4.9. In this context the notion of asymptotic degeneracy, to be introduced in Section 2, plays a crucial role.

Next, let us briefly discuss the relation between the asymptotic distribution of a U-statistic $U_{g,n}$ with that of the V-statistic $V_g(\hat{F}_n)$, where as before $\hat{F}_n$ denotes the empirical df. Since the linear parts of the von Mises decomposition and the Hoeffding decomposition coincide, and the degenerate parts of these decompositions differ only by the term in (4), it can be shown easily that a non-degenerate V-statistic $V_g(\hat{F}_n)$ w.r.t. (g, F) and the corresponding non-degenerate U-statistic $U_{g,n}$ have the same asymptotic distribution if $a_n = \mathrm{o}(n)$ and $\mathbb{E}[|g(X_1, X_1)|] < \infty$; see, for example, Beutner and Zähle [6], Remark 2.5. Hence, in the non-degenerate case the asymptotic distribution of both U-statistics and V-statistics can be derived from (5). On the other hand, in the degenerate case they differ by a constant if a LLN holds for $\sum_{i=1}^n g(X_i, X_i)$. In fact, this follows, because $n(U_{g,n} - V_g(F))$ equals $n(V_g(F_n) - V_g(F)) - \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n g(X_i, X_i) + \frac{n}{n-1}(V_g(F_n) - V_g(F)) + \frac{n}{n-1}V_g(F)$.

We stress that for kernels g that are locally of unbounded variation the asymptotic distribution of the corresponding U-statistic cannot be obtained by the continuous mapping approach; see Beutner and Zähle [7], Remark 1.1. However, if such a kernel is non-degenerate a non-trivial limiting distribution of the corresponding U-statistic can be obtained either by the Hoeffding decomposition or by using the approach of Beutner and Zähle [6] that is based on the concept of quasi-Hadamard differentiability and the Modified Functional Delta Method. In case that the kernel is locally of unbounded variation and degenerate, the approach of Beutner and Zähle [6] also yields little. However, then the traditional approach to degenerate U- and V-statistics that is briefly recalled after Example 3.18 may lead to a non-trivial limiting distribution.

The rest of the article is organized as follows. In Section 2, we will discuss the notion of (non-) degenerate and asymptotically (non-) degenerate U- and V-statistcs. In Section 3, we will first give conditions on the kernel g, the df F and the estimator F_n that ensure that the representation (5) holds (Section 3.1). Thereafter, we will give interesting examples for kernels g

that satisfy these conditions (Section 3.2) and apply the continuous mapping approach to derive weak and strong limit theorems for U- and V-statistics based on weakly dependent data (Sections 3.3 and 3.4). In Section 4, the whole strength of our approach will be illustrated by deriving non-central limit theorems for U- and V-statistics based on strongly dependent data. The representation (5) along with a new non-central limit theorem for the empirical process of a linear long-memory process offers a very simple way to deduce such non-central limit theorems. We will present in particular three astonishing examples. In Example 4.7 both terms of the von Mises decomposition of a non-degenerate V-statistic contribute to the asymptotic distribution whatever the true df F is (in the Gaussian case this example is already known from Dehling and Taqqu [11], Section 3), and in Examples 4.8 and 4.9 the scaling sequences for degenerate V-statistics are given by (a_n^3) and (a_n^4), respectively (and not as usual by the square (a_n^2)) of the scaling sequence (a_n) of $\hat{F}_n - F$. A supplemental article Beutner and Zähle [7] contains a section that discusses some extensions and limitations of our approach.

2. The notions of (non-) degeneracy and asymptotic (non-) degeneracy

In this section, we will recall the notion of (non-) degenerate U- and V-statistics, and we shall introduce the notion of asymptotically (non-) degenerate U- and V-statistics. We will restrict to the case where g, F and F_n admit the von Mises decomposition (3).

The corresponding V-statistic $V_g(F_n)$ will be called *non-degenerate* w.r.t. (g, F), if the linear part of the von Mises decomposition (i.e., the first line on the right-hand side in (3)) does not vanish. The corresponding V-statistic $V_g(F_n)$ will be called *degenerate* w.r.t. (g, F) if the linear part of the von Mises decomposition vanishes, that is, if $\sum_{i=1}^{2} \int g_{i,F} \, \mathrm{d}(F_n - F) = 0$ $\mathbb{P}$-a.s. for every $n \in \mathbb{N}$. This condition holds in particular if $g_{1,F} \equiv g_{2,F} \equiv 0$, or if F_n is a (random) df and both $g_{1,F}$ and $g_{2,F}$ are constant. If the linear part of the von Mises decomposition does (not) vanish when $F_n = \hat{F}_n$, then we also call the corresponding U-statistic $U_{g,n}$ (non-) degenerate w.r.t. (g, F). Recall that it is very common, mainly in the i.i.d. set-up, to call a U-statistic degenerate if $\mathbb{V}\mathrm{ar}[g_{i,F}(X_1)] = 0$ for $i = 1, 2$. Notice that, in this case, this is in line with the convention used here. Indeed, it is easily seen that $\mathbb{V}\mathrm{ar}[g_{i,F}(X_1)] = 0$ is equivalent to $\int g_{i,F} \, \mathrm{d}(\hat{F}_n - F) = 0$ $\mathbb{P}$-a.s. if $\hat{F}_n$ is based on an i.i.d. sequence. Table 1 displays some examples for non-degenerate and degenerate U- and V-statistics.

To introduce the notion of asymptotically (non-) degenerate U- and V-statistics, we let $(a_n) \subset (0, \infty)$ be a scaling sequence such that $a_n(F_n - F)$ converges in distribution to a non-degenerate limit. The representation (7) indicates that for every (non-degenerate) V-statistic $V_g(F_n)$ (w.r.t. (g, F)) only the linear part of the von Mises decomposition may contribute to the limiting distribution of $a_n(V_g(F_n) - V_g(F))$. If there is a non-trivial limiting distribution of the linear part weighted by a_n, then we call the V-statistic $V_g(F_n)$ *asymptotically non-degenerate* w.r.t. $(g, F, (a_n))$, and the analogous terminology is used for U-statistics. Of course, every asymptotically non-degenerate U- or V-statistic w.r.t. $(g, F, (a_n))$ must also be non-degenerate w.r.t. (g, F). However, it might happen that the limiting distribution of the linear part weighted by a_n vanishes. In this case, we call the V-statistic $V_g(F_n)$ *asymptotically degenerate* w.r.t.

Table 1. Examples for non-degenerate and degenerate V-statistics w.r.t. (g, F)

Non-degenerate	Gini's mean difference	Example 3.10
	Variance	Example 3.11
Degenerate	Gini's mean difference (uniform two-point distribution)	Example 3.12
	Variance (4th central = squared 2nd central moment of F)	Example 3.20(i)
	Cramér–von Mises	Example 3.13
	test for symmetry	Example 3.14

$(g, F, (a_n))$, and again the analogous terminology is used for U-statistics. Of course, every degenerate U- or V-statistic w.r.t. (g, F) is also asymptotically degenerate w.r.t. $(g, F, (a_n))$.

For an asymptotically degenerate U- or V-statistic w.r.t. $(g, F, (a_n))$ a non-trivial asymptotic distribution can typically be obtained by weighting the empirical difference by a_n^2 instead of a_n, that is, by considering the limiting distribution of $a_n^2(V_g(F_n) - V_g(F))$. In this context, two different things may occur:

(1) The asymptotic distribution of $a_n^2(V_g(F_n) - V_g(F))$ is non-trivial. In this case, we say that the asymptotically degenerate U- or V-statistic w.r.t. $(g, F, (a_n))$ is of *type 1*.
(2) The asymptotic distribution of $a_n^2(V_g(F_n) - V_g(F))$ is *still* degenerate. In this case, we say that the asymptotically degenerate U- or V-statistic w.r.t. $(g, F, (a_n))$ is of *type 2*.

It seems that behavior (2) only appears in the presence of long-memory sequences; for examples, see Examples 4.8 and 4.9. It is worth pointing out that for an (asymptotically) degenerate U- and V-statistics of type 2 a non-trivial limiting distribution can sometimes be obtained by considering the limiting distribution of $a_n^p(V_g(F_n) - V_g(F))$ for some $p > 2$; see again Examples 4.8 and 4.9. In case (1) we can distinguish between the following three cases:

(1.a) Only the degenerate part of the von Mises decomposition contributes to the limiting distribution of $a_n^2(V_g(F_n) - V_g(F))$. This is in particular the case, if the U- or V-statistic is even degenerate w.r.t. (g, F).
(1.b) Only the linear part contributes to the limiting distribution of $a_n^2(V_g(F_n) - V_g(F))$. This can happen only if the U- or V-statistic is asymptotically degenerate w.r.t. $(g, F, (a_n))$, but non-degenerate w.r.t. (g, F).
(1.c) Both the linear and the degenerate part contribute to the limiting distribution of $a_n^2(V_g(F_n) - V_g(F))$. Again, this can occur only if the U- or V-statistic is asymptotically degenerate w.r.t. $(g, F, (a_n))$, but non-degenerate w.r.t. (g, F).

In the original version of the manuscript, we guessed that the cases (1.b) and (1.c) only appear for U- and V-statistics based on long-memory sequences. However, a referee provided us with the following example that shows that behavior (1.c) also occurs for m-dependent sequences.

Example 2.1. Let (ξ_n) and (δ_n) be independent i.i.d. sequences with $\mathbb{P}[\xi_i = 0] = \mathbb{P}[\xi_i = 1] = 1/2$ and $\mathbb{P}[\delta_i = 0] = \mathbb{P}[\delta_i = 1] = 1/2$. Define the 1-dependent sequence (Z_i) by $Z_i := \xi_i - \xi_{i-1}$

and the 1-dependent sequence (X_i) by

$$X_i := \begin{cases} \sqrt{2}, & Z_i = 1 \text{ and } \delta_i = 1, \\ 1, & Z_i = 0 \text{ and } \delta_i = 1, \\ 0, & Z_i = -1, \\ -1, & Z_i = 0 \text{ and } \delta_i = 0, \\ -\sqrt{2}, & Z_i = 1 \text{ and } \delta_i = 0. \end{cases}$$

We then have $\mu := \mathbb{E}[X_i] = 0$, $\sigma^2 := \mathbb{V}\mathrm{ar}[X_i] = 1$, and $(X_i - \mu)^2 - \sigma^2 = Z_i$. Denote by $\hat{\sigma}_n^2$ the sample variance, which is the V-statistic $V_g(\hat{F}_n)$ with kernel $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$. The Hoeffding decomposition of $\hat{\sigma}_n^2 - \sigma^2$ equals

$$\frac{1}{n}\sum_{i=1}^{n}\big((X_i - \mu)^2 - \sigma^2\big) + \left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right)^2 = \frac{1}{n}(\xi_n - \xi_0) + \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2,$$

where the latter identity follows from $(X_i - \mu)^2 - \sigma^2 = Z_i$, $\sum_{i=1}^n Z_i = \xi_n - \xi_0$, and $\mu = 0$. Thus, we have $\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \frac{1}{\sqrt{n}}(\xi_n - \xi_0) + (\frac{1}{n^{1/4}}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i)^2$, and so we obtain from the central limit theorem along with Slutzky's lemma, and the fact that $\xi_n - \xi_0$ has the same distribution for every $n \in \mathbb{N}$, that $\hat{\sigma}_n^2$ is non-degenerate w.r.t. (g, F), but asymptotically degenerate w.r.t. $(g, F, (\sqrt{n}))$, where F refers to the df of the X_i. On the other hand, we have $n(\hat{\sigma}_n^2 - \sigma) = \xi_n - \xi_0 + (\frac{1}{n}\sum_{i=1}^n X_i)^2$. As already mentioned $\xi_n - \xi_0$ has the same distribution for every $n \in \mathbb{N}$, and $(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i)^2$ converges in distribution to a χ^2 distribution with one degree of freedom. Since $\xi_n - \xi_0$ is the linear part and $(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i)^2$ the degenerate part of the Hoeffding decomposition, the example shows that even for m-dependent sequences both terms of the Hoeffding decomposition may contribute to the limiting distribution.

Table 2 displays some examples for asymptotically non-degenerate and degenerate U- and V-statistics for a linear process with long-memory. It is worth mentioning that there is a difference

Table 2. Examples for asymptotically non-degenerate and asymptotically degenerate U- and V-statistics w.r.t. $(g, F, (a_n))$ for $a_n := n^{p(\beta-1/2)}\ell(n)^{-p}$ with $p(2\beta - 1) < 1$, where the observations are drawn from a linear process $X_t := \sum_{s=0}^{\infty} a_s \varepsilon_{t-s}$ with $a_s = s^{-\beta}\ell(s)$ for some $\beta \in (\frac{1}{2}, 1)$ and some slowly varying ℓ (long-memory)

Asymptotically non-degenerate	Gini's mean difference	$(p = 1)$	Disc. before Corollary 4.3
Asymptotically degenerate – type 1	(1.a) Cramér–von Mises	$(p = 2)$	Disc. before Corollary 4.3
	(1.b) Squared absolute mean	$(p = 2)$	Example 4.6
	(1.c) Variance	$(p = 2)$	Example 4.7
Asymptotically degenerate – type 2	Some artificial kernel	$(p = 3)$	Example 4.8
	test for symmetry	$(p = 4)$	Example 4.9

between the asymptotic degeneracy of the variance in the case of a linear long-memory sequence and the m-dependent sequence of Example 2.1. For a long-memory sequence, the variance is asymptotically degenerate whatever the underlying df is, whereas for an m-dependent sequence it does depend on the underlying df whether the variance is asymptotically degenerate or not.

3. Representation (5): Conditions, examples, and applications

Let $\mathbb{D}$ be the space of all bounded càdlàg functions on $\overline{\mathbb{R}}$. Any metric subspace $(\mathbf{V}, d)$ of $\mathbb{D}$ will be equipped with the σ-algebra $\mathcal{V} := \mathcal{D} \cap \mathbf{V}$ to make it a measurable space, where $\mathcal{D}$ is the σ-algebra generated by the usual coordinate projections $\pi_x : \mathbb{D} \to \mathbb{R}$, $x \in \overline{\mathbb{R}}$. The roles of $\mathbf{V}$ and d will often be played by the space $\mathbb{D}_\phi$ of all $f \in \mathbb{D}$ with $\|f\phi\|_\infty < \infty$ and the weighted sup-metric $d_\phi(f, h) := \|(f - h)\phi\|_\infty$, respectively, where $\phi : \overline{\mathbb{R}} \to [1, \infty]$ is any continuous function being real-valued on $\mathbb{R}$ (henceforth called weight function) and where we use the convention $0 \cdot \infty := 0$. We will frequently work with the particular weight function $\phi_\lambda(x) := (1 + |x|)^\lambda$ for fixed λ.

Further, let $\mathbb{BV}_{\mathrm{loc,rc}}$ be the space of all functions on $\mathbb{R}$ that are right-continuous and locally of bounded variation, and notice that every function in $\mathbb{BV}_{\mathrm{loc,rc}}$ has also left-hand limits. For $\psi \in \mathbb{BV}_{\mathrm{loc,rc}}$, we denote by $d\psi^+$ and $d\psi^-$ the unique positive Radon measures induced by the Jordan decomposition of ψ and we set $|d\psi| := d\psi^+ + d\psi^-$. Analogously, let $\mathbb{BV}^2_{\mathrm{loc,rc}}$ be the space of all functions on $\mathbb{R}^2$ that are upper right-continuous and locally of bounded variation, and for $\tau \in \mathbb{BV}^2_{\mathrm{loc,rc}}$, $d\tau^+$, $d\tau^-$ and $|d\tau|$ are defined analogously to $d\psi^+$, $d\psi^-$ and $|d\psi|$; for details see the discussion subsequent to Remark 3.5 below. We shall interpret integrals as being over the open intervals $(-\infty, \infty)$ and $(-\infty, \infty)^2$, that is, $\int = \int_{(-\infty,\infty)}$ and $\iint = \iint_{(-\infty,\infty)^2}$. Moreover, for a measurable function f we shall say that the integral of f w.r.t. a signed measure μ exists if the four integrals $\int f^+ \,\mathrm{d}\mu^+$, $\int f^- \,\mathrm{d}\mu^+$, $\int f^+ \,\mathrm{d}\mu^-$ and $\int f^- \,\mathrm{d}\mu^-$ are all finite, where f^+ and f^- denote the positive and the negative part of f, and μ^+ and μ^- denote the positive and the negative part of μ. We denote by $\xrightarrow{\mathsf{d}}$ convergence in distribution in the sense of Pollard [26], and the Borel σ-algebra in $\mathbb{R}$ is denoted by $\mathcal{B}(\mathbb{R})$.

3.1. Conditions for the representation (5)

In this section, we provide conditions on g, F and the estimate F_n of F under which the representation (5) holds true. First of all, we impose assumptions on g, F and F_n that ensure that $V_g(F)$ and $V_g(F_n)$ are well defined.

Assumption 3.1. *The integral in* (1) *exists, the estimate F_n of F is a non-decreasing càdlàg process with variation bounded by* 1, *and for all $n \in \mathbb{N}$ we have that $\mathbb{P}$-a.s. $\iint |g(x_1, x_2)| \,\mathrm{d}F_n(x_1) \,\mathrm{d}F_n(x_2) < \infty$.*

A further minimum requirement for the representation (5) is the following.

Assumption 3.2. *For all $n \in \mathbb{N}$ we have that $\mathbb{P}$-a.s. $\iint |g(x_1, x_2)| \,\mathrm{d}F_n(x_1) \,\mathrm{d}F(x_2) < \infty$ and $\iint |g(x_1, x_2)| \,\mathrm{d}F(x_1) \,\mathrm{d}F_n(x_2) < \infty$.*

Notice that the conditions on F_n imposed by Assumptions 3.1–3.2 are always fulfilled if the integral in (1) exists and F_n is the empirical df $\hat{F}_n$. From (3) and Lemmas 3.4 and 3.6 below, we immediately obtain the following theorem.

Theorem 3.3. *If the assumptions of Lemmas 3.4 and 3.6 (below) are fulfilled, then the representation (5) of $V_g(F_n) - V_g(F)$ holds true $\mathbb{P}$-a.s. for every $n \in \mathbb{N}$.*

The following lemma gives conditions that allow to apply almost surely an integration-by-parts formula (see Beutner and Zähle [6], Lemma B.1) to the two summands in the first line on the right-hand side in (3).

Lemma 3.4. *Suppose that*:

(a) *Assumptions 3.1–3.2 hold*,
(b) $g_{i,F} \in \mathbb{BV}_{\mathrm{loc,rc}}$,
(c) $\int |F_n(x-) - F(x-)| |\mathrm{d}g_{i,F}|(x) < \infty$ *$\mathbb{P}$-a.s., for all $n \in \mathbb{N}$ and $i = 1, 2$*,
(d) $\lim_{|x|\to\infty}(F_n - F)(x) g_{i,F}(x) = 0$ *$\mathbb{P}$-a.s., for all $n \in \mathbb{N}$ and $i = 1, 2$.*

Then $\mathbb{P}$-a.s., for every $n \in \mathbb{N}$,

$$\iint g(x_1, x_2)\,\mathrm{d}F_n(x_1)\,\mathrm{d}F(x_2) - \iint g(x_1, x_2)\,\mathrm{d}F(x_1)\,\mathrm{d}F(x_2)$$
$$= -\int \bigl[F_n(x_1-) - F(x_1-)\bigr]\,\mathrm{d}g_{1,F}(x_1),$$
$$\iint g(x_1, x_2)\,\mathrm{d}F(x_1)\,\mathrm{d}F_n(x_2) - \iint g(x_1, x_2)\,\mathrm{d}F(x_1)\,\mathrm{d}F(x_2)$$
$$= -\int \bigl[F_n(x_2-) - F(x_2-)\bigr]\,\mathrm{d}g_{2,F}(x_2).$$

Proof. We only prove the first equation. From Assumptions 3.1 and 3.2, and using Fubini's theorem, we have that the integrals $\int |g_{1,F}(x_1)|\,\mathrm{d}F_n(x_1)$ and $\int |g_{1,F}(x_1)|\,\mathrm{d}F(x_1)$ are finite, and so $\int |g_{1,F}(x_1)||\mathrm{d}(F_n - F)|(x_1)$ exists. Moreover, for every $n \in \mathbb{N}$, we obtain by using Fubini's theorem $\iint g(x_1, x_2)\,\mathrm{d}F_n(x_1)\,\mathrm{d}F(x_2) - \iint g(x_1, x_2)\,\mathrm{d}F(x_1)\,\mathrm{d}F(x_2) = \int g_{1,F}(x_1)\,\mathrm{d}(F_n - F)(x_1)$. Since by assumption (c), we further have that $\int |F_n(x_1-) - F(x_1-)||\,\mathrm{d}g_{1,F}|(x_1)$ exists, the conditions of Lemma B.1 in Beutner and Zähle [6] are fulfilled and the result follows. □

Remark 3.5. (i) For $F_n = \hat{F}_n$ (recall that $\hat{F}_n$ denotes the empirical df) conditions (c) and (d) of Lemma 3.4 boil down to conditions on the tails of F and $g_{i,F}$, $i = 1, 2$.

(ii) More generally, if for $\mathbb{P}$-almost every ω there exist real numbers $x_\ell(\omega)$ and $x_u(\omega)$ such that $F_n(\omega, x) - F(x) = -F(x)$ for all $x \le x_\ell(\omega)$, and $F_n(\omega, x) - F(x) = 1 - F(x)$ for all $x \ge x_u(\omega)$, then again conditions (c) and (d) of Lemma 3.4 boil down to conditions on the tails of F and $g_{i,F}$, $i = 1, 2$.

(iii) Condition (c) holds if $\int \mathrm{d}g_{i,F}$ exists for $i = 1, 2$, and under the conditions of part (ii) of this remark we have that condition (d) holds if $\|g_{i,F}\|_\infty < \infty$ for $i = 1, 2$.

(iv) If for some weight function ϕ we have that $d_\phi(F_n, F)$ is $\mathbb{P}$-a.s. finite for every $n \in \mathbb{N}$, and that $\int 1/\phi |\mathrm{d}g_{i,F}| < \infty$ and $\lim_{|x|\to\infty} g_{i,F}(x)/\phi(x) = 0$ for $i = 1, 2$, then again conditions (c) and (d) of Lemma 3.4 hold. Notice also that under the conditions of part (ii) of this remark, the condition that $d_\phi(F_n, F)$ is $\mathbb{P}$-a.s. finite for all $n \in \mathbb{N}$ is a condition on the tails of F.

(v) If there are $x_{\ell,i} < x_{u,i}$ such that $|g_{i,F}|$ is non-increasing on $(-\infty, x_{\ell,i}]$ and non-decreasing on $[x_{u,i}, \infty)$ for $i = 1, 2$, then part (d) of Lemma 3.4 is already implied by Assumptions 3.1 and 3.2. Indeed, under these assumptions the integrals $\int |g_{1,F}(x_1)|\,\mathrm{d}F_n(x_1)$ and $\int |g_{1,F}(x_1)|\,\mathrm{d}F(x_1)$ exist, and we have for $x \ge x_{u,i}$ that $|g_{i,F}(x)(F_n(x) - F(x))| = |\int_x^\infty g_{i,F}(x)\,\mathrm{d}F(t) - \int_x^\infty g_{i,F}(x)\,\mathrm{d}F_n(t)| \le \int_x^\infty |g_{i,F}(t)|\,\mathrm{d}F(t) + \int_x^\infty |g_{i,F}(t)|\,\mathrm{d}F_n(t)$, $i = 1, 2$. Thus, $\lim_{x\to\infty}(F_n - F)(x)g_{i,F}(x) = 0$ for $i = 1, 2$. Analogously, we obtain $\lim_{x\to-\infty}(F_n - F)(x)g_{i,F}(x) = 0$.

For the functional $\Phi_{3,g}$ to be well defined in the Lebesgue–Stieltjes sense the kernel g must be upper right-continuous and locally of bounded variation. For later use and the reader's convenience we recall the definition of locally bounded variation. For any function $\tau : \mathbb{R}^2 \to \mathbb{R}$, set $\mu_\tau(\mathcal{R}_{(x_1,x_2),(y_1,y_2)}) := \tau(x_2, y_2) - \tau(x_1, y_2) - \tau(x_2, y_1) + \tau(x_1, y_1)$ for every half-open rectangle $\mathcal{R}_{(x_1,x_2),(y_1,y_2)} = (x_1, x_2] \times (y_1, y_2]$ with $(x_1, x_2) \in \mathbb{R}^2$, $x_1 < x_2$, and $(y_1, y_2) \in \mathbb{R}^2$, $y_1 < y_2$. For a fixed half-open rectangle $\mathcal{R} = \mathcal{R}_{(a_1,a_2),(b_1,b_2)} = (a_1, a_2] \times (b_1, b_2]$ in $\mathbb{R}^2$, a pair P of finite sequences $(x_k)_{k=0,\dots,n}$ and $(y_\ell)_{\ell=0,\dots,m}$ is called a grid for $\mathcal{R}$ if $a_1 = x_0 \le x_1 \le \cdots \le x_n = a_2$ and $b_1 = y_0 \le y_1 \le \cdots \le y_m = b_2$. For any grid P, let

$$\begin{aligned}\mathcal{V}(P, \tau) &:= \sum_{i=1}^{n}\sum_{j=1}^{m}\bigl|\mu_\tau(\mathcal{R}_{(x_{i-1},x_i),(y_{i-1},y_i)})\bigr| \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}\bigl|\tau(x_i, y_j) - \tau(x_{i-1}, y_j) - \tau(x_i, y_{j-1}) + \tau(x_{i-1}, y_{j-1})\bigr|.\end{aligned}$$

Moreover, let $\mathcal{V}_\tau(\mathcal{R}) := \sup_{P\in\mathcal{P}} \mathcal{V}(P, \tau)$, where $\mathcal{P}$ is the set of all grids for $\mathcal{R}$. The function τ is said to be locally of bounded variation if for every bounded half-open rectangle $\mathcal{R} \subset \mathbb{R}^2$ we have $\mathcal{V}_\tau(\mathcal{R}) < \infty$, and τ is said to be of bounded total variation if there is a constant $C > 0$ such that $\mathcal{V}_\tau(\mathcal{R}) \le C$ for all bounded half-open rectangles $\mathcal{R} \subset \mathbb{R}^2$. As mentioned earlier, $\mathbb{BV}^2_{\mathrm{loc,rc}}$ denotes the space of all upper right-continuous functions $\tau : \mathbb{R}^2 \to \mathbb{R}$ that are locally of bounded variation, and we use the two-dimensional Jordan decomposition (see, e.g., Ghorpade and Limaye [13], Proposition 1.17) to define $d\tau^+$, $d\tau^-$ and $|d\tau|$ similar as $d\psi^+$, $d\psi^-$ and $|d\psi|$. We can now state the two-dimensional integration-by-parts lemma, which can almost surely be applied to the second line on the right-hand side in (3).

Lemma 3.6. *Suppose that*:

(a) *Assumption 3.1 holds*,
(b) $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$, *and the functions* $g_{x_1}(\cdot) := g(x_1, \cdot)$ *and* $g_{x_2}(\cdot) := g(\cdot, x_2)$ *are locally of bounded variation for every fixed* x_1 *and* x_2, *respectively*,
(c) $\iint |F_n(x_1-) - F(x_1-)||F_n(x_2-) - F(x_2-)||\mathrm{d}g|(x_1, x_2) < \infty$ $\mathbb{P}$-*a.s.*, *for all* $n \in \mathbb{N}$,

(d) *the following limits exist and equal zero $\mathbb{P}$-a.s. for all $n \in \mathbb{N}$:*

$$\lim_{a_1,b_1\to-\infty,a_2,b_2\to\infty}\bigg[(F_n-F)(b_2)\int_{a_1}^{a_2}(F_n-F)(x_1-)\,\mathrm{d}g_{b_2}(x_1) - (F_n-F)(b_1)\int_{a_1}^{a_2}(F_n-F)(x_1-)\,\mathrm{d}g_{b_1}(x_1)\bigg],$$

$$\lim_{a_1,b_1\to-\infty,a_2,b_2\to\infty}\bigg[(F_n-F)(a_2)\int_{b_1}^{b_2}(F_n-F)(x_2-)\,\mathrm{d}g_{a_2}(x_2) - (F_n-F)(a_1)\int_{b_1}^{b_2}(F_n-F)(x_2-)\,\mathrm{d}g_{a_1}(x_2)\bigg],$$

$$\begin{aligned}\lim_{a_1,b_1\to-\infty,a_2,b_2\to\infty}\big[&(F_n-F)(a_2)(F_n-F)(b_2)g(a_2,b_2)\\ &-(F_n-F)(a_1)(F_n-F)(b_2)g(a_1,b_2)\\ &-(F_n-F)(a_2)(F_n-F)(b_1)g(a_2,b_1)\\ &+(F_n-F)(a_1)(F_n-F)(b_1)g(a_1,b_1)\big].\end{aligned}$$

Then $\mathbb{P}$-a.s., for every $n \in \mathbb{N}$,

$$\iint g(x_1,x_2)\,\mathrm{d}(F_n-F)(x_1)\,\mathrm{d}(F_n-F)(x_2) = \iint (F_n-F)(x_1-)(F_n-F)(x_2-)\,\mathrm{d}g(x_1,x_2).$$

In part (d) of the lemma the expression "$\lim_{a_1,b_1\to-\infty,a_2,b_2\to\infty}(\ldots)$" is understood as convergence of a net $(\ldots)_{(n_1,n_2,n_3,n_4)\in\mathbb{N}^4}$, with $(-a_1,a_2,-b_1,b_2)$ playing the role of (n_1,n_2,n_3,n_4), where as usual $\mathbb{N}^4$ is regarded as a directed set w.r.t. the relation $\lhd$, and $(m_1,m_2,m_3,m_4)\lhd(n_1,n_2,n_3,n_4)$ means that $m_i \le n_i$ for $i=1,\ldots,4$. The analogous interpretations are used for the other limits.

Proof of Lemma 3.6. For two functions $f,h \in \mathbb{BV}^2_{\mathrm{loc,rc}}$ and every fixed rectangle $(a_1,a_2]\times(b_1,b_2]$, we have

$$\begin{aligned}\int_{a_1}^{a_2}\int_{b_1}^{b_2} f(x_1,x_2)\,\mathrm{d}h(x_1,x_2) = &\int_{a_1}^{a_2}\int_{b_1}^{b_2} h(x_1-,x_2-)\,\mathrm{d}f(x_1,x_2)\\ &-\int_{a_1}^{a_2} h(x_1-,b_2)\,\mathrm{d}f_{b_2}(x_1) - \int_{b_1}^{b_2} h(a_2,x_2-)\,\mathrm{d}f_{a_2}(x_2)\\ &+\int_{a_1}^{a_2} h(x_1-,a_1)\,\mathrm{d}f_{a_1}(x_1) + \int_{b_1}^{b_2} h(b_1,x_2-)\,\mathrm{d}f_{b_1}(x_2)\\ &+f(a_2,b_2)h(a_2,b_2) - f(a_2,b_1)h(a_2,b_1)\\ &-f(a_1,b_2)h(a_1,b_2) + f(a_1,b_1)h(a_1,b_1);\end{aligned}$$

see Gill, van der Laan and Wellner [14], Lemma 2.2. The remaining part of the proof is then similar to the proof of Lemma B.1 in Beutner, Wu and Zähle [4]. □

Remark 3.7. (i) Again, if for $\mathbb{P}$-almost every ω there are $x_\ell(\omega)$ and $x_u(\omega)$ as in Remark 3.5(ii), then the conditions in part (c) and (d) of Lemma 3.6 reduce to conditions on F and g. Furthermore, if such $x_\ell(\omega)$ and $x_u(\omega)$ exist, then conditions (c) and (d) of Lemma 3.6 hold whenever $\iint |\mathrm{d}g| < \infty$ as well as $\sup_{x_i\in\mathbb{R}} \int |\mathrm{d}g_{x_i}| < \infty$ for $i = 1, 2$.

(ii) If for some weight function ϕ we have that $d_\phi(F_n, F)$ is $\mathbb{P}$-a.s. finite, that the integral $\iint 1/(\phi(x_1)\phi(x_2))|\mathrm{d}g|(x_1, x_2)$ is finite, that $\lim_{x_i\to\pm\infty} 1/\phi(x_i) \int 1/\phi(x)|\mathrm{d}g_{x_i}|(x) = 0$ holds for $i = 1, 2$, and that $g(x_1, x_2)/(\phi(x_1)\phi(x_2))$ converges to zero as $|x_1|, |x_2| \to \infty$, then again conditions (c) and (d) of Lemma 3.6 hold.

(iii) Right-continuity of g_{x_1} and g_{x_2}, which is needed for the integrals in part (d) of Lemma 3.6 to be well defined, is implied by right-continuity of g.

(iv) It is worth pointing out that g_{x_1} and g_{x_2} being locally of bounded variation does not imply that $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$; see Remark 1.1 in Beutner and Zähle [7]. Moreover, $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$ does not imply that g_{x_1} and g_{x_2} are locally of bounded variation. Take, for example, the function $g : [0, 1] \times [0, 1] \to \mathbb{R}$ defined by $g(0, 0) := 0$ and $g(x_1, x_2) := x_1 \sin(\frac{\pi}{x_1})$, $(x_1, x_2) \neq (0, 0)$.

(v) If $g : \mathbb{R}^2 \to \mathbb{R}$ is continuous, the partial derivative $\partial g/\partial x_1$ exists and is continuous, and the mixed partial derivative $\partial^2 g/(\partial x_1\, \partial x_2)$ exists and is bounded on every rectangle $\mathcal{R} \subset \mathbb{R}^2$, then g is locally of bounded variation; see, for instance, Ghorpade and Limaye [13], Proposition 3.59.

(vi) If for all $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$ with $x_1 \le x_2$ and $y_1 \le y_2$ we have that

$$g(x_2, y_2) + g(x_1, y_1) \ge g(x_2, y_1) + g(x_1, y_2), \tag{8}$$

then g is locally of bounded variation. The same claim holds, if we have $\le$ instead of $\ge$ in (8). See, for instance, Ghorpade and Limaye [13], Proposition 1.15.

3.2. Examples for the representation (5)

In this section, we give some examples for set-ups under which the representation (5) holds. In the third, fourth and fifth example the set-up is degenerate because there $g_{1,F} \equiv g_{2,F} \equiv 0$. Before turning to the examples, we state two remarks including some notation needed for the examples.

Remark 3.8. Recall that two functions $f_1, f_2 \in \mathbb{BV}^2_{\mathrm{loc,rc}}$ generate the same measure on $\mathbb{R}^2$ if $f_1(x_1, x_2) = f_2(x_1, x_2) + h_1(x_1) + h_2(x_2)$ for some functions $h_1, h_2 : \mathbb{R} \to \mathbb{R}$.

Remark 3.9. For any positive measure μ on $\mathbb{R}$, and any measurable function $w : \mathbb{R} \to \mathbb{R}_+$, define the measure $\mathcal{H}^1_{w,\mu}$ on $\mathbb{R}^2$ by

$$\mathcal{H}^1_{w,\mu}(A) := \int w(x)\delta_{(x,x)}(A)\mu(\mathrm{d}x), \qquad A \in \mathcal{B}(\mathbb{R}^2), \tag{9}$$

with $\delta_{(x,x)}$ the Dirac measure at $(x,x) \in \mathbb{R}^2$. In particular, for $\mathcal{H}^1_{w,\mu}$-integrable $f : \mathbb{R}^2 \to \mathbb{R}$ we have

$$\iint f(x_1,x_2)\mathcal{H}^1_{w,\mu}\big(\mathrm{d}(x_1,x_2)\big) = \int w(x) f(x,x)\mu(\mathrm{d}x). \tag{10}$$

The $\mathcal{H}^1_{w,\mu}$-measure of the area of a rectangle $\mathcal{R}$ intersecting the diagonal $D = \{(x,x) : x \in \mathbb{R}\}$ is equal to the integral $\int_{\mathcal{R}_\pi} w(x)\mu(\mathrm{d}x)$, where $\mathcal{R}_\pi$ is the projection of $\mathcal{R}$ on one of the axes of that piece of the diagonal D that is contained in the rectangle. So one easily sees that for every $\mathcal{R}_{a,b} = (a_1,a_2] \times (b_1,b_2]$

$$\mathcal{H}^1_{w,\mu}(\mathcal{R}_{a,b}) = \begin{cases} \int_{(\max\{a_1,b_1\},\min\{a_2,b_2\})} w(x)\mu(\mathrm{d}x), & \mathcal{R}_{a,b} \cap D \neq \varnothing, \\ 0, & \text{else.} \end{cases} \tag{11}$$

If $w \equiv 1$ and μ is the Lebesgue measure ℓ on $\mathbb{R}$, then $\mathcal{H}^1_{w,\mu}$ coincides with the one-dimensional Hausdorff measure $\mathcal{H}^1$ in $\mathbb{R}^2$ restricted to the diagonal D and weighted by the constant $1/\sqrt{2}$, that is, with $\mathcal{H}^1(\cdot \cap D)/\sqrt{2}$. In this case, we also write $\mathcal{H}^1_D$ instead of $\mathcal{H}^1_{\mathbb{1},\ell}$. As special cases of (10) and (11) we obtain $\iint f(x_1,x_2)\mathcal{H}^1_D(\mathrm{d}(x_1,x_2)) = \int f(x,x)\,\mathrm{d}x$ and

$$\mathcal{H}^1_D(\mathcal{R}_{a,b}) = \begin{cases} \min\{a_2,b_2\} - \max\{a_1,b_1\}, & \mathcal{R}_{a,b} \cap D \neq \varnothing, \\ 0, & \text{else.} \end{cases}$$

Analogously, we let $\mathcal{H}^1_{\widetilde{D}}$ be the one-dimensional Hausdorff measure $\mathcal{H}^1$ in $\mathbb{R}^2$ restricted to the rotated diagonal $\widetilde{D} = \{(x,-x) : x \in \mathbb{R}\}$ and weighted by the constant $1/\sqrt{2}$, and we note that $\iint f(x_1,x_2)\mathcal{H}^1_{\widetilde{D}}(\mathrm{d}(x_1,x_2)) = \int f(x,-x)\,\mathrm{d}x$ for every $\mathcal{H}^1_{\widetilde{D}}$-integrable $f : \mathbb{R}^2 \to \mathbb{R}$.

Example 3.10 (Gini's mean difference). If $g(x_1,x_2) = |x_1 - x_2|$ and F has a finite first moment, then $V_g(F)$ equals Gini's mean difference $\mathbb{E}[|X_1 - X_2|]$ of two i.i.d. random variables X_1 and X_2 with df F. We will now verify that, if F and the estimator F_n satisfy Assumptions 3.1 and 3.2 for this g, and $d_\phi(F_n,F)$ is $\mathbb{P}$-a.s. finite for all $n \in \mathbb{N}$ and some weight function ϕ satisfying $\int 1/\phi(x)\,\mathrm{d}x < \infty$, then the assumptions of Lemmas 3.4 and 3.6 hold true and we have

$$\mathrm{d}g_{1,F}(x) = \mathrm{d}g_{2,F}(x) = \big(2F(x) - 1\big)\,\mathrm{d}x \quad \text{and} \quad \mathrm{d}g(x_1,x_2) = -2\mathcal{H}^1_D\big(\mathrm{d}(x_1,x_2)\big) \tag{12}$$

with the notation of Remark 3.9. Notice that the left-hand side in (12) shows in particular that the V-statistic corresponding to Gini's mean difference is typically non-degenerate w.r.t. (g,F) in the sense of Section 2.

It was shown in Example 3.1 in Beutner and Zähle [6] that $g_{i,F}(x) = -\mathbb{E}[X_1] + x + 2\int_x^\infty (1 - F(y))\,\mathrm{d}y$ for $i = 1,2$. Therefore, our assumption $\int 1/\phi(x)\,\mathrm{d}x < \infty$ implies $\int 1/\phi(x)|\mathrm{d}g_{i,F}|(x) < \infty$ for $i = 1,2$. From this, and using Remark 3.5(iv), we obtain the validity of assumptions (b)–(d) of Lemma 3.4. It was also established in Example 3.1 in Beutner and Zähle [6] that the left-hand side in (12) holds true. We next focus on the right-hand side in (12) and assumptions (b)–(d) of Lemma 3.6.

As for (b): It was already shown in Example 3.1 in Beutner and Zähle [6] that g_{x_1} and g_{x_2} are locally of bounded variation for the above g. Further, for an arbitrary rectangle $\mathcal{R}_{a,b} = (a_1, a_2] \times (b_1, b_2]$ with $a_2 \le b_1$ we have $|a_2 - b_2| + |a_1 - b_1| - |a_1 - b_2| - |a_2 - b_1| = 0$. The same holds for all rectangles $\mathcal{R}_{a,b} = (a_1, a_2] \times (b_1, b_2]$ with $b_2 \le a_1$. Now, consider a rectangle $\mathcal{R}_{a,b} = (a_1, a_2] \times (b_1, b_2]$ with $a_2 \ge b_2 > a_1 \ge b_1$. Then $|a_2 - b_2| + |a_1 - b_1| - |a_1 - b_2| - |a_2 - b_1| = 2(a_1 - b_2) < 0$. Similar inequalities hold for the remaining cases, that is, for $a_2 \ge b_2 > b_1 > a_1$, $b_2 \ge a_2 > b_1 \ge a_1$, and $b_2 \ge a_2 > a_1 > b_1$. Hence, Remark 3.7(vi) implies $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$.

As for (c): Let μ_g denote the measure generated by Gini's mean difference kernel. We just have seen that $\mu_g(\mathcal{R}_{a,b}) = 0$ for an arbitrary half-open rectangle $\mathcal{R}_{a,b}$ not intersecting the diagonal. Moreover, we have seen that $\mu_g(\mathcal{R}_{a,b}) = 2(a_1 - b_2)$ when $b_1 \le a_1 < b_2 \le a_2$. Taking the other possibilities mentioned above into account, we find that

$$\mu_g(\mathcal{R}_{a,b}) = \begin{cases} 2\bigl(\max\{a_1, b_1\} - \min\{a_2, b_2\}\bigr), & \mathcal{R}_{a,b} \cap D \neq \varnothing, \\ 0, & \text{else.} \end{cases}$$

Thus, in view of Remarks 3.8–3.9, the measure μ_g generated by Gini's mean difference kernel differs from $\mathcal{H}^1_D$ only by the sign and the factor 2, that is, the right-hand side in (12) holds. In view of Remark 3.7(ii), this implies condition (c) of Lemma 3.6, because we assumed $d_\phi(F_n, F) < \infty$ $\mathbb{P}$-a.s. for all $n \in \mathbb{N}$ and some weight function ϕ with $\int 1/\phi(x)\,\mathrm{d}x < \infty$.

As for (d): It was shown in Example 3.1 in Beutner and Zähle [6] that $\mathrm{d}g^+_{x_i}(x) = \mathbb{1}_{(x_i,\infty]}(x)\,\mathrm{d}x$ and $\mathrm{d}g^-_{x_i}(x) = \mathbb{1}_{[-\infty,x_i]}(x)\,\mathrm{d}x$ for $i = 1, 2$. From this and the obvious convergence of $|x_1 - x_2|/(\phi(x_1)\phi(x_2))$ to zero as $|x_1|, |x_2| \to \infty$, along with Remark 3.7(ii), it can be deduced easily that all limits in condition (d) of Lemma 3.6 exist and equal zero $\mathbb{P}$-a.s.

Example 3.11 (Variance). If $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ and F has a finite second moment, then $V_g(F)$ equals the variance of F. We will now verify that, if F and the estimator F_n satisfy Assumptions 3.1 and 3.2 for this g, and $d_\phi(F_n, F)$ is $\mathbb{P}$-a.s. finite for all $n \in \mathbb{N}$ and some weight function ϕ satisfying $\int |x|/\phi(x)\,\mathrm{d}x < \infty$, then the assumptions of Lemmas 3.4 and 3.6 hold true and we have

$$\mathrm{d}g_{1,F}(x) = \mathrm{d}g_{2,F}(x) = \bigl(x - \mathbb{E}[X_1]\bigr)\,\mathrm{d}x \quad \text{and} \quad \mathrm{d}g(x_1, x_2) = -\mathrm{d}x_1\,\mathrm{d}x_2. \tag{13}$$

Notice that the left-hand side of (13) shows in particular that the V-statistic corresponding to the variance is typically non-degenerate w.r.t. (g, F) in the sense of Section 2.

It was already verified in Example 3.2 in Beutner and Zähle [6] that $g_{i,F}(x) = \frac{1}{2}x^2 - x\mathbb{E}[X_1] + \frac{1}{2}\mathbb{E}[X_1^2]$ for $i = 1, 2$. Therefore, our assumption $\int |x|/\phi(x)\,\mathrm{d}x < \infty$ implies $\int 1/\phi(x)|\mathrm{d}g_{i,F}|(x) < \infty$ for $i = 1, 2$. Thus, $g_{i,F} \in \mathbb{BV}_{\mathrm{loc,rc}}$ and condition (c) of Lemma 3.4 follows at once. Moreover, assumption (d) of Lemma 3.4 holds by Remark 3.5(v). It was also established in Example 3.2 in Beutner and Zähle [6] that the left-hand side of (13) holds true. We next focus on the right-hand side of (13) and assumptions (b)–(d) of Lemma 3.6.

As for (b): It was already shown in Example 3.2 in Beutner and Zähle [6] that g_{x_1} and g_{x_2} are locally of bounded variation for the above g. Moreover, we have $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$ by Remark 3.7(v).

As for (c) *and* (d): Notice that $g(x_1, x_2) = -x_1x_2 + \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$ and recall Remark 3.8. Thus, up to the sign, the measure generated by the variance kernel is equal to the Lebesgue measure on $\mathbb{R}^2$, that is, the right-hand side of (13) holds. Moreover, it was verified in Example 3.2 in Beutner and Zähle [6] that $\mathrm{d}g^+_{x_i}(x) = (x - x_i)\mathbb{1}_{(x_i,\infty]}(x)\,\mathrm{d}x$ and $\mathrm{d}g^-_{x_i}(x) = (x_i - x)\mathbb{1}_{[-\infty,x_i]}(x)\,\mathrm{d}x$ for $i = 1, 2$. Thus, we see from Remark 3.7(ii), that conditions (c) and (d) of Lemma 3.6 hold.

Now, let us turn to some examples where the linear part of the representation (5) vanishes.

Example 3.12 (Gini's mean difference, degenerate case). The V-statistic $V_g(\hat{F}_n)$ corresponding to Gini's mean difference kernel $g(x_1, x_2) = |x_1 - x_2|$ is degenerate w.r.t. (g, F) for any df F that assigns probability $1/2$ to two points in $\mathbb{R}$. Indeed, we know from (12) that $\mathrm{d}g_{1,F}(x) = \mathrm{d}g_{2,F}(x) = (2F(x) - 1)\,\mathrm{d}x$ for all $x \in \mathbb{R}$, so that $\int (\hat{F}_n(x) - F(x))\,\mathrm{d}g_{i,F}(x) = 0$, $i = 1, 2$, in this case. Recall that $\hat{F}_n$ refers to the empirical df, and notice that the assumptions of Lemma 3.6 trivially hold, because $(\hat{F}_n - F)(x)$ equals zero for x small and large enough, respectively.

Example 3.13 (Goodness-of-fit test). For a given df F_0 and any measurable (weight) function $w : \mathbb{R} \to \mathbb{R}_+$, the weighted Cramér–von Mises test statistic $T^0_n := \int w(x)(\hat{F}_n(x) - F_0(x))^2\,\mathrm{d}F_0(x)$ was introduced for testing the null hypothesis $F = F_0$, and coincides with the classical Cramér–von Mises test statistic for $w \equiv 1$. The test statistic T^0_n can be expressed as V-statistic $V_g(\hat{F}_n)$ with kernel

$$g(x_1, x_2) := \int w(x)\big(\mathbb{1}_{[x_1,\infty)}(x) - F_0(x)\big)\big(\mathbb{1}_{[x_2,\infty)}(x) - F_0(x)\big)\,\mathrm{d}F_0(x). \tag{14}$$

We will now verify that, if F_0 is continuous and satisfies Assumptions 3.1 and 3.2 for this g and if the integral $\int w(x)\,\mathrm{d}F_0(x)$ is finite, then under the null hypothesis $F = F_0$ the assumptions of Lemmas 3.4 and 3.6 hold true and we have

$$g_{1,F_0} \equiv g_{2,F_0} \equiv 0 \quad \text{and} \quad \mathrm{d}g(x_1, x_2) = \mathcal{H}^1_{w,\mathrm{d}F_0}\big(\mathrm{d}(x_1, x_2)\big) \tag{15}$$

with the notation of Remark 3.9.

The left-hand side in (15) follows by using Fubini's theorem. Hence, the assumptions of Lemma 3.4 trivially hold in this case. We next focus on the right-hand side in (15) and assumptions (b)–(d) of Lemma 3.6. From (14), we easily see that under our assumptions the sections g_{x_1} and g_{x_2} are right-continuous and locally of bounded variation. Moreover, the right-hand side in (15) is known from Dehling and Taqqu [11], Example 3, and so it is apparent that $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$. Hence, (b) holds. From (15), (10), and the fact that we assumed $\int w(x)\,\mathrm{d}F_0(x)$ to exist, we immediately obtain (c). Further, we note that $\mathrm{d}g^+_{x_i}(x) = w(x)F_0(x)\,\mathrm{d}F_0(x)$ and $\mathrm{d}g^-_{x_i}(x) = w(x)\mathbb{1}_{[x_i,\infty)}(x)\,\mathrm{d}F_0(x)$ for $i = 1, 2$. From this and Remark 3.7(i), which can be applied since $\hat{F}_n$ is the empirical df, it can now be easily deduced that all limits in condition (d) of Lemma 3.6 exist and equal zero $\mathbb{P}$-a.s.

Example 3.14 (Test for symmetry). The statistic $T_n := \int_0^\infty (\hat{F}_n(-t) - [1 - \hat{F}_n(t-)])^2\,\mathrm{d}t$ is often used for testing symmetry of F about zero; cf. Arcones and Giné [2], Example 5.1. Using Fubini's theorem, more precisely Theorem 1.15 in Mattila [25] with $d\mu = d\hat{F}_n \times d\hat{F}_n$ and its analogue for negative integrands, T_n can be expressed as V-statistic $V_g(\hat{F}_n)$ with kernel

$$g(x_1,x_2) := \big(|x_1| \wedge |x_2|\big)\big(\mathbb{1}_{\{x_1,x_2>0\}} + \mathbb{1}_{\{x_1,x_2<0\}} - \mathbb{1}_{\{x_1>0,x_2<0\}} - \mathbb{1}_{\{x_1<0,x_2>0\}}\big). \tag{16}$$

We will now verify that, if F satisfies Assumptions 3.1 and 3.2 for this g, and $d_\phi(\hat{F}_n, F_0)$ is $\mathbb{P}$-a.s. finite for all $n \in \mathbb{N}$ and some symmetric weight function ϕ with $\int 1/\phi(x)\,\mathrm{d}x < \infty$, then under the null hypothesis that F be symmetric about zero the assumptions of Lemmas 3.4 and 3.6 hold true and we have

$$g_{1,F} \equiv g_{2,F} \equiv 0 \quad \text{and} \quad \mathrm{d}g(x_1,x_2) = \mathcal{H}^1_D\big(d(x_1,x_2)\big) - \mathcal{H}^1_{\widetilde{D}}\big(d(x_1,x_2)\big) \tag{17}$$

with the notation of Remark 3.9.

The left-hand side of (17) is obvious (under the null hypothesis), and so the assumptions of Lemma 3.4 trivially hold in this case. We next focus on the right-hand side of (17) and assumptions (b)–(d) of Lemma 3.6. From (16) we easily see that the sections g_{x_1} and g_{x_2} are locally of bounded variation. Moreover, we have $\mathrm{d}g^+(x_1,x_2) = \mathcal{H}^1_D(\mathrm{d}(x_1,x_2))$ and $\mathrm{d}g^-(x_1,x_2) = \mathcal{H}^1_{\widetilde{D}}(\mathrm{d}(x_1,x_2))$, and so it is apparent that $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$, and that the right-hand side of (17) holds. Hence, (b) holds. From the right-hand side of (17), (10) and our assumptions we immediately obtain (c). Further, it can be checked easily that for $i = 1, 2$

$$\mathrm{d}g_{x_i}(x) = \begin{cases} \mathbb{1}_{[0,x_i]}(x)\,\mathrm{d}x - \mathbb{1}_{[-x_i,0]}(x)\,\mathrm{d}x, & x_i > 0, \\ -\mathbb{1}_{[x_i,0]}(x)\,\mathrm{d}x + \mathbb{1}_{[0,-x_i]}(x)\,\mathrm{d}x, & x_i < 0. \end{cases}$$

From this, our assumptions and the fact that $\hat{F}_n$ is the empirical df, it can now be easily deduced that the first two limits in condition (d) of Lemma 3.6 exist and equal zero $\mathbb{P}$-a.s. Using our assumption $\int 1/\phi(x)\,\mathrm{d}x < \infty$ and Remark 3.7(ii), it follows that the third limit in condition (d) of Lemma 3.6 exists and equals zero $\mathbb{P}$-a.s.

3.3. Weak (central) limit theorems

In this section, we give a tool for deriving the asymptotic distribution of V-statistics, which is suitable for independent and weakly dependent data. Moreover, in some particular cases it also yields a non-trivial asymptotic distribution for strongly dependent data. Recall that $(\mathbf{V}, d)$ is some metric subspace of $\mathbb{D}$, and that $\mathcal{V} = \mathcal{D} \cap \mathbf{V}$.

Theorem 3.15. *Let (a_n) be a sequence in $(0,\infty)$ with $a_n \to \infty$. Assume that*:

- (a) *the assumptions of Lemmas 3.4 and 3.6 are fulfilled*,
- (b) *on $\mathbf{V}$ the functions $\Phi_{i,g}$, $i = 1, 2, 3$, defined in (6), are well-defined, $(\mathcal{V}, \mathcal{B}(\mathbb{R}))$-measurable and d-continuous*,
- (c) *the process $a_n(F_n - F)$ is a random element of $(\mathbf{V}, \mathcal{V})$ for all $n \in \mathbb{N}$, and there is some random element B° of $(\mathbf{V}, \mathcal{V})$ such that $\mathbb{P}[B^\circ \in S] = 1$ for some d-separable $S \in \mathcal{V}$ consisting of $(\mathcal{V}, d)$-completely regular points (in the sense of Definition* IV.2.6 *in Pollard* [26]) *only, and*

$$a_n(F_n - F) \xrightarrow{\mathrm{d}} B^\circ \qquad \text{in } (\mathbf{V}, \mathcal{V}, d). \tag{18}$$

Then the following assertions hold:

(i) *We always have*

$$a_n\big(V_g(F_n) - V_g(F)\big) \xrightarrow{\mathsf{d}} -\sum_{i=1}^{2} \int B^\circ(x-)\,\mathrm{d}g_{i,F}(x) \qquad in\ \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big). \tag{19}$$

(ii) *If the V-statistic $V_g(F_n)$ is degenerate w.r.t. (g, F), then we additionally have*

$$a_n^2\big(V_g(F_n) - V_g(F)\big) \xrightarrow{\mathsf{d}} \iint B^\circ(x_1-)B^\circ(x_2-)\,\mathrm{d}g(x_1, x_2) \qquad in\ \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big). \tag{20}$$

Proof. (i) Under condition (a), we can apply Lemmas 3.4 and 3.6 to obtain the representation (7). From the Continuous Mapping theorem, which is applicable by conditions (b) and (c), we obtain that $\Phi_{i,g}(a_n(F_n - F)) \xrightarrow{\mathsf{d}} -\int B^\circ(x-)\,\mathrm{d}g_{i,F}(x)$, $i = 1, 2$. From Slutsky's lemma, we obtain that $\sqrt{a_n}(F_n - F)$ converges to zero in probability, and so, according to the Continuous Mapping theorem, $\Phi_{3,g}(\sqrt{a_n}(F_n - F))$ converges in probability to zero as well. Applying once again Slutsky's lemma finishes the proof of part (i).

(ii) If the V-statistic is degenerate, then we obtain analogously by applying Lemmas 3.4 and 3.6 that $a_n^2(V_g(F_n) - V_g(F)) = \Phi_{3,g}(a_n(F_n - F))$. The result is now immediate from the Continuous Mapping theorem. □

Remark 3.16. If, for some weight function ϕ, the integral $\int 1/\phi(x)|\mathrm{d}g_{i,F}|(x)$ is finite for $i = 1, 2$, then the mappings $\Phi_{1,g}$ and $\Phi_{2,g}$ are obviously d_ϕ-continuous. Moreover, if the integral $\int 1/(\phi(x_1)\phi(x_2))|\mathrm{d}g|(x_1, x_2)$ is finite, then $\Phi_{3,g}$ is d_ϕ-continuous, too.

Example 3.17 (I.i.d. data). Suppose $X_1, X_2, \ldots$ are i.i.d. with df F, and let ϕ be a weight function. If $\int \phi^2\,\mathrm{d}F < \infty$, then Theorem 6.2.1 in Shorack and Wellner [30] shows that $\sqrt{n}(\hat{F}_n - F) \xrightarrow{\mathsf{d}} B_F^\circ$ (in $(\mathbb{D}_\phi, \mathcal{D}_\phi, d_\phi)$) with $\hat{F}_n$ the empirical df of $X_1, \ldots, X_n$ and B_F° an F-Brownian bridge, that is, a centered Gaussian process with covariance function $\Gamma(x, y) = F(x \wedge y)\overline{F}(x \vee y)$.

Example 3.18 (Weakly dependent data). Let (X_i) be α-mixing with mixing coefficients satisfying $\alpha(n) = \mathrm{O}(n^{-\theta})$ for some $\theta > 1 + \sqrt{2}$, and let $\lambda \geq 0$. If F has a finite γ-moment for some $\gamma > (2\theta\lambda)/(\theta - 1)$, then it can easily be deduced from Theorem 2.2 in Shao and Yu [29] that $\sqrt{n}(\hat{F}_n - F) \xrightarrow{\mathsf{d}} \widetilde{B}_F^\circ$ (in $(\mathbb{D}_{\phi_\lambda}, \mathcal{D}_{\phi_\lambda}, d_{\phi_\lambda})$) with $\widetilde{B}_F^\circ$ a continuous centered Gaussian process with covariance function $\Gamma(y_0, y_1) = F_0(y_0 \wedge y_1)\overline{F}_0(y_0 \vee y_1) + \sum_{i=0}^{1}\sum_{k=2}^{\infty} \mathbb{C}\mathrm{ov}(\mathbb{1}_{\{X_1 \leq y_i\}}, \mathbb{1}_{\{X_k \leq y_{1-i}\}})$; see Section 3.3 in Beutner and Zähle [5]. If (X_i) is even β- or ρ-mixing, then Lemma 4.1 in Chen and Fan [8] and Theorem 2.3 in Shao and Yu [29] ensure that the mixing condition can be relaxed; see also Section 3.2 in Beutner and Zähle [6].

Part (ii) of Theorem 3.15 also leads to some interesting corollaries that provide convenient alternatives to the common approach to derive the asymptotic distribution of degenerate U-statistics. Before presenting them, it is worth recalling that the common approach to derive the asymptotic distribution of degenerate U-statistics is based on a series expansion of the kernel g of the form $g(x_1, x_2) = \sum_{k=1}^{\infty} \lambda_k \psi_k(x_1)\psi_k(x_2)$, where the λ_k are real numbers and the ψ_k are an orthonormal sequence; see, for example, Serfling [28], Section 5.5. The λ_k and the ψ_k are the eigenvalues and eigenfunctions, respectively, of the operator $A : L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}) \to L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ defined by $A(h(x_1)) = \int g(x_1, x_2)h(x_2)\, \mathrm{d}F(x_2)$. The eigenvalues arise in the asymptotic distribution of $n(U_{g,n} - V_g(F))$ which, in the i.i.d. case, is given by $\sum_{i=1}^{\infty} \lambda_i(\xi_i^2 - 1)$, where the ξ_i are independent and have a χ^2-distribution with 1 degree of freedom.

Corollary 3.19. *Assume that the conditions of Theorem 3.15 hold and that we are in case* (ii) *of this theorem. Moreover, let the* (*possibly signed*) *measure generated by g on $\mathbb{R}^2$ be equal to the product measure of* (*possible signed*) *measures ν_1 and ν_2. Then*

$$a_n^2\big(V_g(F_n) - V_g(F)\big) \xrightarrow{\mathsf{d}} \left(\int B^\circ(x-)\,\mathrm{d}\nu_1(x)\right)\left(\int B^\circ(x-)\,\mathrm{d}\nu_2(x)\right) \qquad \text{in } \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big).$$

The next example shows that two well known kernels are covered by Corollary 3.19.

Example 3.20. (i) The variance kernel $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2 = -x_1x_2 + \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$ is degenerate if and only if the fourth central moment equals the squared second central moment; see, for example, van der Vaart [31], Example 12.12. Moreover, in Example 3.11 we have seen that the measure generated by the variance kernel coincides with the negative Lebesgue measure on $\mathbb{R}^2$. So, in the degenerate case, the variance kernel can be treated by means of Corollary 3.19.

(ii) The kernel $g(x_1, x_2) = x_1x_2$, which corresponds to the characteristic $\mathbb{E}[X_1]^2$ and which is degenerate if the first moment equals zero, obviously generates the Lebesgue measure. In particular, up to the sign, it generates the same measure on $\mathbb{R}^2$ as the variance kernel. So, in the degenerate case, this kernel can be treated by means of Corollary 3.19 as well. Of course, for the corresponding V-statistics the asymptotic distributions can be derived differently, but the continuous mapping approach reveals an interesting relation to the variance kernel.

Corollary 3.21. *Assume that the conditions of Theorem 3.15 hold and that we are in case* (ii) *of this theorem. Moreover, let the measure generated by g be given by $\mathcal{H}^1_{w,\mu}$ defined in* (9). *Then*

$$a_n^2\big(V_g(F_n) - V_g(F)\big) \xrightarrow{\mathsf{d}} \int w(x)\big(B^\circ(x-)\big)^2\mu(\mathrm{d}x) \qquad \text{in } \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big).$$

Here are two examples to which Corollary 3.21 can be applied.

Example 3.22. (i) In Example 3.12, we have seen that Gini's mean difference is degenerate for a df that assigns probability $1/2$ to two points in $\mathbb{R}$. Further, from Example 3.10 we also know that the measure generated by g differs from $\mathcal{H}^1_{\mathbb{1},\ell} = \mathcal{H}^1_D$ only by a constant factor. So, Corollary 3.21 can be applied.

(ii) In Example 3.13, we have seen that the measure generated by the kernel g of the Cramér–von Mises statistic (cf. (14)) equals the measure $\mathcal{H}^1_{w,\mathrm{d}F}$. Thus, the Cramér–von Mises statistic can also be treated by means of Corollray 3.21. For the particular case $w \equiv 1$ see also van der Vaart [31], Corollary 19.21.

Although, the asymptotic distributions in Example 3.22 can be derived differently, the two examples given are appealing from a structural point of view.

3.4. Strong limit theorems

In this section, we focus on almost sure convergence of the plug-in estimator $V_g(F_n)$ to $V_g(F)$. Assume that the representation (5) holds, that the mapping $\Phi : \mathbf{V} \to \mathbb{R}$, $f \mapsto (\sum_{i=1}^2 \Phi_{i,g}(f) + \Phi_{3,g}(f))$, is d-continuous at the null function, and that $F_n - F$ converges $\mathbb{P}$-a.s. to the null function w.r.t. d. Then we immediately obtain almost sure convergence of $V_g(F_n)$ to $V_g(F)$. From the following obvious theorem, we can even deduce the rate of convergence. By local β-Hölder d-continuity of a functional $\Phi : \mathbf{V} \to \mathbb{R}$ at f, we mean that $|\Phi(f_n) - \Phi(f)| = \mathrm{O}(d(f_n, f)^\beta)$ for each sequence $(f_n) \subset \mathbf{V}$ with $d(f_n, f) \to 0$.

Theorem 3.23. *Let $\phi : \mathbb{R} \to [1, \infty)$ be some weight function, let d be homogeneous, and assume that*:

(a) *the assumptions of Lemmas 3.4 and 3.6 are fulfilled*,
(b) *on $\mathbf{V}$ the functions $\Phi_{i,g}$, $i = 1, 2, 3$, defined in (6), are well defined and $(\mathcal{V}, \mathcal{B}(\mathbb{R}))$-measurable, and the function $\sum_{i=1}^3 \Phi_{i,g}$ is locally β-Hölder d-continuous at the null function for some $\beta > 0$*,
(c) *the process $F_n - F$ is a random element of $(\mathbf{V}, \mathcal{V})$ for all $n \in \mathbb{N}$, and, for some sequence (a_n) in $(0, \infty)$*,

$$a_n d(F_n - F, 0) \longrightarrow 0 \qquad \mathbb{P}\text{-}a.s. \tag{21}$$

Then

$$a_n^\beta \big(V_g(F_n) - V_g(F)\big) \longrightarrow 0 \qquad \mathbb{P}\text{-}a.s.$$

Remark 3.24. If, for some weight function ϕ, the integral $\int \phi(x)|\mathrm{d}g_{i,F}|(x)$ is finite for $i = 1, 2$, then the functionals $\Phi_{1,g}$ and $\Phi_{2,g}$ are obviously locally 1-Hölder d_ϕ-continuous at the null function. Moreover, if the integral $\int 1/(\phi(x_1)\phi(x_2))|\mathrm{d}g|(x_1, x_2)$ is finite, then the functional $\Phi_{3,g}$ is obviously 2-Hölder d_ϕ-continuous at the null function. Thus, in this case the functional $\sum_{i=1}^3 \Phi_{i,g}$ is locally 1-Hölder d_ϕ-continuous at the null function, and the rate of convergence of degenerate V-statistics w.r.t. (g, F), that is, of V-statistics with $\sum_{i=1}^2 \Phi_i(F_n - F) = 0$, is twice the rate of non-degenerate V-statistics w.r.t. (g, F).

Example 3.25. (i) Let ϕ be any weight function, and $r \in [0, \frac{1}{2})$. If the sequence (X_i) is i.i.d. and $\int \phi^{1/(1-r)}\,\mathrm{d}F < \infty$, then (21) hold for the weighted sup-metric $d := d_\phi$, the empirical df $F_n := \hat{F}_n$, and $a_n := n^r$; cf. Andersen, Giné and Zinn [1], Theorem 7.3.

(ii) Suppose that $\int \phi\,\mathrm{d}F < \infty$. Further suppose that (X_i) is α-mixing with mixing coefficients $\alpha(n)$, let $\alpha(t) := \alpha(\lfloor t \rfloor)$ be the càdlàg extension of $\alpha(\cdot)$ from $\mathbb{N}$ to $\mathbb{R}_+$, and assume that $\int_0^1 \log(1 + \alpha^{\rightarrow}(s/2))\overline{G}^{\rightarrow}(s)\,\mathrm{d}s < \infty$ for $\overline{G} := 1 - G$, where G denotes the df of $\phi(X_1)$ and $\overline{G}^{\rightarrow}$ the right-continuous inverse of $\overline{G}$. It was shown in Zähle [38] that, under the imposed assumptions, (21) holds for the weighted sup-metric $d := d_\phi$, the empirical df $F_n := \hat{F}_n$, and $a_n := 1$.

Notice that the integrability condition above holds in particular if $\mathbb{E}[\phi(X_1)\log^+\phi(X_1)] < \infty$ and $\alpha(n) = \mathrm{O}(n^{-\vartheta})$ for some arbitrarily small $\vartheta > 0$; cf. Rio [27], Application 5, page 924.

(iii) Suppose that the sequence (X_i) is α-mixing with mixing coefficients $\alpha(n)$. Let $r \in [0, \frac{1}{2})$ and assume that $\alpha(n) \le Kn^{-\vartheta}$ for all $n \in \mathbb{N}$ and some constants $K > 0$ and $\vartheta > 2r$. Then (21) holds for the uniform sup-metric $d := d_\infty$, the empirical df $F_n := \hat{F}_n$, and $a_n := n^r$; cf. Zähle [38].

4. The use of the representation (5) for linear long-memory sequences

As indicated in the Introduction and in Section 2, for sequences exhibiting long-range dependence it may happen that the linear part of the von Mises decomposition degenerates only asymptotically. In such a case, Theorem 3.15 may not yield a non-central limit theorem; for more details see the discussion below just after the proof of Theorem 4.1. Nevertheless, representation (5), the Continuous Mapping theorem, and an "expansion" of the empirical process will lead to a general result to derive non-central limit theorems for U- and V-statistics based on linear long-memory sequences. Thus, in this section, we shall consider a linear process exhibiting long-range dependence (strong dependence), that is,

$$X_t := \sum_{s=0}^{\infty} a_s \varepsilon_{t-s}, \qquad t \in \mathbb{N},$$

where $(\varepsilon_i)_{i\in\mathbb{Z}}$ are i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with zero mean and finite variance, and the coefficients a_s satisfy $\sum_{s=0}^{\infty} a_s^2 < \infty$ (so that $(X_t)_{t\in\mathbb{N}}$ is an L^2-process) and decay sufficiently slowly so that $\sum_{t=1}^{\infty} |\mathbb{C}\mathrm{ov}(X_1, X_t)| = \infty$. The latter divergence gives the precise meaning to the attribute long-range dependence. Notice that if ε_1 has a finite pth moment for some $p \ge 2$, then the same holds for X_1. As before, we denote by F the df of the X_t.

For $n \in \mathbb{N}$ and $p \in \mathbb{N}_0$, assume that the pth moment of F is finite and that F can be differentiated at least p times. Denote the jth derivative of F by $F^{(j)}$, $j = 0, \dots, p$, with the convention $F^{(0)} = F$, and define a stochastic process $\mathcal{E}_{n,p;F}$ with index set $\mathbb{R}$ by

$$\begin{aligned}\mathcal{E}_{n,p;F}(\cdot) &:= \hat{F}_n(\cdot) - \sum_{j=0}^{p} (-1)^j F^{(j)}(\cdot)\left(\frac{1}{n}\sum_{i=1}^{n} A_{j;F}(X_i)\right) \\ &= \hat{F}_n(\cdot) - F(\cdot) - \sum_{j=1}^{p} (-1)^j F^{(j)}(\cdot)\left(\frac{1}{n}\sum_{i=1}^{n} A_{j;F}(X_i)\right),\end{aligned} \tag{22}$$

where $A_{j;F}$ denotes the jth order Appell polynomial associated with F, and we use the convention $\sum_{j=1}^{0}(\cdots) := 0$. Recall that these Appell polynomials are defined by $A_{0;F}(x) := 1$ and for

$j = 1, \dots, p$ recursively by the characteristic conditions

$$\frac{\mathrm{d}}{\mathrm{d}x} A_{j;F}(x) = j A_{j-1;F}(x) \quad \text{and} \quad \int A_{j;F}(y)\,\mathrm{d}F(y) = 0.$$

In particular, $\mathcal{E}_{n,0;F}(\cdot) = (\hat{F}_n(\cdot) - F(\cdot))$ and $\mathcal{E}_{n,1;F}(\cdot) = (\hat{F}_n(\cdot) - F(\cdot)) + F^{(1)}(\cdot)(\frac{1}{n}\sum_{i=1}^n X_i)$. For $p \in \mathbb{N}$, we obviously have

$$\mathcal{E}_{n,p-1;F}(\cdot) = \mathcal{E}_{n,p;F}(\cdot) + (-1)^p F^{(p)}(\cdot)\left(\frac{1}{n}\sum_{i=1}^n A_{p;F}(X_i)\right), \tag{23}$$

and we note that under a suitable re-scaling the limit in distribution, $Z_{p,\beta}$, of the normalized sum $\frac{1}{n}\sum_{i=1}^n A_{p;F}(X_i)$ has been established by Avram and Taqqu [3] for $1 \le p < 1/(2\beta - 1)$ (for the meaning of β see Theorem 4.1 below). So, whenever the process $\mathcal{E}_{n,p;F}(\cdot)$ can be shown to converge in probability to zero under the same re-scaling, we obtain that the limit in distribution of a re-scaled version of the process $\mathcal{E}_{n,p-1;F}(\cdot)$ is given by $(-1)^p F^{(p)}(\cdot) Z_{p,\beta}$. This idea is basically due to Dehling and Taqqu [10] who considered the Gaussian case and the uniform sup-metric d_∞. For the linear process and the uniform sup-metric d_∞ this approach was used by Ho and Hsing [17] and Wu [34] for arbitrary $p \ge 1$, and by Giraitis and Surgailis [15] for $p = 1$. Wu [34] also considered bounds for the second moment of weighted sup-norms of the leading term of $\mathcal{E}_{n,p-1;F}(\cdot)$. For the linear process and the weighted sup-metric d_{ϕ_λ} the approach of Dehling and Taqqu [10] was applied to the case $p = 1$ by Beutner, Wu and Zähle [4]. In the following theorem, we generalize the latter to arbitrary $p \ge 1$.

Theorem 4.1. *Let $p \in \mathbb{N}$, $\lambda \ge 0$, and assume that*:

(a) $a_s = s^{-\beta}\ell(s)$, $s \in \mathbb{N}$, *where $\beta \in (\frac{1}{2}, 1)$ and ℓ is slowly varying at infinity.*
(b) $\mathbb{E}[|\varepsilon_1|^{(4+2\lambda)\vee(2p)}] < \infty$.
(c) *The df G of ε_1 is $p+1$ times differentiable and* $\sum_{j=1}^{p+1}\int_{\mathbb{R}} |G^{(j)}(x)|^2 \phi_{2\lambda}(x)\,\mathrm{d}x < \infty$.
(d) $p(2\beta - 1) < 1$.

Then

$$\{n^{p(\beta-1/2)}\ell(n)^{-p}\}\mathcal{E}_{n,p-1;F}(\cdot) \xrightarrow{\mathrm{d}} (-1)^p F^{(p)}(\cdot) Z_{p,\beta} \qquad \big(\textit{in } (\mathbb{D}_{\phi_\lambda}, \mathcal{D}_{\phi_\lambda}, d_{\phi_\lambda})\big),$$

where

$$Z_{p,\beta} := c_{p,\beta}\int_{-\infty<u_1<\cdots<u_p<1}\left\{\int_0^1 \mathbb{1}_{(u_p,1)}(v)\prod_{j=1}^p (v-u_j)^{-\beta}\,\mathrm{d}v\right\} W(\mathrm{d}u_1)\cdots W(\mathrm{d}u_p)$$

with W a white noise measure (i.e., an additive Gaussian random set function satisfying $\mathbb{E}[W(B)] = 0$ and $\mathbb{E}[W(B) \cap W(B')] = |B \cap B'|$ for all $B, B' \in \mathcal{B}(\mathbb{R})$) and

$$c_{p,\beta} := \left(\frac{p!(1 - p(\beta - 1/2))(1 - p(2\beta - 1))}{\int_0^\infty (x + x^2)^{-\beta}\,\mathrm{d}x}\right)^{1/2}.$$

Remark 4.2. (i) The infinite moving average representation of an ARFIMA(p, d, q) process with fractional difference parameter $d \in (0, 1/2)$ satisfies assumption (a) with $\beta = 1 - d$; see, for instance, Hosking [19], Section 3.

(ii) Here we have chosen to define the stochastic process (22) in terms of $\hat{F}_n$, F and the Appell polynomials of F, because this allows us later to define the statistic (27) in terms of the df of the observables. However, we conjecture that assumption (b) can be relaxed to $\mathbb{E}[|\varepsilon_1|^{(2+2\lambda)}] < \infty$ by replacing as in Wu [34,35] the Appell polynomials $A_{j;F}(X_i)$ by the expressions $A_{j;F}^{=(1,\dots,1)}(X_i)$ (to be introduced at the beginning of the proof of Theorem 4.1) and by the method of proof as in Beutner, Wu and Zähle [4] where this was done for $p = 1$.

(iii) Condition (c) implies in particular that the df F of X_1 is p times differentiable with $F^{(p)} \in \mathbb{D}_{\phi_\lambda}$; cf. inequality (30) in Wu [34] with $n = \infty$, $\kappa = 1$ and $\gamma = 2\lambda$. Further, assumption (c) can be relaxed in that it suffices to require that there is some $m \in \mathbb{N}$ such that the df G_m of $\sum_{s=0}^{m-1} a_s \varepsilon_{m-s}$ is $p+1$ times differentiable and satisfies $\sum_{j=1}^{p+1} \int_{\mathbb{R}} |G_m^{(j)}(x)|^2 \phi_{2\lambda}(x)\,\mathrm{d}x < \infty$. The proof still works in this setting; see also Wu [34].

Proof of Theorem 4.1. It was shown by Avram and Taqqu [3], Theorem 1, that the pth Appell polynomial of F evaluated at X_i has the representation

$$\begin{aligned}
A_{p;F}(X_i) &= \sum_{\ell=1}^{p} \sum_{q(\ell)\in\Pi_{\ell,p}} \frac{p!}{q_1!\cdots q_\ell!} \sum_{m(\ell)\in\Lambda_{q(\ell)}} \prod_{k=1}^{\ell} a_{m_k}^{q_k} A_{q_k;G}(\varepsilon_{i-m_k}) \\
&= \sum_{m(p)\in\Lambda_{q(p)}} p! \prod_{k=1}^{p} a_{m_k} A_{q_k;G}(\varepsilon_{i-m_k}) \\
&\quad + \sum_{\ell=1}^{p-1} \sum_{q(\ell)\in\Pi_{\ell,p}} \frac{p!}{q_1!\cdots q_\ell!} \sum_{m(\ell)\in\Lambda_{q(\ell)}} \prod_{k=1}^{\ell} a_{m_k}^{q_k} A_{q_k;G}(\varepsilon_{i-m_k}) \\
&=: A_{p;F}^{=(1,\dots,1)}(X_i) + A_{p;F}^{\neq(1,\dots,1)}(X_i),
\end{aligned}$$

where $A_{q_k;G}$ denotes the q_kth Appell polynomial of the df G of ε_1, and, for every $\ell \in \{1, \dots, p\}$, $\Pi_{\ell,p}$ is the set of all $q(\ell) = (q_1, \dots, q_\ell) \in \mathbb{N}^\ell$ satisfying $q_1 + \cdots + q_\ell = p$ and $1 \le q_1 \le \cdots \le q_\ell$. Moreover, for a given $q(\ell) = (q_1, \dots, q_\ell)$ we denote by $\Lambda_{q(\ell)}$ the set of all $m(\ell) = (m_1, \dots, m_\ell) \in \mathbb{N}_0^\ell$ such that $m_i \neq m_j$ for $i \neq j$ and, in addition, if $q_i = q_{i+1}$, then $m_i < m_{i+1}$. So, introducing a telescoping sum, we obtain from (23)

$$\begin{aligned}
\mathcal{E}_{n,p-1;F}(\cdot) &= \left\{ \mathcal{E}_{n,p;F}(\cdot) + \sum_{j=1}^{p} (-1)^j F^{(j)}(\cdot) \left(\frac{1}{n} \sum_{i=1}^{n} A_{j;F}^{\neq(1,\dots,1)}(X_i) \right) \right\} \\
&\quad - \sum_{j=1}^{p} (-1)^j F^{(j)}(\cdot) \left(\frac{1}{n} \sum_{i=1}^{n} A_{j;F}^{\neq(1,\dots,1)}(X_i) \right)
\end{aligned}$$

$$
\begin{aligned}
&+ (-1)^p F^{(p)}(\cdot)\left(\frac{1}{n}\sum_{i=1}^n A_{p;F}(X_i)\right)\\
&= \left\{\hat{F}_n(\cdot) - F(\cdot) - \sum_{j=1}^p (-1)^j F^{(j)}(\cdot)\left(\frac{1}{n}\sum_{i=1}^n A_{j;F}^{=(1,\ldots,1)}(X_i)\right)\right\}\\
&\quad - \sum_{j=1}^{p-1}(-1)^j F^{(j)}(\cdot)\left(\frac{1}{n}\sum_{i=1}^n A_{j;F}^{\neq(1,\ldots,1)}(X_i)\right)\\
&\quad + (-1)^p F^{(p)}(\cdot)\left(\frac{1}{n}\sum_{i=1}^n A_{p;F}^{=(1,\ldots,1)}(X_i)\right)\\
&=: S_{n,p}^0(\cdot) + T_{n,p}(\cdot) + U_{n,p}(\cdot).
\end{aligned}
$$

Under assumptions (a), $\mathbb{E}[|\varepsilon_0|^{2p}] < \infty$, and (d), it follows from Step 3 in the proof of Theorem 2 of Avram and Taqqu [3] that the expression

$$
\left\{n^{p(\beta-1/2)}\ell(n)^{-p}\right\}\frac{1}{n}\sum_{i=1}^n A_{j;F}^{\neq(1,\ldots,1)}(X_i) = \frac{1}{n^{1-p(\beta-1/2)}\ell(n)^p}\sum_{i=1}^n A_{j;F}^{\neq(1,\ldots,1)}(X_i)
$$

converges in probability to zero for every $j = 1, \ldots, p$. So, in view of $F^{(j)} \in \mathbb{D}_{\phi_\lambda}$, $j = 1, \ldots, p$, we obtain

$$
\left\{n^{p(\beta-1/2)}\ell(n)^{-p}\right\} \mathrm{d}_{\phi_\lambda}\big(T_{n,p}(\cdot), 0\big) \xrightarrow{\mathsf{p}} 0.
$$

Avram and Taqqu [3], Theorem 2, also showed that under the same assumptions the expression

$$
\left\{n^{p(\beta-1/2)}\ell(n)^{-p}\right\}\frac{1}{n}\sum_{i=1}^n A_{p;F}^{=(1,\ldots,1)}(X_i) = \frac{1}{n^{1-p(\beta-1/2)}\ell(n)^p}\sum_{i=1}^n A_{p;F}^{=(1,\ldots,1)}(X_i)
$$

converges in distribution to $Z_{p,\beta}$; for the shape of the normalizing constant $c_{p,\beta}$ see Ho and Hsing [17], Lemma 6.1. So, in view of $F^{(p)} \in \mathbb{D}_{\phi_\lambda}$, the process $U_{n,p}(\cdot)$ converges in distribution to $(-1)^p F^{(p)}(\cdot) Z_{p,\beta}$ w.r.t. d_{ϕ_λ}. In the remainder of the proof, we will show that

$$
\left\{n^{p(\beta-1/2)}\ell(n)^{-p}\right\} d_{\phi_\lambda}\big(S_{n,p}^0(\cdot), 0\big) \xrightarrow{\mathsf{p}} 0 \tag{24}
$$

so that assertion of Theorem 4.1 will follow from Slutzky's lemma.

For (24) to be true, it suffices to show that $d_{\phi_\lambda}(\frac{S_{n,p}(\cdot)}{\sigma_{n,p}}, 0)$ converges in probability to zero, where $S_{n,p} := nS_{n,p}^0$ and $\sigma_{n,p} := n^{1-p(\beta-1/2)}\ell(n)^p$. Under the assumptions (a), $\mathbb{E}[|\varepsilon_0|^{(4+2\lambda)}] < \infty$, (c) and (d), we have from Theorem 2 and Lemma 5 of Wu [34] that

$$
\mathbb{E}\big[d_{\phi_{2\lambda}}\big(S_{n,p}(\cdot)^2, 0\big)\big] = \mathrm{O}\big(n(\log n)^2 + \Xi_{n,p}\big) \tag{25}
$$

with $\Xi_{n,p} = \mathrm{O}(n^{2-(p+1)(2\beta-1)}\ell(n)^{2(p+1)})$ (notice that there is a typo in Lemma 5 of Wu [34] where it must be $p(2\beta-1) < 1$ instead of $(p+1)(2\beta-1) < 1$). From (25), we obtain by the Markov inequality for some constant $C > 0$ and every $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}\big[\sigma_{n,p}^{-1} d_{\phi_\lambda}\big(S_{n,p}(\cdot), 0\big) > \varepsilon\big] &= \mathbb{P}\big[\sigma_{n,p}^{-2} d_{\phi_{2\lambda}}\big(S_{n,p}(\cdot)^2, 0\big) > \varepsilon^2\big] \\ &\leq \frac{1}{\varepsilon^2} \frac{\mathbb{E}[d_{\phi_{2\lambda}}(S_{n,p}(\cdot)^2, 0)]}{\sigma_{n,p}^2} \\ &\leq C\varepsilon^{-2} \frac{n(\log n)^2 + n^{2-(p+1)(2\beta-1)}\ell(n)^{2(p+1)}}{n^{2-p(2\beta-1)}\ell(n)^{2p}} \\ &\leq C\varepsilon^{-2}\bigg(\frac{(\log n)^2}{n^{1-p(2\beta-1)}} + \frac{1}{n^{2\beta-1}}\bigg)\ell(n)^2.\end{aligned}$$

Due to assumption (d), the latter bound converges to zero as $n \to \infty$. That is, $d_{\phi_\lambda}(\frac{S_{n,p}(\cdot)}{\sigma_{n,p}}, 0)$ indeed converges in probability to zero. $\square$

Combining Theorems 3.15 and 4.1, one can in principle easily derive the asymptotic distribution of non-degenerate and degenerate V-statistics based on linear long-memory sequences. For non-degenerate V-statistics (as, e.g., Gini's mean difference from Example 3.10) one can apply part (i) of Theorem 3.15; see also Hsing [20] who uses a different approach for Gini's mean difference. For degenerate V-statistics (as, e.g., the Cramér–von Mises statistic from Example 3.13), one can apply part (ii) of Theorem 3.15. However, in the long-memory case the situation is often more complex because several V-statistics based on long-memory sequences systematically degenerate asymptotically. For instance, the (sample) variance with corresponding kernel $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ is typically non-degenerate w.r.t. (g, F) (cf. Example 3.11), but in this case the integral on the right-hand side in (19) with $B^\circ(\cdot) = (-1)F^{(1)}(\cdot)Z_{1,\beta}$ equals

$$-\sum_{i=1}^{2} \int B^\circ(x-)\,\mathrm{d}g_{i,F}(x) = Z_{1,\beta} \sum_{i=1}^{2} \int F^{(1)}(x-)\big(x - \mathbb{E}[X_1]\big)\,\mathrm{d}x = 0.$$

Indeed, from Example 3.11 we know that in this case $\mathrm{d}g_{1,F}(x) = \mathrm{d}g_{2,F}(x) = (x - \mathbb{E}[X_1])\,\mathrm{d}x$ holds, and hence $\int F^{(1)}(x-)(x - \mathbb{E}[X_1])\,\mathrm{d}x = \int F^{(1)}(x)(x - \mathbb{E}[X_1])\,\mathrm{d}x = 0$. That is, in the long-memory case the sample variance regarded as a V-statistic is asymptotically degenerate w.r.t. $(g, F, (n^{(\beta-1/2)}\ell(n)^{-1})_n)$ in the sense of Section 2, and so an application of part (i) of Theorem 3.15 yields little. Moreover, part (ii) of Theorem 3.15 is useful neither in this case, because part (ii) of Theorem 3.15 is based on the fact that the linear part in the representation (5) vanishes. However, this is not the case here, since the sample variance is not (finite sample) degenerate w.r.t. (g, F). This is in accordance with the remarkable observation of Dehling and Taqqu [11] that in the long-memory case both terms of the von Mises (resp., Hoeffding) decomposition of the sample variance contribute to the asymptotic distribution. Dehling and Taqqu [11] considered the sample variance based on Gaussian long-memory sequences. From the following Corollary 4.3, we cannot only derive the analogue for linear long-memory sequences (see

Example 4.7), but can also derive the asymptotic distribution of more general asymptotically degenerate U- and V-statistics based on linear long-memory sequences (see, e.g., Examples 4.6, 4.8 and 4.9). We note that recently Lévy-Leduc *et al.* [23] also derived the asymptotic distribution of some asymptotically degenerate U-statistics (with bounded kernels) based on Gaussian long-memory sequences using different techniques. For further applications of their results, see Lévy-Leduc *et al.* [24].

Corollary 4.3. *Let F be a df on the real line, and $g:\mathbb{R}^2 \to \mathbb{R}$ be some measurable function. Assume that the representation* (5) *with $F_n := \hat{F}_n$ holds for F and g, and that*

$$\sum_{i=1}^{2} \int \phi_{-\lambda}(x)|\mathrm{d}g_{i,F}|(x) < \infty \quad \textit{and} \quad \iint \phi_{-\lambda}(x_1)\phi_{-\lambda}(x_2)|\mathrm{d}g|(x_1, x_2) < \infty \tag{26}$$

holds for some $\lambda \geq 0$. Let $p, q, r \in \mathbb{N}$, set

$$\begin{aligned}
&\mathcal{V}_{n,g;p,q,r}(\hat{F}_n) \\
&\quad := V_g(\hat{F}_n) - V_g(F) + \sum_{\ell=1}^{2}\sum_{j=1}^{p-1}(-1)^j\left(\frac{1}{n}\sum_{i=1}^{n} A_{j;F}(X_i)\right)\int F^{(j)}(x-)\,\mathrm{d}g_{\ell,F}(x) \\
&\qquad - \sum_{j=1}^{q-1}(-1)^j\left(\frac{1}{n}\sum_{i=1}^{n} A_{j;F}(X_i)\right)\iint F^{(j)}(x_1-)\big(\hat{F}_n(x_2-) - F(x_2-)\big)\,\mathrm{d}g(x_1, x_2) \\
&\qquad - \sum_{k=1}^{r-1}(-1)^k\left(\frac{1}{n}\sum_{i=1}^{n} A_{k;F}(X_i)\right)\iint \big(\hat{F}_n(x_1-) - F(x_1-)\big)F^{(k)}(x_2-)\,\mathrm{d}g(x_1, x_2) \\
&\qquad + \sum_{j=1}^{q-1}\sum_{k=1}^{r-1}(-1)^{j+k}\left(\frac{1}{n}\sum_{i=1}^{n} A_{j;F}(X_i)\right)\left(\frac{1}{n}\sum_{i=1}^{n} A_{k;F}(X_i)\right) \\
&\qquad\qquad \times \iint F^{(j)}(x_1-)F^{(k)}(x_2-)\,\mathrm{d}g(x_1, x_2)
\end{aligned} \tag{27}$$

(*with the convention $\sum_{j=1}^{0}(\cdots) := 0$*), *and assume that all integrals on the right-hand side in* (27) *are well defined* (*which is in particular the case if F is* $\max\{p, q, r\}$ *times differentiable with $F^{(k)} \in \mathbb{D}_{\phi_\lambda}$ for all $k = 0, \ldots, \max\{p, q, r\}$*).

(i) *Assume $q + r > p$ and that the assumptions* (a)–(c) *of Theorem 4.1 with p replaced by* $\max\{p, q, r\} < \infty$ *hold for the same λ. Then, if in addition $s(2\beta - 1) < 1$ holds for each $s \in \{p, q, r\}$, we have with $Z_{p,\beta}$ defined as in Theorem 4.1*

$$\{n^{p(\beta-1/2)}\ell(n)^{-p}\}\mathcal{V}_{n,g;p,q,r}(\hat{F}_n) \xrightarrow{\mathrm{d}} (-1)^p Z_{p,\beta} \sum_{\ell=1}^{2} \int F^{(p)}(x-)\,\mathrm{d}g_{\ell,F}(x).$$

(ii) *Assume* $q + r = p$ *and that the assumptions* (a)–(d) *of Theorem* 4.1 *hold for the same* λ *and* p. *Then we have with* $Z_{s,\beta}$ *defined as in Theorem* 4.1 *for* $s \in \{p, q, r\}$

$$\begin{aligned}
&\{n^{p(\beta-1/2)}\ell(n)^{-p}\}\mathcal{V}_{n,g;p,q,r}(\hat{F}_n) \\
&\quad\xrightarrow{\mathsf{d}} (-1)^p Z_{p,\beta}\sum_{\ell=1}^{2}\int F^{(p)}(x-)\,\mathrm{d}g_{\ell,F}(x) \\
&\qquad + (-1)^p Z_{q,\beta}Z_{r,\beta}\iint F^{(q)}(x_1-)F^{(r)}(x_2-)\,\mathrm{d}g(x_1,x_2).
\end{aligned}\tag{28}$$

Recall that assumption (c) of Theorem 4.1 implies that F is $\max\{p,q,r\}$ times differentiable and that all derivatives up to the $\max\{p,q,r\}$th derivative lie in $\mathbb{D}_{\phi_\lambda}$.

Remark 4.4. The random variables $Z_{p,\beta}$, $Z_{q,\beta}$ and $Z_{r,\beta}$ in part (ii) of Corollary 4.3 are dependent. The specification of their joint distribution seems to be an open problem. Only in the Gaussian case the joint cumulants of $Z_{1,\beta}^2$ and $Z_{2,\beta}$ are known from the supplementary material to Lévy-Leduc *et al.* [23]. Notice that it is even hard to specify the (Rosenblatt) distribution of $Z_{2,\beta}$; for details see Veillette and Taqqu [32].

Proof of Corollary 4.3. Using the representation (5) of $V_g(\hat{F}_n) - V_g(F)$, we obtain that $\mathcal{V}_{g,n;p,q,r}(F)$ equals

$$\begin{aligned}
&-\sum_{i=1}^{2}\int\big[\hat{F}_n(x-)-F(x-)\big]\,\mathrm{d}g_{i,F}(x) + \iint(\hat{F}_n-F)(x_1-)(F_n-F)(x_2-)\,\mathrm{d}g(x_1,x_2) \\
&\quad+\sum_{\ell=1}^{2}\sum_{j=1}^{p-1}(-1)^j\left(\frac{1}{n}\sum_{i=1}^{n}A_{j;F}(X_i)\right)\int F^{(j)}(x-)\,\mathrm{d}g_{\ell,F}(x) \\
&\quad-\sum_{j=1}^{q-1}(-1)^j\left(\frac{1}{n}\sum_{i=1}^{n}A_{j;F}(X_i)\right)\iint F^{(j)}(x_1-)\big(\hat{F}_n(x_2-)-F(x_2-)\big)\,\mathrm{d}g(x_1,x_2) \\
&\quad-\sum_{k=1}^{r-1}(-1)^k\left(\frac{1}{n}\sum_{i=1}^{n}A_{k;F}(X_i)\right)\iint\big(\hat{F}_n(x_1-)-F(x_1-)\big)F^{(k)}(x_2-)\,\mathrm{d}g(x_1,x_2) \\
&\quad+\sum_{j=1}^{q-1}\sum_{k=1}^{r-1}(-1)^{j+k}\left(\frac{1}{n}\sum_{i=1}^{n}A_{j;F}(X_i)\right)\left(\frac{1}{n}\sum_{i=1}^{n}A_{k;F}(X_i)\right) \\
&\qquad\qquad\times\iint F^{(j)}(x_1-)F^{(k)}(x_2-)\,\mathrm{d}g(x_1,x_2) \\
&\quad=-\sum_{\ell=1}^{2}\int\mathcal{E}_{n,p-1;F}(x-)\,\mathrm{d}g_{\ell,F}(x)+\iint\mathcal{E}_{n,q-1;F}(x_1-)\mathcal{E}_{n,r-1;F}(x_2-)\,\mathrm{d}g(x_1,x_2).
\end{aligned}$$

Moreover, by Theorem 4.1,

$$\big\{n^{s(\beta-1/2)}\ell(n)^{-s}\big\}\mathcal{E}_{n,s-1;F}(\cdot) \xrightarrow{\mathsf{d}} (-1)^s c_{s,\beta} Z_{s,\beta} F^{(s)}(\cdot) \qquad \big(\text{in } (\mathbb{D}_{\phi_\lambda}, \mathcal{D}_{\phi_\lambda}, d_{\phi_\lambda})\big)$$

for $s = p, q, r$. Therefore, assertion (i) follows from the Continuous Mapping theorem and (26) as well as Slutzky's lemma and the assumption $q + r > p$. Moreover, assertion (ii) follows from the Continuous Mapping theorem, (26) and the assumption $p = q + r$. □

It is worth pointing out that, as mentioned at the beginning of this section, $\mathcal{V}_{n,g;p,q,r}(\hat{F}_n)$ is obtained by using the representation (5) and an "expansion" of $\hat{F}_n - F$ in the sense of (22). Obviously, with increasing p, q or r, the expression $\mathcal{V}_{n,g;p,q,r}(F)$ defined in (27) is getting more and more involved. So, for statistical applications one should choose p, q and r as small as possible. On the other hand, p has to be chosen so large so that the limit in distribution of $\{n^{p(\beta-1/2)}\ell(n)^{-p}\}\mathcal{V}_{n,g;p,q,r}(F)$ does not vanish. That is, an application of Corollary 4.3 requires a trade-off between the simplicity of the statistic $\mathcal{V}_{n,g;p,q,r}(F)$ and the benefit of the asymptotic distribution. A particularly favorable situation is the one where some (or preferably all) terms on the right-hand side of (27), which are different from $V_g(\hat{F}_n) - V_g(F)$, vanish. This is the case if the respective integrals $\int F^{(j)}(x-)\,\mathrm{d}g_{\ell,F}(x)$ etc. vanish, for instance, in the case of the sample variance and in the case of the test for symmetry; cf. Examples 4.7 and 4.9. In other situations, the statistic $\mathcal{V}_{n,g;p,q,r}(F)$ might be more complicated than $V_g(\hat{F}_n) - V_g(F)$. Yet, it seems to be among the best achievable results. Finally, notice that in cases where part (i) or part (ii) of Theorem 3.15 already yields a non-trivial asymptotic distribution, the result can also be derived by Corollary 4.3. This is exemplified in the next remark.

Remark 4.5. (i) For Gini's mean difference take $p = q = r = 1$. Then the result obtained from part (i) of Corollary 4.3 coincides with the result we get from part (i) of Theorem 3.15.

(ii) For the weighted Cramér–von Mises statistic take $p = 2$ and $q = r = 1$. Then, under the hypothesis that $F = F_0$, we have from Corollary 4.3(ii) that the asymptotic distribution equals $Z_{1,\beta}Z_{1,\beta}\iint F^{(1)}(x_1-)F^{(1)}(x_2-)\,\mathrm{d}g(x_1,x_2) = (Z_{1,\beta})^2\int w(x)(F^{(1)}(x-))^2\,\mathrm{d}F(x)$, where we used (15). That is, in accordance with Example 3.22.

Gini's mean difference discussed in part (i) of the preceding remark is an example for an asymptotically non-degenerate U- or V-statistic. The weighted Cramér–von Mises statistic discussed in part (ii) of the preceding remark is an example for an asymptotically degenerate U- or V-statistic of type (1.a) in the sense of Section 2. The following two Examples 4.6 and 4.7 provide some asymptotically degenerate U- or V-statistics of type (1.b) and type (1.c), respectively. Examples 4.8 and 4.9 below will provide some asymptotically degenerate U- or V-statistics of type 2 in the sense of Section 2.

Example 4.6 (Squared absolute mean of a symmetric distribution). The kernel $g(x_1, x_2) = x_1 \cdot x_2$ for estimating the squared mean has been investigated repeatedly in the literature. Let us consider here the related kernel $g(x_1, x_2) = |x_1| \cdot |x_2|$ for estimating the squared *absolute* mean of a distribution F having a finite first moment. In this case, we obtain $g_{i,F}(x) = \mathbb{E}[|X_1|] \cdot |x|$ and hence $\mathrm{d}g_{i,F}(x) = \mathbb{E}[|X_1|](-\mathbb{1}_{\{x<0\}} + \mathbb{1}_{\{x\geq 0\}})\,\mathrm{d}x$ for $i = 1, 2$. Moreover, we have $\mathrm{d}g(x_1, x_2) =$

$(\mathbb{1}_{\{x_1\geq 0,x_2\geq 0\}} - \mathbb{1}_{\{x_1<0,x_2\geq 0\}} - \mathbb{1}_{\{x_1\geq 0,x_2<0\}} + \mathbb{1}_{\{x_1<0,x_2<0\}})\,\mathrm{d}x_1\,\mathrm{d}x_2$. It is then easily checked that the conditions of Lemmas 3.4 and 3.6 are fulfilled for any weight function ϕ with $\int 1/\phi(x)\,\mathrm{d}x < \infty$. Now, let us in addition assume that $F^{(1)}$ is symmetric about 0. Then, on one hand, Theorem 3.15(i) with $B^\circ(\cdot) = (-1)F^{(1)}(\cdot)Z_{1,\beta}$ yields that $\{n^{\beta-1/2}\ell(n)^{-1}\}(V_g(\hat{F}_n) - V_g(F))$ converges in distribution to $-\sum_{i=1}^2 \int B^\circ(x-)\,\mathrm{d}g_{i,F}(x) = 2Z_{1,\beta}\mathbb{E}[|X_1|](-\int_{-\infty}^0 F^{(1)}(x-)\,\mathrm{d}x + \int_0^\infty F^{(1)}(x-)\,\mathrm{d}x = 0$. On the other hand, part (ii) of Theorem 3.15 is helpful neither because it only yields that $\{n^{2(\beta-1/2)}\ell(n)^{-2}\}(V_g(\hat{F}_n) - V_g(F))$ converges in distribution to $\iint B^\circ(x_1-)B^\circ(x_2-)\,\mathrm{d}g(x_1,x_2) = Z_{1,\beta}^2 \iint F^{(1)}(x_1)F^{(1)}(x_2)\,\mathrm{d}g(x_1,x_2) = 0$, where for the latter "=" we used the symmetry of $F^{(1)}$. However, if we take $p = 2$ and $q = r = 1$, we have $\mathcal{V}_{n,g;2,1,1}(\hat{F}_n) = V_g(\hat{F}_n) - V_g(F)$ and obtain by Corollary 4.3(ii)

$$\begin{aligned}
&\{n^{2\beta-1}\ell(n)^{-2}\}\big(V_g(\hat{F}_n) - V_g(F)\big)\\
&\quad\xrightarrow{\mathsf{d}} Z_{2,\beta}\sum_{\ell=1}^2 \int F^{(2)}(x-)\,\mathrm{d}g_{\ell,F}(x) + Z_{1,\beta}^2 \iint F^{(1)}(x_1-)F^{(1)}(x_2-)\,\mathrm{d}g(x_1,x_2)\\
&\quad= 2Z_{2,\beta}\bigg(-\int_{-\infty}^0 F^{(2)}(x)\,\mathrm{d}x + \int_0^\infty F^{(2)}(x)\,\mathrm{d}x\bigg) = 4Z_{2,\beta}\bigg(\int_0^\infty F^{(2)}(x)\,\mathrm{d}x\bigg),
\end{aligned}$$

where for the latter "=" we used the antisymmetry of $F^{(2)}$ (i.e., $F^{(2)}(x) = -F^{(2)}(-x)$) which holds by the symmetry of $F^{(1)}$. This shows that in the present case $V_g(\hat{F}_n)$ is an asymptotically degenerate V-statistic w.r.t. $(g, F, (n^{(\beta-1/2)}\ell(n)^{-1}))$ of type (1.b) in the sense of Section 2.

Example 4.7 (Variance). As discussed above, in our long-memory setting we can neither apply part (i) nor part (ii) of Theorem 3.15 to derive a *non-trivial* asymptotic distribution for the sample variance; recall that the sample variance is a V-statistic with corresponding kernel $g(x_1,x_2) = \frac{1}{2}(x_1 - x_2)^2$. However, part (ii) of Corollary 4.3 enables us to derive a non-trivial asymptotic distribution. From Example 3.11, we know that $\mathrm{d}g_{1,F}(x) = \mathrm{d}g_{2,F}(x) = (x - \mathbb{E}[X_1])\,\mathrm{d}x$, which implies $\int F^{(1)}(x-)\,\mathrm{d}g_{\ell,F}(x) = 0$. So we have $\mathcal{V}_{n,g;2,1,1}(F) = V_g(\hat{F}_n) - V_g(F)$ and obtain by Corollary 4.3(ii)

$$\begin{aligned}
&\{n^{2\beta-1}\ell(n)^{-2}\}\big(V_g(\hat{F}_n) - V_g(F)\big)\\
&\quad\xrightarrow{\mathsf{d}} Z_{2,\beta}\sum_{\ell=1}^2 \int F^{(2)}(x-)\,\mathrm{d}g_{\ell,F}(x) + Z_{1,\beta}^2 \iint F^{(1)}(x_1-)F^{(1)}(x_2-)\,\mathrm{d}g(x_1,x_2)\\
&\quad= 2Z_{2,\beta}\int F^{(2)}(x-)\big(x - \mathbb{E}[X_1]\big)\,\mathrm{d}x - \bigg(Z_{1,\beta}\int F^{(1)}(x)\,\mathrm{d}x\bigg)^2 \qquad (29)\\
&\quad= 2Z_{2,\beta}\int F^{(2)}(x-)\big(x - \mathbb{E}[X_1]\big)\,\mathrm{d}x - Z_{1,\beta}^2,
\end{aligned}$$

where for the first "=" we used the fact that $\mathrm{d}g(x_1,x_2)$ is the negative of the Lebesgue measure on $\mathbb{R}^2$; cf. Example 3.11. Notice that $\int F^{(1)}(x-)\,\mathrm{d}g_{\ell,F}(x) = 0$ holds for every (sufficiently smooth) df F, so that (29) is fully satisfactory even in a non-parametric setting. Notice also that in the present case $V_g(\hat{F}_n)$ is an asymptotically degenerate V-statistic w.r.t. $(g, F, (n^{(\beta-1/2)}\ell(n)^{-1}))$ of type (1.c) in the sense of Section 2.

Example 4.8 (Scaling sequence (a_n^3)). Let us consider the kernel $g(x_1, x_2) = x_1(|x_2| - 1)$, and suppose that $F^{(1)}$ is symmetric about zero and that $m := \mathbb{E}[|X_1|] = 1$. This setting is somewhat artificial but it leads to quite an interesting limiting behavior of the corresponding U- or V-statistic. We have $g_{1,F}(x_1) = x_1(m-1) = 0$ and $g_{2,F}(x_2) = \mathbb{E}[X_1](|x_2| - 1) = 0$ due to the assumption $m = 1$ and the symmetry of $F^{(1)}$, respectively. That is, $V_g(\hat{F}_n)$ is a degenerate V-statistic w.r.t. (g, F), and consequently an application of part (i) of Theorem 3.15 does not lead to a non-degenerate limiting distribution. Moreover, it is easily seen that $\mathrm{d}g(x_1, x_2) = (\mathbb{1}_{\{x_2 \geq 0\}} - \mathbb{1}_{\{x_2<0\}})\,\mathrm{d}x_1\,\mathrm{d}x_2$. Therefore, part (ii) of Theorem 3.15 does not provide a tool to derive a non-degenerate limiting distribution, because under the imposed assumption the right-hand side in (20) with $B^\circ(\cdot) = (-1)F^{(1)}(\cdot)Z_{1,\beta}$ vanishes.

On the other hand, part (ii) of Corollary 4.3 enables us to derive a *non-trivial* asymptotic distribution. In contrast to the sample variance in Example 4.7, however, it does not make sense to work with $\mathcal{V}_{n,g;2,1,1}(F)$, because in this case the limit in (28) vanishes. Indeed, the first summand of the limit vanishes since $g_{1,F} \equiv g_{2,F} \equiv 0$, and the second summand of the limit vanishes since under the imposed assumptions the integral $\iint F^{(1)}(x_1-)F^{(1)}(x_2-)\,\mathrm{d}g(x_1, x_2)$ equals zero. As a consequence we need to work with a p larger than 2. For instance, for $(p, q, r) = (3, 1, 2)$ we obtain from Corollary 4.3(ii) and $g_{1,F} \equiv g_{2,F} \equiv 0$ that

$$
\begin{aligned}
&\bigl\{n^{3(\beta-1/2)}\ell(n)^{-3}\bigr\}\mathcal{V}_{n,g;3,1,2}(F) \\
&\quad\xrightarrow{\mathsf{d}} -Z_{3,\beta}\sum_{\ell=1}^{2}\int F^{(3)}(x-)\,\mathrm{d}g_{\ell,F}(x) - Z_{1,\beta}Z_{2,\beta}\iint F^{(1)}(x_1-)F^{(2)}(x_2-)\,\mathrm{d}g(x_1, x_2) \\
&\quad= -Z_{1,\beta}Z_{2,\beta}\biggl(\int_0^\infty\int F^{(1)}(x_1)F^{(2)}(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 - \int_{-\infty}^0\int F^{(1)}(x_1)F^{(2)}(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2\biggr) \\
&\quad= -2Z_{1,\beta}Z_{2,\beta}\int_0^\infty F^{(2)}(x_2)\,\mathrm{d}x_2,
\end{aligned}
$$

which is typically distinct from zero. Notice that above we may replace $\mathcal{V}_{n,g;3,1,2}(F)$ by $V_g(\hat{F}_n) - V_g(F)$ since

$$
\begin{aligned}
&\mathcal{V}_{n,g;3,1,2}(F) \\
&\quad= V_g(\hat{F}_n) - V_g(F) + \sum_{\ell=1}^{2}\sum_{j=1}^{2}(-1)^j\Biggl(\frac{1}{n}\sum_{i=1}^{n}A_{j;F}(X_i)\Biggr)\int F^{(j)}(x-)\,\mathrm{d}g_{\ell,F}(x) \\
&\qquad+ \Biggl(\frac{1}{n}\sum_{i=1}^{n}A_{1;F}(X_i)\Biggr)\iint\bigl(\hat{F}_n(x_1-) - F(x_1-)\bigr)F^{(1)}(x_2-)\,\mathrm{d}g(x_1, x_2)
\end{aligned}
$$

$$
\begin{aligned}
&= V_g(\hat{F}_n) - V_g(F) + \left(\frac{1}{n}\sum_{i=1}^n A_{1;F}(X_i)\right)\left(\int_0^\infty \int \big(\hat{F}_n(x_1-) - F(x_1-)\big) F^{(1)}(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2\right.\\
&\qquad \left. - \int_{-\infty}^0 \int \big(\hat{F}_n(x_1-) - F(x_1-)\big) F^{(1)}(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2\right)\\
&= V_g(\hat{F}_n) - V_g(F),
\end{aligned}
$$

where we used $g_{1,F} \equiv g_{2,F} \equiv 0$, the continuity of $F^{(1)}$, and the symmetry of $F^{(1)}$ about zero. Thus, in the present case $V_g(\hat{F}_n)$ is an asymptotically degenerate V-statistic w.r.t. $(g, F, (n^{(\beta-1/2)}\ell(n)^{-1}))$ of type 2 in the sense of Section 2.

The next example shows that it might even not be sufficient to take the scaling sequence $(n^{3(\beta-1/2)})$ to obtain a non-degenerate limiting distribution.

Example 4.9 (Test for symmetry, scaling sequence (a_n^4)). Let us come back to the test statistic T_n introduced in Example 3.14, which is a V-statistic with kernel given by (16). We restrict to the null hypothesis that the distribution is symmetric about zero. We have seen in Example 3.14 that in this case we obtain $g_{1,F} \equiv g_{2,F} \equiv 0$ and $\mathrm{d}g(x_1,x_2) = \mathcal{H}^1_D(d(x_1,x_2)) - \mathcal{H}^1_{\widetilde{D}}(d(x_1,x_2))$. That is, under the null hypothesis, T_n can be seen as a degenerate V-statistic. So, in principle, we could apply Theorem 3.15(ii) to derive the asymptotic distribution of $T_n = V_g(\hat{F}_n)$. However, the integral on the right-hand side in (20) with $B^\circ(\cdot) = (-1)F^{(1)}(\cdot)Z_{1,\beta}$ equals $\iint B^\circ(x_1)B^\circ(x_2)\,\mathrm{d}g(x_1,x_2) = Z^2_{1,\beta}(\int F^{(1)}(x)F^{(1)}(x)\,\mathrm{d}x - \int F^{(1)}(x)F^{(1)}(-x)\,\mathrm{d}x) = 0$, because $F^{(1)}$ is symmetric about zero. Now one might tend to apply Corollary 4.3 as in Example 4.7, that is, with $(p,q,r) = (3,1,2)$, to obtain a non-trivial limiting distribution. However, the integrals on the right-hand side of (28) equal zero in that case. Indeed, the first one equals zero, because $g_{1,F} \equiv g_{2,F} \equiv 0$. The second one, which is given by $\int F^{(1)}(x)F^{(2)}(x)\,\mathrm{d}x - \int F^{(1)}(x)F^{(2)}(-x)\,\mathrm{d}x$, equals zero, because of the symmetry of $F^{(1)}$ and the antisymmetry of $F^{(2)}$. However, applying part (ii) of Corollary 4.3 with $(p,q,r) = (4,2,2)$ we obtain (using $g_{1,F} \equiv g_{2,F} \equiv 0$)

$$
\begin{aligned}
n^{4\beta-2}\ell(n)^{-4}\mathcal{V}_{n,g;4,2,2}(\hat{F}_n) &\xrightarrow{\mathrm{d}} Z^2_{2,\beta}\left(\int F^{(2)}(x)F^{(2)}(x)\,\mathrm{d}x - \int F^{(2)}(x)F^{(2)}(-x)\,\mathrm{d}x\right)\\
&= 4Z^2_{2,\beta}\int_0^\infty \big(F^{(2)}(x)\big)^2\,\mathrm{d}x
\end{aligned}
$$

by the anti-symmetry of $F^{(2)}$. Using the symmetry of $F^{(1)}$ and once again that $g_{1,F} \equiv g_{2,F} \equiv 0$, it can be easily checked that $\mathcal{V}_{n,g;4,2,2}(\hat{F}_n) = V_g(\hat{F}_n) - V_g(F)$.

Acknowledgement

The authors are very grateful to a referee for providing Example 2.1, for his/her very careful reading, and pointing out a missing word in Example 3.20 that resulted in an incorrect statement.

Supplementary Material

Supplement to paper "Continuous mapping approach to the asymptotics of U- and V-statistics" (DOI: 10.3150/13-BEJ508SUPP; .pdf). The supplement Beutner and Zähle [7] contains a discussion of some extensions and limitations of the approach presented in this paper.

References

[1] Andersen, N.T., Giné, E. and Zinn, J. (1988). The central limit theorem for empirical processes under local conditions: The case of Radon infinitely divisible limits without Gaussian component. *Trans. Amer. Math. Soc.* **308** 603–635. MR0930076
[2] Arcones, M.A. and Giné, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674. MR1165586
[3] Avram, F. and Taqqu, M.S. (1987). Noncentral limit theorems and Appell polynomials. *Ann. Probab.* **15** 767–775. MR0885142
[4] Beutner, E., Wu, W.B. and Zähle, H. (2012). Asymptotics for statistical functionals of long-memory sequences. *Stochastic Process. Appl.* **122** 910–929. MR2891441
[5] Beutner, E. and Zähle, H. (2010). A modified functional delta method and its application to the estimation of risk functionals. *J. Multivariate Anal.* **101** 2452–2463. MR2719874
[6] Beutner, E. and Zähle, H. (2012). Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli* **18** 803–822. MR2948902
[7] Beutner, E. and Zähle, H. (2014). Supplement to "Continuous mapping approach to the asymptotics of U- and V-statistics." DOI:10.3150/13-BEJ508SUPP.
[8] Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *J. Econometrics* **130** 307–335. MR2211797
[9] Dehling, H. (2006). Limit theorems for dependent U-statistics. In *Dependence in Probability and Statistics. Lecture Notes in Statist.* **187** 65–86. New York: Springer. MR2283250
[10] Dehling, H. and Taqqu, M.S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783. MR1026312
[11] Dehling, H. and Taqqu, M.S. (1991). Bivariate symmetric statistics of long-range dependent observations. *J. Statist. Plann. Inference* **28** 153–165. MR1115815
[12] Denker, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics. Advanced Lectures in Mathematics*. Braunschweig: Friedr. Vieweg & Sohn. MR0889896
[13] Ghorpade, S.R. and Limaye, B.V. (2010). *A Course in Multivariable Calculus and Analysis. Undergraduate Texts in Mathematics*. New York: Springer. MR2583676
[14] Gill, R.D., van der Laan, M.J. and Wellner, J.A. (1995). Inefficient estimators of the bivariate survival function for three models. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 545–597. MR1338452
[15] Giraitis, L. and Surgailis, D. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *J. Statist. Plann. Inference* **80** 81–93. MR1713796
[16] Halmos, P.R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17** 34–43. MR0015746
[17] Ho, H.C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024. MR1401834
[18] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325. MR0026294
[19] Hosking, J.R.M. (1981). Fractional differencing. *Biometrika* **68** 165–176. MR0614953
[20] Hsing, T. (2000). Linear processes, long-range dependence and asymptotic expansions. *Stat. Inference Stoch. Process.* **3** 19–29. MR1819283

[21] Koroljuk, V.S. and Borovskich, Y.V. (1994). *Theory of U-statistics*. *Mathematics and Its Applications* **273**. Dordrecht: Kluwer Academic. MR1472486

[22] Lee, A.J. (1990). *U-Statistics*: *Theory and Practice*. *Statistics*: *Textbooks and Monographs* **110**. New York: Dekker. MR1075417

[23] Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M.S. and Reisen, V.A. (2011). Asymptotic properties of U-processes under long-range dependence. *Ann*. *Statist*. **39** 1399–1426. MR2850207

[24] Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M.S. and Reisen, V.A. (2011). Large sample behaviour of some well-known robust estimators under long-range dependence. *Statistics* **45** 59–71. MR2772156

[25] Mattila, P. (1995). *Geometry of Sets and Measures in Euclidean Spaces*: *Fractals and Rectifiability*. *Cambridge Studies in Advanced Mathematics* **44**. Cambridge: Cambridge Univ. Press. MR1333890

[26] Pollard, D. (1984). *Convergence of Stochastic Processes*. *Springer Series in Statistics*. New York: Springer. MR0762984

[27] Rio, E. (1995). A maximal inequality and dependent Marcinkiewicz–Zygmund strong laws. *Ann*. *Probab*. **23** 918–937. MR1334177

[28] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley. MR0595165

[29] Shao, Q.M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann*. *Probab*. **24** 2098–2127. MR1415243

[30] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0838963

[31] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247

[32] Veillette, M.S. and Taqqu, M.S. (2013). Properties and numerical evaluation of the Rosenblatt distribution. *Bernoulli* **19** 982–1005.

[33] von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann*. *Math*. *Statist*. **18** 309–348. MR0022330

[34] Wu, W.B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9** 809–831. MR2047687

[35] Wu, W.B. (2006). Unit root testing for functionals of linear processes. *Econometric Theory* **22** 1–14. MR2212691

[36] Wu, W.B. (2008). Empirical processes of stationary sequences. *Statist*. *Sinica* **18** 313–333. MR2384990

[37] Yukich, J.E. (1992). Weak convergence of smoothed empirical processes. *Scand*. *J*. *Stat*. **19** 271–279. MR1183201

[38] Zähle, H. (2014). Marcinkiewicz-Zygmund and ordinary strong laws for empirical distribution functions and plug-in estimators. *Statistics*. To appear.

Received April 2012 and revised December 2012