Adv. Appl. Probab. **55,** 841–894 (2023)
*doi:*10.1017/apr.2022.51

ASYMPTOTIC NORMALITY FOR *M*-DEPENDENT AND CONSTRAINED *U*-STATISTICS, WITH APPLICATIONS TO PATTERN MATCHING IN RANDOM STRINGS AND PERMUTATIONS

SVANTE JANSON ,* *Uppsala University*

Abstract

We study (asymmetric) U-statistics based on a stationary sequence of m-dependent variables; moreover, we consider constrained U-statistics, where the defining multiple sum only includes terms satisfying some restrictions on the gaps between indices. Results include a law of large numbers and a central limit theorem, together with results on rate of convergence, moment convergence, functional convergence, and a renewal theory version.

Special attention is paid to degenerate cases where, after the standard normalization, the asymptotic variance vanishes; in these cases non-normal limits occur after a different normalization.

The results are motivated by applications to pattern matching in random strings and permutations. We obtain both new results and new proofs of old results.

Keywords: Random words; constrained occurrences; vincular patterns

2020 Mathematics Subject Classification: Primary 60F05
Secondary 05A05; 60C05; 68Q87

1. Introduction

The purpose of the present paper is to present some new results for (asymmetric) U-statistics together with some applications. (See Section 3 for definitions.) The results include a strong law of large numbers and a central limit theorem (asymptotic normality), together with results on rate of convergence, moment convergence, functional convergence, and a renewal theory version.

Many results of these types have been proved for U-statistics under different hypotheses by a large number of authors, from Hoeffding [27] on. The new feature of the results here, which are motivated by applications discussed below, is the combination of the following:

(i) We consider, as in e.g. [35], [37], and [26], but unlike many other authors, *asymmetric U-statistics* and not just the symmetric case. (See Remark 3.3.)

(ii) We consider also *constrained U-statistics*, where the summations are restricted as in (3.2) or (3.3).

Received 21 June 2021; revision received 9 August 2022.
* Postal address: Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden. Email address: svante.janson@math.uu.se

(iii) The U-statistics are based on an underlying sequence that is not necessarily i.i.d. (as is usually assumed); we assume only that the sequence is stationary and *m-dependent*. (This case has been studied earlier by e.g. [58], but not in the present asymmetric case.)

The extension to the m-dependent case might be of interest for some applications, but for us the main motivation is that it allows us to reduce the constrained versions to ordinary U-statistics; hence this extension is implicitly used also when we apply the results for constrained U-statistics based on i.i.d. sequences.

Remark 1.1. The combination of the features (i)–(iii) above is new, but they have each been considered separately earlier.

In particular, constrained U-statistics are special cases of the large class of *incomplete U-statistics* [6]. These are, in turn, special cases of the even more general *weighted U-statistics*; see e.g. [61], [47], [42], [55], [30], [66], and [26]. (These references show asymptotic normality under various conditions; some also study degenerate cases with non-normal limits; [26] includes the asymmetric case.) In view of our applications, we consider here only the constrained case instead of trying to find suitable conditions for general weights.

Similarly, U-statistics have been considered by many authors for more general weakly dependent sequences than m-dependent ones. In particular, asymptotic normality has been shown under various types of mixing conditions by e.g. [59], [64, 65], and [15]. We are not aware of any paper on asymmetric U-statistics with a mixing condition on the variables. Such results might be interesting for future research, but again in view of our applications, we have not pursued this and consider here only the m-dependent case.

There are thus many previous results yielding asymptotic normality for U-statistics under various conditions. One general feature, found already in the first paper [27], is that there are degenerate cases where the asymptotic variance vanishes (typically because of some internal cancellations). In such cases, the theorems only yield convergence to 0 and do not imply asymptotic normality; indeed, typically a different normalization yields a non-normal limit. It is often difficult to calculate the asymptotic variance exactly, and it is therefore of great interest to have simple criteria that show that the asymptotic variance is non-zero. Such a criterion is well known for the standard case of (unconstrained) U-statistics based on i.i.d. variables [27]. We give corresponding (somewhat more complicated) criteria for the m-dependent case studied here, in both the unconstrained and constrained cases. (This is one reason for considering only the m-dependent case in the present paper, and not more general weakly dependent sequences.) We show the applicability of our criteria in some examples.

Like many (but not all) of the references cited above, we base our proof of asymptotic normality on the decomposition method of Hoeffding [27], with appropriate modifications. As pointed out by an anonymous referee, an alternative method is to use dependency graphs together with Stein's method, which under an extra moment assumption yields our main results on asymptotic normality together with an upper bound on the rate of convergence. We do not use this method in the main parts of the paper, partly because it does not seem to yield simple criteria for non-vanishing of the asymptotic variance; however, as a complement, we use this method to give some results on rate of convergence.

1.1. Applications

The background motivating our general results is given by some parallel results for pattern matching in random strings and in random permutations that have previously been shown by

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

different methods, but easily follow from our results; we describe these results here and return to them (and some new results) in Sections 13 and 14. Further applications to pattern matching in random permutations restricted to two classes of permutations are given in [38].

First, consider a random string $\Xi_n = \xi_1 \cdots \xi_n$ consisting of n i.i.d. random letters from a finite alphabet $\mathcal{A}$ (in this context, this is known as a *memoryless source*), and consider the number of occurrences of a given word $\mathbf{w} = w_1 \cdots w_\ell$ as a subsequence; to be precise, an *occurrence* of $\mathbf{w}$ in Ξ_n is an increasing sequence of indices $i_1 < \ldots < i_\ell$ in $[n] = \{1, \ldots, n\}$ such that

$$\xi_{i_1}\xi_{i_2} \cdots \xi_{i_\ell} = \mathbf{w}, \qquad \text{i.e., } \xi_{i_k} = w_k \text{ for every } k \in [\ell]. \tag{1.1}$$

This number, $N_n(\mathbf{w})$ say, was studied by Flajolet, Szpankowski and Vallée [23], who proved that $N_n(\mathbf{w})$ is asymptotically normal as $n \to \infty$.

Flajolet, Szpankowski and Vallée [23] also studied a *constrained* version, where we are also given numbers $d_1, \ldots, d_{\ell-1} \in \mathbb{N} \cup \{\infty\} = \{1, 2, \ldots, \infty\}$ and count only occurrences of w such that

$$i_{j+1} - i_j \leqslant d_j, \qquad 1 \leqslant j < \ell. \tag{1.2}$$

(Thus the jth gap in $i_1, \ldots, i_\ell$ has length strictly less than d_j.) We write $\mathcal{D} := (d_1, \ldots, d_{\ell-1})$, and let $N_n(\mathbf{w}; \mathcal{D})$ be the number of occurrences of $\mathbf{w}$ that satisfy the constraints (1.2). It was shown in [23] that, for any fixed $\mathbf{w}$ and $\mathcal{D}$, $N_n(\mathbf{w}, \mathcal{D})$ is asymptotically normal as $n \to \infty$. See also the book by [Jacquet and Szpankowski [31], Chapter 5].

Remark 1.2. Note that $d_j = \infty$ means no constraint for the jth gap. In particular, $d_1 = \ldots = d_{\ell-1} = \infty$ yields the unconstrained case; we denote this trivial (but important) constraint $\mathcal{D}$ by $\mathcal{D}_\infty$.

In the other extreme case, if $d_j = 1$, then i_j and i_{j+1} have to be adjacent. In particular, in the completely constrained case $d_1 = \cdots = d_{\ell-1} = 1$, $N_n(\mathbf{w}; \mathcal{D})$ counts occurrences of $\mathbf{w}$ as a substring $\xi_i\xi_{i+1} \cdots \xi_{i+\ell-1}$. Substring counts have been studied by many authors; some references with central limit theorems or local limit theorems under varying conditions are [4], [52], [45], and [22, Proposition IX.10, p. 660]. See also [62, Section 7.6.2 and Example 8.8] and [31]; the latter book discusses not only substring and subsequence counts but also other versions of substring matching problems in random strings.

Note also that the case when all $d_i \in \{1, \infty\}$ means that $\mathbf{w}$ is a concatenation $\mathbf{w}_1 \cdots \mathbf{w}_b$ (with $\mathbf{w}$ broken at positions where $d_i = \infty$), such that an occurrence now is an occurrence of each $\mathbf{w}_i$ as a substring, with these substrings in order and non-overlapping, and with arbitrary gaps in between. (This is a special case of the *generalized subsequence problem* in [31, Section 5.6]; the general case can be regarded as a sum of such counts over a set of $\mathbf{w}$.)

There are similar results for random permutations. Let $\mathfrak{S}_n$ be the set of the $n!$ permutations of $[n]$. If $\pi = \pi_1 \cdots \pi_n \in \mathfrak{S}_n$ and $\tau = \tau_1 \cdots \tau_\ell \in \mathfrak{S}_\ell$, then an *occurrence* of the pattern τ in π is an increasing sequence of indices $i_1 < \cdots < i_\ell$ in $[n] = \{1, \ldots, n\}$ such that the order relations in $\pi_{i_1} \cdots \pi_{i_\ell}$ are the same as in $\tau_1 \cdots \tau_\ell$, i.e., $\pi_{i_j} < \pi_{i_k} \iff \tau_j < \tau_k$.

Let $N_n(\tau)$ be the number of occurrences of τ in $\boldsymbol{\pi}$ when $\boldsymbol{\pi} = \boldsymbol{\pi}^{(n)}$ is uniformly random in $\mathfrak{S}_n$. Bóna [7] proved that $N_n(\tau)$ is asymptotically normal as $n \to \infty$, for any fixed τ.

Also for permutations, one can consider, and count, constrained occurrences by again imposing the restriction (1.2) for some $\mathcal{D} = (d_1, \ldots, d_{\ell-1})$. In analogy with strings, we let $N_n(\tau, \mathcal{D})$ be the number of constrained occurrences of τ in $\boldsymbol{\pi}^{(n)}$ when $\boldsymbol{\pi}^{(n)}$ is uniformly random in $\mathfrak{S}_n$. This random number seems to have been studied mainly in the case when each $d_i \in \{1, \infty\}$, i.e., some i_j are required to be adjacent to the next one—in the permutation

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

context, such constrained patterns are known as *vincular* patterns. Hofer [29] proved asymptotic normality of $N_n(\tau, \mathcal{D})$ as $n \to \infty$, for any fixed τ and vincular $\mathcal{D}$. The extreme case with $d_1 = \cdots = d_{\ell-1} = 1$ was earlier treated by Bóna [9]. Another (non-vincular) case that has been studied is that of *d-descents*, given by $\ell = 2$, $\tau = 21$, and $\mathcal{D} = (d)$; Bóna [8] shows asymptotic normality and Pike [50] gives a rate of convergence.

We unify these results by considering U-statistics. It is well known and easy to see that the number $N_n(\mathbf{w})$ of unconstrained occurrences of a given subsequence $\mathbf{w}$ in a random string Ξ_n can be written as an asymmetric U-statistic; see Section 13 and (13.2) for details. There are general results on asymptotic normality of U-statistics that extend the basic result by [27] to the asymmetric case; see e.g. [35, Corollary 11.20] and [37]. Hence, asymptotic normality of $N_n(\mathbf{w})$ follows directly from these general results. Similarly, it is well known that the pattern count $N_n(\tau)$ in a random permutation also can be written as a U-statistic (see Section 14 for details), and again this can be used to prove asymptotic normality. (See [39], with an alternative proof by this method of the result by Bóna [7].)

The constrained case is different, since the constrained pattern counts are not U-statistics. However, they can be regarded as *constrained U-statistics*, which we define in (3.2) below in analogy with the constrained counts above. As stated above, in the present paper we prove general limit theorems for such constrained U-statistics, which thus immediately apply to the constrained pattern counts discussed above in random strings and permutations.

The basic idea in the proofs is that a constrained U-statistic based on a sequence (X_i) can be written (possibly up to a small error) as an unconstrained U-statistic based on another sequence (Y_i) of random variables, where the new sequence (Y_i) is m-dependent (with a different m) if (X_i) is. (However, even if (X_i) is independent, (Y_i) is in general not; this is our main motivation for considering m-dependent sequences.) The unconstrained m-dependent case then is treated by standard methods from the independent case, with appropriate modifications.

Section 2 contains some preliminaries. The unconstrained and constrained U-statistics are defined in Section 3, where also the main theorems are stated. The degenerate case, when the asymptotic variance in the central limit theorem Theorem 3.3, 3.4, or 3.8 vanishes, is discussed later in Section 8, when more notation has been introduced; Theorems 8.1, 8.2, and 8.3, respectively, give criteria that can be used to show that the asymptotic variance is non-zero in an application. On the other hand, Example 8.1 shows that the degenerate case can occur in new ways for constrained U-statistics.

The reduction to the unconstrained case and some other lemmas are given in Section 4, and then the proofs of the main theorems are completed in Sections 5–7 and 9–12. Section 13 gives applications to the problem on pattern matching in random strings discussed above. Similarly, Section 14 gives applications to pattern matching in random permutations. Some further comments and open problems are given in Section 15. The appendix contains some further results on subsequence counts in random strings.

2. Preliminaries

2.1. Some notation

A *constraint* is, as in Section 1, a sequence $\mathcal{D} = (d_1, \dots, d_{\ell-1}) \in (\mathbb{N} \cup \{\infty\})^{\ell-1}$, for some given $\ell \geqslant 1$. Recall that the special constraint $(\infty, \dots, \infty)$ is denoted by $\mathcal{D}_\infty$. Given a constraint $\mathcal{D}$, define $b = b(\mathcal{D})$ by

$$b = b(\mathcal{D}) := \ell - |\{j : d_j < \infty\}| = 1 + |\{j : d_j = \infty\}|. \tag{2.1}$$

We say that b is the number of *blocks* defined by $\mathcal{D}$; see further Section 4 below.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

For a random variable Z, and $p > 0$, we let $\|Z\|_p := (\mathbb{E}[|Z|^p])^{1/p}$.

We use $\xrightarrow{\mathrm{d}}$, $\xrightarrow{\mathrm{p}}$, and $\xrightarrow{\mathrm{a.s.}}$ for convergence of random variables in distribution, in probability, and almost surely (a.s.), respectively. For a sequence of random variables (Z_n), and a sequence $a_n > 0$, we write $Z_n = o_{\mathrm{p}}(a_n)$ when $Z_n/a_n \xrightarrow{\mathrm{p}} 0$.

Unspecified limits are as $n \to \infty$. C denotes an unspecified constant, which may be different at each occurrence. (C may depend on parameters that are regarded as fixed, for example the function f below; this will be clear from the context.)

We use the convention $\binom{n}{k} := 0$ if $n < 0$. (We will always have $k \geqslant 0$.) Some further standard notation: $[n] := \{1, \dots, n\}$, and $\max \emptyset := 0$. All functions are tacitly assumed to be measurable.

2.2. m-dependent variables

For reasons mentioned in the introduction, we will consider U-statistics not only based on sequences of independent random variables, but also based on m-dependent variables.

Recall that a (finite or infinite) sequence of random variables $(X_i)_i$ is *m-dependent* if the two families $\{X_i\}_{i \leqslant k}$ and $\{X_i\}_{i > k+m}$ of random variables are independent of each other for every k. (Here, $m \geqslant 0$ is a given integer.) In particular, 0-dependent is the same as independent; thus the important independent case is included as the special case $m = 0$ below.

It is well known that if $(X_i)_{i \in I}$ is m-dependent, and $I_1, \dots, I_r \subseteq I$ are sets of indices such that $\operatorname{dist}(I_j, I_k) := \inf\{|i - i'| : i \in I_j, i' \in I_k\} > m$ when $j \neq k$, then the families (vectors) of random variables $(X_i)_{i \in I_1}, \dots, (X_i)_{i \in I_r}$ are mutually independent of each other. (To see this, note first that it suffices to consider the case when each I_j is an interval; then use the definition and induction on r.) We will use this property without further comment.

In practice, m-dependent sequences usually occur as *block factors*; i.e. they can be expressed as

$$X_i := h(\xi_i, \dots, \xi_{i+m}) \tag{2.2}$$

for some i.i.d. sequence (ξ_i) of random variables (in some measurable space $\mathcal{S}_0$) and a fixed function h on $\mathcal{S}_0^{m+1}$. (It is obvious that (2.2) then defines a stationary m-dependent sequence.)

3. U-statistics and main results

Let $X_1, X_2, \dots$ be a sequence of random variables, taking values in some measurable space $\mathcal{S}$, and let $f : \mathcal{S}^\ell \to \mathbb{R}$ be a (measurable) function of ℓ variables, for some $\ell \geqslant 1$. Then the corresponding *U-statistic* is the (real-valued) random variable defined for each $n \geqslant 0$ by

$$U_n = U_n(f) = U_n\bigl(f; (X_i)\bigr) := \sum_{1 \leqslant i_1 < \dots < i_\ell \leqslant n} f\bigl(X_{i_1}, \dots, X_{i_\ell}\bigr). \tag{3.1}$$

U-statistics were introduced by Hoeffding [27], who proved a general central limit theorem; the present paper gives an extension of his result that builds on his methods.

Remark 3.1. Of course, for the definition (3.1) it suffices to have a finite sequence $(X_i)_1^n$, but in the present paper we will only consider the initial segments of an infinite sequence.

Remark 3.2. Many authors, including Hoeffding [27], define U_n by dividing the sum in (3.1) by $\binom{n}{\ell}$, the number of terms in it. We find it more convenient for our purposes to use the non-normalized version above.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Remark 3.3. Many authors, including Hoeffding [27], assume that f is a symmetric function of its ℓ variables. In this case, the order of the variables does not matter, and in (3.1) we can sum over all sequences $i_1, \dots, i_\ell$ of ℓ distinct elements of $\{1, \dots, n\}$, up to an obvious factor of $\ell!$. (Hoeffding [27] gives both versions.) Conversely, if we sum over all such sequences, we may without loss of generality assume that f is symmetric. However, in the present paper (as in several earlier papers by various authors) we consider the general case of (3.1) without assuming symmetry; for emphasis, we call this an *asymmetric U-statistic*. (This is essential in our applications to pattern matching.) Note that for independent $(X_i)_1^n$, the asymmetric case can be reduced to the symmetric case by the trick in Remark 11.21, in particular (11.20), of [35]; see also [39, (15)] and (A.18) below. However, this trick does not work in the m-dependent or constrained cases studied here, so we cannot use it here.

As stated in the introduction, we also consider *constrained U-statistics*. Given a constraint $\mathcal{D} = (d_1, \dots, d_{\ell-1})$, we define the constrained U-statistic

$$U_n(f;\mathcal{D}) = U_n(f;\mathcal{D};(X_i)) := \sum_{\substack{1\leqslant i_1<\dots<i_\ell\leqslant n\\ i_{j+1}-i_j\leqslant d_j}} f\bigl(X_{i_1},\dots,X_{i_\ell}\bigr), \qquad n\geqslant 0, \tag{3.2}$$

where we thus impose the constraints (1.2) on the indices.

We define further the *exactly constrained U-statistic*

$$U_n(f;\mathcal{D}{=}) = U_n(f;\mathcal{D}{=};(X_i)) := \sum_{\substack{1\leqslant i_1<\dots<i_\ell\leqslant n\\ i_{j+1}-i_j=d_j \text{ if } d_j<\infty}} f\bigl(X_{i_1},\dots,X_{i_\ell}\bigr), \qquad n\geqslant 0, \tag{3.3}$$

where we thus specify each gap either exactly or (when $d_j=\infty$) not at all. In the vincular case, when all d_j are either 1 or ∞, there is no difference and we have $U_n(f;\mathcal{D}) = U_n(f;\mathcal{D}{=})$.

Note that, trivially, each constrained U-statistic can be written as a sum of exactly constrained U-statistics:

$$U_n(f;\mathcal{D}) = \sum_{\mathcal{D}'} U_n(f;\mathcal{D}'{=}), \tag{3.4}$$

where we sum over all constraints $\mathcal{D}' = (d'_1, \dots, d'_\ell)$ with

$$\begin{cases} 1\leqslant d'_j\leqslant d_j, & d_j<\infty,\\ d'_j=\infty, & d_j=\infty. \end{cases} \tag{3.5}$$

Remark 3.4. As stated in the introduction, the [exactly] constrained U-statistics thus belong to the large class of *incomplete U-statistics* [6], where the summation in (3.1) is restricted to some, in principle arbitrary, subset of the set of all ℓ-tuples $(i_1, \dots, i_\ell)$ in $[n]$.

The standard setting, in [27] and many other papers, is to assume that the underlying random variables X_i are i.i.d. In the present paper we consider a more general case, and we will assume only that $X_1, X_2, \dots$ is an infinite stationary *m-dependent* sequence, for some fixed integer $m\geqslant 0$; see Section 2.2 for the definition, and recall in particular that the special case $m=0$ yields the case of independent variables X_i.

We will consider limits as $n\to\infty$. The sequence $X_1, X_2, \dots$ (and thus the space $\mathcal{S}$ and the integer m) and the function f (and thus ℓ) will be fixed, and do not depend on n.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

We will throughout assume the following moment condition for $p = 2$; in a few places (always explicitly stated) we also assume it for some larger p:

$$(\mathrm{A}_p) \quad \mathbb{E}|f(X_{i_1}, \dots, X_{i_\ell})|^p < \infty \quad \text{for every } i_1 < \dots < i_\ell.$$

Note that in the independent case ($m = 0$), it suffices to verify (A_p) for a single sequence $i_1, \dots, i_\ell$, for example $1, \dots, \ell$. In general, it suffices to verify (A_p) for all sequences with $i_1 = 1$ and $i_{j+1} - i_j \leqslant m + 1$ for every $j \leqslant \ell - 1$, since the stationarity and m-dependence imply that every larger gap can be reduced to $m + 1$ without changing the distribution of $f(X_{i_1}, \dots, X_{i_\ell})$. Since there is only a finite number of such sequences, it follows that (A_2) is equivalent to the uniform bound

$$\mathbb{E}|f(X_{i_1}, \dots, X_{i_\ell})|^2 \leqslant C \qquad \text{for every } i_1 < \dots < i_\ell, \tag{3.6}$$

and similarly for (A_p).

3.1. Expectation and law of large numbers

We first make an elementary observation on the expectations $\mathbb{E}U_n(f; \mathcal{D})$ and $\mathbb{E}U_n(f; \mathcal{D}=)$. These can be calculated exactly by taking the expectation inside the sums in (3.2) and (3.3). In the independent case, all terms have the same expectation, so it remains only to count the number of them. In general, because of the m-dependence of (X_i), the expectations of the terms in (3.3) are not all equal, but most of them coincide, and it is still easy to find the asymptotics.

Theorem 3.1. *Let $(X_i)_1^\infty$ be a stationary m-dependent sequence of random variables with values in a measurable space $\mathcal{S}$, let $\ell \geqslant 1$, and let $f : \mathcal{S}^\ell \to \mathbb{R}$ satisfy (A_2). Then, as $n \to \infty$, with μ given by (5.1) below,*

$$\mathbb{E}U_n(f) = \binom{n}{\ell}\mu + O\bigl(n^{\ell-1}\bigr) = \frac{n^\ell}{\ell!}\mu + O\bigl(n^{\ell-1}\bigr). \tag{3.7}$$

More generally, let $\mathcal{D} = (d_1, \dots, d_{\ell-1})$ be a constraint, and let $b := b(\mathcal{D})$. Then, as $n \to \infty$, for some real numbers $\mu_{\mathcal{D}}$ and $\mu_{\mathcal{D}=}$ given by (5.5) and (5.4),

$$\mathbb{E}U_n(f; \mathcal{D}) = \frac{n^b}{b!}\mu_{\mathcal{D}} + O\bigl(n^{b-1}\bigr), \tag{3.8}$$

$$\mathbb{E}U_n(f; \mathcal{D}=) = \frac{n^b}{b!}\mu_{\mathcal{D}=} + O\bigl(n^{b-1}\bigr). \tag{3.9}$$

If $m = 0$, i.e., the sequence $(X_i)_1^\infty$ is i.i.d., then, moreover,

$$\mu = \mu_{\mathcal{D}=} = \mathbb{E}f(X_1, \dots, X_\ell), \tag{3.10}$$

$$\mu_{\mathcal{D}} = \mu \prod_{j\,:\,d_j<\infty} d_j = \prod_{j\,:\,d_j<\infty} d_j \cdot \mathbb{E}f(X_1, \dots, X_\ell). \tag{3.11}$$

The straightforward proof is given in Section 5, where we also give formulas for $\mu_{\mathcal{D}}$ and $\mu_{\mathcal{D}=}$ in the general case, although in an application it might be simpler to find the leading term of the expectation directly.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Next, we have a corresponding strong law of large numbers, proved in Section 7. This extends well-known results in the independent case; see [28, 37, 57].

Theorem 3.2. *Let $(X_i)_1^\infty$ be a stationary m-dependent sequence of random variables with values in a measurable space $\mathcal{S}$, let $\ell \geqslant 1$, and let $f : \mathcal{S}^\ell \to \mathbb{R}$ satisfy (A_2). Then, as $n \to \infty$, with μ given by (5.1),*

$$n^{-\ell} U_n(f) \xrightarrow{\text{a.s.}} \frac{1}{\ell!}\mu. \tag{3.12}$$

More generally, let $\mathcal{D} = (d_1, \dots, d_{\ell-1})$ be a constraint, and let $b := b(\mathcal{D})$. Then, as $n \to \infty$,

$$n^{-b} U_n(f; \mathcal{D}) \xrightarrow{\text{a.s.}} \frac{1}{b!}\mu_{\mathcal{D}}, \tag{3.13}$$

$$n^{-b} U_n(f; \mathcal{D}=) \xrightarrow{\text{a.s.}} \frac{1}{b!}\mu_{\mathcal{D}=}, \tag{3.14}$$

where $\mu_{\mathcal{D}}$ and $\mu_{\mathcal{D}=}$, as in Theorem 3.1, are given by (5.5) and (5.4).

Equivalently,

$$n^{-\ell}\big[U_n(f) - \mathbb{E}U_n(f)\big] \xrightarrow{\text{a.s.}} 0, \tag{3.15}$$

$$n^{-b}\big[U_n(f; \mathcal{D}) - \mathbb{E}U_n(f; \mathcal{D})\big] \xrightarrow{\text{a.s.}} 0, \tag{3.16}$$

$$n^{-b}\big[U_n(f; \mathcal{D}=) - \mathbb{E}U_n(f; \mathcal{D}=)\big] \xrightarrow{\text{a.s.}} 0. \tag{3.17}$$

Remark 3.5. For convenience, we assume (A_2) in Theorem 3.2 as in the rest of the paper, which leads to a simple proof. We conjecture that the theorem holds assuming only (A_1) (i.e., finite first moments) instead of (A_2), as in [28, 37] for the independent case.

3.2. Asymptotic normality

We have the following theorems yielding asymptotic normality. The proofs are given in Section 6.

The first theorem is for the unconstrained case, and extends the basic theorem by Hoeffding [27] for symmetric U-statistics based on independent $(X_i)_1^\infty$ to the asymmetric and m-dependent case. Note that both these extensions have previously been treated, but separately. For symmetric U-statistics in the m-dependent setting, asymptotic normality was proved by Sen [58] (at least assuming a third moment); moreover, bounds on the rate of convergence (assuming a moment condition) were given by Malevich and Abdalimov [43]. The asymmetric case with independent $(X_i)_1^\infty$ has been treated e.g. in [35, Corollary 11.20] and [37]; furthermore, as stated in Remark 3.3, for independent (X_i), the asymmetric case can be reduced to the symmetric case by the method in [35, Remark 11.21].

Theorem 3.3. *Let $(X_i)_1^\infty$ be a stationary m-dependent sequence of random variables with values in a measurable space $\mathcal{S}$, let $\ell \geqslant 1$, and let $f : \mathcal{S}^\ell \to \mathbb{R}$ satisfy (A_2). Then, as $n \to \infty$,*

$$\operatorname{Var}\big[U_n(f)\big]/n^{2\ell-1} \to \sigma^2 \tag{3.18}$$

for some $\sigma^2 = \sigma^2(f) \in [0, \infty)$, and

$$\frac{U_n(f) - \mathbb{E}U_n(f)}{n^{\ell-1/2}} \xrightarrow{\text{d}} \mathsf{N}\big(0, \sigma^2\big). \tag{3.19}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

The second theorem extends Theorem 3.3 to the constrained cases.

Theorem 3.4. *Let $(X_i)_1^\infty$ be a stationary m-dependent sequence of random variables with values in a measurable space $\mathcal{S}$, let $\ell \geqslant 1$, and let $f : \mathcal{S}^\ell \to \mathbb{R}$ satisfy* (A_2). *Let $\mathcal{D} = (d_1, \dots, d_{\ell-1})$ be a constraint, and let $b := b(\mathcal{D})$. Then, as $n \to \infty$,*

$$\operatorname{Var}\bigl[U_n(f;\mathcal{D})\bigr]/n^{2b-1} \to \sigma^2 \tag{3.20}$$

for some $\sigma^2 = \sigma^2(f;\mathcal{D}) \in [0, \infty)$, and

$$\frac{U_n(f;\mathcal{D}) - \mathbb{E}U_n(f;\mathcal{D})}{n^{b-1/2}} \xrightarrow{\mathrm{d}} \mathsf{N}\bigl(0, \sigma^2\bigr). \tag{3.21}$$

The same holds, with some (generally different) $\sigma^2 = \sigma^2(f;\mathcal{D}=)$, for the exactly constrained $U_n(f;\mathcal{D}=)$.

Remark 3.6. It follows immediately by the Cramér–Wold device [25, Theorem 5.10.5] (i.e., considering linear combinations), that Theorem 3.3 extends in the obvious way to joint convergence for any finite number of different $f : \mathcal{S}^\ell \to \mathbb{R}$, with σ^2 now a covariance matrix. Moreover, the proof shows that this holds also for a family of different f with (possibly) different $\ell \geqslant 1$.

Similarly, Theorem 3.4 extends to joint convergence for any finite number of different f (possibly with different ℓ and $\mathcal{D}$); this follows by the proof below, which reduces the results to Theorem 3.3.

Remark 3.7. The asymptotic variance σ^2 in Theorems 3.3 and 3.4 can be calculated explicitly; see Remark 6.2.

Remark 3.8. Note that it is possible that the asymptotic variance $\sigma^2 = 0$ in Theorems 3.3 and 3.4; in this case, (3.19) and (3.21) just give convergence in probability to 0. This degenerate case is discussed in Section 8.

Remark 3.9. We do not consider extensions to triangular arrays where f or X_i (or both) depend on n. In the symmetric m-dependent case, such a result (with fixed ℓ but possibly increasing m, under suitable conditions) has been shown by [43], with a bound on the rate of convergence. In the independent case, results for triangular arrays are given by e.g. [56] and [32]; see also [40] for the special case of substring counts $N_n(\mathbf{w})$ with $\mathbf{w}$ depending on n (and growing in length). It seems to be an interesting (and challenging) open problem to formulate useful general theorems for constrained U-statistics in such settings.

3.3. Rate of convergence

Under stronger moment assumptions on f, an alternative method of proof (suggested by a referee) yields the asymptotic normality in Theorems 3.3 and 3.4 together with an upper bound on the rate of convergence, provided $\sigma^2 > 0$.

In the following theorem of Berry–Esseen type, we assume for simplicity that f is bounded (as it is in our applications in Sections 13–14); see further Remark 9.1. Let d_K denote the Kolmogorov distance between distributions; recall that for two distributions $\mathcal{L}_1$, $\mathcal{L}_2$ with distribution functions $F_1(x)$ and $F_2(x)$, $d_K = d_K(\mathcal{L}_1, \mathcal{L}_2) := \sup_x |F_1(x) - F_2(x)|$; we use also the notation $d_K(X, \mathcal{L}_2) := d_K(\mathcal{L}(X), \mathcal{L}_2)$ for a random variable X.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Theorem 3.5. *Suppose in addition to the hypotheses in Theorem 3.3 or 3.4 that $\sigma^2 > 0$ and that f is bounded. Then*

$$d_K\Big(\frac{U_n - \mathbb{E}U_n}{\sqrt{\operatorname{Var} U_n}}, \mathsf{N}(0,1)\Big) = O\big(n^{-1/2}\big), \tag{3.22}$$

where U_n denotes $U_n(f)$, $U_n(f,\mathcal{D})$, or $U_n(f;\mathcal{D}{=})$.

In the symmetric and unconstrained case, this (and more) was shown by Malevich and Abdalimov [43]. The proof of Theorem 3.5 is given in Section 9, together with further remarks.

3.4. Moment convergence

Theorems 3.3 and 3.4 include convergence of the first (trivially) and second moments in (3.19) and (3.21). This extends to higher moments under a corresponding moment condition on f. (The unconstrained case with independent X_i was shown in [37, Theorem 3.15].)

Theorem 3.6. *Suppose in addition to the hypotheses in Theorems 3.3 or 3.4 that (A_p) holds for some real $p \geqslant 2$. Then all absolute and ordinary moments of order up to p converge in (3.19) or (3.21).*

The proof is given in Section 10, where we also give related estimates for maximal functions.

3.5. Functional limit theorems

We can extend Theorem 3.4 to functional convergence. For unconstrained U-statistics, this was done by Miller and Sen [44] in the classical case of independent X_i and symmetric f; the asymmetric case is [37, Theorem 3.2]; furthermore, Yoshihara [65] proved the case of dependent X_i satisfying a suitable mixing condition (assuming a technical condition on f besides symmetry).

Theorem 3.7. *Suppose that (A_2) holds. Then as $n \to \infty$, with $b = b(\mathcal{D})$, in $D[0,\infty)$,*

$$\frac{U_{\lfloor nt\rfloor}(f;\mathcal{D}) - \mathbb{E}U_{\lfloor nt\rfloor}(f;\mathcal{D})}{n^{b-1/2}} \xrightarrow{\mathrm{d}} Z(t), \qquad t \geqslant 0, \tag{3.23}$$

where $Z(t)$ is a continuous centered Gaussian process. Equivalently, in $D[0,\infty)$,

$$\frac{U_{\lfloor nt\rfloor}(f;\mathcal{D}) - (\mu_{\mathcal{D}}/b!)n^b t^b}{n^{b-1/2}} \xrightarrow{\mathrm{d}} Z(t), \qquad t \geqslant 0. \tag{3.24}$$

The same holds for exact constraints. Moreover, the results hold jointly for any finite set of f and $\mathcal{D}$ (possibly with different ℓ and b), with limits $Z(t)$ depending on f and $\mathcal{D}$.

The proof is given in Section 11.

Remark 3.10. A comparison between (3.23) and (3.21) yields $Z(t) \sim \mathsf{N}\big(0, t^{2b-1}\sigma^2\big)$, with σ^2 as in Theorem 3.4. Equivalently, $\operatorname{Var} Z(t) = t^{2b-1}\sigma^2$, which can be calculated by Remark 6.2. Covariances $\operatorname{Cov}(Z(s), Z(t))$ can be calculated by the same method and (11.20) in the proof; we leave the details to the reader. Note that these covariances determine the distribution of the process Z.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

3.6. Renewal theory

Assume further that $h : \mathcal{S} \to \mathbb{R}$ is another (fixed) measurable function, with

$$\nu := \mathbb{E}h(X_1) > 0. \tag{3.25}$$

We define

$$S_n = S_n(h) := \sum_{i=1}^{n} h(X_i), \tag{3.26}$$

and, for $x > 0$,

$$N_-(x) := \sup\{n \geqslant 0 : S_n \leqslant x\}, \tag{3.27}$$

$$N_+(x) := \inf\{n \geqslant 0 : S_n > x\}. \tag{3.28}$$

$N_-(x)$ and $N_+(x)$ are finite a.s. by the law of large numbers for S_n (12.1); see further Lemma 12.1. We let $N_\pm(x)$ denote either $N_-(x)$ or $N_+(x)$, in statements and formulas that are valid for both.

Remark 3.11. In [37], instead of $h(x)$ we consider more generally a function of several variables, and we define $N_\pm$ using the corresponding U-statistic instead of S_n. We believe that the results of the present paper can be extended to that setting, but we have not pursued this, and leave it as an open problem.

Remark 3.12. If $h(X_1) \geqslant 0$ a.s., which is often assumed in renewal theory, then $N_+(x) = N_-(x) + 1$. However, if h may be negative (still assuming (3.25)), then $N_-(x)$ may be larger than $N_+(x)$. Nevertheless, the difference is typically small, and we obtain the same asymptotic results for both N_+ and N_-. (We can also obtain the same results if we instead use $S_n < x$ or $S_n \geqslant x$ in the definitions.)

In this situation, we have the following limit theorems, which extend results in [37]. Proofs are given in Section 12. For an application, see [38].

Theorem 3.8. *With the assumptions and notation of Theorem 3.4, assume* (A_2)*, and suppose also that* $\nu := \mathbb{E}h(X_1) > 0$ *and* $\mathbb{E}h(X_1)^2 < \infty$*. Then, with notation as above, as* $x \to \infty$*,*

$$\frac{U_{N_\pm(x)}(f;\mathcal{D}) - \mu_{\mathcal{D}}\nu^{-b}b!^{-1}x^b}{x^{b-1/2}} \xrightarrow{\mathrm{d}} \mathsf{N}\bigl(0, \gamma^2\bigr), \tag{3.29}$$

for some $\gamma^2 = \gamma^2(f;h;\mathcal{D}) \geqslant 0$*.*

The same holds for exact constraints. Moreover, the results hold jointly for any finite set of f *and* $\mathcal{D}$ *(possibly with different* ℓ *and* b*).*

Theorem 3.9. *Suppose in addition to the hypotheses in Theorem 3.8 that* $h(X_1)$ *is integer-valued and that* $(X_i)_1^\infty$ *are independent. Then (3.29) holds also conditioned on* $S_{N_-(x)} = x$ *for integers* $x \to \infty$*.*

We consider here tacitly only x such that $\mathbb{P}(S_{N_-(x)} = x) > 0$.

Remark 3.13. We prove Theorem 3.9 only for independent X_i (which, in any case, is our main interest, as stated in the introduction). It seems likely that the result can be extended to at least

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

some m-dependent (X_i), using a modification of the proof below and the m-dependent renewal theorem (under some conditions) [1, Corollary 4.2], but we have not pursued this.

Theorem 3.10. *Suppose in addition to the hypotheses in Theorem 3.8 that* (A_p) *holds and* $\mathbb{E}[|h(X_1)|^p] < \infty$ *for every* $p < \infty$. *Then all moments converge in* (3.29).

Under the additional hypothesis in Theorem 3.9, this holds also conditioned on $S_{N_-(x)} = x$.

Remark 3.14. In Theorem 3.10, unlike Theorem 3.6, we assume pth moments for all p, and conclude convergence of all moments. If we only want to show convergence for a given p, some sufficient moment conditions on f and h can be derived from the proof, but we do not know any sharp results and have not pursued this. Cf. [37, Remark 6.1] and the references there.

4. Some lemmas

We give here some lemmas that will be used in the proofs in later sections. In particular, they will enable us to reduce the constrained cases to the unconstrained one.

Let $\mathcal{D} = (d_1, \dots, d_{\ell-1})$ be a given constraint. Recall that $b = b(\mathcal{D})$ is given by (2.1), and let $1 = \beta_1 < \dots < \beta_b$ be the indices in $[\ell]$ just after the unconstrained gaps; in other words, β_j are defined by $\beta_1 := 1$ and $d_{\beta_j - 1} = \infty$ for $j = 2, \dots, b$. For convenience we also define $\beta_{b+1} := \ell + 1$. We say that the constraint $\mathcal{D}$ separates the index set $[\ell]$ into the b *blocks* $B_1, \dots, B_b$, where $B_k := \{\beta_k, \dots, \beta_{k+1} - 1\}$. Note that the constraints (1.2) thus are constraints on i_j for j in each block separately.

Lemma 4.1. *Let* $(X_i)_1^\infty$ *be a stationary* m*-dependent sequence of random variables with values in* $\mathcal{S}$, *let* $\ell \geqslant 1$, *and let* $f : \mathcal{S}^\ell \to \mathbb{R}$ *satisfy* (A_2). *Let* $\mathcal{D} = (d_1, \dots, d_{\ell-1})$ *be a constraint. Then*

$$\operatorname{Var}\bigl[U_n(f;\mathcal{D})\bigr] = O\bigl(n^{2b(\mathcal{D})-1}\bigr), \qquad n \geqslant 1. \tag{4.1}$$

Furthermore,

$$\operatorname{Var}\bigl[U_n(f;\mathcal{D}) - U_{n-1}(f;\mathcal{D})\bigr] = O\bigl(n^{2b(\mathcal{D})-2}\bigr), \qquad n \geqslant 1. \tag{4.2}$$

Moreover, the same estimates hold for $U_n(f;\mathcal{D}=)$.

Proof. The definition (3.2) yields

$$\operatorname{Var}\bigl[U_n(f;\mathcal{D})\bigr] = \sum_{\substack{1\leqslant i_1<\dots<i_\ell\leqslant n\\ i_{k+1}-i_k\leqslant d_k}} \sum_{\substack{1\leqslant j_1<\dots<j_\ell\leqslant n\\ j_{k+1}-j_k\leqslant d_k}} \operatorname{Cov}\bigl(f\bigl(X_{i_1},\dots,X_{i_\ell}\bigr), f\bigl(X_{j_1},\dots,X_{j_\ell}\bigr)\bigr). \tag{4.3}$$

Let d_* be the largest finite d_j in the constraint $\mathcal{D}$, i.e.,

$$d_* := \max_j \bigl\{d_j : d_j < \infty\bigr\}. \tag{4.4}$$

The constraints imply that for each block B_q and all indices $k \in B_q$, coarsely,

$$0 \leqslant i_k - i_{\beta_q} \leqslant d_*\ell \qquad \text{and} \qquad 0 \leqslant j_k - j_{\beta_q} \leqslant d_*\ell. \tag{4.5}$$

It follows that if $|i_{\beta_r} - j_{\beta_s}| > d_*\ell + m$ for all $r, s \in [b]$, then $|i_\alpha - j_\beta| > m$ for all $\alpha, \beta \in [\ell]$. Since $(X_i)_1^\infty$ is m-dependent, this implies that the two random vectors $(X_{i_1}, \dots, X_{i_\ell})$ and $(X_{j_1}, \dots, X_{j_\ell})$ are independent, and thus the corresponding term in (4.3) vanishes.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Consequently, we only have to consider terms in the sum in (4.3) such that

$$|i_{\beta_r} - j_{\beta_s}| \leqslant d_*\ell + m \tag{4.6}$$

for some $r, s \in [b]$. For each of the $O(1)$ choices of r and s, we can choose $i_{\beta_1}, \dots, i_{\beta_b}$ in at most n^b ways; then j_{β_s} in $O(1)$ ways such that (4.6) holds; then the remaining j_{β_q} in $O\bigl(n^{b-1}\bigr)$ ways; then, finally, all remaining i_k and j_k in $O(1)$ ways because of (4.5). Consequently, the number of non-vanishing terms in (4.3) is $O\bigl(n^{2b-1}\bigr)$. Moreover, each term is $O(1)$ by (3.6) and the Cauchy–Schwarz inequality, and thus (4.1) follows.

For (4.2), we note that $U_n(f;\mathcal{D}) - U_{n-1}(f;\mathcal{D})$ is the sum in (3.2) with the extra restriction $i_\ell = n$. Hence, its variance can be expanded as in (4.3), with the extra restrictions $i_\ell = j_\ell = n$. We then argue as above, but note that (4.5) and $i_\ell = n$ imply that there are only $O(1)$ choices of i_b, and hence $O\bigl(n^{b-1}\bigr)$ choices of $i_1, \dots, i_b$. We thus obtain $O\bigl(n^{2b-2}\bigr)$ non-vanishing terms in the sum, and (4.2) follows.

The argument for the exactly constrained $U_n(f;\mathcal{D}{=})$ is the same (and slightly simpler). (Alternatively, we could do this case first, and then use (3.4) to obtain the results for $U_n(f;\mathcal{D})$.) □

The next lemma is the central step in the reduction to the unconstrained case.

Lemma 4.2. *Let $(X_i)_1^\infty$, $f : \mathcal{S}^\ell \to \mathbb{R}$, and $\mathcal{D} = (d_1, \dots, d_{\ell-1})$ be as in Lemma 4.1, and let*

$$D := \sum_{j\,:\,d_j<\infty} d_j. \tag{4.7}$$

Let $M > D$ and define

$$Y_i := (X_i, X_{i+1}, \dots, X_{i+M-1}) \in \mathcal{S}^M, \qquad i \geqslant 1. \tag{4.8}$$

Then there exists a function $g = g_{\mathcal{D}=} : (\mathcal{S}^M)^b \to \mathbb{R}$ such that for every $n \geqslant 0$,

$$U_n(f;\mathcal{D}{=};(X_i)) = \sum_{j_1<\dots<j_b\leqslant n-D} g\bigl(Y_{j_1}, \dots, Y_{j_b}\bigr) = U_{n-D}\bigl(g;(Y_i)\bigr), \tag{4.9}$$

with $U_{n-D}(g) := 0$ when $n < D$. Furthermore,

$$\mathbb{E}\bigl|g\bigl(Y_{j_1}, \dots, Y_{j_b}\bigr)\bigr|^2 < \infty, \tag{4.10}$$

for every $j_1 < \dots < j_b$.

Proof. For each block $B_q = \{\beta_q, \dots, \beta_{q+1} - 1\}$ defined by $\mathcal{D}$, let

$$\ell_q := |B_q| = \beta_{q+1} - \beta_q, \tag{4.11}$$

$$t_{qr} := \sum_{j=1}^{r-1} d_{\beta_q+j-1}, \qquad r = 1, \dots, \ell_q, \tag{4.12}$$

$$u_q := t_{q,\ell_q} = \sum_{j=1}^{\beta_{q+1}-\beta_q-1} d_{\beta_q+j-1}, \tag{4.13}$$

$$v_q := \sum_{k<q} u_k. \tag{4.14}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Note that $t_{q1} = 0$ for every q and that t_{qr}, $u_q < \infty$. (We stop the summation in (4.13) just before the next infinite d_j, which occurs for $j = \beta_{q+1} - 1$ provided $q < b$.) Note also that

$$u_b + v_b = \sum_{k \leqslant b} u_k = D. \tag{4.15}$$

We then rewrite (3.3) as follows, letting $k_q := i_{\beta_q}$ and grouping the arguments of f according to the blocks of $\mathcal{D}$ (using an obvious notation for this):

$$U_n(f; \mathcal{D}{=}) = \sum_{\substack{1 \leqslant k_1 < k_2 < \ldots < k_b \leqslant n - u_b, \\ k_{q+1} > k_q + u_q}} f\Big(\big(X_{k_1 + t_{1r}}\big)_{r=1}^{\ell_1}, \ldots, \big(X_{k_b + t_{br}}\big)_{r=1}^{\ell_b} \Big). \tag{4.16}$$

Change summation variables by $k_q = j_q + v_q$. Then, recalling (4.14)–(4.15), (4.16) yields

$$U_n(f; \mathcal{D}{=}) = \sum_{1 \leqslant j_1 < j_2 < j_b \leqslant n - D} f\Big(\big(X_{j_1 + v_1 + t_{1r}}\big)_{r=1}^{\ell_1}, \ldots, \big(X_{j_b + v_b + t_{br}}\big)_{r=1}^{\ell_b} \Big). \tag{4.17}$$

Define, for $y_i = (y_{ik})_{k=1}^M \in \mathcal{S}^M$,

$$g(y_1, \ldots, y_b) = f\Big(\big(y_{1, v_1 + t_{1r} + 1}\big)_{r=1}^{\ell_1}, \ldots, \big(y_{b, v_b + t_{br} + 1}\big)_{r=1}^{\ell_b} \Big). \tag{4.18}$$

(Note that $v_j + t_{jr} + 1 \leqslant v_j + u_j + 1 \leqslant D + 1 \leqslant M$.) We have $Y_j = (X_{j+k-1})_{k=1}^M$, and thus (4.18) yields

$$g\big(Y_{j_1}, \ldots, Y_{j_b}\big) = f\Big(\big(X_{j_1 + v_1 + t_{1r}}\big)_{r=1}^{\ell_1}, \ldots, \big(X_{j_b + v_b + t_{br}}\big)_{r=1}^{\ell_b} \Big). \tag{4.19}$$

Consequently, (4.9) follows from (4.17) and (4.19).

Furthermore, (4.10) follows from (4.19) and (A_2). □

Lemma 4.3. *Let $(X_i)_1^\infty$ and $\mathcal{D} = (d_1, \ldots, d_{\ell-1})$ be as in Lemma 4.1, and let M and Y_i be as in Lemma 4.2. For every $f : \mathcal{S}^\ell \to \mathbb{R}$ such that (A_2) holds, there exist functions $g_{\mathcal{D}}, g_{\mathcal{D}=} : (\mathcal{S}^M)^b \to \mathbb{R}$ such that (4.10) holds for both, and*

$$\operatorname{Var}\Big[U_n\big(f; \mathcal{D}; (X_i)\big) - U_n\big(g_{\mathcal{D}}; (Y_i)\big) \Big] = O\big(n^{2b(\mathcal{D})-2}\big), \tag{4.20}$$

$$\operatorname{Var}\Big[U_n(f; \mathcal{D}{=}; (X_i)) - U_n\big(g_{\mathcal{D}=}; (Y_i)\big) \Big] = O\big(n^{2b(\mathcal{D})-2}\big). \tag{4.21}$$

Proof. First, letting $g_{\mathcal{D}=}$ be as in Lemma 4.2, we have by (4.9)

$$\begin{aligned} U_n(f; \mathcal{D}{=}; (X_i)) - U_n\big(g_{\mathcal{D}=}; (Y_i)\big) &= U_{n-D}\big(g_{\mathcal{D}=}\big) - U_n\big(g_{\mathcal{D}=}\big) \\ &= -\sum_{k=1}^{q} \Big(U_{n-k+1}\big(g_{\mathcal{D}=}\big) - U_{n-k}\big(g_{\mathcal{D}=}\big) \Big). \end{aligned} \tag{4.22}$$

Thus (4.21) follows by (4.2) in Lemma 4.1 applied to $g_{\mathcal{D}=}$, the trivial constraint $\mathcal{D}_\infty$ (i.e., no constraint), and $(Y_i)_1^\infty$.

Next, we recall (3.4) and define

$$g_{\mathcal{D}} = \sum_{\mathcal{D}'} g_{\mathcal{D}'=}, \tag{4.23}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

again summing over all constraints $\mathcal{D}'$ satisfying (3.5). This is a finite sum, and by (3.4) and (4.23),

$$U_n(f;\mathcal{D};(X_i)) - U_n\bigl(g_{\mathcal{D}};(Y_i)\bigr) = \sum_{\mathcal{D}'} \bigl(U_n(f;\mathcal{D}'{=};(X_i)) - U_n\bigl(g_{\mathcal{D}'=};(Y_i)\bigr)\bigr); \tag{4.24}$$

thus (4.20) follows from (4.21). □

To avoid some of the problems caused by dependencies between the X_i, we follow Sen [58] and introduce another type of constrained U-statistics, where we require the gaps between the summation indices to be large, instead of small as in (3.2). We need only one case, and define

$$U_n(f;>m) := \sum_{\substack{1\leqslant i_1<\ldots<i_\ell\leqslant n\\ i_{j+1}-i_j>m}} f\bigl(X_{i_1},\ldots,X_{i_\ell}\bigr), \qquad n\geqslant 0, \tag{4.25}$$

summing only over terms where all gaps $i_{j+1}-i_j>m, j=1,\ldots,\ell-1$. (The advantage is that in each term in (4.25), the variables $X_{i_1},\ldots,X_{i_\ell}$ are independent.)

Lemma 4.4. *Let $(X_i)_1^\infty$ and $f:\mathcal{S}^\ell\to\mathbb{R}$ be as in Lemma 4.1. Then*

$$\operatorname{Var}\bigl(U_n(f)-U_n(f;>m)\bigr) = O\bigl(n^{2\ell-3}\bigr). \tag{4.26}$$

Proof. We can express the type of constrained U-statistic in (4.25) as a combination of constrained U-statistics of the previous type by the following inclusion–exclusion argument:

$$\begin{aligned}
U_n(f;>m) &= \sum_{1\leqslant i_1<\ldots<i_\ell\leqslant n} f\bigl(X_{i_1},\ldots,X_{i_\ell}\bigr)\prod_{j=1}^{\ell-1}\mathbf{1}\{i_{j+1}-i_j>m\}\\
&= \sum_{1\leqslant i_1<\ldots<i_\ell\leqslant n} f\bigl(X_{i_1},\ldots,X_{i_\ell}\bigr)\prod_{j=1}^{\ell-1}\Bigl(1-\mathbf{1}\{i_{j+1}-i_j\leqslant m\}\Bigr)\\
&= \sum_{J\subseteq[\ell-1]}(-1)^{|J|}\sum_{1\leqslant i_1<\ldots<i_\ell\leqslant n} f\bigl(X_{i_1},\ldots,X_{i_\ell}\bigr)\prod_{j\in J}\mathbf{1}\{i_{j+1}-i_j\leqslant m\}\\
&= \sum_{J\subseteq[\ell-1]}(-1)^{|J|}U_n(f;\mathcal{D}_J),
\end{aligned} \tag{4.27}$$

where we sum over the $2^{\ell-1}$ subsets J of $[\ell-1]$, and use the constraints

$$\mathcal{D}_J := (d_{Jj})_{j=1}^{\ell-1} \qquad \text{with} \qquad d_{Jj} = \begin{cases} m, & j\in J,\\ \infty, & j\notin J. \end{cases} \tag{4.28}$$

We have $b(\mathcal{D}_J)=\ell-|J|$, and thus $b(\mathcal{D}_J)<\ell$ unless $J=\emptyset$. Moreover, $\mathcal{D}_\emptyset=(\infty,\ldots,\infty)=\mathcal{D}_\infty$, and this means no constraint, so $U_n(f;\mathcal{D}_\emptyset)=U_n(f)$, the unconstrained U-statistic. Consequently, by (4.27) and Lemma 4.1,

$$\operatorname{Var}\bigl(U_n(f)-U_n(f;>m)\bigr) = \operatorname{Var}\Bigl(\sum_{J\neq\emptyset}(-1)^{|J|-1}U_n(f;\mathcal{D}_J)\Bigr) = O\bigl(n^{2\ell-3}\bigr), \tag{4.29}$$

which proves the estimate (4.26). □

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

4.1. Triangular arrays

We will also use a central limit theorem for m-dependent triangular arrays satisfying the Lindeberg condition, which we state as Theorem 4.1 below. The theorem is implicit in Orey [48]; it follows from his theorem there exactly as his corollary does, although the latter is stated for a sequence and not for a triangular array. See also [Peligrad [49], Theorem 2.1], which contains the theorem below (at least for $\sigma^2 > 0$; the case $\sigma^2 = 0$ is trivial), and is much more general in that it only assumes strong mixing instead of m-dependence.

Recall that a *triangular array* is an array $(\xi_{ni})_{1\leqslant i\leqslant n<\infty}$ of random variables, such that the variables $(\xi_{ni})_{i=1}^n$ in a single row are defined on a common probability space. (As usual, it is only for convenience that we require that the nth row has length n; the results extend to arbitrary lengths N_n.) We are here mainly interested in the case when each row is an m-dependent sequence; in this case, we say that (ξ_{ni}) is an *m-dependent triangular array*. (We make no assumption on the relationships between variables in different rows; these may even be defined on different probability spaces.)

Theorem 4.1. (Orey [48].) *Let $(\xi_{ni})_{1\leqslant i\leqslant n<\infty}$ be an m-dependent triangular array of real-valued random variables with $\mathbb{E}\xi_{ni} = 0$. Let $\widehat{S}_n := \sum_{i=1}^n \xi_{ni}$. Assume that, as $n \to \infty$,*

$$\operatorname{Var} \widehat{S}_n \to \sigma^2 \in [0, \infty), \tag{4.30}$$

that the ξ_{ni} satisfy the Lindeberg condition

$$\sum_{i=1}^{n} \mathbb{E}\bigl[\xi_{ni}^2 \mathbf{1}\{|\xi_{ni}| > \varepsilon\}\bigr] \to 0, \qquad \text{for every } \varepsilon > 0, \tag{4.31}$$

and that

$$\sum_{i=1}^{n} \operatorname{Var}\xi_{ni} = O(1). \tag{4.32}$$

Then, as $n \to \infty$,

$$\widehat{S}_n \xrightarrow{\mathrm{d}} \mathsf{N}\bigl(0, \sigma^2\bigr). \tag{4.33}$$

Note that Theorem 4.1 extends the standard Lindeberg–Feller central limit theorem for triangular arrays with row-wise independent variables (see e.g. [25, Theorem 7.2.4]), to which it reduces when $m = 0$.

Remark 4.1. In fact, the assumption (4.32) is not needed in Theorem 4.1; see [63, Theorem 2.1]. However, it is easily verified in our case (and many other applications), so we need only this classical result.

5. The expectation

The expectation of a (constrained) U-statistic, and in particular its leading term, is easily found from the definition. Nevertheless, we give a detailed proof of Theorem 3.1, for completeness and for later reference.

Proof of Theorem 3.1. Consider first the unconstrained case. We take expectations in (3.1). The sum in (3.1) has $\binom{n}{\ell}$ terms. We consider first the terms that satisfy the restriction

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

$i_{j+1} > i_j + m$ for every $j \in [\ell - 1]$ (i.e., the terms in (4.25)). As noted above, in each such term, the variables $X_{j_1}, \dots, X_{j_\ell}$ are independent. Hence, let $(\widehat{X}_i)_1^\ell$ be an *independent* sequence of random variables in $\mathcal{S}$, each with the same distribution as X_1 (and thus as each X_j), and define

$$\mu := \mathbb{E} f(\widehat{X}_1, \dots, \widehat{X}_\ell). \tag{5.1}$$

Then

$$\mu = \mathbb{E} f\big(X_{i_1}, \dots, X_{i_\ell}\big) \tag{5.2}$$

for every sequence of indices $i_1, \dots, i_\ell$ with $i_{j+1} > i_j + m$ for all $j \in [\ell - 1]$. Moreover, the number of terms in (3.1) that do *not* satisfy these constraints is $O(n^{\ell-1})$, and their expectations are uniformly $O(1)$ as a consequence of (3.6). Thus, (3.7) follows from (3.1).

Next, consider the exactly constrained case. We use Lemma 4.2 and then apply the unconstrained case just treated to g and (Y_i); this yields

$$\mathbb{E} U_n\big(f; \mathcal{D}=\big) = \mathbb{E} U_{n-D}\big(g; (Y_i)\big) = \binom{n-D}{b} \mathbb{E} g(\widehat{Y}_1, \dots, \widehat{Y}_b) + O\big(n^{b-1}\big) \tag{5.3}$$

with $\widehat{Y}_1, \dots, \widehat{Y}_b \overset{\mathrm{d}}{=} Y_1$ independent. Using (4.19), and the notation there, this yields (3.9) with

$$\mu_{\mathcal{D}=} := \mathbb{E} g\big(Y_{j_1}, \dots, Y_{j_b}\big) = \mathbb{E} f\big(\big(X_{j_1+v_1+t_{1r}}\big)_{r=1}^{\ell_1}, \dots, \big(X_{j_b+v_b+t_{br}}\big)_{r=1}^{\ell_b}\big), \tag{5.4}$$

for any sequence $j_1, \dots, j_b$ with $j_{k+1} - j_k \geqslant m + M$ for all $k \in [b-1]$. (Note that $(Y_i)_1^\infty$ is $(m + M - 1)$-dependent.)

Finally, the constrained case (3.8) follows by (3.9) and the decomposition (3.4), with

$$\mu_{\mathcal{D}} := \sum_{\mathcal{D}'} \mu_{\mathcal{D}'=}, \tag{5.5}$$

summing over all $\mathcal{D}'$ satisfying (3.5).

In the independent case $m = 0$, the results above simplify. First, for the unconstrained case, the formula for μ in (3.10) is a special case of (5.2). Similarly, in the exactly unconstrained case, (5.4) yields the formula for $\mu_{\mathcal{D}=}$ in (3.10). Finally, (3.10) shows that $\mu_{\mathcal{D}=}$ does not depend on $\mathcal{D}$, and thus all terms in the sum in (5.5) are equal to μ. Furthermore, it follows from (3.5) that the number of terms in the sum is $\prod_{d_j < \infty} d_j$, and (3.11) follows.

Alternatively, in the independent case, all terms in the sums in (3.1), (3.2), and (3.3) have the same expectation μ given by (3.10), and the result follows by counting the number of terms. In particular, exactly,

$$\mathbb{E} U_n(f) = \binom{n}{\ell} \mu \tag{5.6}$$

and, with D given by (4.7),

$$\mathbb{E} U_n(f; \mathcal{D}=) = \binom{n-D}{b} \mu. \tag{5.7}$$

□

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

6. Asymptotic normality

The general idea of the proof of Theorem 3.3 is to use the projection method of Hoeffding [27], together with modifications as in [58] to treat m-dependent variables and modifications as in e.g. [37] to treat the asymmetric case. We then obtain the constrained version, Theorem 3.4, by reduction to the unconstrained case.

Proof of Theorem 3.3. We first note that by Lemma 4.4, it suffices to prove (3.18)–(3.19) for $U_n(f;\,>m)$. (This uses standard arguments with Minkowski's inequality and the Cramér–Slutsky theorem [25, Theorem 5.11.4], respectively; we omit the details. The same arguments are used several times below without comment.)

As remarked above, the variables inside each term in the sum in (4.25) are independent; this enables us to use Hoeffding's decomposition for the independent case, which (in the present, asymmetric case) we define as follows.

As in Section 5, let $(\widehat{X}_i)_1^\ell$ be an independent sequence of random variables in $\mathcal{S}$, each with the same distribution as X_1. Recall μ defined in (5.1), and, for $i=1,\dots,\ell$, define the function f_i as the one-variable projection

$$f_i(x) := \mathbb{E} f\bigl(\widehat{X}_1,\dots,\widehat{X}_{i-1},x,\widehat{X}_{i+1},\dots,\widehat{X}_\ell\bigr)-\mu. \tag{6.1}$$

Equivalently,

$$f_i(\widehat{X}_i) = \mathbb{E}\bigl(f(\widehat{X}_1,\dots,\widehat{X}_\ell)\mid \widehat{X}_i\bigr)-\mu. \tag{6.2}$$

(In general, f_i is defined only $\mathcal{L}(\widehat{X}_i)$-almost everywhere (a.e.), but it does not matter which version we choose.) Define also the residual function

$$f_*(x_1,\dots,x_d) := f(x_1,\dots,x_d)-\mu-\sum_{j=1}^{\ell} f_j(x_j). \tag{6.3}$$

Note that the variables $f_i(X_j)$ are centered by (5.1) and (6.2):

$$\mathbb{E} f_i(X_j) = \mathbb{E} f_i(\widehat{X}_i) = 0. \tag{6.4}$$

Furthermore, (A_2) implies that $f_i(\widehat{X}_i)$, and thus each $f_i(X_j)$, is square integrable.

The essential property of f_* is that, as an immediate consequence of the definitions and (6.4), its one-variable projections vanish:

$$\mathbb{E}\bigl(f_*(\widehat{X}_1,\dots,\widehat{X}_\ell)\mid \widehat{X}_i=x\bigr) = \mathbb{E} f_*\bigl(\widehat{X}_1,\dots,\widehat{X}_{i-1},x,\widehat{X}_{i+1},\dots,\widehat{X}_\ell\bigr)=0. \tag{6.5}$$

We assume from now on for simplicity that $\mu=0$; the general case follows by replacing f by $f-\mu$. Then (4.25) and (6.3) yield, by counting the terms where $i_j=k$ for given j and k,

$$\begin{aligned} U_n(f;\,>m) &= \sum_{\substack{1\leqslant i_1<\dots<i_\ell\leqslant n\\ i_{j+1}-i_j>m}} \Bigl(\sum_{j=1}^{\ell} f_j(X_{i_j}) + f_*\bigl(X_{i_1},\dots,X_{i_\ell}\bigr)\Bigr) \\ &= \sum_{j=1}^{\ell}\sum_{k=1}^{n} \binom{k-1-(j-1)m}{j-1}\binom{n-k-(\ell-j)m}{\ell-j} f_j(X_k) + U_n(f_*;\,>m). \end{aligned} \tag{6.6}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Let us first dispose of the last term in (6.6). Let $i_1 < \ldots < i_\ell$ and $j_1 < \ldots < j_\ell$ be two sets of indices such that the constraints $i_{k+1} - i_k > m$ and $j_{k+1} - j_k > m$ in (4.25) hold. First, as in the proof of Lemma 4.1, if also $|i_\alpha - j_\beta| > m$ for all $\alpha, \beta \in [\ell]$, then all X_{i_α} and X_{j_β} are independent; thus $f_*(X_{i_1}, \ldots, X_{i_\ell})$ and $f_*(X_{j_1}, \ldots, X_{j_\ell})$ are independent, and

$$\mathbb{E}\big[f_*(X_{i_1}, \ldots, X_{i_\ell}) f_*(X_{j_1}, \ldots, X_{j_\ell})\big] = \mathbb{E} f_*(X_{i_1}, \ldots, X_{i_\ell}) \mathbb{E} f_*(X_{j_1}, \ldots, X_{j_\ell}) = 0. \tag{6.7}$$

Moreover, suppose that $|i_\alpha - j_\beta| > m$ for all but one pair $(\alpha, \beta) \in [\ell]^2$, say for $(\alpha, \beta) \neq (\alpha_0, \beta_0)$. Then the pair $(X_{i_{\alpha_0}}, X_{j_{\beta_0}})$ is independent of all the variables $\{X_{i_\alpha} : \alpha \neq \alpha_0\}$ and $\{X_{j_\beta} : \beta \neq \beta_0\}$, and all these are mutually independent. Hence, recalling (6.5), a.s.

$$\begin{aligned} &\mathbb{E}\big[f_*(X_{i_1}, \ldots, X_{i_\ell}) f_*(X_{j_1}, \ldots, X_{j_\ell}) \mid X_{i_{\alpha_0}}, X_{j_{\beta_0}}\big] \\ &\quad = \mathbb{E}\big[f_*(X_{i_1}, \ldots, X_{i_\ell}) \mid X_{i_{\alpha_0}}\big] \mathbb{E}\big[f_*(X_{j_1}, \ldots, X_{j_\ell}) \mid X_{j_{\beta_0}}\big] = 0. \end{aligned} \tag{6.8}$$

Thus, taking the expectation, we find that unconditionally

$$\mathbb{E}\big[f_*(X_{i_1}, \ldots, X_{i_\ell}) f_*(X_{j_1}, \ldots, X_{j_\ell})\big] = 0. \tag{6.9}$$

Consequently, if we expand $\mathrm{Var}[U_n(f_*; > m)]$ in analogy with (4.3), then all terms where $|i_\alpha - j_\beta| \leqslant m$ for at most one pair (α, β) will vanish. The number of remaining terms, i.e., those with at least two such pairs (α, β), is $O(n^{2\ell-2})$, and each term is $O(1)$, by (A_2) and the Cauchy–Schwarz inequality. Consequently,

$$\mathrm{Var}\big[U_n(f_*; > m)\big] = O\big(n^{2\ell-2}\big). \tag{6.10}$$

Hence, we may ignore the final term $U_n(f_*; > m)$ in (6.6).

We turn to the main terms in (6.6), i.e., the double sum; we denote it by $\widehat{U}_n$ and write it as

$$\widehat{U}_n = \sum_{j=1}^{\ell} \sum_{k=1}^{n} a_{j,k,n} f_j(X_k), \tag{6.11}$$

where we thus define

$$\begin{aligned} a_{j,k,n} &:= \binom{k-1-(j-1)m}{j-1} \binom{n-k-(\ell-j)m}{\ell-j} \\ &:= \frac{1}{(j-1)!\,(\ell-j)!} k^{j-1} (n-k)^{\ell-j} + O(n^{\ell-2}), \end{aligned} \tag{6.12}$$

where the O is uniform over all $k \leqslant n$ and $j \leqslant \ell$. For $j = 1, \ldots, \ell$, define the polynomial functions

$$\psi_j(x) := \frac{1}{(j-1)!\,(\ell-j)!} x^{j-1} (1-x)^{\ell-j}, \qquad x \in \mathbb{R}. \tag{6.13}$$

Then (6.12) yields, again uniformly for all $k \leqslant n$ and $j \leqslant \ell$,

$$a_{j,k,n} = n^{\ell-1} \psi_j(k/n) + O\big(n^{\ell-2}\big). \tag{6.14}$$

The expansion (6.11) yields

$$\mathrm{Var}\, \widehat{U}_n = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \sum_{k=1}^{n} \sum_{q=1}^{n} a_{i,k,n} a_{j,q,n} \mathrm{Cov}\big[f_i(X_k), f_j(X_q)\big], \tag{6.15}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

where all terms with $|k-q|>m$ vanish because the sequence (X_i) is m-dependent. Hence, with $r_-:=\max\{-r,0\}$ and $r_+:=\max\{r,0\}$,

$$\operatorname{Var}\widehat{U}_n=\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\sum_{r=-m}^{m}\sum_{k=1+r_-}^{n-r_+}a_{i,k,n}a_{j,k+r,n}\operatorname{Cov}\bigl[f_i(X_k),f_j(X_{k+r})\bigr]. \tag{6.16}$$

The covariance in (6.16) is independent of k; we thus define, for any $k>r_-$,

$$\gamma_{i,j,r}:=\operatorname{Cov}\bigl[f_i(X_k),f_j(X_{k+r})\bigr] \tag{6.17}$$

and obtain

$$\operatorname{Var}\widehat{U}_n=\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\sum_{r=-m}^{m}\gamma_{i,j,r}\sum_{k=1+r_-}^{n-r_+}a_{i,k,n}a_{j,k+r,n}. \tag{6.18}$$

Furthermore, by (6.14),

$$\begin{aligned}
n^{2-2\ell}\sum_{k=1+r_-}^{n-r_+}a_{i,k,n}a_{j,k+r,n}&=\sum_{k=1+r_-}^{n-r_+}\bigl(\psi_i(k/n)+O\bigl(n^{-1}\bigr)\bigr)\bigl(\psi_j(k/n)+O\bigl(n^{-1}\bigr)\bigr)\\
&=\sum_{k=1+r_-}^{n-r_+}\bigl(\psi_i(k/n)\psi_j(k/n)+O\bigl(n^{-1}\bigr)\bigr)\\
&=\sum_{k=1}^{n}\psi_i(k/n)\psi_j(k/n)+O(1)\\
&=\int_0^n\psi_i(x/n)\psi_j(x/n)\,\mathrm{d}x+O(1)\\
&=n\int_0^1\psi_i(t)\psi_j(t)\,\mathrm{d}t+O(1).
\end{aligned} \tag{6.19}$$

Consequently, (6.18) yields

$$n^{1-2\ell}\operatorname{Var}\widehat{U}_n=\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\sum_{r=-m}^{m}\gamma_{i,j,r}\int_0^1\psi_i(t)\psi_j(t)\,\mathrm{d}t+O\bigl(n^{-1}\bigr). \tag{6.20}$$

Since (4.26), (6.6), and (6.10) yield

$$\operatorname{Var}\bigl[U_n(f)-\widehat{U}_n\bigr]=O\bigl(n^{2\ell-2}\bigr), \tag{6.21}$$

the result (3.18) follows from (6.20), with

$$\sigma^2=\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\sum_{r=-m}^{m}\gamma_{i,j,r}\int_0^1\psi_i(t)\psi_j(t)\,\mathrm{d}t. \tag{6.22}$$

Next, we use (6.11) and write

$$n^{\frac12-\ell}\widehat{U}_n=\sum_{k=1}^{n}Z_{kn}, \tag{6.23}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

with

$$Z_{kn} := \sum_{j=1}^{\ell} n^{\frac{1}{2}-\ell} a_{j,k,n} f_j(X_k). \tag{6.24}$$

Since Z_{kn} is a function of X_k, it is evident that (Z_{kn}) is an m-dependent triangular array with centered variables. Furthermore, $\mathbb{E}Z_{kn} = 0$ as a consequence of (6.4).

We apply Theorem 4.1 to (Z_{kn}), so $\widehat{S}_n = n^{\frac{1}{2}-\ell}\widehat{U}_n$ by (6.23), and verify first its conditions. The condition (4.30) holds by (6.20) and (6.22). Write $Z_{kn} = \sum_{j=1}^{\ell} Z_{jkn}$ with

$$Z_{jkn} := n^{\frac{1}{2}-\ell} a_{j,k,n} f_j(X_k). \tag{6.25}$$

Since (6.12) yields $|a_{j,k,n}| \leqslant n^{\ell-1}$, we have, for $\varepsilon \geqslant 0$,

$$\mathbb{E}\big[Z_{jkn}^2 \mathbf{1}\{|Z_{jkn}| > \varepsilon\}\big] \leqslant n^{-1}\mathbb{E}\big[|f_j(X_k)|^2 \mathbf{1}\{|f_j(X_k)| > \varepsilon n^{1/2}\}\big]. \tag{6.26}$$

The distribution of $f_j(X_k)$ does not depend on k, and thus the Lindeberg condition (4.31) for each triangular array $(Z_{jkn})_{k,n}$ follows from (6.26). The Lindeberg condition (4.31) for $(Z_{nk})_{k,n}$ then follows easily. Finally, taking $\varepsilon = 0$ in (6.26) yields $\mathbb{E}Z_{jkn}^2 \leqslant Cn^{-1}$, and thus $\mathbb{E}Z_{kn}^2 \leqslant Cn^{-1}$, which shows (4.32).

We have shown that Theorem 4.1 applies, and thus, recalling (6.23) and (6.4),

$$n^{\frac{1}{2}-\ell}\big(\widehat{U}_n - \mathbb{E}\widehat{U}_n\big) = n^{\frac{1}{2}-\ell}\widehat{U}_n = \sum_{k=1}^{n} Z_{kn} \xrightarrow{\mathrm{d}} \mathsf{N}\big(0, \sigma^2\big). \tag{6.27}$$

The result (3.19) now follows from (6.27) and (6.21). □

Proof of Theorem 3.4. Lemma 4.3 implies that it suffices to consider $U_n(g; (Y_i))$ instead of $U_n(f; \mathcal{D})$ or $U_n(f; \mathcal{D}{=})$. Note that the definition (4.8) implies that $(Y_i)_1^\infty$ is a stationary m'-dependent sequence, with $m' := m + M - 1$. Hence, the result follows from Theorem 3.3 applied to g and $(Y_i)_1^\infty$. □

Remark 6.1. The integrals in (6.22) are standard beta integrals [46, 5.12.1]; we have

$$\begin{aligned}\int_0^1 \psi_i(t)\psi_j(t)\,\mathrm{d}t &= \frac{1}{(i-1)!\,(j-1)!\,(\ell-i)!\,(\ell-j)!}\int_0^1 t^{i+j-2}(1-t)^{2\ell-i-j}\,\mathrm{d}t \\ &= \frac{(i+j-2)!\,(2\ell-i-j)!}{(i-1)!\,(j-1)!\,(\ell-i)!\,(\ell-j)!\,(2\ell-1)!}.\end{aligned} \tag{6.28}$$

Remark 6.2. In the unconstrained case Theorem 3.3, the asymptotic variance σ^2 is given by (6.22) together with (6.17), (6.1), and (6.28).

In the constrained cases, the proof above shows that σ^2 is given by (6.22) applied to the function g given by Lemma 4.3 and $(Y_i)_1^\infty$ given by (4.8) (with $M = D + 1$ for definiteness); note that this also entails replacing ℓ by b and m by $m + M - 1 = m + D$ in the formulas above. In particular, in the exactly constrained case (3.3), it follows from (6.1) and (4.18) that, with $y = (x_1, \dots, x_M) \in \mathcal{S}^M$ and other notation as in (4.11)–(4.14) and (5.4),

$$g_i(x_1, \dots, x_M) = \mathbb{E}f\Big(\big(X_{j_1+v_1+t_{1r}}\big)_{r=1}^{\ell_1}, \dots, \big(x_{1+v_i+t_{ir}}\big)_{r=1}^{\ell_i}, \dots, \big(X_{j_b+v_b+t_{br}}\big)_{r=1}^{\ell_b}\Big) - \mu_{\mathcal{D}=}, \tag{6.29}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

where the ith group of variables consists of the given x_i, the other $b-1$ groups contain variables X_i, and $j_1, \dots, j_b$ is any sequence of indices that has large enough gaps: $j_{i+1} - j_i > m + M - 1 = m + D$.

In the constrained case (3.2), $g = g_{\mathcal{D}}$ is obtained as the sum (4.23), and thus each g_i is a similar sum of functions that can be obtained as (6.29). (Note that $M := D + 1$ works in Lemma 4.2 for all terms by (3.5).) Then, σ^2 is given by (6.22) (with substitutions as above).

7. Law of large numbers

Proof of Theorem 3.2. Note first that if R_n is any sequence of random variables such that

$$\mathbb{E}R_n^2 = O\bigl(n^{-2}\bigr), \tag{7.1}$$

then Markov's inequality and the Borel–Cantelli lemma show that $R_n \xrightarrow{\text{a.s.}} 0$.

We begin with the unconstrained case, $\mathcal{D} = \mathcal{D}_\infty = (\infty, \dots, \infty)$. We may assume, as in the proof of Theorem 3.3, that $\mu = 0$. Then (6.21) holds, and thus by the argument just given, and recalling that $\mathbb{E}\widehat{U}_n = 0$ by (6.11) and (6.4),

$$n^{-\ell}\bigl[U_n(f) - \mathbb{E}U_n(f) - \widehat{U}_n\bigr] \xrightarrow{\text{a.s.}} 0. \tag{7.2}$$

Hence, to prove (3.15), it suffices to prove $n^{-\ell}\widehat{U}_n \xrightarrow{\text{a.s.}} 0$.

For simplicity, we fix $j \in [\ell]$ and define, with f_j as above given by (6.1),

$$S_{jn} = S_{jn}(f) := S_n(f_j) := \sum_{k=1}^{n} f_j(X_k) \tag{7.3}$$

and, using partial summation,

$$\widehat{U}_{jn} := \sum_{k=1}^{n} a_{j,k,n} f_j(X_k) = \sum_{k=1}^{n-1} (a_{j,k,n} - a_{j,k+1,n}) S_{jk} + a_{j,n,n} S_{jn}. \tag{7.4}$$

The sequence $(f_j(X_k))_k$ is m-dependent, stationary, and with $\mathbb{E}|f_j(X_k)| < \infty$. As is well known, the strong law of large numbers holds for stationary m-dependent sequences with finite means. (This follows by considering the subsequences $(X_{(m+1)n+q})_{n\geqslant 0}$, which for each fixed $q \in [m+1]$ is an i.i.d. sequence.) Thus, by (7.3) and (6.4),

$$S_{jn}/n \xrightarrow{\text{a.s.}} \mathbb{E}f_j(X_k) = 0. \tag{7.5}$$

In other words, a.s. $S_{jn} = o(n)$, and thus also

$$\max_{1\leqslant k\leqslant n} |S_{jk}| = o(n) \qquad \text{a.s.} \tag{7.6}$$

Moreover, (6.12) implies $a_{j,k,n} - a_{j,k+1,n} = O(n^{\ell-2})$. Hence, (7.4) yields

$$n^{-\ell}\widehat{U}_{jn} = \sum_{k=1}^{n-1} O\bigl(n^{-2}\bigr) \cdot S_{jk} + O\bigl(n^{-1}\bigr) \cdot S_{jn}, \tag{7.7}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

and thus, using (7.6),

$$\bigl|n^{-\ell}\widehat{U}_{jn}\bigr| \leqslant Cn^{-1}\max_{k\leqslant n}|S_{jk}| = o(1) \qquad \text{a.s.} \tag{7.8}$$

Consequently,

$$n^{-\ell}\widehat{U}_n = \sum_{j=1}^{\ell} n^{-\ell}\widehat{U}_{jn} \xrightarrow{\text{a.s.}} 0, \tag{7.9}$$

which together with (7.2) yields the desired result (3.15).

Next, for an exact constraint $\mathcal{D}=$, we use Lemma 4.2. Then (4.9) together with the result just shown applied to g and (Y_i) yields

$$n^{-b}\bigl[U_n(f;\mathcal{D}=) - \mathbb{E}U_n(f;\mathcal{D}=)\bigr] = n^{-b}\bigl[U_{n-D}(g) - \mathbb{E}U_{n-D}(g)\bigr] \xrightarrow{\text{a.s.}} 0. \tag{7.10}$$

This proves (3.17), and (3.16) follows by (3.4).

Finally, using Theorem 3.1, (3.12)–(3.14) are equivalent to (3.15)–(3.17). □

8. The degenerate case

As is well known, even in the original symmetric and independent case studied in [27], the asymptotic variance σ^2 in Theorem 3.3 may vanish also in non-trivial cases. In such cases, (3.19) is still valid, but says only that the left-hand side converges to 0 in probability. In the present section, we characterize this degenerate case in Theorems 3.3 and 3.4. Note that in applications, it is frequently natural to guess that $\sigma^2 > 0$, but this is sometimes surprisingly difficult to prove. One purpose of the theorems below is to assist in showing $\sigma^2 > 0$; see the applications in Sections 13 and 14.

For an unconstrained U-statistic and an independent sequence $(X_i)_1^\infty$ (the case $m = 0$ of Theorem 3.3), it is known, and not difficult to see, that $\sigma^2 = 0$ if and only if every projection $f_i(X_1)$ defined by (6.1) vanishes a.s.; see [37, Corollary 3.5]. (This is included in the theorem below by taking $m = 0$ in (iii), and it is also the correct interpretation of (vi) when $m = 0$.) In the m-dependent case, the situation is similar, but somewhat more complicated, as shown by the following theorem. Note that $S_n(f_j)$ defined in (8.8) below equals S_{jn}; for later applications we find this change of notation convenient.

Theorem 8.1. *With assumptions and notation as in Theorem 3.3, define f_i by (6.1), $\gamma_{i,j,r}$ by (6.17) and S_{jn} by (7.3). Then the following are equivalent:*

(i)
$$\sigma^2 = 0. \tag{8.1}$$

(ii)
$$\operatorname{Var} U_n = O\bigl(n^{2\ell-2}\bigr). \tag{8.2}$$

(iii)
$$\sum_{r=-m}^{m} \gamma_{i,j,r} = 0, \qquad \forall i, j \in [\ell]. \tag{8.3}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

(iv)

$$\operatorname{Cov}\bigl[S_{in}, S_{jn}\bigr]/n \to 0 \text{ as } n \to \infty \quad \forall i, j \in [\ell]. \tag{8.4}$$

(v)

$$\operatorname{Var}\bigl[S_{jn}\bigr]/n \to 0 \text{ as } n \to \infty \quad \forall j \in [\ell]. \tag{8.5}$$

(vi) *For each $j \in [\ell]$ there exists a stationary sequence $(Z_{j,k})_{k=0}^{\infty}$ of $(m-1)$-dependent random variables such that a.s.*

$$f_j(X_k) = Z_{j,k} - Z_{j,k-1}, \qquad k \geqslant 1. \tag{8.6}$$

Moreover, suppose that the sequence $(X_k)_1^{\infty}$ is a block factor given by (2.2) for some function h and i.i.d. ξ_i, and that $\sigma^2 = 0$. Then, in (vi), we may take $Z_{j,k}$ as block factors

$$Z_{j,k} = \varphi_j(\xi_{k+1}, \dots, \xi_{k+m}), \tag{8.7}$$

for some functions $\varphi_j : \mathcal{S}_0^m \to \mathbb{R}$. Hence, for every $j \in [\ell]$ and $n \geqslant 1$,

$$S_n(f_j) := \sum_{k=1}^{n} f_j(X_k) = Z_{j,n} - Z_{j,0} = \varphi_j(\xi_{n+1}, \dots, \xi_{n+m}) - \varphi_j(\xi_1, \dots, \xi_m), \tag{8.8}$$

and thus $S_n(f_j)$ is independent of $\xi_{m+1}, \dots, \xi_n$ for every $j \in [\ell - 1]$ and $n > m$.

To prove Theorem 8.1, we begin with a well-known algebraic lemma; for completeness we include a proof.

Lemma 8.1. *Let $A = (a_{ij})_{i,j=1}^{\ell}$ and $B = (b_{ij})_{i,j=1}^{\ell}$ be symmetric real matrices such that A is positive definite and B is positive semidefinite. Then*

$$\sum_{i,j=1}^{\ell} a_{ij} b_{ij} = 0 \iff b_{ij} = 0 \quad \forall i, j \in [\ell]. \tag{8.9}$$

Proof. Since A is positive definite, there exists an orthonormal basis $(v_k)_1^{\ell}$ in $\mathbb{R}^{\ell}$ consisting of eigenvectors of A, in other words satisfying $Av_k = \lambda_k v_k$; furthermore, the eigenvalues $\lambda_k > 0$. Write $v_k = (v_{ki})_{i=1}^{\ell}$. We then have

$$a_{ij} = \sum_{k=1}^{\ell} \lambda_k v_{ki} v_{kj}. \tag{8.10}$$

Thus

$$\sum_{i,j=1}^{\ell} a_{ij} b_{ij} = \sum_{k=1}^{\ell} \lambda_k \sum_{i,j=1}^{\ell} b_{ij} v_{ki} v_{kj} = \sum_{k=1}^{\ell} \lambda_k \langle v_k, B v_k \rangle. \tag{8.11}$$

Since B is positive semidefinite, all terms in the last sum are $\geqslant 0$, so the sum is 0 if and only if every term is, and thus

$$\sum_{i,j=1}^{\ell} a_{ij} b_{ij} = 0 \iff \langle v_k, B v_k \rangle = 0 \quad \forall k \in [\ell]. \tag{8.12}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

By the Cauchy–Schwarz inequality for the semidefinite bilinear form $\langle v, Bw\rangle$ (or, alternatively, by using $\langle v_k \pm v_n, B(v_k \pm v_n)\rangle \geqslant 0$), it follows that this condition implies $\langle v_k, Bv_n\rangle = 0$ for any $k, n \in [\ell]$, and thus

$$\sum_{i,j=1}^{\ell} a_{ij}b_{ij} = 0 \iff \langle v_k, Bv_n\rangle = 0 \quad \forall k, n \in [\ell]. \tag{8.13}$$

Since $(v_k)_1^\ell$ is a basis, this is further equivalent to $\langle v, Bw\rangle = 0$ for any $v, W \in \mathbb{R}^\ell$, and thus to $B = 0$. This yields (8.9). □

Proof of Theorem 8.1. The ℓ polynomials ψ_j, $j = 1, \dots, \ell$, of degree $\ell - 1$ defined by (6.13) are linearly independent (e.g., since the matrix of their coefficients in the standard basis $\{1, x, \dots, x^{\ell-1}\}$ is upper triangular with non-zero diagonal elements). Hence, the Gram matrix $A = (a_{ij})_{i,j}$ with

$$a_{ij} := \int_0^1 \psi_i(t)\psi_j(t)\,\mathrm{d}t \tag{8.14}$$

is positive definite.

We have by (7.3), similarly to (6.15)–(6.18),

$$\begin{aligned}\operatorname{Cov}\bigl(S_{in}, S_{jn}\bigr) &= \sum_{k=1}^{n}\sum_{q=1}^{n} \operatorname{Cov}\bigl[f_i(X_k), f_j(X_q)\bigr] = \sum_{r=-m}^{m}\sum_{k=1+r_-}^{n-r_+} \operatorname{Cov}\bigl[f_i(X_k), f_j(X_{k+r})\bigr] \\ &= \sum_{r=-m}^{m} (n-|r|)\operatorname{Cov}\bigl[f_i(X_k), f_j(X_{k+r})\bigr] = \sum_{r=-m}^{m} (n-|r|)\gamma_{i,j,r},\end{aligned} \tag{8.15}$$

and thus, as $n \to \infty$,

$$\operatorname{Cov}\bigl(S_{in}, S_{jn}\bigr)/n \to \sum_{r=-m}^{m} \gamma_{i,j,r} =: b_{ij}. \tag{8.16}$$

Note that (6.22) can be written as

$$\sigma^2 = \sum_{i,j=1}^{\ell} b_{ij}a_{ij}. \tag{8.17}$$

The covariance matrices $(\operatorname{Cov}(S_{in}, S_{jn}))_{i,j=1}^{\ell}$ are positive semidefinite, and thus so is the limit $B = (b_{ij})$ defined by (8.16). Hence Lemma 8.1 applies and yields, using (8.17) and the definition of b_{ij} in (8.16), the equivalence (i) $\iff$ (iii).

Furthermore, (8.16) yields (iii) $\iff$ (iv).

The implication (iv) $\implies$ (v) is trivial, and the converse follows by the Cauchy–Schwarz inequality.

If (iii) holds, then (6.20) yields $\operatorname{Var} \widehat{U}_n = O\bigl(n^{2\ell-2}\bigr)$, and (ii) follows by (6.21). Conversely, (ii) $\implies$ (i) by (3.18).

Moreover, for $m \geqslant 1$, (v) $\iff$ (vi) holds by [36, Theorem 1], recalling $\mathbb{E}f_j(X_k) = 0$ by (6.4). (Recall also that any stationary sequence $(W_k)_1^\infty$ of real random variables can be extended

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

to a doubly-infinite stationary sequence $(W_k)_{-\infty}^{\infty}$.) The case $m=0$ is trivial, since then (v) is equivalent to $\operatorname{Var} f_j(X_k)=0$ and thus $f_j(X_k)=0$ a.s. by (6.4), while (vi) should be interpreted to mean that (8.6) holds for some non-random $Z_{j,k}=z_j$.

Finally, suppose that $(X_i)_1^\infty$ is a block factor. In this case, [36, Theorem 2] shows that $Z_{j,k}$ can be chosen as in (8.7). (Again, the case $m=0$ is trivial.) Then (8.8) is an immediate consequence of (8.6)–(8.7). □

Remark 8.1. It follows from the proof in [36] that in (vi), we can choose Z_{jk} such that also the random vectors $(Z_{jk})_{j=1}^{\ell}$, $k\geqslant 0$, form a stationary $(m-1)$-dependent sequence.

Theorem 8.2. *With assumptions and notation as in Theorem 3.4, define also g_i, $i\in[b]$, as in Remark 6.2, i.e., by (6.29) in the exactly constrained case and otherwise as a sum of such terms over all $\mathcal{D}'$ given by (3.5). Also (again as in Remark 6.2), let D be given by (4.7) and Y_k by (4.8) with $M=D+1$. Then $\sigma^2=0$ if and only if for every $j\in[b]$, there exists a stationary sequence $(Z_{j,k})_{k=0}^{\infty}$ of $(m+D-1)$-dependent random variables such that a.s.*

$$g_j(Y_k)=Z_{j,k}-Z_{j,k-1},\qquad k\geqslant 1. \tag{8.18}$$

Moreover, if the sequence $(X_i)_1^\infty$ is independent and $\sigma^2=0$, then there exist functions $\varphi_j:\mathcal{S}^D\to\mathbb{R}$ such that (8.18) holds with

$$Z_{j,k}=\varphi_j(X_{k+1},\dots,X_{k+D}), \tag{8.19}$$

and consequently a.s.

$$S_n(g_j):=\sum_{k=1}^{n} g_j(Y_k)=\varphi_j(X_{n+1},\dots,X_{n+D})-\varphi_j(X_1,\dots,X_D); \tag{8.20}$$

thus $S_n(g_j)$ is independent of $X_{D+1},\dots,X_n$ for every $j\in[\ell-1]$ and $n>D$.

Proof. As in the proof of Theorem 3.4, it suffices to consider $U_n(g)$ with g given by Lemma 4.3 (with $M=D+1$). The first part then is an immediate consequence of Theorem 8.1(i)⇔(vi) applied to g and $Y_i:=(X_i,\dots,X_{i+D})$, with appropriate substitutions $\ell\mapsto b$ and $m\mapsto m+D$.

The second part follows similarly by the last part of Theorem 8.1, with $\xi_i=X_i$; note that then (Y_i) is a block factor as in (2.2), with m replaced by D. □

Remark 8.2. Of course, under the assumptions of Theorem 8.2, the other equivalences in Theorem 8.1 hold as well, with the appropriate interpretations, substituting g for f and so on.

We give an example of a constrained U-statistic where $\sigma^2=0$ in a somewhat non-trivial way.

Example 8.1. Let $(X_i)_1^\infty$ be an infinite i.i.d. symmetric random binary string; i.e., $\mathcal{S}=\{0,1\}$ and $X_i\sim \mathrm{Be}(1/2)$ are i.i.d. Let

$$f(x,y,z):=\mathbf{1}\{xyz=101\}-\mathbf{1}\{xyz=011\} \tag{8.21}$$

and consider the constrained U-statistic

$$U_n(f;\mathcal{D})=\sum_{1\leqslant i<i+1<j\leqslant n} f\bigl(X_i,X_{i+1},X_j\bigr), \tag{8.22}$$

which thus has constraint $\mathcal{D}=(1,\infty)$. (In this case, $U_n(f,\mathcal{D})=U_n(f;\mathcal{D}{=})$.) Note that (8.22) is a difference of two constrained subsequence counts.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Although the function (8.21) might look non-trivial and innocuous at first glance, this turns out to be a degenerate case. In fact, it is easily verified that

$$f(x, y, z) = (x - y)z, \qquad x, y, z \in \{0, 1\}. \tag{8.23}$$

Hence, with $m = 0$, $D = 1$, and $M = D + 1 = 2$, (5.4) yields

$$\mu_{\mathcal{D}} = \mu_{\mathcal{D}=} = \mathbb{E}g(Y_1, Y_3) = \mathbb{E}f(X_1, X_2, X_4) = 0, \tag{8.24}$$

while (6.29) yields

$$g_1(x, y) = \mathbb{E}f(x, y, X_4) = \mathbb{E}\big[(x - y)X_4\big] = \tfrac{1}{2}(x - y), \tag{8.25}$$

$$g_2(x, y) = \mathbb{E}f(X_1, X_2, y) = \mathbb{E}\big[(X_1 - X_2)y\big] = 0. \tag{8.26}$$

Thus g_2 vanishes but not g_1. Nevertheless, $g_1(Y_k) = g_1(X_k, X_{k+1}) = \frac{1}{2}(X_k - X_{k+1})$ is of the type in (8.18)–(8.19) (with $Z_{1,k} := -\frac{1}{2}X_{k+1}$). Hence, Theorem 8.2 shows that $\sigma^2 = 0$, and thus Theorem 3.4 and (3.8) yield $n^{-3/2}U_n(f; \mathcal{D}=) \xrightarrow{\mathrm{p}} 0$.

In fact, in this example we have by (8.23), for $n \geqslant 3$,

$$\begin{aligned} U_n(f; \mathcal{D}) &= \sum_{j=3}^{n} \sum_{i=1}^{j-2} (X_i - X_{i+1})X_j = \sum_{j=3}^{n} X_j(X_1 - X_{j-1}) \\ &= X_1 \sum_{j=3}^{n} X_j - \sum_{j=3}^{n} X_{j-1}X_j. \end{aligned} \tag{8.27}$$

Hence, by the law of large numbers for stationary m-dependent sequences,

$$n^{-1}U_n(f; \mathcal{D}) \xrightarrow{\text{a.s.}} X_1\mathbb{E}X_2 - \mathbb{E}\big[X_2X_3\big] = \tfrac{1}{2}X_1 - \tfrac{1}{4} = \tfrac{1}{2}\big(X_1 - \tfrac{1}{2}\big). \tag{8.28}$$

As a consequence, $n^{-1}U_n(f; \mathcal{D})$ has a non-degenerate limiting distribution. Note that this example differs in several respects from the degenerate cases that may occur for standard U-statistics, i.e. unconstrained U-statistics based on independent (X_i). In this example, (8.28) shows that the asymptotic distribution is a linear transformation of a Bernoulli variable, and is thus neither normal, nor of the type that appears as limits of degenerate standard U-statistics. (The latter are polynomials in independent normal variables, in general infinitely many; see e.g. Theorem A.2 and, in general, [56] and [35, Chapter 11].) Moreover, the a.s. convergence to a non-degenerate limit is unheard of for standard U-statistics, where the limit is mixing.

8.1. The degenerate case in renewal theory

In the renewal theory setting in Theorem 3.8, the degenerate case is characterized by a modified version of the conditions above.

Theorem 8.3. *With the assumptions and notation of Theorem 3.8, let g_i, $i \in [b]$, be as in Theorem 8.2 and Remark 6.2. Then $\gamma^2 = 0$ if and only if for every $j \in [b]$, the function*

$$\widetilde{g}_j(y) := g_j(y) + \mu_{\mathcal{D}} - \frac{\mu_{\mathcal{D}}}{\nu}h(y_1), \qquad y = (y_1, \dots, y_b) \in \mathcal{S}^b, \tag{8.29}$$

satisfies the condition (8.18). Moreover, if the sequence $(X_i)_1^\infty$ is independent and $\gamma^2 = 0$, then the functions $\widetilde{g}_j$ also satisfy (8.19)–(8.20).

The proof is given in Section 12. Note that $\mathbb{E}\widetilde{g}_j(Y_1) = 0$ for each $j \in [b]$ by (6.4) and (3.25).

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

9. Rate of convergence

Here we use a different method from that used in the rest of the paper.

Proof of Theorem 3.5. We consider $U_n(f,\mathcal{D})$; the argument for $U_n(f;\mathcal{D}=)$ is identical, and $U_n(f)$ is a special case.

Let $\mathcal{I}$ denote the set of all indices $(i_1,\dots,i_\ell)$ in the sum (3.2); thus (3.2) can be written $U_n(f;\mathcal{D})=\sum_{I\in\mathcal{I}} Z_I$, where $Z_{i_1,\dots,i_\ell}:=f(X_{i_1},\dots,X_{i_\ell})$. Note that the size $|\mathcal{I}|\sim Cn^b$ for some $C>0$, where $b=b(\mathcal{D})$.

We define a graph $\widehat{\mathcal{I}}$ with vertex set $\mathcal{I}$ by putting an edge between $I=(i_1,\dots,i_\ell)$ and $I'=(i'_1,\dots,i'_\ell)$ if and only if $|i_j-i'_k|\leqslant m$ for some $j,k\in\{1\dots,\ell\}$. Let Δ be $1+$ the maximum degree of the graph $\widehat{\mathcal{I}}$; it is easy to see that $\Delta=O\bigl(n^{b-1}\bigr)$. Moreover, it follows from the m-dependence of (X_i) that $\widehat{\mathcal{I}}$ is a *dependency graph* for the random variables $(Z_I)_I$, meaning that if A and B are two disjoint subsets of $\mathcal{I}$ such that there is no edge between A and B, then the two random vectors $(Z_I)_{I\in A}$ and $(Z_I)_{I\in B}$ are independent.

The result now follows from [54, Theorem 2.2], which in our notation yields the following bound, with $\sigma_n^2:=\operatorname{Var}U_n\sim\sigma^2n^{2b-1}$ and $B:=2\sup|f|$, which implies $|Z_I-\mathbb{E}Z_I|\leqslant B$ a.s. for every $I\in\mathcal{I}$:

$$\begin{aligned} d_K &\leqslant \frac{1}{\sigma_n}\biggl\{(2\pi)^{-1/2}\Delta B+16\Bigl(\frac{|\mathcal{I}|\Delta}{\sigma_n^2}\Bigr)^{1/2}\Delta B^2+10\Bigl(\frac{|\mathcal{I}|\Delta}{\sigma_n^2}\Bigr)\Delta B^3\biggr\}\\ &\leqslant C\frac{\Delta}{\sigma_n}\leqslant Cn^{-1/2}, \end{aligned} \tag{9.1}$$

since $|\mathcal{I}|\Delta\leqslant Cn^{b+b-1}\leqslant C\sigma_n^2$ and B is a constant. (Alternatively, one could use the similar bound in [20, Theorem 2.1].) □

Remark 9.1. The assumption in Theorem 3.5 that f is bounded can be relaxed to the 6th moment condition (A_6) by using Theorem 2.1 instead of Theorem 2.2 of [54], together with Hölder's inequality and straightforward estimates.

The similar bound [2, Corollary 2] gives the weaker estimate $d_K=O(n^{-1/4})$, assuming again that f is bounded; this can be relaxed to (A_4) by instead using [2, Theorem 1].

If, instead of the Kolmogorov distance, we use the Wasserstein distance d_W (see e.g. [14, pp. 63–64] for several equivalent definitions, and for several alternative names), the estimate $d_W=O(n^{-1/2})$ follows similarly from [3, Theorem 1], assuming only the third moment condition (A_3); we omit the details. (Actually, [3] does not state the result for the Wasserstein distance but for a weaker version called bounded Wasserstein distance; however, the same proof yields estimates for d_W.) See also [51, Theorem 3 and Remark 3], which yield the same estimate under (A_3), and furthermore imply convergence in distribution assuming only (A_2). (This thus yields an alternative proof of Theorems 3.3 and 3.4.)

Returning to the Kolmogorov distance, we do not believe that the moment assumption (A_6) in Remark 9.1 is the best possible. For unconstrained and symmetric U-statistics, [Malevich and Abdalimov [43], Theorem 2] have shown bounds for the Kolmogorov distance, which in particular show that then (A_3) is sufficient to yield $d_K=O(n^{-1/2})$; we conjecture that the same holds in our, more general, setting.

Conjecture 9.1. *Theorem 3.5 holds assuming only (A_3) (instead of f bounded).*

Remark 9.2. If we do not care about the rate of convergence, for bounded f we can alternatively obtain convergence in distribution in (3.22), and thus in (3.19) and (3.21), by [34, Theorem 2], using the dependency graph $\widehat{\mathcal{I}}$ in the proof of Theorem 3.5. This can easily

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

be extended to any f satisfying the second moment condition (A_2) by a standard truncation argument.

10. Higher moments and maximal functions

To prove Theorem 3.6, we will show estimates for maximal functions that will also be used in Sections 11 and 12. Let $p \geqslant 2$ be fixed throughout the section; explicit and implicit constants may thus depend on p. We let

$$U_n^*(f) := \max_{j \leqslant n} |U_j(f)|, \tag{10.1}$$

and use similar notation for maximal functions of other sequences of random variables.

We use another decomposition of f and $U_n(f)$ which was used in [37] for the independent case ($m = 0$); unlike Hoeffding's decomposition in Section 6, it focuses on the order of the arguments.

Recall from Section 5 that $(\widehat{X}_i)_1^\ell$ are i.i.d. with the same distribution as X_1. Let $\widehat{F}_0 := \mu$ defined in (5.1) and, for $1 \leqslant k \leqslant \ell$,

$$\widehat{F}_k(x_1, \dots, x_k) := \mathbb{E} f\big(x_1, \dots, x_k, \widehat{X}_{k+1}, \dots, \widehat{X}_\ell\big), \tag{10.2}$$

$$F_k(x_1, \dots, x_k) := \widehat{F}_k\big(x_1, \dots, x_k)\big) - \widehat{F}_{k-1}(x_1, \dots, x_{k-1}. \tag{10.3}$$

(These are defined at least for $\mathcal{L}(X_1)$-a.e. $x_1, \dots, x_k \in \mathcal{S}$, which is enough for our purposes.) In other words, a.s.,

$$\widehat{F}_k(\widehat{X}_1, \dots, \widehat{X}_k) = \mathbb{E}\big(f(\widehat{X}_1, \dots, \widehat{X}_\ell) \mid \widehat{X}_1, \dots, \widehat{X}_k\big), \tag{10.4}$$

and thus $\widehat{F}_k(\widehat{X}_1, \dots, \widehat{X}_k)$, $k = 0, \dots, \ell$, is a martingale, with the martingale differences $F_k(\widehat{X}_1, \dots, \widehat{X}_k)$, $k = 1, \dots, \ell$. Hence, or directly from (10.2)–(10.3), for a.e. $x_1, \dots, x_{k-1}$,

$$\mathbb{E} F_k(x_1, \dots, x_{k-1}, \widehat{X}_k) = 0. \tag{10.5}$$

Furthermore, if (A_p) holds, then by (10.4) and Jensen's inequality,

$$\|\widehat{F}_k(\widehat{X}_1, \dots, \widehat{X}_k)\|_p \leqslant \|f(\widehat{X}_1, \dots, \widehat{X}_\ell)\|_p \leqslant C, \tag{10.6}$$

and thus by (10.3),

$$\|F_k(\widehat{X}_1, \dots, \widehat{X}_k)\|_p \leqslant 2\|f(\widehat{X}_1, \dots, \widehat{X}_\ell)\|_p \leqslant C. \tag{10.7}$$

Lemma 10.1. *Suppose that* (A_p) *holds for some* $p \geqslant 2$, *and that* $\mu = 0$. *Then*

$$\big\| U_n^*(f; > m) \big\|_p \leqslant C n^{\ell - 1/2}. \tag{10.8}$$

Proof. We argue as in [37, Lemmas 4.4 and 4.7] with some minor differences. By (10.2)–(10.3), $f(x_1, \dots, x_\ell) = \widehat{F}_\ell(x_1, \dots, x_\ell) = \sum_{k=1}^\ell F_k(x_1, \dots, x_k)$ for a.e. $x_1, \dots, x_\ell$, and thus, a.s.,

$$\begin{aligned} U_n(f; > m) &= \sum_{k=1}^{\ell} \sum_{\substack{1 \leqslant i_1 < \dots < i_k \leqslant n \\ i_{j+1} - i_j > m}} \binom{n - i_k - (\ell - k)m}{\ell - k} F_k(X_{i_1}, \dots, X_{i_k}) \\ &= \sum_{k=1}^{\ell} \sum_{i=1}^{n} \binom{n - i - (\ell - k)m}{\ell - k} \big(U_i(F_k; > m) - U_{i-1}(F_k; > m)\big) \\ &= U_n(F_\ell; > m) + \sum_{k=1}^{\ell-1} \sum_{i=1}^{n-1} \binom{n - i - (\ell - k)m - 1}{\ell - k - 1} U_i(F_k; > m), \end{aligned} \tag{10.9}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

using a summation by parts and the identity $\binom{n'}{\ell-k}-\binom{n'-1}{\ell-k}=\binom{n'-1}{\ell-k-1}$. In particular,

$$\begin{aligned}|U_n(f;\,>m)| &\leqslant |U_n(F_\ell;\,>m)| + \sum_{k=1}^{\ell-1}\sum_{i=1}^{n-1}\binom{n-i-(\ell-k)m-1}{\ell-k-1}U_n^*(F_k;\,>m)\\ &= |U_n(F_\ell;\,>m)| + \sum_{k=1}^{\ell-1}\binom{n-(\ell-k)m-1}{\ell-k}U_n^*(F_k;\,>m)\\ &\leqslant \sum_{k=1}^{\ell} n^{\ell-k}U_n^*(F_k;\,>m). \end{aligned}\tag{10.10}$$

Since the right-hand side is weakly increasing in n, it follows that, a.s.,

$$U_n^*(f;\,>m)\leqslant \sum_{k=1}^{\ell} n^{\ell-k}U_n^*(F_k;\,>m).\tag{10.11}$$

We thus may consider each F_k separately. Let $1\leqslant k\leqslant \ell$, and let

$$\Delta U_n(F_k;\,>m) := U_n(F_k;\,>m)-U_{n-1}(F_k;\,>m).\tag{10.12}$$

By the definition (4.25), $\Delta U_n(F_k;\,>m)$ is a sum of $\binom{n-(k-1)m-1}{k-1}\leqslant n^{k-1}$ terms $F_k\bigl(X_{i_1},\dots,X_{i_{k-1}},X_n\bigr)$ that all have the same distribution as $F_k(\widehat{X}_1,\dots,\widehat{X}_k)$, and thus by Minkowski's inequality and (10.7),

$$\|\Delta U_n(F_k;\,>m)\|_p\leqslant n^{k-1}\|F_k(\widehat{X}_1,\dots,\widehat{X}_k)\|_p\leqslant Cn^{k-1}.\tag{10.13}$$

Furthermore, in each such term $F_k\bigl(X_{i_1},\dots,X_{i_{k-1}},X_n\bigr)$ we have $i_{k-1}\leqslant n-m-1$. Hence, if we let $\mathcal{F}_i$ be the σ-field generated by $X_1,\dots,X_i$, then, by m-dependence, X_n is independent of $\mathcal{F}_{i_{k-1}}$, whence (10.5) implies

$$\begin{aligned}\mathbb{E}\bigl(F_k(X_{i_1},\dots,X_{i_{k-1}},X_n)\mid\mathcal{F}_{n-m-1}\bigr) &= \mathbb{E}\bigl(F_k(X_{i_1},\dots,X_{i_{k-1}},X_n)\mid X_{i_1},\dots,X_{i_{k-1}}\bigr)\\ &=0.\end{aligned}\tag{10.14}$$

Consequently,

$$\mathbb{E}\bigl(\Delta U_n(F_k;\,>m)\mid\mathcal{F}_{n-m-1}\bigr)=0.\tag{10.15}$$

In the independent case $m=0$ treated in [37], this means that $U_n(F_k)$ is a martingale. In general, we may as a substitute split $U_n(F_k;\,>m)$ as a sum of $m+1$ martingales. For $j=1,\dots,m+1$ and $i\geqslant 1$, let

$$\Delta M_i^{(k,j)} := \Delta U_{(i-1)(m+1)+j}(F_k;\,>m),\tag{10.16}$$

$$M_i^{(k,j)} := \sum_{q=1}^{i}\Delta M_q^{(k,j)} = \sum_{q=1}^{i}\Delta U_{q(m+1)+j}(F_k;\,>m).\tag{10.17}$$

Then (10.15) implies that $(M_i^{(k,j)})_{i\geqslant 0}$ is a martingale, for each k and j. Hence, Burkholder's inequality [25, Theorem 10.9.5] yields, for the maximal function $M_n^{(k,j)*}$,

$$\|M_n^{(k,j)*}\|_p\leqslant C\biggl\|\Bigl(\sum_{i=1}^{n}\bigl|\Delta M_i^{(k,j)}\bigr|^2\Bigr)^{1/2}\biggr\|_p = C\Bigl\|\sum_{i=1}^{n}\bigl|\Delta M_i^{(k,j)}\bigr|^2\Bigr\|_{p/2}^{1/2}.\tag{10.18}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Furthermore, Minkowski's inequality yields (since $p/2 \geqslant 1$), using also (10.16) and (10.13), for $n \geqslant 1$,

$$\left\| \sum_{i=1}^{n} \left| \Delta M_i^{(k,j)} \right|^2 \right\|_{p/2} \leqslant \sum_{i=1}^{n} \left\| \left| \Delta M_i^{(k,j)} \right|^2 \right\|_{p/2} = \sum_{i=1}^{n} \left\| \Delta M_i^{(k,j)} \right\|_p^2 \leqslant C n^{1+2(k-1)}. \tag{10.19}$$

Combining (10.18) and (10.19) yields

$$\| M_n^{(k,j)*} \|_p \leqslant C n^{k-1/2}. \tag{10.20}$$

It follows from (10.16)–(10.17) that

$$U_n(F_k; \, > m) = \sum_{j=1}^{m+1} M^{(k,j)}_{\lfloor (n-j)/(m+1) \rfloor + 1}. \tag{10.21}$$

Hence (coarsely),

$$U_n^*(F_k; \, > m) \leqslant \sum_{j=1}^{m+1} M_n^{(k,j)*}, \tag{10.22}$$

and thus (10.20) and Minkowski's inequality yield

$$\left\| U_n^*(F_k; \, > m) \right\|_p \leqslant C n^{k-1/2}, \tag{10.23}$$

for $k = 1, \ldots, \ell$.

The result (10.8) now follows from (10.23) and (10.11) by a final application of Minkowski's inequality. □

Theorem 10.1. *Suppose that* (A_p) *holds for some* $p \geqslant 2$. *Then, with* $b = b(\mathcal{D})$,

$$\left\| \max_{j \leqslant n} \left| U_j(f; \mathcal{D}) - \mathbb{E} U_j(f; \mathcal{D}) \right| \right\|_p = O\big(n^{b-1/2}\big), \tag{10.24}$$

$$\left\| \max_{j \leqslant n} \left| U_j(f; \mathcal{D}) - \frac{\mu_{\mathcal{D}}}{b!} j^b \right| \right\|_p = O\big(n^{b-1/2}\big), \tag{10.25}$$

$$\left\| U_n^*(f; \mathcal{D}) \right\|_p = O\big(n^b\big). \tag{10.26}$$

The same results hold for an exact constraint $\mathcal{D}=$.

Proof. We use induction on b. We split the induction step into three cases.
Case 1: no constraint, i.e., $\mathcal{D} = \mathcal{D}_\infty$ *and* $b = \ell$ By (4.27)–(4.28),

$$U_n(f) = U_n(f; \mathcal{D}_\emptyset) = U_n(f; \, > m) - \sum_{J \neq \emptyset} (-1)^{|J|} U_n(f; \mathcal{D}_J). \tag{10.27}$$

Thus,

$$U_n^*(f) \leqslant U_n^*(f; \, > m) + \sum_{J \neq \emptyset} U_n^*(f; \mathcal{D}_J). \tag{10.28}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Suppose first that $\mu = 0$; then Lemma 10.1 applies to $U_n^*(f; > m)$. Furthermore, the induction hypothesis applies to each term in the sum in (10.28), since $b(\mathcal{D}_J) = \ell - |J| \leqslant \ell - 1 = b - 1$. Hence, Minkowski's inequality yields

$$\bigl\|U_n^*(f)\bigr\|_p \leqslant \bigl\|U_n^*(f; > m)\bigr\|_p + \sum_{J\neq\emptyset} \bigl\|U_n^*(f; \mathcal{D}_J)\bigr\|_p \leqslant Cn^{\ell-1/2} + Cn^{\ell-1}. \tag{10.29}$$

When $\mu = 0$, (3.7) yields

$$\mathbb{E}U_n(f) = O\bigl(n^{\ell-1}\bigr). \tag{10.30}$$

Now (10.24) follows from (10.29) and (10.30), which shows (10.24) when $\mu = 0$. The general case follows by considering $f - \mu$; this does not affect $U_n(f) - \mathbb{E}U_n(f)$.

Finally, both (10.25) and (10.26) follow from (10.24) and (3.7).

Case 2: an exact constraint $\mathcal{D}_=$, $b < \ell$. This is an immediate consequence of (4.9) in Lemma 4.2 and Case 1 applied to g; note that g too satisfies (A_p) by (4.19).

Case 3: a constraint $\mathcal{D}$, $b < \ell$. This is a consequence of Case 2 by (3.4) and (5.5). □

Lemma 10.2. *Suppose that (A_p) holds for some $p \geqslant 2$. Let $b := b(\mathcal{D})$. Then the sequences*

$$n^{1/2-b} \max_{j\leqslant n} \bigl|U_j(f; \mathcal{D}) - \mathbb{E}U_j(f; \mathcal{D})\bigr| \qquad \text{and} \qquad n^{-b}U_n^*(f; \mathcal{D}) \qquad (n \geqslant 1) \tag{10.31}$$

are uniformly pth-power integrable.

The same holds for an exact constraint $\mathcal{D}_=$.

Proof. We consider the second sequence in (10.31); the proof for the first sequence differs only notationally.

We have so far let f be fixed, so the constants above may depend on f. However, it is easy to see that the proof of (10.26) yields

$$\bigl\|U_n^*(f; \mathcal{D})\bigr\|_p \leqslant C_p \max_{i_1<\ldots<i_\ell} \|f(X_{i_1}, \ldots, X_{i_\ell})\|_p n^b, \tag{10.32}$$

with C_p independent of f (but depending on p). (Note that we only have to consider a finite set of indices $(i_1, \ldots, i_\ell)$, as discussed above (3.6).)

Truncate f, and for $B > 0$ define $f_B(\mathbf{x}) := f(\mathbf{x})\mathbf{1}\{|f(\mathbf{x})| \leqslant B\}$. Then (10.32) yields

$$\bigl\|n^{-b}U_n^*(f - f_B; \mathcal{D})\bigr\|_p \leqslant C_p\varepsilon(B), \tag{10.33}$$

where

$$\varepsilon(B) := \max_{i_1<\ldots<i_\ell} \|f(X_{i_1}, \ldots, X_{i_\ell})\mathbf{1}\{|f(X_{i_1}, \ldots, X_{i_\ell})| > B\}\|_p \to 0 \tag{10.34}$$

as $B \to \infty$.

Let $q := 2p$. Since f_B is bounded, we may apply (10.32) (or Theorem 10.1) with p replaced by q and obtain

$$\sup_n \bigl\|n^{-\ell}U_n^*(f_B; \mathcal{D})\bigr\|_{2p} < \infty. \tag{10.35}$$

Hence, for any B, the sequence $n^{-\ell}U_n^*(f_B; \mathcal{D})$ is uniformly pth-power integrable. Since $U_n^*(f; \mathcal{D}) \leqslant U_n^*(f_B; \mathcal{D}) + U_n^*(f - f_B; \mathcal{D})$, the result now follows from the following simple observation. □

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Lemma 10.3. *Let $1 \leqslant p < \infty$. Let $(\xi_n)_{n\geqslant 1}$ be a sequence of random variables. Suppose that for every $\varepsilon > 0$, there exist random variables η_n^ε and ζ_n^ε, $n \geqslant 1$, such that*

(i) $|\xi_n| \leqslant \eta_n^\varepsilon + \zeta_n^\varepsilon$,

(ii) *the sequence $(|\eta_n^\varepsilon|^p)_n$ is uniformly integrable, and*

(iii) $\|\zeta_n^\varepsilon\|_p \leqslant \varepsilon$.

Then $(|\xi_n|^p)_n$ is uniformly integrable.

Proof. Since (i) implies $|\xi_n|^p \leqslant 2^p|\eta_n^\varepsilon|^p + 2^p|\zeta_n^\varepsilon|^p$, it suffices (by appropriate substitutions) to consider the case $p = 1$. This is a simple exercise, using for example [25, Theorem 5.4.1]. □

Proof of Theorem 3.6. This is an immediate consequence of Theorems 3.3–3.4 and the uniform integrability given by Lemma 10.2. □

11. Functional convergence

We begin by improving (10.8) in a special situation. (We consider only $p = 2$.)

Lemma 11.1. *Suppose that (A_2) holds and that $\mu = 0$ and $f_i(X_i) = 0$ a.s. for every $i = 1, \dots, \ell$. Then*

$$\big\| U_n^*(f; > m)\big\|_2 \leqslant Cn^{\ell-1}. \tag{11.1}$$

Proof. Note that (6.6) and (6.10) immediately give this estimate for $\|U_n(f; > m)\|_2$. To extend it to the maximal function $U_n^*(f; > m)$, we reuse the proof of Lemma 10.1 (with $p = 2$), and analyze the terms $U_n^*(F_k; > m)$ further. First, by (10.4), (6.2), and the assumptions, for every $k \in [\ell]$,

$$\mathbb{E}\big(\widehat{F}_k(\widehat{X}_1, \dots, \widehat{X}_k) \,|\, \widehat{X}_k\big) = \mathbb{E}\big(f(\widehat{X}_1, \dots, \widehat{X}_\ell) \,|\, \widehat{X}_k\big) = f_k(\widehat{X}_k) + \mu = 0. \tag{11.2}$$

In particular, for $k = 1$, (11.2) yields $\widehat{F}_1(\widehat{X}_1) = 0$ a.s., and thus

$$U_n^*(F_1; > m) = 0 \quad \text{a.s.} \tag{11.3}$$

For $k \geqslant 2$, as stated in the proof of Lemma 10.1, $\Delta U_n(F_k; > m)$ is a sum of $\leqslant n^{k-1}$ terms $F_k\big(X_{i_1}, \dots, X_{i_{k-1}}, X_n\big)$. Consider two such terms $F_k\big(X_{i_1}, \dots, X_{i_{k-1}}, X_n\big)$ and $F_k\big(X_{i'_1}, \dots, X_{i'_{k-1}}, X_n\big)$, and suppose that $|i_j - i'_{j'}| > m$ for all $j, j' \in [k-1]$. Then all variables X_{ij}, $X_{i'_{j'}}$, and X_n are independent, and thus a.s.

$$\begin{aligned}&\mathbb{E}\big[F_k\big(X_{i_1}, \dots, X_{i_{k-1}}, X_n\big)F_k\big(X_{i'_1}, \dots, X_{i'_{k-1}}, X_n\big) \,|\, X_n\big]\\ &\quad = \mathbb{E}\big[F_k\big(X_{i_1}, \dots, X_{i_{k-1}}, X_n\big) \,|\, X_n\big]\mathbb{E}\big[F_k\big(X_{i'_1}, \dots, X_{i'_{k-1}}, X_n\big) \,|\, X_n\big] = 0,\end{aligned} \tag{11.4}$$

by (11.2). Hence, taking the expectation,

$$\mathbb{E}\Big[F_k\big(X_{i_1}, \dots, X_{i_{k-1}}, X_n\big)F_k\big(X_{i'_1}, \dots, X_{i'_{k-1}}, X_n\big)\Big] = 0, \tag{11.5}$$

unless $|i_j - i'_{j'}| \leqslant m$ for some pair (j, j'). For each $(i_1, \dots, i_{k-1})$, there are only $O(n^{k-2})$ such $(i'_1, \dots, i'_{k-1})$, and for each of these, the expectation in (11.5) is $O(1)$ by (A_2) and the

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Cauchy–Schwarz inequality. Consequently, summing over all appearing $(i_1, \dots, i_{k-1})$ and $(i'_1, \dots, i'_{k-1})$,

$$\begin{aligned}\mathbb{E}\bigl[\bigl|\Delta U_n(F_k; >m)\bigr|^2\bigr] &= \sum_{i_1,\dots,i_{k-1},i'_1,\dots,i'_{k-1}} \mathbb{E}\bigl[F_k\bigl(X_{i_1},\dots,X_{i_{k-1}},X_n\bigr)F_k\bigl(X_{i'_1},\dots,X_{i'_{k-1}},X_n\bigr)\bigr] \\ &= O\bigl(n^{k-1}\cdot n^{k-2}\bigr) = O\bigl(n^{2k-3}\bigr). \end{aligned} \tag{11.6}$$

We have gained a factor of n compared to (10.13). Hence, recalling (10.16) and using (11.6) in (10.18)–(10.19) (which for $p=2$ is essentially Doob's inequality), we improve (10.20) to

$$\|M_n^{(k,j)*}\|_2 \leqslant Cn^{k-1}. \tag{11.7}$$

Finally, (11.7) and (10.22) yield

$$\bigl\|U_n^*(F_k; >m)\bigr\|_2 \leqslant Cn^{k-1}, \tag{11.8}$$

for $2 \leqslant k \leqslant \ell$; this holds trivially for $k=1$ too by (11.3). The result follows by (10.11) and (11.8). □

Proof of Theorem 3.7. We prove (3.23); then (3.24) follows by (3.8). By replacing f by $f-\mu$, we may assume that $\mu=0$.

Consider first the unconstrained case. We argue as in [37], with minor modifications. We use (6.6), which we write as follows (cf. (6.11)):

$$U_n(f; >m) = \sum_{j=1}^{\ell}\sum_{i=1}^{n} a_{j,i,n} f_j(X_i) + U_n(f_*; >m), \tag{11.9}$$

with, as in (6.12),

$$a_{j,i,n} := \binom{i-1-(j-1)m}{j-1}\binom{n-i-(\ell-j)m}{\ell-j}. \tag{11.10}$$

Lemma 11.1 applies to f_* and shows that

$$\|U_n^*(f_*; >m)\|_2 = O\bigl(n^{\ell-1}\bigr) = o\bigl(n^{\ell-1/2}\bigr), \tag{11.11}$$

which implies that the last term in (11.9) is negligible, so we concentrate on the sum. Define $\Delta a_{j,i,n} := a_{j,i+1,n} - a_{j,i,n}$ and, using a summation by parts,

$$\widehat{U}_{n,j} := \sum_{i=1}^{n} a_{j,i,n} f_j(X_i) = a_{j,n,n} S_n(f_j) - \sum_{i=1}^{n-1} \Delta a_{j,i,n} S_i(f_j). \tag{11.12}$$

Donsker's theorem extends to m-dependent stationary sequences [5], and thus, as $n\to\infty$,

$$n^{-1/2} S_{\lfloor nt\rfloor}(f_j) \xrightarrow{\mathrm{d}} W_j(t) \qquad \text{in } D[0,\infty), \tag{11.13}$$

for a continuous centered Gaussian process W_j (a suitable multiple of Brownian motion); furthermore, as is easily seen, this holds jointly for $j=1,\dots,\ell$. Moreover, define

$$\psi_j(s,t) := \frac{1}{(j-1)!\,(\ell-j)!} s^{j-1}(t-s)^{\ell-j}. \tag{11.14}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

(Thus $\psi(s,1)=\psi(s)$ defined in (6.13); the present homogeneous version is more convenient here.) Let $\psi_j'(s,t):=\frac{\partial}{\partial s}\psi(s,t)$. Then straightforward calculations (as in [37, Lemma 4.2]) show that, extending (6.12),

$$a_{j,i,n}=\psi_j(i,n)+O(n^{\ell-2}), \tag{11.15}$$

$$\Delta a_{j,i,n}=\psi_j'(i,n)+O\bigl(n^{\ell-3}+n^{\ell-2}\mathbf{1}\{i\leqslant m \text{ or } i\geqslant n-m\}\bigr) \tag{11.16}$$

uniformly for all n, j, i that are relevant; moreover, the error terms with negative powers, i.e., $n^{\ell-2}$ for $\ell=1$ and $n^{\ell-3}$ for $\ell\leqslant 2$, vanish.

By the Skorokhod coupling theorem [41, Theorem 4.30], we may assume that the convergences (11.13) hold a.s., and similarly (see (11.11)), a.s.

$$n^{1/2-\ell}U_n^*(f_*;\,>m)\to 0. \tag{11.17}$$

It then follows from (11.12)–(11.16) and the homogeneity of ψ_j that a.s., uniformly for $t\in[0,T]$ for any fixed T,

$$\begin{aligned}
n^{1/2-\ell}\widehat{U}_{\lfloor nt\rfloor,j}&=n^{1-\ell}\psi_j(\lfloor nt\rfloor,\lfloor nt\rfloor)W_j(t)-\sum_{i=1}^{\lfloor nt\rfloor-1}n^{1-\ell}\psi_j'(i,\lfloor nt\rfloor)W_j(i/n)+o(1)\\
&=\psi_j(t,t)W_j(t)-\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor-1}\psi_j'(i/n,t)W_j(i/n)+o(1)\\
&=\psi_j(t,t)W_j(t)-\int_0^t\psi_j'(s,t)W_j(s)\,\mathrm{d}s+o(1).
\end{aligned} \tag{11.18}$$

Summing over $j\in[\ell]$, we obtain by (11.9), (11.12), (11.18), and (11.17), a.s. uniformly for $t\in[0,T]$ for any T,

$$n^{1/2-\ell}U_{\lfloor nt\rfloor}(f;\,>m)=Z(t)+o(1), \tag{11.19}$$

where

$$Z(t):=\sum_{j=1}^{\ell}\Bigl(\psi_j(t,t)W_j(t)-\int_0^t\psi_j'(s,t)W_j(s)\,\mathrm{d}s\Bigr), \tag{11.20}$$

which obviously is a centered Gaussian process. We can rewrite (11.19) as

$$n^{1/2-\ell}U_{\lfloor nt\rfloor}(f;\,>m)\to Z(t)\qquad\text{in } D[0,\infty). \tag{11.21}$$

Finally we use (10.27), which implies

$$\max_{k\leqslant n}\bigl|U_k(f)-U_k(f;\,>m)\bigr|\leqslant\sum_{J\neq\emptyset}U_n^*(f;\mathcal{D}_J), \tag{11.22}$$

and thus, by Theorem 10.1, recalling $b(\mathcal{D}_J)=\ell-|J|\leqslant\ell-1$,

$$\Bigl\|\max_{k\leqslant n}\bigl|U_k(f)-U_k(f;\,>m)\bigr|\Bigr\|_2\leqslant\sum_{J\neq\emptyset}\bigl\|U_n^*(f;\mathcal{D}_J)\bigr\|_2\leqslant\sum_{J\neq\emptyset}Cn^{b(\mathcal{D}_J)}\leqslant Cn^{\ell-1}. \tag{11.23}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

It follows that, in each $D[0,T]$ and thus in $D[0,\infty)$,

$$n^{1/2-\ell}\bigl(U_{\lfloor nt\rfloor}(f)-U_{\lfloor nt\rfloor}(f;>m)\bigr)\xrightarrow{\mathrm{p}}0. \tag{11.24}$$

Furthermore, recalling the assumption $\mu=0$, $\mathbb{E}U_n(f)=O(n^{\ell-1})$ by (3.7), and thus

$$n^{1/2-\ell}\mathbb{E}U_{\lfloor nt\rfloor}(f)\to 0 \qquad \text{in } \mathcal{D}[0,\infty). \tag{11.25}$$

The result (3.23) in the unconstrained case follows from (11.21), (11.24), and (11.25).

Joint convergence for several f (in the unconstrained case) follows by the same proof.

Finally, as usual, the exactly constrained case follows by (4.9) in Lemma 4.2 and the constrained case then follows by (3.4), using joint convergence for all $g_{\mathcal{D}'=}$, with notation as in (3.4) and Lemma 4.2. To obtain joint convergence for several f and $\mathcal{D}$, we only have to choose M in (4.8) large enough to work for all of them. □

12. Renewal theory

Note first that by the law of large numbers for m-dependent sequences,

$$S_n/n=S_n(h)/n\xrightarrow{\text{a.s.}}\mathbb{E}h(X_1)=\nu \qquad \text{as } n\to\infty. \tag{12.1}$$

As a simple consequence, we have the following (the case N_+ is in [33, Theorems 2.1 and 2.2]), which extends the well-known case of independent X_i; see e.g. [24, Sections 3.4 and 3.10].

Lemma 12.1. *Assume that* $\nu:=\mathbb{E}h(X_1)>0$ *and that* $\mathbb{E}|h(X_1)|^2<\infty$. *As* $x\to\infty$,

$$N_\pm(x)/x\xrightarrow{\text{a.s.}}1/\nu, \tag{12.2}$$

$$S_{N_\pm(x)}(h)=x+o_{\mathrm{p}}\bigl(x^{1/2}\bigr). \tag{12.3}$$

Proof. Note that, by the definitions (3.27)–(3.28),

$$S_{N_-(x)}(h)\leqslant x<S_{N_-(x)+1}(h) \qquad \text{and} \qquad S_{N_+(x)-1}(h)\leqslant x<S_{N_+(x)}(h). \tag{12.4}$$

Then (12.2) follows easily from (12.1). Furthermore, for any $\varepsilon>0$,

$$\mathbb{P}\Bigl[\max_{0\leqslant N<2x/\nu}|S_{N+1}(h)-S_N(h)|>\varepsilon x^{1/2}\Bigr]\leqslant\lceil 2x/\nu\rceil\mathbb{P}\bigl[|h(X_1)|>\varepsilon x^{1/2}\bigr]\to 0 \tag{12.5}$$

as $x\to\infty$, and (12.3) follows from (12.4), (12.5), and (12.2). We omit the standard details. □

Moreover, still assuming $\mathbb{E}|h(X_1)|^2<\infty$, Donsker's theorem for m-dependent sequences [5] yields, as in (11.13),

$$n^{-1/2}S_{\lfloor nt\rfloor}(h-\nu)=n^{-1/2}\bigl(S_{\lfloor nt\rfloor}(h)-\lfloor nt\rfloor\nu\bigr)\xrightarrow{\mathrm{d}}W_h(t) \qquad \text{in } D[0,\infty) \tag{12.6}$$

for a continuous centered Gaussian process $W_h(t)$.

Proof of Theorem 3.8. Note that (12.6) is the special (unconstrained) case $f=h$, $\mathcal{D}=()$, $\ell=b=1$ of (3.23). By joint convergence in Theorem 3.7 for $(f,\mathcal{D})$ and h, we thus have (3.24) jointly with (12.6). We use again the Skorokhod coupling theorem and assume that (3.24), (12.6), and (12.3), as well as (12.2), hold a.s.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Take $n := \lceil x \rceil$ and $t := N_\pm(x)/n$, and let $x \to \infty$. Then $t \to 1/\nu$ a.s. by (12.2), and thus (3.24) implies, a.s.,

$$\begin{aligned} U_{N_\pm(x)}(f;\mathcal{D}) &= \frac{\mu_\mathcal{D}}{b!} N_\pm(x)^b + Z(\nu^{-1}) n^{b-1/2} + o\bigl(n^{b-1/2}\bigr) \\ &= \frac{\mu_\mathcal{D}}{b!} N_\pm(x)^b + Z(\nu^{-1}) x^{b-1/2} + o\bigl(x^{b-1/2}\bigr). \end{aligned} \tag{12.7}$$

Similarly, (12.6) implies, a.s.,

$$S_{N_\pm(x)}(h) = N_\pm(x)\nu + W_h(\nu^{-1}) x^{1/2} + o\bigl(x^{1/2}\bigr). \tag{12.8}$$

By (12.8) and (12.3), we have a.s.

$$\nu N_\pm(x) = S_{N_\pm(x)}(h) - W_h(\nu^{-1}) x^{1/2} + o\bigl(x^{1/2}\bigr) = x - W_h(\nu^{-1}) x^{1/2} + o\bigl(x^{1/2}\bigr). \tag{12.9}$$

Thus, by the binomial theorem, a.s.,

$$(\nu N_\pm(x))^b = x^b - b W_h(\nu^{-1}) x^{b-1/2} + o\bigl(x^{b-1/2}\bigr). \tag{12.10}$$

Hence, (12.7) yields, a.s.,

$$U_{N_\pm(x)}(f;\mathcal{D}) = \frac{\mu_\mathcal{D}}{\nu^b b!}\bigl(x^b - b W_h(\nu^{-1}) x^{b-1/2}\bigr) + Z(\nu^{-1}) x^{b-1/2} + o\bigl(x^{b-1/2}\bigr), \tag{12.11}$$

which yields (3.29) with

$$\gamma^2 = \mathrm{Var}\Bigl[Z(\nu^{-1}) - \frac{\mu_\mathcal{D}}{\nu^b (b-1)!} W_h(\nu^{-1})\Bigr]. \tag{12.12}$$

The exactly constrained case and joint convergence follow similarly. □

Proof of Theorem 8.3. We may as in the proof of Theorem 3.8 assume that (3.24), (12.6), and (12.10) hold a.s. Recall that the proofs above use the decomposition (3.4) and Lemma 4.2 applied to every $\mathcal{D}' =$ there, with $b := b(\mathcal{D})$, D given by (4.7), $M := D + 1$ (for definiteness), and Y_i defined by (4.8). Furthermore, (4.20) holds with $g = g_\mathcal{D} : \mathcal{S}^b \to \mathbb{R}$ given by (4.23); in (3.23)–(3.24), we thus have the same limit $Z(t)$ for $U_n(f;\mathcal{D};(X_i))$ and $U_n(g;(Y_i))$. We may assume that this limit holds a.s. also for g.

Recall that $h : \mathcal{S} \to \mathbb{R}$. We abuse notation and extend it to $\mathcal{S}^M$ by $h(x_1, \dots, x_M) := h(x_1)$; thus $h(Y_i) = h(X_i)$. In particular, $S_n(h;(X_i)) = S_n(h;(Y_i))$, so we may write $S_n(h)$ without ambiguity. We define $H : (\mathcal{S}^M)^b \to \mathbb{R}$ by

$$H(y_1, \dots, y_b) := \sum_{j=1}^{b} h(y_j). \tag{12.13}$$

Note that (5.1) and (6.1) applied to the function H yield

$$\mu_H := \mathbb{E} H\bigl(\widehat{Y}_1, \dots, \widehat{Y}_b\bigr) = b\nu, \tag{12.14}$$

$$H_j(y) := h(y) + (b-1)\nu - \mu_H = h(y) - \nu. \tag{12.15}$$

(Since H is symmetric, H_j is the same for every j.)

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

In the unconstrained sum (3.1), there are $\binom{n-1}{\ell-1}$ terms that contain X_i, for each $i \in [n]$. Applying this to H and (Y_i), we obtain by (12.13)

$$U_n(H;(Y_i)) = \binom{n-1}{b-1} S_n(h). \tag{12.16}$$

Hence, for each fixed $t > 0$, by (12.6), a.s.,

$$\begin{aligned} U_{\lfloor nt\rfloor}(H;(Y_i)) - \mathbb{E}U_{\lfloor nt\rfloor}(H;(Y_i)) &= \binom{\lfloor nt\rfloor - 1}{b-1} S_{\lfloor nt\rfloor}(h-\nu) \\ &= \frac{(nt)^{b-1}}{(b-1)!} n^{1/2} W_h(t) + o\bigl(n^{b-1/2}\bigr). \end{aligned} \tag{12.17}$$

Combining (3.23) (for g and (Y_i)) and (12.17), we obtain that, a.s.,

$$\frac{U_{\lfloor nt\rfloor}\bigl(g - \frac{\mu_{\mathcal{D}}}{\nu} H\bigr) - \mathbb{E}U_{\lfloor nt\rfloor}\bigl(g - \frac{\mu_{\mathcal{D}}}{\nu} H\bigr)}{n^{b-1/2}} = Z(t) - \frac{\mu_{\mathcal{D}} t^{b-1}}{\nu(b-1)!} W_h(t) + o(1). \tag{12.18}$$

Taking $t = \nu^{-1}$, we see that this converges to the random variable in (12.12). Let $G := g - \mu_{\mathcal{D}}\nu^{-1}H$. Then a comparison with Theorem 3.3 (applied to G) shows that

$$\gamma^2 = t^{2b-1}\sigma^2(G) = \nu^{1-2b}\sigma^2(G). \tag{12.19}$$

In particular, $\gamma^2 = 0$ if and only if $\sigma^2(G) = 0$, and the result follows by Theorem 8.2, noting that $G_j := g_j - \mu_{\mathcal{D}}\nu^{-1}H_j = g_j + \mu_{\mathcal{D}} - \mu_{\mathcal{D}}\nu^{-1}h$ by (12.15). □

Proof of Theorem 3.9. This can be proved like [37, Theorem 3.13], by first stopping at $N_+(x_-)$, with $x_- := \lfloor x - \ln x\rfloor$, and then continuing to $N_-(x)$; we therefore only sketch the details. Let $R(x) := S_{N_+(x)} - x > 0$ be the overshoot at x, and let $\Delta(x) := x - S_{N_+(x_-)} = x - x_- - R(x_-)$. It is well known (see e.g. [24, Theorem 2.6.2]) that $R(x)$ converges in distribution as $x \to \infty$. In particular, $\Delta(x) \xrightarrow{\mathrm{p}} +\infty$ and thus $\mathbb{P}[\Delta(x) > 0] \to 1$. Since $N_+(x_-)$ is a stopping time and (X_i) are independent, the increments of the random walk S_n after $N_+(x_-)$ are independent of $U_{N_+(x_-)}(f)$, and it follows that the overshoot $R(x-1)$ is asymptotically independent of $U_{N_+(x_-)}(f)$. The event $\{S_{N_-(x)} = x\}$ equals $\{R(x-1) = 1\}$, and thus the asymptotic distribution of $U_{N_+(x_-)}(f)$ conditioned on $S_{N_-(x)} = x$ is the same as without conditioning, and given by (3.29). Finally, the difference between $U_{N_+(x_-)}(f)$ and $U_{N_\pm(x)}(f)$ is negligible, e.g. as a consequence of (3.24). □

In the remainder of the section, we prove moment convergence. Since we here consider different exponents p simultaneously, we let C_p denote constants that may depend on p. (By our standing convention, they may also depend on e.g. f and h.)

Lemma 12.2. *Assume that $\nu := \mathbb{E}h(X_1) > 0$ and that $\mathbb{E}|h(X_1)|^p < \infty$ for every $p < \infty$. Then, for every $p < \infty$, $A \geqslant 2/\nu$, and $x \geqslant 1$, we have*

$$\mathbb{P}\bigl[N_\pm(x) \geqslant Ax\bigr] \leqslant C_p (Ax)^{-p}. \tag{12.20}$$

Proof. We have $N_+(x) \leqslant N_-(x) + 1$; hence it suffices to consider $N_-(x)$.
Let $A \geqslant 2/\nu$. If $N_-(x) \geqslant Ax$, then (12.4) implies

$$S_{N_-(x)}(h-\nu) \leqslant x - N_-(x)\nu \leqslant (1 - A\nu)x \leqslant -(A\nu/2)x. \tag{12.21}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Hence, for any $p \geqslant 2$, using (10.24) for h (which is a well-known consequence of Doob's and Burkholder's, or Rosenthal's, inequalities),

$$\begin{aligned}\mathbb{P}\big[Ax \leqslant N_-(x) \leqslant 2Ax\big] &\leqslant (Ax\nu/2)^{-p}\mathbb{E}\big[\big|S_{N_-(x)}(h-\nu)\big|^p \mathbf{1}\{N_-(x) \leqslant 2Ax\}\big] \\ &\leqslant (A\nu x/2)^{-p}\mathbb{E}\big[\big|S^*_{\lfloor 2Ax\rfloor}(h-\nu)\big|^p\big] \leqslant C_p(Ax)^{-p}(2Ax)^{p/2} \\ &\leqslant C_p(Ax)^{-p/2}. \end{aligned} \tag{12.22}$$

We replace p by $2p$ and A by 2^kA in (12.22), and sum for $k \geqslant 0$; this yields (12.20). □

Lemma 12.3. *Assume that* (A_p) *and* $\mathbb{E}|h(X_1)|^p < \infty$ *hold for every* $p < \infty$, *and that* $\nu := \mathbb{E}h(X_1) > 0$. *Then, for every* $p \geqslant 1$ *and* $x \geqslant 1$,

$$\Big\| U_{N_\pm(x)}(f;\mathcal{D}) - \frac{\mu_\mathcal{D}}{\nu^b b!}x^b \Big\|_p \leqslant C_p x^{b-1/2}. \tag{12.23}$$

Proof. Let $V_n := U_n(f;\mathcal{D}) - \frac{\mu_\mathcal{D}}{b!}n^b$, let $B := 2/\nu$, and choose $q := 2bp$. Then the Cauchy–Schwarz inequality, (10.25), and Lemma 12.2 yield, with $V^*_x := \sup_{n\leqslant x}|V_n|$,

$$\begin{aligned}\mathbb{E}\Big|U_{N_\pm(x)}(f;\mathcal{D}) - \frac{\mu_\mathcal{D}}{b!}N_\pm(x)^b\Big|^p &= \mathbb{E}|V_{N_\pm(x)}|^p \\ &= \mathbb{E}\big[|V_{N_\pm(x)}|^p\mathbf{1}\{N_\pm(x) \leqslant Bx\}\big] + \sum_{k=1}^\infty \mathbb{E}\big[|V_{N_\pm(x)}|^p\mathbf{1}\{2^{k-1}Bx < N_\pm(x) \leqslant 2^kBx\}\big] \\ &\leqslant \mathbb{E}\big[|V^*_{Bx}|^p\big] + \sum_{k=1}^\infty \mathbb{E}\big[|V^*_{2^kBx}|^p\mathbf{1}\{N_\pm(x) > 2^{k-1}Bx\}\big] \\ &\leqslant \mathbb{E}\big[|V^*_{Bx}|^p\big] + \sum_{k=1}^\infty \mathbb{E}\big[|V^*_{2^kBx}|^{2p}\big]^{1/2}\mathbb{P}\big[N_\pm(x) > 2^{k-1}Bx\big]^{1/2} \\ &\leqslant C_px^{p(b-1/2)} + \sum_{k=1}^\infty C_p(2^kx)^{p(b-1/2)}C_q(2^{k-1}x)^{-q/2} \leqslant C_px^{p(b-1/2)}. \end{aligned} \tag{12.24}$$

In other words,

$$\Big\| U_{N_\pm(x)}(f;\mathcal{D}) - \frac{\mu_\mathcal{D}}{b!}N_\pm(x)^b \Big\|_p \leqslant C_px^{b-1/2}. \tag{12.25}$$

We may here replace $U_n(f;\mathcal{D})$ by $S_n(h)$ (and thus b by 1 and $\mu_\mathcal{D}$ by ν). This yields

$$\big\| S_{N_\pm(x)}(h) - \nu N_\pm(x)\big\|_p \leqslant C_px^{1/2}. \tag{12.26}$$

By the same proof, this holds also if we replace $N_\pm(x)$ by $N_\pm(x) \mp 1$. Using (12.4), it follows that

$$\big\| x - \nu N_\pm(x)\big\|_p \leqslant C_px^{1/2}. \tag{12.27}$$

In particular, by Minkowski's inequality,

$$\big\| \nu N_\pm(x)\big\|_p \leqslant x + \big\| x - \nu N_\pm(x)\big\|_p \leqslant C_px. \tag{12.28}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Consequently, by Minkowski's and Hölder's inequalities, (12.27) and (12.28),

$$\begin{aligned}\bigl\|(\nu N_\pm(x))^b - x^b\bigr\|_p &= \Biggl\|\sum_{k=0}^{b-1}(\nu N_\pm(x)-x)(\nu N_\pm(x))^k x^{b-1-k}\Biggr\|_p \\ &\leqslant \sum_{k=0}^{b-1}\bigl\|\nu N_\pm(x)-x\bigr\|_{(k+1)p}\bigl\|\nu N_\pm(x)\bigr\|_{(k+1)p}^k x^{b-1-k} \\ &\leqslant C_p x^{b-1/2}. \end{aligned} \tag{12.29}$$

Combining (12.25) and (12.29), we obtain (12.23). □

Proof of Theorem 3.10. We have shown that the left-hand side of (3.29) is uniformly bounded in L^p for $x \geqslant 1$. By replacing p with $2p$, say, this implies that these left-hand sides are uniformly pth-power integrable, for every $p < \infty$, which implies convergence of all moments in (3.29).

The proof of Theorem 3.9 shows that, under the assumptions there, $\mathbb{P}(S_{N_-(x)} = x)$ converges to a positive limit as $x \to \infty$; hence $\mathbb{P}(S_{N_-(x)} = x) > c$ for some $c > 0$ and all large x. This implies that the uniform pth-power integrability also holds after conditioning (for large x), and thus all moments also converge in (3.29) after conditioning. □

13. Constrained pattern matching in words

As stated in Section 1, Flajolet, Szpankowski and Vallée [23] studied the following problem; see also [Jacquet and Szpankowski [31], Chapter 5]. Consider a random string $\Xi_n = \xi_1 \cdots \xi_n$, where the letters ξ_i are i.i.d. random elements in some finite alphabet $\mathcal{A}$. (We may regard Ξ_n as the initial part of an infinite string $\xi_1\xi_2\ldots$ of i.i.d. letters.) Consider also a fixed word $\mathbf{w} = w_1 \cdots w_\ell$ from the same alphabet. (Thus, $\ell \geqslant 1$ denotes the length of w; we keep $\mathbf{w}$ and ℓ fixed.) Let $N_n(\mathbf{w})$ be the (random) number of occurrences of $\mathbf{w}$ in Ξ_n. More generally, for any constraint $\mathcal{D} = (d_1, \ldots, d_{\ell-1})$, let $N_n(\mathbf{w};\mathcal{D})$ be the number of constrained occurrences. This is a special case of the general setting in (3.1)–(3.2), with $X_i = \xi_i$, and (cf. (1.1))

$$f(x_1, \ldots, x_\ell) = \mathbf{1}\{x_1, \ldots, x_\ell = \mathbf{w}\} = \mathbf{1}\{x_i = w_i \,\forall i \in [\ell]\}. \tag{13.1}$$

Consequently,

$$N_n(\mathbf{w};\mathcal{D}) = U_n(f;\mathcal{D};(\xi_i)) \tag{13.2}$$

with f given by (13.1).

Denote the distribution of the individual letters by

$$p(x) := \mathbb{P}(\xi_1 = x), \qquad x \in \mathcal{A}. \tag{13.3}$$

We will, without loss of generality, assume $p(x) > 0$ for every $x \in \mathcal{A}$. Then (13.2) and the general results above yield the following result from [23], with $b = b(\mathcal{D})$ given by (2.1). The unconstrained case (also in [23]) is a special case. Moreover, the theorem also holds for the exactly constrained case, with $\mu_{\mathcal{D}=} = \prod_i p(w_i)$ and some $\sigma^2(\mathbf{w};\mathcal{D}=)$; we leave the detailed

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

statement to the reader. A formula for σ^2 is given in [23, (14)]; we show explicitly that $\sigma^2 > 0$ except in trivial (non-random) cases, which seems to have been omitted from [23].

Theorem 13.1. (Flajolet, Szpankowski and Vallée [23].) *With notation as above, as* $n \to \infty$,

$$\frac{N_n(\mathbf{w}; \mathcal{D}) - \frac{\mu_{\mathcal{D}}}{b!} n^b}{n^{b-1/2}} \xrightarrow{\mathrm{d}} N\big(0, \sigma^2\big) \tag{13.4}$$

for some $\sigma^2 = \sigma^2(\mathbf{w}; \mathcal{D}) \geqslant 0$, *with*

$$\mu_{\mathcal{D}} := \prod_{d_j < \infty} d_j \cdot \prod_{i=1}^{\ell} p(w_i). \tag{13.5}$$

Furthermore, all moments converge in (13.4).

Moreover, if $|\mathcal{A}| \geqslant 2$, *then* $\sigma^2 > 0$.

Proof. By (13.2), the convergence (13.4) is an instance of (3.21) in Theorem 3.4 together with (3.8) in Theorem 3.1. The formula (13.5) follows from (3.11) since, by (13.1) and independence,

$$\mu := \mathbb{E} f(\xi_1, \dots, \xi_\ell) = \mathbb{P}\big(\xi_1 \cdots \xi_\ell = w_1 \cdots w_\ell\big) = \prod_{i=1}^{\ell} \mathbb{P}(\xi_i = w_i). \tag{13.6}$$

Moment convergence follows by Theorem 3.6; note that (A_p) is trivial, since f is bounded.

Finally, assume $|\mathcal{A}| \geqslant 2$ and suppose that $\sigma^2 = 0$. Then Theorem 8.2 says that (8.20) holds, and thus, for each $n > D$, the sum $S_n(g_j)$ is independent of $\xi_{D+1}, \dots, \xi_n$. We consider only $j = 1$. Choose $a \in \mathcal{A}$ with $a \neq w_1$. Consider first an exact constraint $\mathcal{D}=$. Then $g_{\mathcal{D}=}$ is given by (4.18). Since $f(x_1, \dots, x_\ell) = 0$ whenever $x_1 = a$, it follows from (4.18) that $g(y_1, \dots, y_b) = 0$ whenever $y_1 = (y_{1k})_{k=1}^{M}$ has $y_{11} = a$. Hence, (6.1) shows that

$$g_1(y_1) = -\mu_{\mathcal{D}=} = -\mu, \qquad \text{if } y_{11} = a. \tag{13.7}$$

Consequently, on the event $\xi_1 = \ldots = \xi_n = a$, recalling (4.8) and $M = D + 1$, we have $g_1(Y_k) = g_1(\xi_k, \dots, \xi_{k+D}) = -\mu$ for every $k \in [n]$. Thus,

$$S_n(g_1) = -n\mu \qquad \text{if} \quad \xi_1 = \cdots = \xi_n = a. \tag{13.8}$$

On the other hand, as noted above, the assumption $\sigma^2 = 0$ implies that $S_n(g_1)$ is independent of $\xi_{D+1}, \dots, \xi_n$. Consequently, (13.8) implies

$$S_n(g_1) = -n\mu \qquad \text{if} \quad \xi_1 = \cdots = \xi_D = a, \tag{13.9}$$

regardless of $\xi_{D+1} \dots, \xi_{n+D}$. This is easily shown to lead to contradiction. For example, we have, by (6.4),

$$\mathbb{E} g_1(Y_k) = \mathbb{E} g_1(\xi_k, \dots, \xi_{k+D}) = 0, \tag{13.10}$$

and thus, conditioning on $\xi_1, \dots, \xi_D$,

$$\begin{aligned} \mathbb{E}\big(S_n(g_1) \mid \xi_1 = \cdots = \xi_D = a\big) &= \sum_{k=1}^{n} \mathbb{E}\big(g_1(\xi_k, \dots, \xi_{k+D}) \mid \xi_1 = \cdots = \xi_D = a\big) \\ &= O(1), \end{aligned} \tag{13.11}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

since all terms with $k > D$ are unaffected by the conditioning and thus vanish by (13.10); this contradicts (13.9) for large n, since $\mu > 0$. This contradiction shows that $\sigma^2 > 0$ for an exact constraint $\mathcal{D}=$.

Alternatively, instead of using the expectation as in (13.11), one might easily show that if $n > D + \ell$, then $\xi_{D+1}, \dots, \xi_n$ can be chosen such that (13.9) does not hold.

For a constraint $\mathcal{D}$, $g = g_{\mathcal{D}}$ is given by a sum (4.23) of exactly constrained cases $\mathcal{D}'=$. Hence, by summing (13.7) for these $\mathcal{D}'=$, it follows that (13.7) also holds for $g_{\mathcal{D}}$ (with μ replaced by $\mu_{\mathcal{D}}$). This leads to a contradiction exactly as above. □

Theorem 13.1 shows that, except in trivial cases, the asymptotic variance $\sigma^2 > 0$ for a subsequence count $N_n(\mathbf{w}; \mathcal{D})$, and thus (13.4) yields a non-degenerate limit, so that it really shows asymptotic normality. By the same proof (see also Remark 3.6), Theorem 13.1 extends to linear combinations of different subsequence counts (in the same random string Ξ_n), but in this case, it may happen that $\sigma^2 = 0$, and then (13.4) has a degenerate limit and thus yields only convergence in probability to 0. (We consider only linear combinations with coefficients not depending on n.) One such degenerate example with constrained subsequence counts is discussed in Example 8.1. There are also degenerate examples in the unconstrained case. In fact, the general theory of degenerate (in this sense) U-statistics based on independent $(X_i)_1^\infty$ is well understood; for symmetric U-statistics this case was characterized by [27] and studied in detail by [56], and their results were extended to the asymmetric case relevant here in [35, Chapter 11.2]. In Appendix A we apply these general results to string matching and give a rather detailed treatment of the degenerate cases of linear combinations of unconstrained subsequence counts. See also [19] for further algebraic aspects of both non-degenerate and degenerate cases.

Problem 13.1. Appendix A considers only the unconstrained case. Example 8.1 shows that for linear combinations of constrained pattern counts, there are further possibilities to have $\sigma^2 = 0$. It would be interesting to extend the results in Appendix A and characterize these cases, and also to obtain limit theorems for such cases, extending Theorem A.2 (in particular the case $k = 2$); note again that the limit in Example 8.1 is of a different type than the ones occurring in unconstrained cases (Theorem A.2). We leave these as open problems.

14. Constrained pattern matching in permutations

Consider now random permutations. As usual, we generate a random permutation $\boldsymbol{\pi} = \boldsymbol{\pi}^{(n)} \in \mathfrak{S}_n$ by taking a sequence $(X_i)_1^n$ of i.i.d. random variables with a uniform distribution $X_i \sim U(0, 1)$, and then replacing the values $X_1, \dots, X_n$, in increasing order, by $1, \dots, n$. Then the number $N_n(\tau)$ of occurrences of a fixed permutation $\tau = \tau_1 \cdots \tau_\ell$ in $\boldsymbol{\pi}$ is given by the U-statistic $U_n(f)$ defined by (3.1) with

$$f(x_1, \dots, x_\ell) := \prod_{1 \leqslant i < j \leqslant \ell} \mathbf{1}\{x_i < x_j \iff \tau_i < \tau_j\}. \tag{14.1}$$

Similarly, for any constraint $\mathcal{D} = (d_1, \dots, d_{\ell-1})$, we have for the number of constrained occurrences of τ, with the same f as given by (14.1),

$$N_n(\tau; \mathcal{D}) = U_n(f; \mathcal{D}). \tag{14.2}$$

Hence, Theorems 3.3 and 3.4 yield the following result showing asymptotic normality of the number of (constrained) occurrences. As stated in the introduction, the unconstrained case

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

was shown by Bóna [7], the case $d_1 = \cdots = d_{\ell-1} = 1$ by Bóna [9], and the general vincular case by Hofer [29]; we extend the result to general constrained cases. The fact that $\sigma^2 > 0$ was shown in [29] (in vincular cases); we give a shorter proof based on Theorem 8.2. Again, the theorem also holds for the exactly constrained case, with $\mu_{\mathcal{D}=} = 1/\ell!$ and some $\sigma^2(\tau; \mathcal{D}=)$.

Theorem 14.1. (Largely Bóna [7, 9] and Hofer [29].) *For any fixed permutation $\tau \in \mathfrak{S}_\ell$ and constraint $\mathcal{D} = (d_1, \dots, d_{\ell-1})$, as $n \to \infty$,*

$$\frac{N_n(\tau; \mathcal{D}) - \frac{\mu_{\mathcal{D}}}{b!} n^b}{n^{b-1/2}} \xrightarrow{\mathrm{d}} N\bigl(0, \sigma^2\bigr) \tag{14.3}$$

for some $\sigma^2 = \sigma^2(\tau; \mathcal{D}) \geqslant 0$ and

$$\mu_{\mathcal{D}} := \frac{1}{\ell!} \prod_{d_j < \infty} d_j. \tag{14.4}$$

Furthermore, all moments converge in (14.3).

Moreover, if $\ell \geqslant 2$, then $\sigma^2 > 0$.

Proof. This is similar to the proof of Theorem 13.1. By (14.2), the convergence (14.3) is an instance of (3.21) together with (3.8). The formula (14.4) follows from (3.11), since $\mu := \mathbb{E} f(X_1, \dots, X_\ell)$ by (14.1) is the probability that $X_1, \dots, X_\ell$ have the same order as $\tau_1, \dots, \tau_\ell$, i.e., $1/\ell!$. Moment convergence follows by Theorem 3.6.

Finally, suppose that $\ell \geqslant 2$ but $\sigma^2 = 0$. Then Theorem 8.2 says that (8.20) holds, and thus, for each j and each $n > D$, the sum $S_n(g_j)$ is independent of $X_{D+1}, \dots, X_n$; we want to show that this leads to a contradiction. We again choose $j = 1$, but we now consider two cases separately.

Case 1: $d_1 < \infty$ Recall the notation in (4.11)–(4.14), and note that in this case

$$\ell_1 > 1, \qquad t_{11} = 0, \qquad t_{12} = d_1, \qquad v_1 = 0. \tag{14.5}$$

Assume for definiteness that $\tau_1 > \tau_2$. (Otherwise, we may exchange $<$ and $>$ in the argument below.) Then (14.1) implies that

$$f(x_1, \dots, x_\ell) = 0 \qquad \text{if} \quad x_1 < x_2. \tag{14.6}$$

Consider first the exact constraint $\mathcal{D}=$. Then $g_{\mathcal{D}=}$ is given by (4.18). Hence, (14.6) and (14.5) imply that

$$g_{\mathcal{D}=}(y_1, \dots, y_b) = 0 \qquad \text{if} \quad y_1 = (y_{1,k})_{k=1}^{M} \text{ with } y_{1,1} < y_{1,1+d_1}. \tag{14.7}$$

In particular,

$$g_{\mathcal{D}=}(y_1, \dots, y_b) = 0 \qquad \text{if} \quad y_{1,1} < y_{1,2} < \cdots < y_{1,M}. \tag{14.8}$$

By (4.23), the same holds for the constraint $\mathcal{D}$. Hence, (6.1) shows that, for $g = g_{\mathcal{D}}$,

$$g_1(y_1) = -\mu_{\mathcal{D}} \qquad \text{if} \quad y_{1,1} < y_{1,2} < \cdots < y_{1,M}. \tag{14.9}$$

Consequently, on the event $X_1 < \cdots < X_{n+D}$, recalling (4.8) and $M = D + 1$, we have $g_1(Y_k) = g_1(X_k, \dots, X_{k+D}) = -\mu_{\mathcal{D}}$ for every $k \in [n]$, and thus

$$S_n(g_1) = -n\mu_{\mathcal{D}} \qquad \text{if} \quad X_1 < \cdots < X_{n+D}. \tag{14.10}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

On the other hand, as noted above, the assumption $\sigma^2=0$ implies that $S_n(g_1)$ is independent of $X_{D+1},\dots,X_n$. Consequently, (14.10) implies that a.s.

$$S_n(g_1)=-n\mu_{\mathcal{D}} \qquad \text{if} \quad X_1<\dots<X_D<X_{n+1}<\dots<X_{n+D}. \tag{14.11}$$

However, in analogy with (13.10)–(13.11), we have $\mathbb{E}g_1(Y_k)=0$ by (6.4), and thus

$$\begin{aligned}&\mathbb{E}\bigl(S_n(g_1)\mid X_1<\dots<X_D<X_{n+1}<\dots<X_{n+D}\bigr)\\&\quad=\sum_{k=1}^{n}\mathbb{E}\bigl(g_1(X_k,\dots,X_{k+D})\mid X_1<\dots<X_D<X_{n+1}<\dots<X_{n+D}\bigr)\\&\quad=O(1),\end{aligned} \tag{14.12}$$

since all terms with $D<k\leqslant n-D$ are unaffected by the conditioning and thus vanish. But (14.12) contradicts (14.11) for large n, since $\mu_{\mathcal{D}}>0$. This contradiction shows that $\sigma^2>0$ when $\ell\geqslant 2$ and $d_1<\infty$.

Case 2: $d_1=\infty$ In this case, $\ell_1=1$. Consider again first $\mathcal{D}=$. Since (X_i) are i.i.d., (4.18) and (6.1) yield, choosing $j_i:=(D+1)i$, say,

$$g_1(y_1)=\mathbb{E}g\bigl(y_1,Y_{j_2},\dots,Y_{j_b}\bigr)-\mu=\mathbb{E}f\bigl(y_{11},X_2,\dots,X_\ell\bigr)-\mu=f_1(y_{11}) \tag{14.13}$$

(with $\mu=1/\ell!$). Thus, recalling (4.8), we have

$$S_n(g_1)=\sum_{k=1}^{n}g_1(Y_k)=\sum_{k=1}^{n}f_1(X_k). \tag{14.14}$$

By Theorem 8.2, the assumption $\sigma^2=0$ thus implies that the final sum in (14.14) is independent of X_{D+1}, for any $n\geqslant D+1$. Since (X_i) are independent, this is possible only if $f_1(X_{D+1})=c$ a.s. for some constant c, i.e., if $f_1(x)=c$ for a.e. $x\in(0,1)$.

However, by (14.1), $f(x,X_2,\dots,X_\ell)=1$ if and only if τ_1-1 prescribed X_j are in $(0,x)$ and in a specific order, and the remaining $\ell-\tau_1$ ones are in $(x,1)$ and in a specific order. Hence, (6.1) yields

$$f_1(x)=\frac{1}{(\tau_1-1)!\,(\ell-\tau_1)!}x^{\tau_1-1}(1-x)^{\ell-\tau_1}-\mu. \tag{14.15}$$

Since $\ell\geqslant 2$, $f_1(x)$ is a non-constant polynomial in x.

This is a contradiction, and shows that $\sigma^2>0$ also when $d_1=\infty$. □

Remark 14.1. Although $\sigma^2>0$ for each pattern count $N_n(\tau;\mathcal{D})$ with $\ell>1$, non-trivial linear combinations might have $\sigma^2=0$, and thus variance of lower order, even in the unconstrained case. (This is similar to the case of patterns in strings in Section 13 and Appendix A.) In fact, for the unconstrained case, it is shown in [39] that for permutations τ of a given length ℓ, the $\ell!$ counts $N_n(\tau)$ converge jointly, after normalization as above, to a multivariate normal distribution of dimension only $(\ell-1)^2$, meaning that there is a linear space of dimension $\ell!-(\ell-1)^2$ of linear combinations that have $\sigma^2=0$. This is further analyzed in [18], where the spaces of linear combinations of $N_n(\tau)$ having variance $O(n^{2\ell-r})$ are characterized for each $r=1,\dots,\ell-1$, using the representation theory of the symmetric group. In particular, the highest degeneracy, with variance $\Theta(n^{\ell+1})$, is obtained for the sign statistic $U_n(\mathrm{sgn})$,

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

where $\operatorname{sgn}(x_1, \dots, x_\ell)$ is the sign of the permutation defined by the order of $(x_1, \dots, x_n)$; in other words, $U_n(\operatorname{sgn})$ is the sum of the signs of the $\binom{n}{\ell}$ subsequences of length ℓ of a random permutation $\boldsymbol{\pi}^{(n)} \in \mathfrak{S}_n$. For $\ell = 3$, the asymptotic distribution of $n^{-(\ell+1)/2}U_n(\operatorname{sgn})$ is of the type in (A.19); see [21] and [39, Remark 2.7]. For larger ℓ, using the methods in Appendix A, the asymptotic distribution can be expressed as a polynomial of degree $\ell - 1$ in infinitely many independent normal variables, as in (A.17); however, we do not know any concrete such representation.

We expect that, in analogy with Example 8.1, for linear combinations of constrained pattern counts, there are further possibilities to have $\sigma^2 = 0$. We have not pursued this, and we leave it as an open problem to characterize these cases with $\sigma^2 = 0$; moreover, it would also be interesting to extend the results of [18] characterizing cases with higher degeneracies to constrained cases.

15. Further comments

We discuss here briefly some possible extensions of the present work. We have not pursued them, and they are left as open problems.

15.1. Mixing and Markov input

We have in this paper studied U-statistics based on a sequence (X_i) that is allowed to be dependent, but only under the rather strong assumption of m-dependence (partly motivated by our application to constrained U-statistics). It would be interesting to extend the results to weaker assumptions on (X_i), for example that it is stationary with some type of mixing property. (See e.g. [12] for various mixing conditions and central limit theorems under some of them.)

Alternatively (or possibly as a special case of mixing conditions), it would be interesting to consider (X_i) that form a stationary Markov chain (under suitable assumptions).

In particular, it seems interesting to study constrained U-statistics under such assumptions, since the mixing or Markov assumptions typically imply strong dependence for sets of variables X_i with small gaps between the indices, but not if the gaps are large.

Markov models are popular models for random strings. Substring counts, i.e., the completely constrained case of subsequence counts (see Remark 1.2), have been treated for Markov sources by e.g. [52], [45], and [31].

A related model for random strings is a probabilistic dynamic source; see e.g. [31, Section 1.1]. For substring counts, asymptotic normality has been shown by [11]. For (unconstrained or constrained) subsequence counts, asymptotic results on mean and variance are special cases of [10] and [31, Theorem 5.6.1]; we are not aware of any results on asymptotic normality in this setting.

15.2. Generalized U-statistics

Generalized U-statistics (also called *multi-sample U-statistics*) are defined similarly to (3.1), but are based on two (for simplicity) sequences $(X_i)_1^{n_1}$ and $(Y_j)_1^{n_2}$ of random variables, with the sum in (3.1) replaced by a sum over all $i_1 < \dots < i_{\ell_1} \leqslant n_1$ and $j_1 < \dots < j_{\ell_2} \leqslant n_2$, and f now a function of $\ell_1 + \ell_2$ variables. Limit theorems, including asymptotic normality, under suitable conditions are shown in [60], and extensions to asymmetric cases are sketched in [35, Example 11.24]. We do not know any extensions to m-dependent or constrained cases, but we expect that such extensions are straightforward.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Appendix A. Linear combinations for unconstrained subsequence counts

As promised in Section 13, we consider here unconstrained subsequence counts in a random string Ξ_n with i.i.d. letters, normalized as in Theorem 13.1, and study further the case of linear combinations of such normalized counts (with coefficients not depending on n); in particular, we study in some detail such linear combinations that are degenerate in the sense that the asymptotic variance $\sigma^2 = 0$.

The results are based on the orthogonal decomposition introduced in the symmetric case by Hoeffding [28] (see also Rubin and Vitale [56]); this is extended to the asymmetric case in [35, Chapter 11.2], but the treatment there uses a rather heavy formalism, and we therefore give here a direct treatment in the present special case. (This case is somewhat simpler than the general case since we only have to consider finite-dimensional vector spaces below, but otherwise the general case is similar.) See also [19], which contains a much deeper algebraic study of the asymptotic variance $\sigma^2(f)$ and the vector spaces below, and in particular a spectral decomposition that refines (A.9).

Fix $\mathcal{A}$ and the random string $(\xi_i)_1^\infty$. Assume, as in Section 13, that $p(x) > 0$ for every $x \in \mathcal{A}$. Let $A := |\mathcal{A}|$, the number of different letters.

We fix also $\ell \geqslant 1$ and consider all unconstrained subsequence counts $N_n(\mathbf{w})$ with $|\mathbf{w}| = \ell$. There are A^ℓ such words $\mathbf{w}$, and it follows from (13.2) and (13.1) that the linear combinations of these counts are precisely the asymmetric U-statistics (3.1) for all $f : \mathcal{A}^\ell \to \mathbb{R}$, by the relation

$$\sum_{\mathbf{w}\in\mathcal{A}^\ell} f(\mathbf{w}) N_n(\mathbf{w}) = U_n(f). \tag{A.1}$$

Note that Theorem 3.3 applies to every $U_n(f)$, and thus (3.18) and (3.19) hold for some $\sigma^2 = \sigma^2(f) \geqslant 0$. (As stated above, this case of Theorem 3.3 with i.i.d. X_i, i.e., the case $m = 0$, is also treated in [35, Corollary 11.20] and [37].)

Let V be the linear space of all functions $f : \mathcal{A}^\ell \to \mathbb{R}$. Thus $\dim V = A^\ell$. Similarly, let W be the linear space of all functions $h{:}\mathcal{A} \to \mathbb{R}$, i.e., all functions of a single letter; thus $\dim W = A$. Then V can be identified with the tensor product $W^{\otimes\ell}$, with the identification

$$h_1 \otimes \cdots \otimes h_\ell(x_1, \dots, x_\ell) = \prod_1^\ell h_i(x_i). \tag{A.2}$$

We regard V as a (finite-dimensional) Hilbert space with inner product

$$\langle f, g\rangle_V := \mathbb{E}\bigl[f(\Xi_n)g(\Xi_n)\bigr], \tag{A.3}$$

and, similarly, W as a Hilbert space with inner product

$$\langle h, k\rangle_W := \mathbb{E}\bigl[h(\xi_1)k(\xi_1)\bigr]. \tag{A.4}$$

Let W_0 be the subspace of W defined by

$$W_0 := \{1\}^\perp = \{h \in W{:}\langle h, 1\rangle_W = 0\} = \{h \in W{:}\mathbb{E}h(\xi_1) = 0\}. \tag{A.5}$$

Thus, $\dim W_0 = A - 1$.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

For a subset $\mathcal{B} \subseteq \mathcal{A}$, let $V_{\mathcal{B}}$ be the subspace of V spanned by all functions $h_1 \otimes \cdots \otimes h_\ell$ as in (A.2) such that $h_i \in W_0$ if $i \in \mathcal{B}$, and $h_i = 1$ if $i \notin \mathcal{B}$. In other words, if for a given $\mathcal{B}$ we define $W'_i := W_0$ when $i \in \mathcal{B}$ and $W'_i = \mathbb{R}$ when $i \notin \mathcal{B}$, then

$$V_{\mathcal{B}} = W'_1 \otimes \cdots \otimes W'_\ell \cong W_0^{\otimes|\mathcal{B}|}. \tag{A.6}$$

It is easily seen that these 2^A subspaces of V are orthogonal, and that we have an orthogonal decomposition

$$V = \bigoplus_{\mathcal{B}\subseteq\mathcal{A}} V_{\mathcal{B}}. \tag{A.7}$$

Furthermore, for $k = 0, \ldots, \ell$, define

$$V_k := \bigoplus_{|\mathcal{B}|=k} V_{\mathcal{B}}. \tag{A.8}$$

Thus, we also have an orthogonal decomposition (as in [28] and [56])

$$V = \bigoplus_{k=0}^{\ell} V_k. \tag{A.9}$$

Note that, by (A.6) and (A.8),

$$\dim V_{\mathcal{B}} = (A-1)^{|\mathcal{B}|}, \qquad \dim V_k = \binom{\ell}{k}(A-1)^k. \tag{A.10}$$

Let $\Pi_{\mathcal{B}}$ and $\Pi_k = \sum_{|\mathcal{B}|=k} \Pi_{\mathcal{B}}$ be the orthogonal projections of V onto $V_{\mathcal{B}}$ and V_k. Then, for any $f \in V$, we may consider its components $\Pi_k f \in V_k$.

First, V_0 is the 1-dimensional space of constant functions in V. Trivially, if $f \in V_0$, then $U_n(f)$ is non-random, so $\operatorname{Var} U_n(f) = 0$ for every n, and $\sigma^2(f) = 0$. More interesting is that for any $f \in V$, we have

$$\Pi_0 f = \mathbb{E} f(\xi_1, \ldots, \xi_\ell) = \mu. \tag{A.11}$$

Next, it is easy to see that taking $\mathcal{B} = \{i\}$ yields the projection f_i defined by (6.1), except that $\Pi_{\{i\}} f$ is defined as a function on $\mathcal{A}^\ell$; to be precise,

$$\Pi_{\{i\}} f(x_1, \ldots, x_\ell) = f_i(x_i). \tag{A.12}$$

Recalling (A.1), this leads to the following characterization of degenerate linear combinations of unconstrained subsequence counts.

Theorem A.1. *With notation and assumptions as above, if $f : \mathcal{A}^\ell \to \mathbb{R}$, then the following are equivalent:*

(i) $\sigma^2(f) = 0$.

(ii) $f_i = 0$ *for every* $i = 1, \ldots, \ell$.

(iii) $\Pi_1 f = 0$.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Proof. The equivalence (i) $\iff$ (ii) is a special case of Theorem 8.1; as stated above, this special case is also given in [37, Corollary 3.5].

The equivalence (ii) $\iff$ (iii) follows by (A.8) and (A.12), which give

$$\Pi_1 f = 0 \iff \Pi_{\{i\}} f = 0 \,\forall i \iff f_i = 0 \,\forall i. \tag{A.13}$$

□

Note that (A.8) and (A.12) also yield

$$\Pi_1 f(x_1, \dots, x_\ell) = \sum_{i=1}^{\ell} \Pi_{\{i\}} f(x_1, \dots, x_\ell) = \sum_{i=1}^{\ell} f_i(x_i). \tag{A.14}$$

Corollary A.1. *The A^ℓ different unconstrained subsequence counts $N_n(\mathbf{w})$ with $\mathbf{w} \in \mathcal{A}^\ell$ converge jointly, after normalization as in (13.4), to a centered multivariate normal distribution in $\mathbb{R}^{A^\ell}$ whose support is a subspace of dimension $\ell(A-1)$.*

Proof. By Remark 3.6, we have joint convergence in Theorem 13.1 to some centered multivariate normal distribution in $V = \mathbb{R}^{A^\ell}$. Let L be the support of this distribution; then L is a subspace of V. Let $f \in V$. Then, by Theorems 13.1 and A.1,

$$\begin{aligned} f \perp L &\iff \sum_{\mathbf{w}\in\mathcal{A}^\ell} f(\mathbf{w}) \frac{N_n(\mathbf{w}) - \mathbb{E}N_n(\mathbf{w})}{n^{\ell-1/2}} \xrightarrow{\mathrm{d}} 0 \iff \sigma^2(f) = 0 \\ &\iff \Pi_1 f = 0 \iff f \perp V_1. \end{aligned} \tag{A.15}$$

Hence $L = V_1$, and the result follows by (A.10). □

What happens in the degenerate case when $\Pi_1 f = 0$ and thus $\sigma^2(f) = 0$? For symmetric U-statistics, this was considered by Hoeffding [27] (variance) and Rubin and Vitale [56] (asymptotic distribution); see also Dynkin and Mandelbaum [16]. Their results extend to the present asymmetric situation as follows. We make a final definition of a special subspace of V: let

$$V_{\geqslant k} := \bigoplus_{i=k}^{\ell} V_i = \{f \in V : \Pi_i f = 0 \text{ for } i = 0, \dots, k-1\}. \tag{A.16}$$

In particular, $V_{\geqslant 1}$ consists of all f with $\mathbb{E}f(\Xi_n) = 0$. Note also that f_* in (6.3), by (A.11) and (A.14), equals $f - \Pi_0 f - \Pi_1 f \in V_{\geqslant 2}$.

Lemma A.1. *Let $0 \leqslant k \leqslant \ell$. If $f \in V_{\geqslant k}$, then $\mathbb{E}U_n(f)^2 = O(n^{2\ell-k})$. Moreover, if $f \in V_{\geqslant k} \setminus V_{\geqslant k+1}$, then $\mathbb{E}U_n(f)^2 = \Theta(n^{2\ell-k})$.*

Proof. This is easily seen using the expansion (4.3) without the constraint $\mathcal{D}$, and is similar to the symmetric case in [27]; cf. also (in the more complicated m-dependent case) the cases $k = 1$ in (4.1) and $k = 2$ in (6.10). We omit the details. □

We can now state a general limit theorem that also includes degenerate cases.

Theorem A.2. *Let $k \geqslant 1$ and suppose that $f \in V_{\geqslant k}$. Then*

$$n^{k/2-\ell} U_n(f) \xrightarrow{\mathrm{d}} Z, \tag{A.17}$$

where Z is some polynomial of degree k in independent normal variables (possibly infinitely many). Moreover, Z is not degenerate unless $f \in V_{\geqslant k+1}$.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

Proof. This follows by [35, Theorem 11.19]. As noted in [35, Remark 11.21], it can also be reduced to the symmetric case in [56] by the following trick. Let $(\eta_i)_1^\infty$ be an i.i.d. sequence, independent of $(\xi_i)_1^\infty$, with $\eta_i \sim U(0,1)$; then

$$U_n\big(f;(\xi_i)\big) \stackrel{\mathrm{d}}{=} \sum_{i_1,\dots,i_\ell \leqslant n}^{*} f\big(\xi_{i_1},\dots,\xi_{i_\ell}\big)\mathbf{1}\{\eta_{i_1} < \dots < \eta_{i_\ell}\}, \tag{A.18}$$

where $\sum^*$ denotes summation over all distinct $i_1,\dots,i_\ell \in [n]$, and the sum in (A.18) can be regarded as a symmetric U-statistic based on $(\xi_i,\eta_i)_1^\infty$. The result (A.17) then follows by [56]. □

Remark A.1. The case $k=1$ in Theorem A.2 is just a combination of Theorem 13.1 (in the unconstrained case) and Theorem A.1; then Z is simply a normal variable. When $k=2$, there is a canonical representation (where the number of terms is finite or infinite)

$$Z = \frac{1}{2(\ell-2)!}\sum_i \lambda_i(\zeta_i^2 - 1), \tag{A.19}$$

where ζ_i are i.i.d. $\mathsf{N}(0,1)$ random variables and λ_i are the non-zero eigenvalues (counted with multiplicity) of a compact self-adjoint integral operator on $L^2(\mathcal{A}\times[0,1], \nu\times \mathrm{d}t)$, where $\nu := \mathcal{L}(\xi_1)$ is the distribution of a single letter and $\mathrm{d}t$ is Lebesgue measure; the kernel K of this integral operator can be constructed from f by applying [35, Corollary 11.5(iii)] to the symmetric U-statistic in (A.18). We omit the details, but note that in the particular case $k=\ell=2$, this kernel K is given by

$$K\big((x,t),(y,u)\big) = f(x,y)\mathbf{1}\{t<u\} + f(y,x)\mathbf{1}\{t>u\}, \tag{A.20}$$

and thus the integral operator is

$$h \mapsto Th(x,t) := \mathbb{E}\int_0^t f(\xi_1,x)h(\xi_1,u)\,\mathrm{d}u + \mathbb{E}\int_t^1 f(x,\xi_1)h(\xi_1,u)\,\mathrm{d}u. \tag{A.21}$$

When $k \geqslant 3$, the limit Z can be represented as a multiple stochastic integral [35, Theorem 11.19], but we do not know any canonical representation of it. See also [56] and [16].

We give two simple examples of limits in degenerate cases; in both cases $k=2$. The second example shows that although the space V has finite dimension, the representation (A.19) might require infinitely many terms. (Note that the operator T in (A.21) acts in an infinite-dimensional space.)

Example A.1. Let Ξ_n be a symmetric binary string, i.e., $\mathcal{A}=\{0,1\}$ and $p(0)=p(1)=1/2$. Consider

$$N_n(00) + N_n(11) - N_n(01) - N_n(10) = U_n(f), \tag{A.22}$$

with

$$\begin{aligned} f(x,y) &:= \mathbf{1}\{xy=00\} + \mathbf{1}\{xy=11\} - \mathbf{1}\{xy=01\} - \mathbf{1}\{xy=10\} \\ &:= \big(\mathbf{1}\{x=1\} - \mathbf{1}\{x=0\}\big)\big(\mathbf{1}\{y=1\} - \mathbf{1}\{y=0\}\big). \end{aligned} \tag{A.23}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

For convenience, we change notation and consider instead the letters $\hat{\xi}_i := 2\xi_i - 1 \in \{\pm1\}$; then f corresponds to

$$\hat{f}(\hat{x}, \hat{y}) := \hat{x}\hat{y}. \tag{A.24}$$

Thus

$$\begin{aligned} U_n\bigl(f;(\xi_i)\bigr) = U_n\bigl(\hat{f};(\hat{\xi}_i)\bigr) &= \sum_{1\leqslant i<j\leqslant n} \hat{\xi}_i\hat{\xi}_j = \frac{1}{2}\left(\left(\sum_{i=1}^n \hat{\xi}_i\right)^2 - \sum_{i=1}^n \hat{\xi}_i^2\right) \\ &= \frac{1}{2}\left(\sum_{i=1}^n \hat{\xi}_i\right)^2 - \frac{n}{2}. \end{aligned} \tag{A.25}$$

By the central limit theorem, $n^{-1/2}\sum_{i=1}^n \hat{\xi}_i \xrightarrow{\mathrm{d}} \zeta \sim \mathsf{N}(0,1)$, and thus (A.25) implies

$$n^{-1}U_n(f) \xrightarrow{\mathrm{d}} \tfrac{1}{2}(\zeta^2 - 1). \tag{A.26}$$

This is an example of (A.17), with $k = \ell = 2$ and limit given by (A.19), in this case with a single term in the sum and $\lambda_1 = 1$.

Note that in this example, the function f is symmetric, so (A.22) is an example of a symmetric U-statistic, and thus the result (A.26) is also an example of the limit result in [56].

Example A.2. Let $\mathcal{A} = \{a, b, c, d\}$, with ξ_i having the symmetric distribution $p(x) = 1/4$ for each $x \in \mathcal{A}$. Consider

$$N_n(ac) - N_n(ad) - N_n(bc) + N_n(bd) = U_n(f), \tag{A.27}$$

with, writing $\mathbf{1}_y(x) := \mathbf{1}\{x = y\}$,

$$f(x, y) := \bigl(\mathbf{1}_a(x) - \mathbf{1}_b(x)\bigr)\bigl(\mathbf{1}_c(x) - \mathbf{1}_d(x)\bigr). \tag{A.28}$$

Then $\Pi_0 f = \Pi_1 f = 0$ by symmetry, so $f \in V_{\geqslant 2} = V_2$ (since $\ell = 2$).

Consider the integral operator T on $L^2(\mathcal{A} \times [0,1])$ defined by (A.21). Let h be an eigenfunction with eigenvalue $\lambda \neq 0$, and write $h_x(t) := h(x, t)$. The eigenvalue equation $Th = \lambda h$ then is equivalent to the following, using (A.21) and (A.28):

$$\lambda h_a(t) = \frac{1}{4}\int_t^1 \bigl(h_c(u) - h_d(u)\bigr)\,\mathrm{d}u, \tag{A.29}$$

$$\lambda h_b(t) = \frac{1}{4}\int_t^1 \bigl(-h_c(u) + h_d(u)\bigr)\,\mathrm{d}u, \tag{A.30}$$

$$\lambda h_c(t) = \frac{1}{4}\int_0^t \bigl(h_a(u) - h_b(u)\bigr)\,\mathrm{d}u, \tag{A.31}$$

$$\lambda h_d(t) = \frac{1}{4}\int_0^t \bigl(-h_a(u) + h_b(u)\bigr)\,\mathrm{d}u. \tag{A.32}$$

These equations hold a.e., but we can redefine $h_x(t)$ by these equations so that they hold for every $t \in [0,1]$. Moreover, although originally we assume only $h_x \in L^2[0,1]$, it follows from (A.29)–(A.32) that the functions $h_x(t)$ are continuous in t, and then by induction that they are

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

infinitely differentiable on [0, 1]. Note also that (A.29) and (A.30) yield $h_b(t) = -h_a(t)$, and similarly $h_d(t) = -h_c(t)$. Hence, we may reduce the system to

$$\lambda h_a(t) = \frac{1}{2}\int_t^1 h_c(u)\,\mathrm{d}u, \tag{A.33}$$

$$\lambda h_c(t) = \frac{1}{2}\int_0^t h_a(u)\,\mathrm{d}u. \tag{A.34}$$

By differentiation, for $t \in (0, 1)$,

$$h_a{}'(t) = -\frac{1}{2\lambda}h_c(t), \tag{A.35}$$

$$h_c{}'(t) = \frac{1}{2\lambda}h_a(t). \tag{A.36}$$

Hence, with $\omega := 1/(2\lambda)$,

$$h_c{}''(t) = -\omega^2 h_c(t). \tag{A.37}$$

Furthermore, (A.34) yields $h_c(0) = 0$, and thus (A.37) has the solution (up to a constant factor that we may ignore)

$$h_c(t) = \sin \omega t = \sin \frac{t}{2\lambda}. \tag{A.38}$$

By (A.36), we then obtain

$$h_a(t) = \cos \omega t = \cos \frac{t}{2\lambda}. \tag{A.39}$$

However, (A.33) also yields $h_a(1) = 0$ and thus we must have $\cos(1/2\lambda) = 0$; hence

$$\lambda = \frac{1}{(2N+1)\pi}, \qquad N \in \mathbb{Z}. \tag{A.40}$$

Conversely, for every λ of the form (A.40), the argument can be reversed to find an eigenfunction h with eigenvalue λ. It follows also that all these eigenvalues are simple. Consequently, Theorem A.2 and (A.19) yield

$$n^{-1}U_n(f) \xrightarrow{\mathrm{d}} \frac{1}{2\pi}\sum_{N=-\infty}^{\infty}\frac{1}{2N+1}(\zeta_N^2 - 1) = \frac{1}{2\pi}\sum_{N=0}^{\infty}\frac{1}{2N+1}(\zeta_N^2 - \zeta_{-N-1}^2), \tag{A.41}$$

where, as above, ζ_N are i.i.d. and $\mathsf{N}(0, 1)$. A simple calculation, using the product formula for cosine (see [17, Section 12], [46, 4.22.2]), shows that the moment generating function of the limit distribution Z in (A.41) is

$$\mathbb{E}e^{sZ} = \frac{1}{\cos^{1/2}(s/2)}, \qquad |\mathrm{Re}\, s| < \pi. \tag{A.42}$$

It can be shown that $Z \overset{\mathrm{d}}{=} \frac{1}{2}\int_0^1 B_1(t)\,\mathrm{d}B_2(t)$ if $B_1(t)$ and $B_2(t)$ are two independent standard Brownian motions; this is, for example, a consequence of (A.42) and the following calculation, using [13] or [53, p. 445] for the final equality:

$$\begin{aligned}\mathbb{E}e^{s\int_0^1 B_1(t)\,\mathrm{d}B_2(t)} &= \mathbb{E}\mathbb{E}\big[e^{s\int_0^1 B_1(t)\,\mathrm{d}B_2(t)} \mid B_1\big] = \mathbb{E}e^{\frac{s^2}{2}\int_0^1 B_1(t)^2\,\mathrm{d}t}\\ &= \cos^{-1/2}(s), \qquad |\mathrm{Re}\, s| < \pi/2.\end{aligned} \tag{A.43}$$

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

We omit the details, but note that this representation of the limit Z is related to the special form (A.28) of f; we may, intuitively at least, interpret B_1 and B_2 as limits (by Donsker's theorem) of partial sums of $\mathbf{1}_a(\xi_i) - \mathbf{1}_b(\xi_i)$ and $\mathbf{1}_c(\xi_i) - \mathbf{1}_d(\xi_i)$. In fact, in this example it is possible to give a rigorous proof of $n^{-1}U_n(f) \xrightarrow{\mathrm{d}} Z$ by this approach; again we omit the details.

Funding information

This work was supported by the Knut and Alice Wallenberg Foundation.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

Acknowledgements

I thank Wojciech Szpankowski for stimulating discussions on patterns in random strings over the last 20 years. Furthermore, I thank Andrew Barbour and Nathan Ross for help with references, and the anonymous referees for helpful comments and further references; in particular, Theorem 3.5 is based on suggestions from a referee.

References

[1] Alsmeyer, G. and Hoefs, V. (2001). Markov renewal theory for stationary m-block factors. *Markov Process. Relat. Fields* **7**, 325–348.
[2] Baldi, P. and Rinott, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Prob.* **17**, 1646–1650.
[3] Barbour, A. D., Karoński, M. and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combinatorial Theory B* **47**, 125–145.
[4] Bender, E. A. and Kochman, F. (1993). The distribution of subword counts is usually normal. *Europ. J. Combinatorics* **14**, 265–275.
[5] Billingsley, P. (1956). The invariance principle for dependent random variables. *Trans. Amer. Math. Soc.* **83**, 250–268.
[6] Blom, G. (1976). Some properties of incomplete U-statistics. *Biometrika* **63**, 573–580.
[7] Bóna, M. (2007). The copies of any permutation pattern are asymptotically normal. Preprint. Available at https://arxiv.org/0712.2792.
[8] Bóna, M. (2008). Generalized descents and normality. *Electron. J. Combinatorics* **15**, paper no. 21, 8 pp.
[9] Bóna, M. (2010). On three different notions of monotone subsequences. In *Permutation Patterns*, Cambridge University Press, pp. 89–114.
[10] Bourdon, J. and Vallée, B. (2002). Generalized pattern matching statistics. In *Mathematics and Computer Science II: Algorithms, Trees, Combinatorics and Probabilities*, Birkhäuser, Basel, pp. 249–265.
[11] Bourdon, J. and Vallée, B. (2006). Pattern matching statistics on correlated sources. In *LATIN 2006: Theoretical Informatics*, Springer, Berlin, pp. 224–237.
[12] Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions*, Vol. 1–3. Kendrick Press, Heber City, UT.
[13] Cameron, R. H. and Martin, W. T. (1945). Transformations of Wiener integrals under a general class of linear transformations. *Trans. Amer. Math. Soc.* **58**, 184–219.
[14] Chen, L. H. Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method*. Springer, Berlin.
[15] Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *J. Multivariate Anal.* **101**, 126–137.
[16] Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.* **11**, 739–745.
[17] Euler, L. (1743). De summis serierum reciprocarum ex potestatibus numerorum naturalium ortarum dissertatio altera, in qua eaedem summationes ex fonte maxime diverso derivantur. *Miscellanea Berolinensia* **7**, 172–192. Reprinted in *Opera Omnia*, Ser. 1, Vol. 14, Teubner, Leipzig, 1925, pp. 138–155.
[18] Even-Zohar, C. (2020). Patterns in random permutations. *Combinatorica* **40**, 775–804.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

[19] EVEN-ZOHAR, C., LAKREC, T. AND TESSLER, R. J. (2021). Spectral analysis of word statistics. *Sém. Lothar. Combin.* **85B**, paper no. 81, 12 pp.
[20] FANG, X. (2016). A multivariate CLT for bounded decomposable random vectors with the best known rate. *J. Theoret. Prob.* **29**, 1510–1523.
[21] FISHER, N. I. AND LEE, A. J. (1982). Nonparametric measures of angular–angular association. *Biometrika* **69**, 315–321.
[22] FLAJOLET, P. AND SEDGEWICK, R. (2009). *Analytic Combinatorics*. Cambridge University Press.
[23] FLAJOLET, P., SZPANKOWSKI, W. AND VALLÉE, B. (2006). Hidden word statistics. *J. Assoc. Comput. Mach.* **53**, 147–183.
[24] GUT, A. (2009). *Stopped Random Walks*, 2nd edn. Springer, New York.
[25] GUT, A. (2013). *Probability: A Graduate Course*, 2nd edn. Springer, New York.
[26] HAN, F. AND QIAN, T. (2018). On inference validity of weighted *U*-statistics under data heterogeneity. *Electron. J. Statist.* **12**, 2637–2708.
[27] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293–325.
[28] HOEFFDING, W. (1961). The strong law of large numbers for *U*-statistics. Tech. Rep., University of North Carolina, Institute of Statistics, Mimeograph Series No. 302. Available at https://repository.lib.ncsu.edu/handle/1840.4/2128.
[29] HOFER, L. (2018). A central limit theorem for vincular permutation patterns. *Discrete Math. Theoret. Comput. Sci.* **19**, paper no. 9, 26 pp.
[30] HSING, T. AND WU, W. B. (2004). On weighted *U*-statistics for stationary processes. *Ann. Prob.* **32**, 1600–1631.
[31] JACQUET, P. AND SZPANKOWSKI, W. (2015). *Analytic Pattern Matching: From DNA to Twitter*. Cambridge University Press.
[32] JAMMALAMADAKA, S. R. AND JANSON, S. (1986). Limit theorems for a triangular scheme of *U*-statistics with applications to inter-point distances. *Ann. Prob.* **14**, 1347–1358.
[33] JANSON, S. (1983). Renewal theory for *M*-dependent variables. *Ann. Prob.* **11**, 558–568.
[34] JANSON, S. (1988). Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs. *Ann. Prob.* **16**, 305–312.
[35] JANSON, S. (1997). *Gaussian Hilbert Spaces*. Cambridge University Press.
[36] JANSON, S. (2015). On degenerate sums of *m*-dependent variables. *J. Appl. Prob.* **52**, 1146–1155.
[37] JANSON, S. (2018). Renewal theory for asymmetric *U*-statistics. *Electron. J. Prob.* **23**, paper no. 129, 27 pp.
[38] JANSON, S. (2022). The number of occurrences of patterns in a random tree or forest permutation. Preprint. Available at https://arxiv.org/2203.04182.
[39] JANSON, S., NAKAMURA, B. AND ZEILBERGER, D. (2015). On the asymptotic statistics of the number of occurrences of multiple permutation patterns. *J. Combinatorics* **6**, 117–143.
[40] JANSON, S. AND SZPANKOWSKI, W. (2021). Hidden words statistics for large patterns. *Electron. J. Combinatorics* **28**, paper no. P2.36, 26 pp.
[41] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.
[42] MAJOR, P. (1994). Asymptotic distributions for weighted *U*-statistics. *Ann. Prob.* **22**, 1514–1535.
[43] MALEVICH, T. L. AND ABDALIMOV, B. (1982). Refinement of the central limit theorem for *U*-statistics of *m*-dependent variables. *Teor. Veroyat. Primen.* **27**, 369–373 (in Russian). English translation: *Theory Prob. Appl.* **27**, 391–396.
[44] MILLER, R. G., Jr. AND SEN, P. K. (1972). Weak convergence of *U*-statistics and von Mises' differentiable statistical functions. *Ann. Math. Statist.* **43**, 31–41.
[45] NICODÈME, P., SALVY, B. AND FLAJOLET, P. (2002). Motif statistics. *Theoret. Comput. Sci.* **287**, 593–617.
[46] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. AND CLARK, C. W. (eds) (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press.
[47] O'NEIL, K. A. AND REDNER, R. A. (1993). Asymptotic distributions of weighted *U*-statistics of degree 2. *Ann. Prob.* **21**, 1159–1169.
[48] OREY, S. (1958). A central limit theorem for *m*-dependent random variables. *Duke Math. J.* **25**, 543–546.
[49] PELIGRAD, M. (1996). On the asymptotic normality of sequences of weak dependent random variables. *J. Theoret. Prob.* **9**, 703–715.
[50] PIKE, J. (2011). Convergence rates for generalized descents. *Electron. J. Combinatorics* **18**, paper no. 236, 14 pp.
[51] RAIČ, M. (2004). A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Prob.* **17**, 573–603.
[52] RÉGNIER, M. AND SZPANKOWSKI, W. (1998). On pattern frequency occurrences in a Markovian sequence. *Algorithmica* **22**, 631–649.
[53] REVUZ, D. AND YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd edn. Springer, Berlin.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press

[54] RINOTT, Y. (1994). On normal approximation rates for certain sums of dependent random variables. *J. Comput. Appl. Math.* **55**, 135–143.
[55] RINOTT, Y. AND ROTAR, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted *U*-statistics. *Ann. Appl. Prob.* **7**, 1080–1105.
[56] RUBIN, H. AND VITALE, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8**, 165–170.
[57] SEN, P. K. (1960). On some convergence properties of *U*-statistics. *Calcutta Statist. Assoc. Bull.* **10**, 1–18.
[58] SEN, P. K. (1963). On the properties of *U*-statistics when the observations are not independent. I. Estimation of non-serial parameters in some stationary stochastic process. *Calcutta Statist. Assoc. Bull.* **12**, 69–92.
[59] SEN, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary mixing processes. *Z. Wahrscheinlichkeitsth.* **25**, 71–82.
[60] SEN, P. K. (1974). Weak convergence of generalized *U*-statistics. *Ann. Prob.* **2**, 90–102.
[61] SHAPIRO, C. P. AND HUBERT, L. (1979). Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. *Ann. Statist.* **7**, 788–794.
[62] SZPANKOWSKI, W. (2001). *Average Case Analysis of Algorithms on Sequences*. John Wiley, New York.
[63] UTEV, S. A. (1990). Central limit theorem for dependent random variables. In *Probability Theory and Mathematical Statistics: Proc. Fifth Vilnius Conference, June 25–July 1, 1990*, Vol. II, Mokslas, Vilnius, pp. 519–528.
[64] YOSHIHARA, K. (1976). Limiting behavior of *U*-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitsth.* **35**, 237–252.
[65] YOSHIHARA, K. (1992). Limiting behavior of *U*-statistics for strongly mixing sequences. *Yokohama Math. J.* **39**, 107–113.
[66] ZHOU, Z. (2014). Inference of weighted *V*-statistics for nonstationary time series and its applications. *Ann. Statist.* **42**, 87–114.

https://doi.org/10.1017/apr.2022.51 Published online by Cambridge University Press