Electronic Journal of Probability

Electron. J. Probab. **23** (2018), no. 129, 1–27.
ISSN: 1083-6489 https://doi.org/10.1214/18-EJP252

Renewal theory for asymmetric U-statistics*

Svante Janson†

Abstract

We extend a functional limit theorem for symmetric U-statistics [Miller and Sen, 1972] to asymmetric U-statistics, and use this to show some renewal theory results for asymmetric U-statistics.

Some applications are given.

Keywords: U-statistics; renewal theory; functional limit theorems.
AMS MSC 2010: Primary 60F05, Secondary 60F17; 60K05.
Submitted to EJP on April 16, 2018, final version accepted on November 27, 2018.
Supersedes arXiv:1804.05509.

1 Introduction

Let $X, X_1, X_2, \dots,$ be an i.i.d. sequence of random variables taking values in an arbitrary measurable space $S = (S, \mathcal{S})$. (In most cases, $S = \mathbb{R}$ or perhaps $\mathbb{R}^k$, or a Borel subset of one of these, but we can just as well consider the general case.) Furthermore, let $d \geqslant 1$ and let $f : S^d \to \mathbb{R}$ be a given measurable function. We then define the (real-valued) random variables

$$U_n = U_n(f) := \sum_{1 \leqslant i_1 < \dots < i_d \leqslant n} f\bigl(X_{i_1}, \dots, X_{i_d}\bigr), \qquad n \geqslant 0. \tag{1.1}$$

We call U_n a U*-statistic*, following Hoeffding [16].

Remark 1.1. Many authors, including Hoeffding [16], normalize U_n by dividing the sum in (1.1) by $\binom{n}{d}$, the number of terms in it; the traditional definition (which assumes $n \geqslant d$) is thus in our notation $U_n/\binom{n}{d}$. We find it more convenient for our purposes to use the unnormalized version above.

It is common, following Hoeffding [16], to assume that f is a symmetric function of its d variables. In this case, the order of the variables does not matter, and we can in (1.1) sum over all sequences $i_1, \dots, i_d$ of d distinct elements of $\{1, \dots, n\}$, up to an obvious factor of $d!$. ([16] 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

*Partly supported by the Knut and Alice Wallenberg Foundation
†Uppsala University, Sweden. E-mail: svante.janson@math.uu.se

paper we consider the general case of (1.1) without assuming symmetry, which we for emphasis may call *asymmetric U-statistics*. One of the purposes of this paper is to generalize a result by [26] on functional convergence from the symmetric case to the general, asymmetric case. We then use this result to derive some renewal theory results for the sequence U_n. One motivation for this is some applications to random restricted permutations, see Section 5.

Univariate limit results, i.e., limits in distribution of U_n after suitable normalization, are well-known also in the asymmetric case, see e.g. [20, Chapter 11.2]. The possibility of functional limits is briefly mentioned in [20, Remark 11.25], and a special case ($d = 2$ and f antisymmetric) was studied in [24], see Example 5.1; However, we are not aware of functional limit theorems in the generality of the present paper.

The main results are stated in Section 3. The proofs are given in Section 4; they use standard methods, in particular the decomposition and projection method of Hoeffding [16], but some complications arise in the asymmetric case; this includes applications to random restricted permutations that gave the initial motivation to write the present paper. Some examples and applications are discussed in Section 5. We end with some further comments and open problems in Section 6; this includes more comments on the relation between the symmetric and asymmetric cases.

The results in the present paper focus on the non-degenerate case, where the covariance matrix $\Sigma = (\sigma_{ij})$ defined by (3.2) below is non-zero. In the degenerate case when $\Sigma = 0$, the result still holds but are less interesting, since the obtained limits in e.g. Theorem 3.2 are degenerate. See Remark 6.3 for further comments on the degenerate case.

2 Some notation

We consider as in the introduction, unless otherwise said, some given i.i.d. random variables $X_i \in S$ and a given function $f : S^d \to \mathbb{R}$. In particular, $d \geqslant 1$ is fixed, and we therefore often omit it from the notation. (When we consider two U-statistics of possibly different degrees, we denote them by d and $\tilde{d}$.)

We assume throughout $f(X_1, \dots, X_d) \in L^1$ (and usually L^2), and define

$$\mu := \mathbb{E} f(X_1, \dots, X_d). \tag{2.1}$$

We study $U_n = U_n(f)$ defined by (1.1). Let

$$U_n^* = U_n^*(f) := \max_{1 \leqslant m \leqslant n} |U_m(f)|. \tag{2.2}$$

We use $\| \|_p$ for the L^p-norm: $\|Y\|_p := (\mathbb{E}\, Y^p)^{1/p}$ for any random variable Y and $p > 0$, and $\|f\|_p := \|f(X_1, \dots, X_d)\|_p$ (and similarly for other functions).

$\mathcal{F}_n$ is the σ-field generated by $X_1, \dots, X_n$.

If we consider a limit as $n \to \infty$, and a_n is a given sequence, then $o_{\text{a.s.}}(a_n)$ denotes a sequence of random variables R_n such that $R_n/a_n \xrightarrow{\text{a.s.}} 0$. This extends to other limits such as $x \to \infty$, *mutatis mutandis*.

C denotes positive constants that may change from one occurence to the next; they may depend on d but not on f or n or other variables. Similarly, C_f denote constants that may depend on f, C_p denotes constants that may depend on the parameter p (and d), and so on.

3 Main results

3.1 Limit theorems

For completeness, we begin with the law of large numbers, extending the result by Hoeffding [17] to the asymmetric case.

Theorem 3.1. *Suppose that $f(X_1,\dots,X_d)\in L^1$. Then, as $n\to\infty$,*

$$U_n/\binom{n}{d}\xrightarrow{\text{a.s.}}\mu. \tag{3.1}$$

Next we state a functional limit theorem, extending the theorem by Miller and Sen [26] for the symmetric case. We use the space $D[0,\infty)$ with the usual Skorohod topology, see e.g. [25, Appendix A2] or [2]; recall that convergence in $D[0,\infty)$ to a continuous limit is equivalent to uniform convergence on any compact interval $[0,T]$. We assume $f(X_1,\dots,X_d)\in L^2$ and define the $d\times d$ matrix $\Sigma=(\sigma_{ij})$ by

$$\sigma_{ij}:=\operatorname{Cov}\bigl(f_i(X),f_j(X)\bigr)=\mathbb{E}\bigl(f_i(X)f_j(X)\bigr),\qquad i,j=1,\dots,d, \tag{3.2}$$

where the functions f_i are the one-variable projections

$$\begin{aligned}f_i(x)&:=\mathbb{E}\bigl(f(X_1,\dots,X_d)\mid X_i=x\bigr)-\mu\\&=\mathbb{E}\,f\bigl(X_1,\dots,X_{i-1},x,X_{i+1},\dots,X_d\bigr)-\mu.\end{aligned} \tag{3.3}$$

(In general, these are defined only a.e., but that is no problem.) Let $\mathbf{W}(t):=\bigl(W_1(t),\dots,W_d(t)\bigr)$, $t\geqslant 0$, be a continuous d-dimensional Gaussian process with $\mathbf{W}(0)=0$ and stationary independent increments

$$\mathbf{W}(s+t)-\mathbf{W}(s)\sim N\bigl(0,t\Sigma\bigr). \tag{3.4}$$

Note that each component W_j is a standard Brownian motion up to a factor $\sigma_{jj}^{1/2}$, and that we can represent $\mathbf{W}$ as $\mathbf{W}(t)=\Sigma^{1/2}\mathbf{B}(t)$, where $\mathbf{B}(t)$ is a d-dimensional standard Brownian motion. Define also the functions

$$\psi_j(s,t)=\psi_{j;d}(s,t):=\frac{1}{(j-1)!\,(d-j)!}s^{j-1}(t-s)^{d-j}. \tag{3.5}$$

We extend U_n defined by (1.1) to a function of a real variable by $U_x:=U_{\lfloor x\rfloor}$, $x>0$. (We tacitly do the same for other sequences later, and similarly allow upper summation limits to be non-integer.)

Theorem 3.2. *Suppose that $f(X_1,\dots,X_d)\in L^2$. Then, as $n\to\infty$,*

$$\frac{U_{nt}-n^dt^d\mu/d!}{n^{d-1/2}}\xrightarrow{\text{d}}Z_t,\qquad t\geqslant 0, \tag{3.6}$$

in $D[0,\infty)$, where Z_t is a continuous centered Gaussian process that can be defined as

$$Z_t:=\sum_{j=1}^d\int_0^t\psi_j(s,t)\,\mathrm{d}W_j(s). \tag{3.7}$$

Equivalently, Z_t has the covariance function, for $0\leqslant s\leqslant t$,

$$\operatorname{Cov}(Z_s,Z_t)=\sum_{i,j=1}^d\sigma_{ij}\int_0^s\psi_i(u,s)\psi_j(u,t)\,\mathrm{d}u$$

$$= \sum_{i,j=1}^{d} \frac{\sigma_{ij}}{(i-1)!\,(j-1)!\,(d-i)!\,(d-j)!} \int_0^s u^{i+j-2}(s-u)^{d-i}(t-u)^{d-j}\,\mathrm{d}u. \tag{3.8}$$

Moreover, (3.6) *holds jointly for several functions $f^{(k)}$, possibly with different $d^{(k)}$, with limits given by* (3.7)*, where the corresponding $W_j^{(k)}$ together form a Gaussian process with stationary independent increments given by the covariances*

$$\operatorname{Cov}\bigl(W_i^{(k)}(s), W_j^{(\ell)}(t)\bigr) = \operatorname{Cov}\bigl(f_i^{(k)}(X), f_j^{(\ell)}(X)\bigr)\cdot(s\wedge t). \tag{3.9}$$

The Itô integrals appearing in (3.7) can by (3.5) be written as linear combinations of $t^k \int_0^t s^{d-1-k}\,\mathrm{d}W_j(s)$ with $0 \leqslant k \leqslant d-j$; thus Z_t is well-defined and continuous for $t \geqslant 0$, with $Z_0 = 0$. These stochastic integrals can also by integration by parts be expressed as Riemann integrals of continuous stochastic processes, see (4.17).

Note that the final integral in (3.8) is elementary, for any given i, j, d, and that the covariance function in (3.8) is a homogeneous polynomial in s and t of degree $2d-1$.

Example 3.3. In the case $d = 2$, we obtain from (3.8), still for $0 \leqslant s \leqslant t$,

$$\operatorname{Cov}(Z_s, Z_t) = \tfrac{1}{2}\bigl(\sigma_{11}+\sigma_{12}\bigr)s^2 t + \tfrac{1}{6}\bigl(2\sigma_{22}-\sigma_{11}-\sigma_{12}\bigr)s^3. \tag{3.10}$$

Remark 3.4. By (3.5) and the binomial theorem,

$$\sum_{j=1}^{d} \psi_j(s,t) = \frac{t^{d-1}}{(d-1)!}. \tag{3.11}$$

In the symmetric case, all f_i are equal and thus all σ_{ij} are equal, see (3.2). Hence, (3.8) simplifies by (3.11) to

$$\operatorname{Cov}(Z_s, Z_t) = \sigma_{11} \int_0^s \frac{s^{d-1}}{(d-1)!}\,\frac{t^{d-1}}{(d-1)!}\,\mathrm{d}u = \frac{\sigma_{11}}{(d-1)!^2} s^d t^{d-1}. \tag{3.12}$$

Equivalently, $t^{-(d-1)}Z_t$ is $\sigma_{11}^{1/2}(d-1)!^{-1}B_t$ for a standard Brownian motion B_t. This recovers the result by Miller and Sen [26] for the symmetric case. Note that our general result Theorem 3.2 is similar to the symmetric case, with a continuous Gaussian limit process, but that the covariance function in general is more complicated, as seen for $d = 2$ in (3.10), and that the limit thus is not a Brownian motion.

By restricting attention to $t = 1$, we obtain the following univariate limit, shown in [20, Corollary 11.20].

Corollary 3.5. *Suppose that $f(X_1, \dots, X_d) \in L^2$. Then, as $n \to \infty$,*

$$\frac{U_n - \binom{n}{d}\mu}{n^{d-1/2}} \overset{\mathrm{d}}{\longrightarrow} N\bigl(0, \sigma^2\bigr), \tag{3.13}$$

where

$$\begin{aligned} \sigma^2 &:= \lim_{n\to\infty} \frac{\operatorname{Var}(U_n)}{n^{2d-1}} = \operatorname{Var}(Z_1) \\ &= \sum_{i,j=1}^{d} \frac{(i+j-2)!\,(2d-i-j)!}{(i-1)!\,(j-1)!\,(d-i)!\,(d-j)!\,(2d-1)!}\sigma_{ij}. \end{aligned} \tag{3.14}$$

Moreover,

$$\sigma^2 = 0 \iff f_i(X) = 0 \text{ a.s. for every } i = 1, \dots, d. \tag{3.15}$$

Example 3.6. For $d = 1$, Corollary 3.5 reduces to the Central Limit Theorem; indeed, (3.14) then yields $\sigma^2 = \sigma_{11}$.

For $d = 2$, (3.14) yields

$$\sigma^2 = \frac{\sigma_{11}+\sigma_{12}+\sigma_{22}}{3}. \tag{3.16}$$

3.2 Renewal theory

For $x > 0$, let

$$N_-(x) := \sup\{n \geqslant 0 : U_n \leqslant x\}, \tag{3.17}$$

$$N_+(x) := \inf\{n \geqslant 0 : U_n > x\}. \tag{3.18}$$

Note that if $f \geqslant 0$, then $N_+(x) = N_-(x) + 1$, but if f attains negative values, then $N_-(x) > N_+(x)$ is possible. Most of our results apply to both N_+ and N_-; we then use $N_\pm$ to denote any of them.

The results above easily imply some renewal theorems for U-statistics generalizing well-known results for S_n (i.e., the case $d = 1$). We begin with a law of large numbers.

Theorem 3.7. *Suppose that $f(X_1, \dots, X_d) \in L^1$ and $\mu > 0$. Then a.s. $N_\pm(x) < \infty$ for every $x < \infty$, and*

$$\frac{N_\pm(x)}{x^{1/d}} \xrightarrow{\text{a.s.}} \Bigl(\frac{d!}{\mu}\Bigr)^{1/d} \qquad \text{as } x \to \infty. \tag{3.19}$$

Assuming $f \in L^2$, we obtain also a central limit theorem for $N_\pm$.

Theorem 3.8. *Suppose that $f(X_1, \dots, X_d) \in L^2$ and $\mu > 0$. Then, as $x \to \infty$,*

$$\frac{N_\pm(x) - (d!/\mu)^{1/d} x^{1/d}}{x^{1/2d}} \xrightarrow{\text{d}} N\Bigl(0, \bigl(d!/\mu\bigr)^{2+1/d} d^{-2} \sigma^2\Bigr), \tag{3.20}$$

where σ^2 is given by (3.14).

A situation that is common in application is to stop when when one process (such as our U_n) reaches a threshold, and then look at the value of another process, say $\widetilde{U}_n$. For standard renewal theory, i.e. the case $d = 1$ in our setting, this was studied in [14]; we extend the main result there to (asymmetric) U-statistics. We consider as above an i.i.d. sequence $X_1, X_2, \dots$ with values in S, but we now have two functions $f : S^d \to \mathbb{R}$ and $\tilde{f} : S^{\tilde{d}} \to \mathbb{R}$, where the numbers of variables d and $\tilde{d}$ may be different. We use notations as above for both f and $\tilde{f}$, with $\tilde{}$ to denote variables defined by $\tilde{f}$, for example $\widetilde{U}_n := U_n(\tilde{f})$ and $\tilde{\mu} := \mathbb{E}\,\tilde{f}$; we furthermore assume that the Gaussian processes $W_i(t)$ and $\widetilde{W}_j(t)$ have the joint distribution specified by (3.9) (with obvious notational changes), and thus (3.6) holds jointly for f and $\tilde{f}$ with limits Z_t and $\tilde{Z}_t$.

Theorem 3.9. (i) *Suppose that $f(X_1, \dots, X_d) \in L^1$, $\tilde{f}(X_1, \dots, X_{\tilde{d}}) \in L^1$ and $\mu > 0$. Then, as $x \to \infty$,*

$$\frac{\widetilde{U}_{N_\pm(x)}}{x^{\tilde{d}/d}} \xrightarrow{\text{a.s.}} \frac{\tilde{\mu}}{\tilde{d}!}\Bigl(\frac{d!}{\mu}\Bigr)^{\tilde{d}/d}. \tag{3.21}$$

(ii) *Suppose that $f(X_1, \dots, X_d) \in L^2$, $\tilde{f}(X_1, \dots, X_{\tilde{d}}) \in L^2$ and $\mu > 0$. Then, as $x \to \infty$,*

$$\frac{\widetilde{U}_{N_\pm(x)} - \bigl(\frac{d!}{\mu}\bigr)^{\tilde{d}/d} \frac{\tilde{\mu}}{\tilde{d}!} x^{\tilde{d}/d}}{x^{\tilde{d}/d - 1/2d}} \xrightarrow{\text{d}} N\bigl(0, \gamma^2\bigr), \tag{3.22}$$

where, with $(Z_1, \tilde{Z}_1)$ as in Theorem 3.2,

$$\gamma^2 := \Bigl(\frac{d!}{\mu}\Bigr)^{(2\tilde{d}-1)/d} \operatorname{Var}\Bigl(\tilde{Z}_1 - \frac{(d-1)!\,\tilde{\mu}}{(\tilde{d}-1)!\,\mu} Z_1\Bigr). \tag{3.23}$$

(iii) *Assume the conditions in* (ii). *If $\tilde{d} \geqslant d$, then $\gamma^2 = 0$ if and only if*

$$\tilde{f}_i(X) = \frac{\tilde{\mu}}{\mu} \sum_j \frac{\binom{d-1}{j-1}\binom{\tilde{d}-d}{i-j}}{\binom{\tilde{d}-1}{i-1}} f_j(X) \text{ a.s.}, \quad i = 1, \dots, \tilde{d}. \tag{3.24}$$

If $\tilde d < d$, then $\gamma^2 = 0$ if and only if (3.24) *holds with f, d, μ and $\tilde f, \tilde f, \tilde\mu$ interchanged (and $\tilde\mu \neq 0$ unless all $\tilde f_j(X) = 0$ a.s.)*

In particular, if $\tilde d = d$, then

$$\gamma^2 = 0 \iff \mu \tilde f_i(X) = \tilde\mu f_i(X) \quad a.s., \qquad i = 1, \dots, d, \tag{3.25}$$

and if $d = 1$, then

$$\gamma^2 = 0 \iff \mu \tilde f_i(X) = \tilde\mu f_1(X) \quad a.s., \qquad i = 1, \dots, \tilde d. \tag{3.26}$$

Remark 3.10. Theorem 3.8 can be regarded as a special case of Theorem 3.9 with $\tilde d = 1$ and $\tilde f(X) \equiv 1$.

The asymptotic variance γ^2 in Theorem 3.9 can easily be calculated exactly using (3.7), (3.9) and (3.5), but a general formula seems more messy than illuminating, and we state only the special case $d = 1$. (In this case, U_n is the standard partial sum $\sum_{i=1}^n f(X_i)$.)

Theorem 3.11. *Suppose that $f(X) \in L^2$, $\tilde f(X_1, \dots, X_{\tilde d}) \in L^2$ and $\mu > 0$. Then, as $x \to \infty$,*

$$\frac{\widetilde U_{N_\pm(x)} - \mu^{-\tilde d}\tilde\mu \tilde d!^{-1} x^{\tilde d}}{x^{\tilde d - 1/2}} \xrightarrow{\mathrm{d}} N\bigl(0, \gamma^2\bigr), \tag{3.27}$$

where

$$\begin{aligned}\gamma^2 := \mu^{1-2\tilde d} \sum_{i,j=1}^{\tilde d} &\frac{(i+j-2)!\,(2\tilde d - i - j)!}{(i-1)!\,(j-1)!\,(\tilde d - i)!\,(\tilde d - j)!\,(2\tilde d - 1)!} \operatorname{Cov}\bigl(\tilde f_i(X), \tilde f_j(X)\bigr) \\ &- 2\frac{\mu^{-2\tilde d}\tilde\mu}{(\tilde d - 1)!\,\tilde d!} \sum_{i=1}^{\tilde d} \operatorname{Cov}\bigl(f(X), \tilde f_i(X)\bigr) \\ &+ \frac{\mu^{-2\tilde d - 1}\tilde\mu^2}{(\tilde d - 1)!^2} \operatorname{Var}\bigl(f(X)\bigr).\end{aligned} \tag{3.28}$$

Moreover,

$$\gamma^2 = 0 \iff \mu \tilde f_i(X) = \tilde\mu (f(X) - \mu) \quad a.s., \qquad i = 1, \dots, \tilde d. \tag{3.29}$$

Continue to assume that $d = 1$, and assume for simplicity that $Y := f(X) \geqslant 0$ a.s. Thus

$$U_n(f) = S_n(f) := \sum_1^n Y_i \tag{3.30}$$

is a renewal process, and its *overshoot* (*residual life time*) is

$$R(x) := U_{N_+(x)} - x > 0. \tag{3.31}$$

A classical result, see e.g. [12, Section 2.6], says that if $0 < \mu < \infty$, then $R(x)$ converges in distribution. Recall that (the distribution of) Y has *span* $d > 0$ if $Y \in d\mathbb{Z}$ a.s., and d is maximal with this property, and that (the distribution of) Y is nonarithmetic if no such d exists.

Proposition 3.12 (e.g. [12, Theorem 2.6.2])**.** *Let $R(x)$ be given by* (3.31)*, and assume that $f(X) \geqslant 0$ a.s. and $0 < \mathbb{E} f(X) < \infty$.*

(i) *If $f(X)$ is nonarithmetic, then $R(x) \xrightarrow{\mathrm{d}} R_\infty$ as $x \to \infty$, with*

$$\mathbb{P}\bigl(R_\infty \leqslant y\bigr) = \frac{1}{\mu}\int_0^y \mathbb{P}\bigl(f(X) > s\bigr)\,\mathrm{d}s, \qquad y \geqslant 0. \tag{3.32}$$

(ii) *If $f(X)$ has span $d > 0$, then $R(x) \xrightarrow{\mathrm{d}} R_\infty$ as $x \to \infty$ with $x \in d\mathbb{Z}$, with*

$$\mathbb{P}\big(R_\infty = kd\big) = \frac{d}{\mu}\,\mathbb{P}\big(f(X) \geqslant kd\big), \qquad k \geqslant 1. \tag{3.33}$$

□

This classical result may be combined with Theorem 3.11 as follows.

Theorem 3.13. *Suppose in addition to the assumptions of Theorem 3.11 that $f(X) \geqslant 0$ a.s. Let R_∞ be as in Proposition 3.12.*

(i) *If $f(X)$ is nonarithmetic, then* (3.27) *and $R(x) \xrightarrow{\mathrm{d}} R_\infty$ hold jointly, with independent limits, as $x \to \infty$.*

(ii) *If $f(X)$ has span $d > 0$, then* (3.27) *and $R(x) \xrightarrow{\mathrm{d}} R_\infty$ hold jointly, with independent limits, as $x \to \infty$ with $x \in d\mathbb{Z}$.*

(iii) *If $f(X)$ is integer-valued, then for every fixed integer $k \geqslant 1$,* (3.27) *holds also conditioned on $R(x) = k$, for integers $x = n \to \infty$. Moreover,* (3.27) *holds also conditioned on $U_{N_-(x)} = x$, as $x = n \to \infty$. (We consider only x such that we condition on an event of positive probability.)*

Note that in (iii), the event $U_{N_-(x)} = x$ holds if and only if some partial sum $U_n := \sum_1^n f(X_i) = x$.

Remark 3.14. If $d = \tilde{d} = 1$, (3.28) reduces to $\gamma^2 = \mu^{-3}\operatorname{Var}\big(\mu\tilde{f}(X) - \tilde{\mu}f(X)\big)$, as shown in [14, Theorem 3].

3.3 Moment convergence

In Corollary 3.5, we have convergence of the second moment in (3.13), and trivially also of the first moment. We have also convergence of higher moments, provided we assume the corresponding integrability of f.

Theorem 3.15. *Suppose that $f(X_1, \dots, X_d) \in L^p$ with $p \geqslant 2$. Then,* (3.13) *holds with convergence of all moments and absolute moments of order $\leqslant p$.*

For moment convergence in the renewal theory theorems, we assume for simplicity that f and $\tilde{f}$ have finite moments of all orders; see also Remark 6.1. (For the case $d = \tilde{d} = 1$, see e.g. [19], [14], and [12, Section 3.8 and Theorem 4.2.3].)

Theorem 3.16. *Suppose that $f(X_1, \dots, X_d) \in L^p$ for every $p < \infty$, and that $\mu > 0$. Then,* (3.19) *and* (3.20) *hold with convergence of all moments and absolute moments. In particular, as $x \to \infty$,*

$$\mathbb{E}\,N_\pm(x) \sim \Big(\frac{d!}{\mu}\Big)^{1/d} x^{1/d}, \tag{3.34}$$

$$\operatorname{Var} N_\pm(x) \sim \big(d!/\mu\big)^{2+1/d} d^{-2}\sigma^2 x^{1/d}. \tag{3.35}$$

Theorem 3.17. *Suppose that $f(X_1, \dots, X_d) \in L^p$ and $\tilde{f}(X_1, \dots, X_{\tilde{d}}) \in L^p$ for every $p < \infty$, and that $\mu > 0$. Then,* (3.21) *and* (3.22) *hold with convergence of all moments and absolute moments. In particular, as $x \to \infty$,*

$$\mathbb{E}\,\widetilde{U}_{N_\pm(x)} \sim \frac{\tilde{\mu}}{\tilde{d}!}\Big(\frac{d!}{\mu}\Big)^{\tilde{d}/d} x^{\tilde{d}/d}, \tag{3.36}$$

$$\operatorname{Var} \widetilde{U}_{N_\pm(x)} \sim \gamma^2 x^{(2\tilde{d}-1)/d}. \tag{3.37}$$

Theorem 3.18. *Let $d = 1$. Suppose that $f(X) \in L^p$ and $\tilde{f}(X_1, \dots, X_{\tilde{d}}) \in L^p$ for every $p < \infty$, and that $\mu > 0$.*

(i) *Then,* (3.27) *holds with convergence of all moments and absolute moments.*

(ii) *If furthermore $f(X)$ is integer-valued and $f(X) \geqslant 0$, then* (i) *holds also conditioned on $R(x) = k$ or on $U_{N_-(x)} = x$ as in Theorem 3.13*(iii).

4 Proofs

4.1 Limit theorems

The method used by Hoeffding [16] and many later papers is a decomposition, which in the asymmetric case is as follows. Assume that $f(X_1, \dots, X_d) \in L^2$. Define μ and f_i by (2.1) and (3.3), and

$$f_*(x_1, \dots, x_d) := f(x_1, \dots, x_d) - \mu - \sum_{j=1}^{d} f_j(x_j). \tag{4.1}$$

Then, by the definition (1.1),

$$\begin{aligned} U_n(f) &= \binom{n}{d}\mu + \sum_{j=1}^{d} \sum_{1 \leqslant i_1 < \dots < i_d \leqslant n} f_j\bigl(X_{i_j}\bigr) + U_n(f_*) \\ &= \binom{n}{d}\mu + \sum_{j=1}^{d}\sum_{i=1}^{n} \binom{i-1}{j-1}\binom{n-i}{d-j} f_j\bigl(X_i\bigr) + U_n(f_*). \end{aligned} \tag{4.2}$$

We consider the three terms in (4.2) separately. The first is a constant, and we shall see that the third term is negligible, so the main term is the second term.

Remark 4.1. The decomposition (4.2) may be continued to higher terms by expanding f_* further, see e.g. [16] for the symmetric case and [20, Chapter 11.2] in general; this is important when treating degenerate cases, see Remark 6.3, but for our purposes we have no need of this.

For the second term, we define for convenience, for $1 \leqslant j \leqslant d$ and $n \geqslant 1$,

$$a_{n,j}(i) := \binom{i-1}{j-1}\binom{n-i}{d-j}, \qquad 1 \leqslant i \leqslant n, \tag{4.3}$$

$$\Delta a_{n,j}(i) := a_{n,j}(i+1) - a_{n,j}(i), \qquad 1 \leqslant i < n. \tag{4.4}$$

Recall $\psi(s,t)$ defined in (3.5), and let $\psi'(s,t)$ denote $\frac{\partial}{\partial s}\psi(s,t)$.

Lemma 4.2. *Uniformly for all n, j, i such that the variables are defined,*

$$a_{n,j}(i) = \psi_j(i,n) + O\bigl(n^{d-2}\bigr), \tag{4.5}$$

$$\Delta a_{n,j}(i) = \psi'_j(i,n) + O\bigl(n^{d-3}\bigr). \tag{4.6}$$

In particular, $a_{n,j}(i) = O\bigl(n^{d-1}\bigr)$ and $\Delta a_{n,j}(i) = O\bigl(n^{d-2}\bigr)$. Furthermore (for $d \leqslant 2$), any error term $O(n^{-1})$ or $O(n^{-2})$ here vanishes identically.

Proof. By (4.3), for $1 \leqslant i \leqslant n$,

$$\begin{aligned} a_{n,j} &= \frac{i^{j-1} + O(n^{j-2})}{(j-1)!} \cdot \frac{(n-i)^{d-j} + O(n^{d-j-1})}{(d-j)!} \\ &= \frac{i^{j-1}(n-i)^{d-j}}{(j-1)!\,(d-j)!} + O\bigl(n^{d-2}\bigr), \end{aligned} \tag{4.7}$$

which is (4.5). Similarly, for $1 \leqslant i < n$, with $\binom{k}{-1} = 0$,

$$\begin{aligned}\Delta a_{n,j} &= \left(\binom{i}{j-1} - \binom{i-1}{j-1}\right)\binom{n-i}{d-j} \\ &\qquad + \binom{i}{j-1}\left(\binom{n-i-1}{d-j} - \binom{n-i}{d-j}\right) \\ &= \binom{i-1}{j-2}\binom{n-i}{d-j} - \binom{i}{j-1}\binom{n-i-1}{d-j-1} \\ &= \frac{(j-1)i^{j-2}(n-i)^{d-j} - (d-j)i^{j-1}(n-i)^{d-j-1} + O\bigl(n^{d-3}\bigr)}{(j-1)!\,(d-j)!} \\ &= \psi_j'(i,n) + O\bigl(n^{d-3}\bigr).\end{aligned} \tag{4.8}$$

□

We now take care of the second term in (4.2).

Lemma 4.3. *Let*

$$\widehat{U}_{n,j} := \sum_{i=1}^{n} \binom{i-1}{j-1}\binom{n-i}{d-j} f_j\bigl(X_i\bigr) = \sum_{i=1}^{n} a_{n,j}(i) f_j\bigl(X_i\bigr). \tag{4.9}$$

Then, as $n \to \infty$, with W_j as in (3.4)*,*

$$n^{-(d-1/2)}\widehat{U}_{nt,j} \xrightarrow{\mathrm{d}} \int_0^t \psi_j(u,t)\,\mathrm{d}W_j(u), \qquad t \geqslant 0, \tag{4.10}$$

in $D[0,\infty)$, jointly for $j = 1,\dots,d$.

Proof. Let for any function $g : S \to \mathbb{R}$,

$$S_n(g) := U_n(g) := \sum_{i=1}^{n} g(X_i). \tag{4.11}$$

Then, by (4.11), (4.4), and a summation by parts,

$$\begin{aligned}\widehat{U}_{n,j} &= \sum_{i=1}^{n} a_{n,j}(i) f_j(X_i) = \sum_{i=1}^{n} a_{n,j}(i)\bigl(S_i(f_j) - S_{i-1}(f_j)\bigr) \\ &= a_{n,j}(n)S_n(f_j) - \sum_{i=1}^{n-1} \Delta a_{n,j}(i) S_i(f_j).\end{aligned} \tag{4.12}$$

By (3.3), $\mathbb{E}\, f_j(X) = 0$, and furthermore $f_j(X) \in L^2$. Hence, by Donsker's theorem,

$$n^{-1/2} S_{nt}(f_j) \xrightarrow{\mathrm{d}} W_j(t), \tag{4.13}$$

in $D[0,\infty)$, jointly for $j = 1,\dots,d$, where W_j are continuous centered Gaussian processes as in (3.4).

By the Skorohod coupling theorem [25, Theorem 4.30], we may assume that the convergence in (4.13) holds a.s. The convergence is in $D[0,\infty)$, which since the limit is continuous means uniform convergence of any compact interval (cf. [2, Section 14]); thus as $n \to \infty$,

$$n^{-1/2} S_{nt}(f_j) = W_j(t) + o_{\mathrm{a.s.}}(1), \tag{4.14}$$

uniformly for $t \in [0,T]$ and all j, for every fixed $T < \infty$. (Note that the error term here, $R_{n,j,t}$ say, is random; the uniformity means that $\sup_{j \leqslant d,\, t \leqslant T} |R_{n,j,t}| \xrightarrow{\mathrm{a.s.}} 0$ for every T.)

Fix T, and let $m = ns$ with $s \leqslant T$. Then, by (4.12), (4.14) and Lemma 4.2, uniformly for $s \in [0,T]$,

$$\begin{aligned} n^{-1/2}\widehat{U}_{m,j} &= a_{m,j}(m)W_j(s) - \sum_{i=1}^{m-1} \Delta a_{m,j}(i)W_j(i/n) + o_{\text{a.s.}}\bigl(n^{d-1}\bigr) \\ &= \psi_j(m,m)W_j(s) - \sum_{i=1}^{m-1} \psi_j'(i,m)W_j(i/n) + o_{\text{a.s.}}\bigl(n^{d-1}\bigr). \end{aligned} \tag{4.15}$$

Furthermore, since W_j is bounded and uniformly continuous on $[0,T]$, and $\psi_j'(s,t) = O(t^{d-2})$, $\psi_j''(s,t) = O(t^{d-3})$ for $0 \leqslant s \leqslant t$,

$$\begin{aligned} \sum_{i=1}^{m-1} \psi_j'(i,m)W_j(i/n) &= \int_0^m \psi_j'(x,m)W_j(x/n)\,\mathrm{d}x + o\bigl(m^{d-1}\bigr) + O(1) \\ &= n\int_0^s \psi_j'(nu,ns)W_j(u)\,\mathrm{d}u + o\bigl(n^{d-1}\bigr) \\ &= n^{d-1}\int_0^s \psi_j'(u,s)W_j(u)\,\mathrm{d}u + o\bigl(n^{d-1}\bigr). \end{aligned} \tag{4.16}$$

An integration by parts yields (with stochastic integrals), since $W_j(0) = 0$,

$$\int_0^s \psi_j'(u,s)W_j(u)\,\mathrm{d}u = \psi_j(s,s)W_j(s) - \int_0^s \psi_j(u,s)\,\mathrm{d}W_j(u) \tag{4.17}$$

and combining (4.15), (4.16) and (4.17) yields, using $\psi_j(m,m) = n^{d-1}\psi_j(s,s)$,

$$n^{-1/2}\widehat{U}_{ns,j} = n^{-1/2}\widehat{U}_{m,j} = n^{d-1}\int_0^s \psi_j(u,s)\,\mathrm{d}W_j(u) + o_{\text{a.s.}}\bigl(n^{d-1}\bigr), \tag{4.18}$$

uniformly for $0 \leqslant s \leqslant T$. Since T is arbitrary, this yields (4.10), jointly for all j. ☐

To show that the final term in (4.2) is negligible, we give another lemma. Cf. [32] for similar results in the symmetric case.

Lemma 4.4. *Suppose that $f(X_1,\dots,X_d) \in L^2$.*

(i) *Then*

$$\mathbb{E}\,|U_n^*(f-\mu)|^2 \leqslant Cn^{2d-1}\|f\|_2^2. \tag{4.19}$$

(ii) *If furthermore $f_i = 0$ for $i = 1,\dots,d$, then*

$$\mathbb{E}\,|U_n^*(f-\mu)|^2 \leqslant Cn^{2d-2}\|f\|_2^2. \tag{4.20}$$

Proof. (i): We introduce another decomposition of f and U_n, which unlike the one in (3.3)–(4.2) focusses on the order of the arguments. Let $\widehat{F}_0 := \mu$ and, for $1 \leqslant k \leqslant d$,

$$\widehat{F}_k(x_1,\dots,x_k) := \mathbb{E}\,f(x_1,\dots,x_k,X_{k+1},\dots,X_d), \tag{4.21}$$

$$F_k(x_1,\dots,x_k) := \widehat{F}_k(x_1,\dots,x_k) - \widehat{F}_{k-1}(x_1,\dots,x_{k-1}). \tag{4.22}$$

In other words, $\widehat{F}_k(X_1,\dots,X_k) := \mathbb{E}\bigl(f(X_1,\dots,X_d) \mid X_1,\dots,X_k\bigr)$, and thus $\widehat{F}_k(X_1,\dots,X_k)$, $k = 0,\dots,d$, is a martingale, with the martingale differences $F_k(X_1,\dots,X_k)$, $k = 1,\dots,d$. Hence,

$$\mathbb{E}\,F_k(x_1,\dots,x_{k-1},X_k) = 0. \tag{4.23}$$

By (4.21)–(4.22), $f(x_1,\dots,x_d)-\mu=\sum_{k=1}^d F_k(x_1,\dots,x_k)$, and thus

$$\begin{aligned} U_n(f-\mu) &= \sum_{k=1}^d \sum_{i_1<\dots<i_k\leqslant n} \binom{n-i_k}{d-k} F_k(X_{i_1},\dots,X_{i_k}) \\ &= \sum_{k=1}^d \sum_{i=1}^n \binom{n-i}{d-k}\bigl(U_i(F_k)-U_{i-1}(F_k)\bigr) \\ &= U_n(F_d)+\sum_{k=1}^{d-1}\sum_{i=1}^{n-1}\binom{n-i-1}{d-k-1}U_i(F_k), \end{aligned} \tag{4.24}$$

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

$$\begin{aligned} |U_n(f-\mu)| &\leqslant |U_n(F_d)|+\sum_{k=1}^{d-1}\sum_{i=1}^{n-1}\binom{n-i-1}{d-k-1}U_n^*(F_k) \\ &= |U_n(F_d)|+\sum_{k=1}^{d-1}\binom{n-1}{d-k}U_n^*(F_k)\leqslant \sum_{k=1}^d n^{d-k}U_n^*(F_k). \end{aligned} \tag{4.25}$$

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

$$U_n^*(f-\mu)\leqslant \sum_{k=1}^d n^{d-k}U_n^*(F_k). \tag{4.26}$$

By the definition (1.1), $\Delta U_n(F_k) := U_n(F_k)-U_{n-1}(F_k)$ is a sum of $\binom{n-1}{k-1}$ terms $F_k(X_{i_1},\dots,X_{i_{k-1}},X_n)$ that all have the same distribution, and thus by Minkowski's inequality,

$$\mathbb{E}\,|\Delta U_n(F_k)|^2=\|\Delta U_n(F_k)\|_2^2\leqslant \binom{n-1}{k-1}^2\|F_k\|_2^2\leqslant n^{2k-2}\|f\|_2^2. \tag{4.27}$$

Furthermore, it follows from (4.23) that $\mathbb{E}\bigl(U_n(F_k)-U_{n-1}(F_k)\mid \mathcal{F}_{n-1}\bigr)=0$, and thus $U_n(F_k)$, $n\geqslant 0$, is a martingale. Consequently, using (4.27),

$$\mathbb{E}\,|U_n(F_k)|^2=\sum_{i=1}^n \mathbb{E}\,|\Delta U_i(F_k)|^2\leqslant n^{2k-1}\|f\|_2^2 \tag{4.28}$$

and Doob's inequality yields

$$\|U_n^*(F_k)\|_2\leqslant C\|U_n(F_k)\|_2\leqslant Cn^{k-1/2}\|f\|_2. \tag{4.29}$$

Finally, (4.26), (4.29) and Minkowski's inequality yield

$$\|U_n^*(f-\mu)\|_2\leqslant \sum_{k=1}^d n^{d-k}\|U_n^*(F_k)\|_2\leqslant Cn^{d-1/2}\|f\|_2, \tag{4.30}$$

which yields (4.19) by squaring.

(ii): By (4.21)–(4.22) and (3.3),

$$\mathbb{E}\bigl(F_k(X_1,\dots,X_k)\mid X_k\bigr)=\mathbb{E}\bigl(f(X_1,\dots,X_d)\mid X_k\bigr)-\mathbb{E}\,f=f_k(X_k). \tag{4.31}$$

Hence, assuming $f_k=0$,

$$\mathbb{E}\bigl(F_k(X_1,\dots,X_k)\mid X_k\bigr)=0. \tag{4.32}$$

It was seen in the proof of (i) that $\Delta U_n(F_k)$ is a sum of $\binom{n-1}{k-1}$ terms $F_k(X_{i_1},\dots,X_{i_{k-1}},X_n)$. It now follows from (4.32) that if $\{i_1,\dots,i_{k-1}\}$ and $\{j_1,\dots,j_{k-1}\}$ are two disjoint sets of indices, then, by first conditioning on X_n,

$$\mathbb{E}\bigl(F_k(X_{i_1},\dots,X_{i_{k-1}},X_n)F_k(X_{j_1},\dots,X_{j_{k-1}},X_n)\bigr)=0. \tag{4.33}$$

Hence, only the $O\bigl(n^{2k-3}\bigr)$ pairs of index sets $\{i_1,\dots,i_{k-1}\}$ and $\{j_1,\dots,j_{k-1}\}$ with at least one common element contribute to $\mathbb{E}\bigl(\Delta U_n(F_k)\bigr)^2$, and we obtain, for $1\leqslant k\leqslant d$, that (4.27) is improved to

$$\mathbb{E}\,|\Delta U_n(F_k)|^2\leqslant Cn^{2k-3}\|f\|_2^2. \tag{4.34}$$

(For $k=1$, $F_1=f_1=0$, and (4.34) still holds.) The result now follows as in (i), see (4.28)–(4.30), by (4.34), Doob's inequality, (4.26) and Minkowski's inequality. ☐

Proof of Theorem 3.2. We use the decomposition (4.2), with n replaced by $\lfloor nt\rfloor$. For the constant term, note that $\binom{\lfloor nt\rfloor}{d}\mu=n^dt^d\mu/d!+O\bigl(n^{d-1}\bigr)$ when $t=O(1)$.

The second term in (4.2) is $\sum_{j=1}^{d}\widehat{U}_{nt,j}$, using the notation in (4.9), and we use Lemma 4.3; (4.10) shows that this term divided by $n^{d-1/2}$ converges in $D[0,\infty)$ to Z_t defined in (3.7).

For the third term, we apply Lemma 4.4 to f_*. It follows from the definition (4.1) that $\mu_*:=\mathbb{E}\,f_*(X_1,\dots,X_d)=0$ and that, applying (3.3) to f_*, $(f_*)_i=0$ for every $i\leqslant d$. Hence, Lemma 4.4(ii) applies to f_* and yields

$$\mathbb{E}\,|U_n^*(f_*)|^2\leqslant Cn^{2d-2}\|f_*\|_2^2\leqslant Cn^{2d-2}\|f\|_2^2. \tag{4.35}$$

Let $T>0$ be fixed. Applying (4.35) to nT, we see in particular that $n^{-(d-1/2)}U_{nt}(f_*)\overset{\mathrm{p}}{\longrightarrow}0$ uniformly on $[0,T]$.

Consequently, (3.6) follows from (4.2).

Joint convergence for several functions $f^{(k)}$, with limits given by (3.9), follows by the same proof, using joint convergence for all $f_i^{(k)}$ in (4.13). ☐

Proof of Theorem 3.1. We do this in several steps.

Step 1. First, suppose that $f(X_1,\dots,X_d)\in L^2$. We may assume $\mu=0$, and then Lemma 4.4(i) implies, for any $N\geqslant 1$,

$$\mathbb{E}\sup_{N\leqslant n\leqslant 2N}\bigl(|U_n|/n^d\bigr)^2\leqslant N^{-2d}\,\mathbb{E}(U_{2N}^*)^2\leqslant CN^{-1}\|f\|_2^2. \tag{4.36}$$

Summing over all $N=2^m$, $m=0,1,\dots$, we find

$$\mathbb{E}\sum_{m=0}^{\infty}\sup_{2^m\leqslant n\leqslant 2^{m+1}}\bigl(|U_n|/n^d\bigr)^2<\infty. \tag{4.37}$$

Hence, a.s. the terms in the sum tend to 0, which implies $U_n/n^d\to 0$ and thus $U_n/\binom{n}{d}\to 0=\mu$. This proves (3.1) for $f\in L^2$.

Step 2. Assume now $f\in L^1$ and $f\geqslant 0$. Define the truncation $f_M:=f\wedge M$. Then $f_M\in L^2$ and Step 1 shows that for every $M<\infty$, a.s.,

$$\liminf_{n\to\infty}\frac{U_n(f)}{\binom{n}{d}}\geqslant\liminf_{n\to\infty}\frac{U_n(f_M)}{\binom{n}{d}}=\mathbb{E}\,f_M(X_1,\dots,X_d). \tag{4.38}$$

Letting $M\to\infty$ yields $\liminf_{n\to\infty}U_n(f)/\binom{n}{d}\geqslant\mu$ a.s.

Step 3. Continue to assume $f\in L^1$ and $f\geqslant 0$. For every permutation $\pi\in\mathfrak{S}_d$, let $f_\pi(X_1,\dots,X_d):=f(X_{\pi(1)},\dots,X_{\pi(d)})$, and let $F:=\sum_{\pi\in\mathfrak{S}}f_\pi$ and $g:=F-f=\sum_{\pi\neq id}f_\pi$.

Note that $f, g \in L^1$ with $f, g \geqslant 0$; thus Step 2 applies to both f and g. Furthermore, $F = f + g$ is symmetric, so we have $U_n(F)/\binom{n}{d} \xrightarrow{\text{a.s.}} \mathbb{E}\, F := \mathbb{E}\, F(X_1, \dots, X_d)$ by the theorem by Hoeffding [17] for the symmetric case. (This case has a simple reverse martingale proof, see Remark 6.6.) Consequently, a.s.,

$$\limsup_{n\to\infty} \frac{U_n(f)}{\binom{n}{d}} = \lim_{n\to\infty} \frac{U_n(F)}{\binom{n}{d}} - \liminf_{n\to\infty} \frac{U_n(g)}{\binom{n}{d}} \leqslant \mathbb{E}\, F - \mathbb{E}\, g = \mu. \tag{4.39}$$

Combined with Step 2, this shows (3.1) for every $f \in L^1$ with $f \geqslant 0$.

Step 4. The general case follows by linearity. □

We used for convenience the known symmetric case in this proof. An alternative would be to use suitable truncations, similarly to the original proof of the symmetric case by Hoeffding [17].

Lemma 4.5. *Suppose that $f(X_1, \dots, X_d) \in L^2$. Then, as $n \to \infty$, with Z_1 defined by* (3.7)*,*

$$\begin{aligned} \frac{\operatorname{Var} U_n}{n^{2d-1}} &\to \sigma^2 := \operatorname{Var} Z_1 \\ &= \sum_{i,j=1}^{d} \frac{(i+j-2)!\,(2d-i-j)!}{(i-1)!\,(j-1)!\,(d-i)!\,(d-j)!\,(2d-1)!}\sigma_{ij}. \end{aligned} \tag{4.40}$$

Proof. We may assume $\mu = 0$. Then

$$\operatorname{Var} U_n = \mathbb{E}\, U_n^2 = \sum_{i_1<\dots<i_d} \sum_{j_1<\dots<j_d} \mathbb{E}\big(f(X_{i_1}, \dots, X_{i_d}) f(X_{j_1}, \dots, X_{j_d})\big), \tag{4.41}$$

where all terms with $\{i_1, \dots, i_d\} \cap \{j_1, \dots, j_d\} = \emptyset$ vanish. There are only $O\big(n^{2d-2}\big)$ terms with $|\{i_1, \dots, i_d\} \cap \{j_1, \dots, j_d\}| \geqslant 2$, so we concentrate on the case when, say, $i_k = j_\ell = i$, and all other indices are distinct. Thus, using (3.3) and the notation (4.3) together with (3.2) and Lemma 4.2,

$$\begin{aligned} \mathbb{E}\, U_n^2 &= \sum_{k=1}^{d} \sum_{\ell=1}^{d} \sum_{i=1}^{n} a_{n,k}(i) a_{n,\ell}(i)\, \mathbb{E}\big(f_k(X_i) f_\ell(X_i)\big) + O\big(n^{2d-2}\big) \\ &= \sum_{k=1}^{d} \sum_{\ell=1}^{d} \sum_{i=1}^{n} \psi_k(i,n)\psi_\ell(i,n)\sigma_{k\ell} + O\big(n^{2d-2}\big) \\ &= \sum_{k=1}^{d} \sum_{\ell=1}^{d} \sigma_{k\ell} \int_0^n \psi_k(x,n)\psi_\ell(x,n)\,\mathrm{d}x + O\big(n^{2d-2}\big) \\ &= n^{2d-1} \sum_{k=1}^{d} \sum_{\ell=1}^{d} \sigma_{k\ell} \int_0^1 \psi_k(u,1)\psi_\ell(u,1)\,\mathrm{d}u + O\big(n^{2d-2}\big). \end{aligned} \tag{4.42}$$

Consequently, by (3.8),

$$\frac{\operatorname{Var} U_n}{n^{2d-1}} \to \sum_{k=1}^{d} \sum_{\ell=1}^{d} \sigma_{k\ell} \int_0^1 \psi_k(u,1)\psi_\ell(u,1)\,\mathrm{d}u = \operatorname{Var}(Z_1). \tag{4.43}$$

Furthermore, this equals the sum in (4.40), as is seen by taking $s = t = 1$ in (3.8) and evaluating the resulting Beta integral. □

Remark 4.6. Similarly, it follows more generally that $\operatorname{Cov}\big(U_{ns}, U_{nt}\big)/n^{2d-1} \to \operatorname{Cov}(Z_s, Z_t)$ given by (3.8), for any fixed $s, t \geqslant 0$. In other words, (3.6) holds with convergence of second moments.

Proof of Corollary 3.5. The functional limit (3.6) implies, since Z_t is continuous, convergence (in distribution) for each fixed $t \geqslant 0$. Taking $t = 1$ we obtain (3.13) with $\sigma^2 = \operatorname{Var} Z_1$, which is evaluated by Lemma 4.5.

By (3.8) and (3.2),

$$\begin{aligned}\operatorname{Var}(Z_1) &= \sum_{i,j=1}^{d} \operatorname{Cov}\bigl(f_i(X), f_j(X)\bigr) \int_0^1 \psi_i(s,1)\psi_j(s,1)\,\mathrm{d}s \\ &= \int_0^1 \operatorname{Var}\Bigl(\sum_{i=1}^{d} \psi_i(s,1) f_i(X)\Bigr)\,\mathrm{d}s\end{aligned} \tag{4.44}$$

Hence, $\sigma^2 = 0 \iff \sum_{i=1}^d \psi_i(s,1) f_i(X) = 0$ a.s. for (almost) every $s \in [0,1]$, which is equivalent to $f_i(X) = 0$ a.s. for every i since the polynomials $\psi_i(s,1)$ are linearly independent. □

4.2 Renewal theory

Proof of Theorem 3.7. Consider first N_-. Note that Theorem 3.1 and $\mu > 0$ imply $U_n \to \infty$ a.s., and then $N_-(x) < \infty$ for every x.

Furthermore, it is trivial that $N_-(x) \to \infty$ as $x \to \infty$. Thus we may substitute $n = N_-(x)$ in (3.1) and obtain

$$\frac{U_{N_-(x)}}{N_-(x)^d} = \frac{U_{N_-(x)}}{\binom{N_-(x)}{d}} \cdot \frac{\binom{N_-(x)}{d}}{N_-(x)^d} \xrightarrow{\text{a.s.}} \frac{\mu}{d!} \qquad \text{as } x \to \infty. \tag{4.45}$$

Furthermore, we also have, again by (3.1),

$$\frac{U_{N_-(x)+1}}{N_-(x)^d} = \frac{U_{N_-(x)+1}}{\binom{N_-(x)+1}{d}} \cdot \frac{\binom{N_-(x)+1}{d}}{N_-(x)^d} \xrightarrow{\text{a.s.}} \frac{\mu}{d!}. \tag{4.46}$$

By the definition of $N_-(x)$, $U_{N_-(x)} \leqslant x < U_{N_-(x)+1}$, and thus (4.45)–(4.46) imply

$$\frac{x}{N_-(x)^d} \xrightarrow{\text{a.s.}} \frac{\mu}{d!} \qquad \text{as } x \to \infty. \tag{4.47}$$

which is equivalent to (3.19) for N_-.

The proof for N_+ is the same, using $U_{N_+(x)-1} \leqslant x < U_{N_+(x)}$. □

Proof of Theorem 3.8. Again, we consider N_-; the argument for N_+ is the same. Let

$$n(x) := (d!/\mu)^{1/d} x^{1/d}, \tag{4.48}$$

$$T(x) := N_-(x)/\lfloor n(x) \rfloor. \tag{4.49}$$

As $x \to \infty$, $n(x) \to \infty$ and thus (3.6) implies

$$\frac{U_{\lfloor n(x) \rfloor t} - (\lfloor n(x) \rfloor t)^d \mu/d!}{n(x)^{d-1/2}} \xrightarrow{\text{d}} Z_t \qquad \text{in } D[0,\infty). \tag{4.50}$$

Furthermore, (3.19) implies

$$T(x) \to 1 \tag{4.51}$$

a.s., and thus in probability. Hence, (4.50) and (4.51) hold jointly in distribution [2, Theorem 4.4]. Now, $(F,t) \mapsto F(t)$ is a measurable mapping $D[0,\infty) \times [0,\infty) \to \mathbb{R}$ that is continuous at every (F,t) with F continuous. Hence, by [2, Theorem 5.1], it follows from the joint convergence in (4.50) and (4.51), together with continuity of Z_t, that we may substitute $t = T(x)$ in (4.50) and obtain, as $x \to \infty$,

$$\frac{U_{N_-(x)} - N_-(x)^d \mu/d!}{n(x)^{d-1/2}} \xrightarrow{\mathrm{d}} Z_1. \tag{4.52}$$

Taking instead $t = T_1(x) := (N_-(x)+1)/\lfloor n(x) \rfloor$, we similarly obtain

$$\frac{U_{N_-(x)+1} - N_-(x)^d \mu/d!}{n(x)^{d-1/2}} = \frac{U_{N_-(x)+1} - (N_-(x)+1)^d \mu/d! + O\big(N_-(x)^{d-1}+1\big)}{n(x)^{d-1/2}} \xrightarrow{\mathrm{d}} Z_1, \tag{4.53}$$

using $(N_-(x)^{d-1}+1)/n(x)^{d-1/2} \xrightarrow{\mathrm{p}} 0$ by (3.19) and (4.48). Since $U_{N_-(x)} \leqslant x < U_{N_-(x)+1}$, (4.52) and (4.53) together imply, as $x \to \infty$,

$$\frac{x - N_-(x)^d \mu/d!}{n(x)^{d-1/2}} \xrightarrow{\mathrm{d}} Z_1. \tag{4.54}$$

Hence, recalling (4.48),

$$\frac{x}{n(x)^{d-1/2}} \left(\left(\frac{N_-(x)}{n(x)} \right)^d - 1 \right) = \frac{N_-(x)^d \mu/d! - x}{n(x)^{d-1/2}} \xrightarrow{\mathrm{d}} -Z_1. \tag{4.55}$$

Furthermore, letting $T_2(x) := N_-(x)/n(x)$, we have $T_2(x) \xrightarrow{\mathrm{a.s.}} 1$ by (3.19), and thus, interpreting the quotients as d when $T_2(x) = 1$,

$$\frac{\big(N_-(x)/n(x)\big)^d - 1}{\big(N_-(x)/n(x)\big) - 1} = \frac{T_2(x)^d - 1}{T_2(x) - 1} \xrightarrow{\mathrm{a.s.}} d. \tag{4.56}$$

Dividing (4.55) by (4.56) yields

$$\frac{x}{n(x)^{d-1/2}} \left(\frac{N_-(x)}{n(x)} - 1 \right) \xrightarrow{\mathrm{d}} -\frac{1}{d} Z_1. \tag{4.57}$$

Since

$$\frac{N_-(x) - n(x)}{x^{1/2d}} = \left(\frac{n(x)}{x^{1/d}} \right)^{d+1/2} \frac{x}{n(x)^{d-1/2}} \left(\frac{N_-(x)}{n(x)} - 1 \right), \tag{4.58}$$

(4.57) and (4.48) imply

$$\frac{N_-(x) - n(x)}{x^{1/2d}} \xrightarrow{\mathrm{d}} -\left(\frac{d!}{\mu} \right)^{1+1/2d} d^{-1} Z_1, \tag{4.59}$$

which yields (3.20), since $Z_1 \sim N(0, \sigma^2)$ by Lemma 4.5. □

Proof of Theorem 3.9. (i): By Theorem 3.1 for $\tilde{f}$ and (3.19),

$$\frac{\widetilde{U}_{N_\pm(x)}}{x^{\tilde{d}/d}} = \frac{\widetilde{U}_{N_\pm(x)}}{N_\pm(x)^{\tilde{d}}} \frac{N_\pm(x)^{\tilde{d}}}{x^{\tilde{d}/d}} \xrightarrow{\mathrm{a.s.}} \frac{\tilde{\mu}}{\tilde{d}!} \Big(\frac{d!}{\mu} \Big)^{\tilde{d}/d}. \tag{4.60}$$

(ii): Define again $n(x)$ and $T(x)$ by (4.48)–(4.49). We have joint convergence in (3.6) for f and $\tilde{f}$, and thus, as $x \to \infty$, (4.50) holds jointly with

$$\frac{\widetilde{U}_{\lfloor n(x) \rfloor t} - (\lfloor n(x) \rfloor t)^{\tilde{d}} \tilde{\mu}/\tilde{d}!}{n(x)^{\tilde{d}-1/2}} \xrightarrow{\mathrm{d}} \tilde{Z}_t \qquad \text{in } D[0,\infty). \tag{4.61}$$

By (4.51) and the argument in the proof of Theorem 3.8, now using the mapping $(F,\tilde F,t)\mapsto(F(t),\tilde F(t))$ that maps $D[0,\infty)\times D[0,\infty)\times[0,\infty)\to\mathbb{R}^2$, it follows that (4.52) holds jointly with

$$\frac{\widetilde U_{N_-(x)}-N_-(x)^{\tilde d}\tilde\mu/\tilde d!}{n(x)^{\tilde d-1/2}}\xrightarrow{\mathrm{d}}\tilde Z_1. \tag{4.62}$$

Furthermore, (4.52) and (4.54) together with $U_{N_-(x)}\leqslant x$ imply

$$\frac{x-U_{N_-(x)}}{n(x)^{d-1/2}}\xrightarrow{\mathrm{p}}0. \tag{4.63}$$

Consequently, (4.54) and (4.62) hold jointly. The argument in the proof of Theorem 3.8 now holds with every convergence in distribution holding jointly with (4.62). Hence, (4.62) holds jointly with (4.59), which implies, see (4.56) and (4.48),

$$\frac{N_-(x)^{\tilde d}-n(x)^{\tilde d}}{n(x)^{\tilde d-1/2}}=\frac{\big(N_-(x)/n(x)\big)^{\tilde d}-1}{\big(N_-(x)/n(x)\big)-1}\,\frac{N_-(x)-n(x)}{n(x)^{1/2}}\xrightarrow{\mathrm{d}}-\tilde d\,\frac{d!}{\mu}\,d^{-1}Z_1. \tag{4.64}$$

Consequently, (4.62) and (4.64) hold jointly, and thus

$$\frac{\widetilde U_{N_-(x)}-n(x)^{\tilde d}\tilde\mu/\tilde d!}{n(x)^{\tilde d-1/2}}\xrightarrow{\mathrm{d}}\widehat Z:=\tilde Z_1-\frac{(d-1)!\,\tilde\mu}{(\tilde d-1)!\,\mu}Z_1. \tag{4.65}$$

We obtain (3.22)–(3.23) by substituting the definition (4.48) of $n(x)$.

(iii): By (3.23), $\gamma^2=0\iff\operatorname{Var}\big(\mu(\tilde d-1)!\,\tilde Z_1-\tilde\mu(d-1)!\,Z_1\big)=0$, and arguing as in (4.44), and recalling (3.9), this is equivalent to

$$\operatorname{Var}\Big(\mu(\tilde d-1)!\sum_{i=1}^{\tilde d}\psi_{i;\tilde d}(s,1)\tilde f_i(X)-\tilde\mu(d-1)!\sum_{j=1}^{d}\psi_{j;d}(s,1)f_j(X)\Big)=0 \tag{4.66}$$

for (almost) every $s\in[0,1]$, and by the definition (3.5), this is the same as

$$\mu\sum_{i=1}^{\tilde d}\binom{\tilde d-1}{i-1}s^{i-1}(1-s)^{\tilde d-i}\tilde f_i(X)=\tilde\mu\sum_{j=1}^{d}\binom{d-1}{j-1}s^{j-1}(1-s)^{d-j}f_j(X) \tag{4.67}$$

a.s., for every s.

If $\tilde d\geqslant d$, multiply the right-hand side of (4.67) by $(s+1-s)^{\tilde d-d}=\sum_{k=0}^{\tilde d-d}\binom{\tilde d-d}{k}s^k(1-s)^{\tilde d-d-k}$, which equals 1, and identify the coefficients of $s^{i-1}(1-s)^{\tilde d-i}$ on both sides; this yields (3.24). Conversely, (3.24) implies (4.67) by the same argument.

The case $\tilde d<d$ follows by the symmetry in (4.67).

The special cases (3.25) and (3.26) are immediate consequences of (3.24). □

Proof of Theorem 3.11. Take $d=1$ in Theorem 3.9(ii). To obtain the formula (3.28) for γ^2, we use (3.23) and note first that $\operatorname{Var}(\tilde Z_1)$ is given by (3.14), mutatis mutandis, which yields the first term on the right-hand side of (3.28). Furthermore, (3.14) yields also, with $d=1$, $\operatorname{Var}(Z_1)=\sigma_{11}=\operatorname{Var}(f(X))$, yielding the third term. Finally, note that when $d=1$, (3.5) yields $\psi_{1;1}(s,t)=1$, and thus (3.7) yields $Z_t=W(t)$; consequently, using

(3.7) and (3.9) and a standard Beta integral,

$$\begin{aligned}\operatorname{Cov}\bigl(\tilde{Z}_1, Z_1\bigr) &= \sum_{j=1}^{\tilde{d}} \operatorname{Cov}\Bigl(\int_0^1 \psi_{j;\tilde{d}}(s,1)\,\mathrm{d}\widetilde{W}_j(s), \int_0^1 \mathrm{d}W(s)\Bigr) \\ &= \sum_{j=1}^{\tilde{d}} \int_0^1 \psi_{j;\tilde{d}}(s,1)\operatorname{Cov}\bigl(\tilde{f}_j(X), f(X)\bigr)\,\mathrm{d}s \\ &= \sum_{j=1}^{\tilde{d}} \frac{1}{\tilde{d}!}\operatorname{Cov}\bigl(\tilde{f}_j(X), f(X)\bigr).\end{aligned} \tag{4.68}$$

This yields the second term on the right-hand side, and completes the proof. □

Proof of Theorem 3.13. Let (for $x \geqslant 2$, say) $x_- := x - \ln x$ in the nonarithmetic case, and $x_- := d\lfloor (x - \ln x)/d \rfloor$ if $f(X)$ has span $d > 0$; also, in the latter case, consider only $x \in d\mathbb{Z}$.

First, run the process U_n until the stopping time $N_+(x_-)$. Let

$$\Delta x := x - U_{N_+(x_-)} = x - x_- - R(x_-). \tag{4.69}$$

As $x \to \infty$, $R(x_-) \xrightarrow{\mathrm{d}} R_\infty$ by Proposition 3.12, and $x - x_- \geqslant \ln x \to \infty$; hence $\Delta x \xrightarrow{\mathrm{p}} \infty$. In particular, with probability tending to 1 as $x \to \infty$, $\Delta x \geqslant 0$.

Restart the process U_n after $N_+(x_-)$ and continue until $N_+(x)$. Since $d = 1$ so (3.30) holds, and $N_+(x_-)$ is a stopping time, in this continuation $U_n - U_{N_+(x_-)}$ is independent of what happened up to $N_+(x_-)$, and thus it can be regarded as a renewal process S_n^* starting at 0 and running to $N_+(\Delta x)$; in particular, the overshoot $R^*(\Delta x)$ of this renewal process equals the overshoot $R(x)$ of the original one. Here Δx is random, but independent of the renewal process S_n^*, and since $\Delta x \xrightarrow{\mathrm{p}} \infty$, Proposition 3.12 implies that the overshoot $R(x) = R^*(\Delta x) \xrightarrow{\mathrm{d}} R_\infty$. Furthermore, this holds conditioned on any events $\mathcal{E}(x_-)$ that depend on the original process up to $N_+(x_-)$, provided $\liminf_{x\to\infty} \mathbb{P}(\mathcal{E}(x_-)) > 0$.

Denote the left-hand side of (3.27) by $\tilde{V}(x)$. By (3.27), $\tilde{V}(x_-) \xrightarrow{\mathrm{d}} N(0,\gamma^2)$ as $x \to \infty$. Fix $a, b \in \mathbb{R}$ and let $\mathcal{E}(x_-) := \{\tilde{V}(x_-) \leqslant a\}$. It then follows from the argument above that, as $x \to \infty$,

$$\begin{aligned}\mathbb{P}\bigl(\tilde{V}(x_-) \leqslant a,\, R(x) \leqslant b\bigr) &= \mathbb{P}\bigl(R(x) \leqslant b \mid \tilde{V}(x_-) \leqslant a\bigr)\,\mathbb{P}\bigl(\tilde{V}(x_-) \leqslant a\bigr) \\ &\xrightarrow{\mathrm{d}} \mathbb{P}\bigl(R_\infty \leqslant b\bigr)\,\mathbb{P}\bigl(N(0,\gamma^2) \leqslant a\bigr).\end{aligned} \tag{4.70}$$

Consequently, $\tilde{V}(x_-)$ and $R(x)$ converge jointly, with independent limits given by (3.27) and (3.32)–(3.33).

It remains only to replace by $\tilde{V}(x_-)$ by $\tilde{V}(x)$. First, since $x_- = x - O(\ln x)$ it follows that $\tilde{V}(x_-) \xrightarrow{\mathrm{d}} N(0,\gamma^2)$ is equivalent to

$$\frac{\widetilde{U}_{N_\pm(x_-)} - \mu^{-\tilde{d}}\tilde{\mu}\tilde{d}!^{-1}x^{\tilde{d}}}{x^{\tilde{d}-1/2}} \xrightarrow{\mathrm{d}} N\bigl(0,\gamma^2\bigr), \tag{4.71}$$

Hence, (4.71) and $R(x) \xrightarrow{\mathrm{d}} R_\infty$ hold jointly, with independent limits.

Next, suppose first that $\tilde{f}(X_1,\dots,X_{\tilde{d}}) \geqslant 0$. Then, $\widetilde{U}_{N_\pm(x)} \geqslant \widetilde{U}_{N_\pm(x_-)}$ a.s., and thus (3.27) and (4.71) imply

$$\frac{\widetilde{U}_{N_\pm(x)} - \widetilde{U}_{N_\pm(x_-)}}{x^{\tilde{d}-1/2}} \xrightarrow{\mathrm{p}} 0. \tag{4.72}$$

By linearity, (4.72) holds for arbitrary $\tilde{f} \in L^2$. Finally, (4.72) and (4.71) imply (3.27), and hence (4.72) and the joint convergence of (4.71) and $R(x) \xrightarrow{\mathrm{d}} R_\infty$ imply the joint convergence of (3.27) and $R(x) \xrightarrow{\mathrm{d}} R_\infty$, proving (i) and (ii).

For (iii), let d be the span of $f(X)$, and assume first $d = 1$. Note that $\mathbb{P}(R_\infty = k) = 0 \iff \mathbb{P}\big(f(X) \geqslant k\big) = 0$ by (3.33), and then (3.31) implies $R(x) \leqslant f(X_{N_+(x)}) < k$ a.s. for every x; hence we only consider k such that $\mathbb{P}(R_\infty = k) > 0$, and the first part of (iii) follows from (ii).

If the span $d > 1$, then $R(x) = k$ implies $x + k = U_{N_+(x)} \equiv 0 \pmod{d}$ and thus $x \equiv -k \pmod{d}$, so we consider only $x \in -k + d\mathbb{Z}$. Let $k_0 := d\lceil k/d \rceil$ and $\Delta := k_0 - k \in [0, d-1]$. Then $x - \Delta \equiv x + k \equiv 0 \pmod{d}$, and thus, since $S_n(f) \in d\mathbb{Z}$, $N_+(x) = N_+(x - \Delta)$ and $R(x - \Delta) = U_{N_+(x)} - x + \Delta = R(x) + \Delta$; hence

$$R(x) = k \iff R(x - \Delta) = k + \Delta = k_0. \tag{4.73}$$

Hence, we may replace x and k by $x - \Delta$ and k_0, and thus it suffices to consider $x, k \in d\mathbb{Z}$, but then we can reduce to the case $d = 1$ by replacing $f(X)$ by $f(X)/d$.

Finally, for an integer n, $U_{N_-(n)} = n \iff R(n-1) = 1$. Hence, the part just proved shows that (3.27) with $x = n - 1$ holds as $n \to \infty$, also conditioned on $U_{N_-(n)} = n$. The argument above showing (4.72) shows also that $(\widetilde{U}_{N_\pm(n)} - \widetilde{U}_{N_\pm(n-1)})/n^{\tilde{d}-1/2} \xrightarrow{\mathrm{p}} 0$ as $n \to \infty$, and it follows that (3.27) with $x = n$ holds as $n \to \infty$, conditioned on $U_{N_-(n)} = n$. □

4.3 Moment convergence

We turn to proving the theorems on moment convergence in Section 3.3, and begin by extending Lemma 4.4 to higher absolute moments.

Lemma 4.7. *Suppose that $f(X_1, \dots, X_d) \in L^p$ with $p \geqslant 2$. Then*

$$\mathbb{E}\,|U_n^*(f - \mu)|^p \leqslant C_p n^{p(d-1/2)} \|f\|_p^p. \tag{4.74}$$

Proof. We use the same decomposition as in the proof of Lemma 4.4. Note that, by Jensen's inequality, $\|\widehat{F}_k\|_p \leqslant \|f\|_p$, and thus,

$$\|F_k\|_p \leqslant 2\|f\|_p, \qquad 1 \leqslant k \leqslant d. \tag{4.75}$$

Hence, Minkowski's inequality yields, as in (4.27),

$$\mathbb{E}\,|\Delta U_n(F_k)|^p = \|\Delta U_n(F_k)\|_p^p \leqslant \binom{n-1}{k-1}^p \|F_k\|_p^p \leqslant C_p n^{pk-p} \|f\|_p^p. \tag{4.76}$$

Consequently, the Burkholder inequalities [13, Theorem 10.9.5(i)] applied to the martingale $U_n(F_k)$ yield, using also Hölder's inequality,

$$\begin{aligned} \mathbb{E}\,|U_n^*(F_k)|^p &\leqslant C_p\, \mathbb{E}\Big(\sum_{i=1}^n |\Delta U_i(F_k)|^2\Big)^{p/2} \leqslant C_p\, \mathbb{E}\Big(n^{p/2-1} \sum_{i=1}^n |\Delta U_i(F_k)|^p\Big) \\ &= C_p n^{p/2-1} \sum_{i=1}^n \mathbb{E}\,|\Delta U_i(F_k)|^p \leqslant C_p n^{pk-p/2} \|f\|_p^p. \end{aligned} \tag{4.77}$$

Equivalently,

$$\|U_n^*(F_k)\|_p \leqslant C_p n^{k-1/2} \|f\|_p. \tag{4.78}$$

Finally, (4.26), (4.78) and Minkowski's inequality yield

$$\|U_n^*(f - \mu)\|_p \leqslant \sum_{k=1}^d n^{d-k} \|U_n^*(F_k)\|_p \leqslant C_p n^{d-1/2} \|f\|_p, \tag{4.79}$$

which is (4.74). □

We shall also use the following standard result, stated in detail and proved for convenience and completeness.

Lemma 4.8. *Let $\{V_\alpha : \alpha \in \mathcal{A}\}$ be a set of random variables, and let $0 < p < q$. Suppose that for every $\varepsilon > 0$ there exist decompositions $V_\alpha = V'_\alpha + V''_\alpha$ and a $B_\varepsilon < \infty$ such that, for every $\alpha \in \mathcal{A}$, $\|V'_\alpha\|_q \leqslant B_\varepsilon$ and $\|V''_\alpha\|_p \leqslant \varepsilon$. Then the set $\{|V_\alpha|^p\}$ is uniformly integrable.*

Proof. If $\delta > 0$ and $\mathcal{E}$ is any event with $\mathbb{P}(\mathcal{E}) \leqslant \delta$, then, using Hölder's inequality,

$$\begin{aligned} \mathbb{E}\big(|V_\alpha|^p \mathbf{1}_\mathcal{E}\big) &\leqslant C_p\, \mathbb{E}\big(|V'_\alpha|^p \mathbf{1}_\mathcal{E}\big) + C_p\, \mathbb{E}\big(|V''_\alpha|^p \mathbf{1}_\mathcal{E}\big) \\ &\leqslant C_p \|V'_\alpha\|_q^p\, \mathbb{P}(\mathcal{E})^{1-p/q} + C_p \|V''_\alpha\|_p^p \\ &\leqslant C_p B_\varepsilon^p \delta^{1-p/q} + C_p \varepsilon^p. \end{aligned} \tag{4.80}$$

Since ε is arbitrary, this can be made arbitrarily small, uniformly in α, by choosing first choosing ε and then δ small. □

Proof of Theorem 3.15. Denote the left-hand side of (3.13) by V_n. Then $\mathbb{E}\,|V_n|^p$ is bounded by Lemma 4.7. This implies convergence of all moments and absolute moments of order $< p$ in (3.13) by standard arguments, but is not by itself enough to include moments of order p. Thus we use a truncation: let $M > 0$ and let $f = f' + f''$ with $f' := f\mathbf{1}\{|f| \leqslant M\}$. This yields a corresponding decomposition $V_n = V'_n + V''_n$. Let $\varepsilon_M := \|f''\|_p$. Then

$$\varepsilon_M := \|f\mathbf{1}\{|f| > M\}\|_p \to 0 \quad \text{as } M \to \infty. \tag{4.81}$$

Lemma 4.7 yields

$$\|V''_n\|_p \leqslant C_p \|f''\|_p = C_p \varepsilon_M \tag{4.82}$$

and also, using $2p$ instead of p,

$$\|V'_n\|_{2p}^{2p} \leqslant C_p \|f'\|_{2p}^{2p} = C_p\, \mathbb{E}\,|f'|^{2p} \leqslant C_p M^p\, \mathbb{E}\,|f|^p. \tag{4.83}$$

(4.81)–(4.83) show that the conditions of Lemma 4.8 are satisfied; hence, $\{|V_n|^p\}$ is uniformly integrable, and the result follows from (3.13). □

We use another simple lemma.

Lemma 4.9. *Suppose that, for each $x \geqslant 1$, $V(x)$ is a non-negative random variable and $v(x) > 0$ is deterministic.*

(i) *If $p \geqslant 1$, $q \geqslant 1$ and, for some function $h(x) > 0$,*

$$\mathbb{E}\,|V(x)^q - v(x)^q|^p = O\big(v(x)^{pq} h(x)^p\big), \qquad x \geqslant 1, \tag{4.84}$$

then

$$\mathbb{E}\,|V(x) - v(x)|^p = O\big(v(x)^p h(x)^p\big), \qquad x \geqslant 1. \tag{4.85}$$

(ii) *Conversely, if* (4.85) *holds for every $p \geqslant 1$ and $h(x) \leqslant 1$, then* (4.84) *holds for every $p, q \geqslant 1$.*

Proof. (i): If $a > b \geqslant 0$, then

$$a^q - b^q = a^q\big(1 - (b/a)^q\big) \geqslant a^q\big(1 - (b/a)\big) = a^{q-1}(a-b) = \max\{a,b\}^{q-1}(a-b). \tag{4.86}$$

Hence, by symmetry, for all $a, b \geqslant 0$,

$$|a^q - b^q| \geqslant \max\{a,b\}^{q-1}|a-b|. \tag{4.87}$$

In particular,

$$|V(x)^q - v(x)^q| \geqslant v(x)^{q-1}|V(x) - v(x)|, \tag{4.88}$$

and thus (4.84) implies (4.85).

(ii): If $V(x) \leqslant 2v(x)$, then, by the mean value theorem,

$$|V(x)^q - v(x)^q| \leqslant C_q v(x)^{q-1} |V(x) - v(x)|. \tag{4.89}$$

Thus, using (4.85),

$$\begin{aligned}\mathbb{E}\big(|V(x)^q - v(x)^q|^p \mathbf{1}\{V(x) \leqslant 2v(x)\}\big) \\ \leqslant C_{p,q} v(x)^{pq-p}\,\mathbb{E}\,|V(x) - v(x)|^p = O\big(v(x)^{pq} h(x)^p\big).\end{aligned} \tag{4.90}$$

On the other hand, if $V(x) > 2v(x)$, then $|V(x)^q - v(x)^q| \leqslant V(x)^q \leqslant 2^q |V(x) - v(x)|^q$. Thus, using (4.85) with p replaced by pq,

$$\begin{aligned}\mathbb{E}\big(|V(x)^q - v(x)^q|^p \mathbf{1}\{V(x) > 2v(x)\}\big) \\ \leqslant C_{p,q}\,\mathbb{E}\,|V(x) - v(x)|^{pq} = O\big(v(x)^{pq} h(x)^{pq}\big).\end{aligned} \tag{4.91}$$

The result follows by (4.90) and (4.91). □

Proof of Theorem 3.16. As usual, we consider for definiteness $N_-(x)$. By the definition (3.17), $U_{N_-(x)} \leqslant x < U_{N_-(x)+1}$. Hence,

$$\begin{aligned}-U^*_{N_-(x)}(f-\mu) \leqslant U_{N_-(x)}(f-\mu) \leqslant x - \binom{N_-(x)}{d}\mu \\ \leqslant U^*_{N_-(x)+1}(f-\mu) + C_f N_-(x)^{d-1}\end{aligned} \tag{4.92}$$

and thus

$$\Big|x - N_-(x)^d \frac{\mu}{d!}\Big| \leqslant U^*_{N_-(x)+1}(f-\mu) + C_f N_-(x)^{d-1}. \tag{4.93}$$

Suppose throughout $x \geqslant 1$, and let again $n(x) := (d!/\mu)^{1/d} x^{1/d}$. By (4.93) and Lemma 4.7, for any $p > 0$ and any $A \geqslant 1$,

$$\begin{aligned}&\mathbb{E}\Big(\Big|x - N_-(x)^d \frac{\mu}{d!}\Big|^p \mathbf{1}\{N_-(x) \leqslant An(x)\}\Big) \\ &\leqslant C_p\,\mathbb{E}\,|U^*_{An(x)+1}(f-\mu)|^p + C_{p,f}\big(An(x)\big)^{p(d-1)} \\ &\leqslant C_{p,f}\big(An(x)\big)^{p(d-1/2)} = C_{p,f} A^{p(d-1/2)} x^{p(1-1/2d)}.\end{aligned} \tag{4.94}$$

Furthermore, for any constant $A \geqslant 2$, $N_-(x) \geqslant An(x)$ implies $N_-(x)^d \frac{\mu}{d!} - x \geqslant (A^d - 1)x \geqslant \frac{1}{2} A^d x$. Hence, for any $p \geqslant 0$ and $q > 0$, using (4.94),

$$\begin{aligned}&\mathbb{E}\Big(\Big|x - N_-(x)^d \frac{\mu}{d!}\Big|^p \mathbf{1}\{An(x) < N_-(x) \leqslant 2An(x)\}\Big) \\ &\leqslant C_q A^{-dq} x^{-q}\,\mathbb{E}\Big(\Big|x - N_-(x)^d \frac{\mu}{d!}\Big|^{p+q} \mathbf{1}\{N_-(x) \leqslant 2An(x)\}\Big) \\ &\leqslant C_{p,q,f} A^{(p+q)(d-1/2)-dq} x^{(p+q)(1-1/2d)-q} \\ &= C_{p,q,f} A^{p(d-1/2)-q/2} x^{p(1-1/2d)-q/2d}.\end{aligned} \tag{4.95}$$

Choosing $q := 2dp$, we obtain by summing (4.94) with $A = 2$ and (4.95) with $A = 2^k$, $k = 1, 2, \dots$, for every $p > 0$,

$$\begin{aligned}\mathbb{E}\big|n(x)^d - N_-(x)^d\big|^p = C_{p,f}\,\mathbb{E}\Big|x - N_-(x)^d \frac{\mu}{d!}\Big|^p \\ \leqslant C_{p,f} x^{p(1-1/2d)} + C_{p,f} \sum_{k=1}^{\infty} 2^{-kp/2} x^{-p/2d}\end{aligned}$$

$$\leqslant C_{p,f} x^{p(1-1/2d)}. \tag{4.96}$$

By Lemma 4.9(i), with $q = d$, $h(x) := x^{-1/2d}$ and $v(x) := n(x)$, (4.96) implies, for $p \geqslant 1$,

$$\mathbb{E}\big|n(x) - N_-(x)\big|^p \leqslant C_{p,f} x^{p/2d}. \tag{4.97}$$

This shows that if $Y(x)$ denotes the left-hand side of (3.20), then $\mathbb{E}\,|Y(x)|^p \leqslant C_{p,f}$ for $x \geqslant 1$. By standard arguments [13, Chapter 5.4–5], this implies uniform integrability of $|Y(x)|^r$ for any $r < p$, and thus by (3.20) convergence of moments of order $< p$. Since p is arbitrary, convergence of arbitrary moments in (3.20) follows.

Moment convergence in (3.19) is an immediate corollary. Alternatively, (4.96) implies

$$\mathbb{E}\big(N_-(x)^{dp}\big) = O\big(x^p\big), \qquad x \geqslant 1, \tag{4.98}$$

for every fixed $p > 0$, which implies moment convergence in (3.19) by the same uniform integrability argument. □

Proof of Theorem 3.17. Recall again the definition (4.48) of $n(x)$, and suppose again $x \geqslant 1$. We decompose the numerator in (3.22):

$$\widetilde{U}_{N_\pm(x)} - \Big(\frac{d!}{\mu}\Big)^{\tilde{d}/d} \frac{\tilde{\mu}}{\tilde{d}!} x^{\tilde{d}/d} = U_{N_\pm(x)}(\tilde{f} - \tilde{\mu}) + \frac{\tilde{\mu}}{\tilde{d}!}\big(N_\pm(x)^{\tilde{d}} - n(x)^{\tilde{d}}\big) + O\big(N_\pm(x)^{\tilde{d}-1}\big). \tag{4.99}$$

For the first term on the right-hand side of (4.99), we argue similarly to the proof of Theorem 3.16. First, for any $A \geqslant 2$, by Lemma 4.7,

$$\begin{aligned}\mathbb{E}\Big(\big|U_{N_\pm(x)}(\tilde{f} - \tilde{\mu})\big|^p \mathbf{1}\{N_\pm(x) \leqslant An(x)\}\Big) &\leqslant \mathbb{E}\big|U^*_{An(x)}(\tilde{f} - \tilde{\mu})\big|^p \\ &\leqslant C_{p,\tilde{f}}\big(An(x)\big)^{p(\tilde{d}-1/2)} = C_{p,f,\tilde{f}}\big(Ax^{1/d}\big)^{p(\tilde{d}-1/2)}. \end{aligned} \tag{4.100}$$

Furthermore, for any $q > 0$, taking $p = 0$ in (4.95),

$$\mathbb{P}\big(An(x) < N(x) \leqslant 2An(x)\big) \leqslant C_{q,f}\big(Ax^{1/d}\big)^{-q/2}. \tag{4.101}$$

Consequently, using the Cauchy–Schwarz inequality, (4.100)–(4.101), and choosing $q := 4(p\tilde{d} + 1)$,

$$\begin{aligned}&\mathbb{E}\Big(\big|\widetilde{U}_{N_\pm(x)}(\tilde{f} - \tilde{\mu})\big|^p \mathbf{1}\{An(x) < N_\pm(x) \leqslant 2An(x)\}\Big) \\ &\qquad\leqslant \Big(\mathbb{E}\Big(\big|\widetilde{U}_{N_\pm(x)}(\tilde{f} - \tilde{\mu})\big|^{2p} \mathbf{1}\{N_\pm(x) \leqslant 2An(x)\}\Big)\Big)^{1/2} \\ &\qquad\qquad \times \mathbb{P}\big(An(x) < N_\pm(x) \leqslant 2An(x)\big)^{1/2} \\ &\qquad\leqslant C_{p,f,\tilde{f}}\big(Ax^{1/d}\big)^{p(\tilde{d}-1/2)-q/4} \leqslant C_{p,f,\tilde{f}} A^{-1} x^{p(\tilde{d}-1/2)/d}. \end{aligned} \tag{4.102}$$

Summing (4.100) for $A = 2$ and (4.102) for $A = 2^k$, $k = 1, 2, \dots$, we obtain

$$\mathbb{E}\big|\widetilde{U}_{N_\pm(x)}(\tilde{f} - \tilde{\mu})\big|^p \leqslant C_{p,f,\tilde{f}} x^{p(\tilde{d}-1/2)/d}\Big(1 + \sum_{k=1}^{\infty} 2^{-k}\Big) = C_{p,f,\tilde{f}} x^{p(\tilde{d}-1/2)/d}. \tag{4.103}$$

For the second term on the right-hand side of (4.99), we use (4.97) and Lemma 4.9(ii), with $q = \tilde{d}$ and, again, $h(x) := x^{-1/2d}$ and $v(x) := n(x)$, and conclude, for every $p \geqslant 1$,

$$\mathbb{E}\big|N_\pm(x)^{\tilde{d}} - n(x)^{\tilde{d}}\big|^p \leqslant C_{p,f,\tilde{f}} x^{p(\tilde{d}-1/2)/d}. \tag{4.104}$$

Finally, by Theorem 3.16 we have moment convergence in (3.19) and thus

$$\mathbb{E}\big(N_\pm(x)^{p(\tilde{d}-1)}\big) = O\big(x^{p(\tilde{d}-1)/d}\big), \tag{4.105}$$

which also follows from (4.98) (changing p).

It follows from (4.99) and (4.103)–(4.105) that

$$\mathbb{E}\left|\frac{\widetilde{U}_{N_\pm(x)} - \left(\frac{d!}{\mu}\right)^{\tilde{d}/d}\frac{\tilde{\mu}}{\tilde{d}!}x^{\tilde{d}/d}}{x^{(\tilde{d}-1/2)d}}\right|^p \leqslant C_{p,f,\tilde{f}}. \tag{4.106}$$

Since p is arbitrary, this implies convergence of arbitrary moments in (3.22) by the same standard argument as in the proof of Theorem 3.16.

Moment convergence in (3.21) is a corollary. □

Proof of Theorem 3.18. (i): This is a special case of Theorem 3.17.

(ii): Denote the left-hand side of (3.27) by $V(x)$, for integers $x \geqslant 1$, and let $p > 0$. It follows from (i) that the family $|V(x)|^p$, $x \geqslant 1$, is uniformly integrable. This property is preserved by the conditioning, since we condition on a sequence of events $\mathcal{E}_x$ with $\liminf_{x\to\infty} \mathbb{P}(\mathcal{E}_x) > 0$ by the proof of Theorem 3.13; hence the result follows from Theorem 3.13. □

5 Examples and applications

Example 5.1. Let $d = 2$, and let f be anti-symmetric: $f(y,x) = -f(x,y)$; this case was studied in [24]. We have $\mu = 0$ and $f_2(x) = \mathbb{E}\, f(X,x) = -\mathbb{E}\, f(x,X) = -f_1(x)$; hence $\sigma_{11} = -\sigma_{12} = \sigma_{22}$ and (3.4) implies $W_2(t) = -W_1(t) = \sigma B(t)$, where $\sigma := \|f_1\|_2 \geqslant 0$ and $B(t)$ is a standard Brownian motion.

For $d = 2$, (3.5) yields $\psi_1(s,t) = t - s$ and $\psi_2(s,t) = s$. Hence, (3.6), (3.7) and integration by parts, see (4.17), yield

$$\begin{aligned}\frac{U_{nt}}{n^{3/2}} \xrightarrow{\mathrm{d}} Z_t &= \int_0^t (t-2s)\,\mathrm{d}W_1(s) = -tW_1(t) + 2\int_0^t W_1(s)\,\mathrm{d}s \\ &= \sigma t B(t) - 2\sigma\int_0^t B(s)\,\mathrm{d}s\end{aligned} \tag{5.1}$$

in $D[0,\infty)$, as shown in [24] (where also the degenerate case $\sigma = 0$ is studied further).

Example 5.2 (Substrings)**.** Consider a random string $X_1 \cdots X_n$ of length n from a finite alphabet $\mathcal{A}$, with the letters X_i i.i.d. with some distribution $\mathbb{P}(X_i = a) = p_a$, $a \in \mathcal{A}$. Fix a *pattern* $\mathcal{W} = w_1 \cdots w_m$; this is an arbitrary string in $\mathcal{A}^m$, for some $m \geqslant 1$. A *substring* of $X_1 \cdots X_n$ is any string $X_{i_1} \cdots X_{i_k}$ with $1 \leqslant i_1 < \cdots < i_k \leqslant n$, and we let $N_n = N_{\mathcal{W}}(X_1 \cdots X_n)$ be the number of substrings that have the pattern $\mathcal{W}$. Obviously, this is an asymmetric U-statistic as in (1.1) with $\mathcal{S} = \mathcal{A}$, $d = m$ and

$$f(x_1,\dots,x_m) := \mathbf{1}\{x_1 \cdots x_m = w_1 \cdots w_m\} = \prod_{i=1}^m \mathbf{1}\{x_i = w_i\}. \tag{5.2}$$

Corollary 3.5 yields asymptotic normality of N_n as $n \to \infty$, as shown by Flajolet, Szpankowski and Vallée [10].

For example, let $\mathcal{A} := \{0,1\}$, let $X_i \sim \mathrm{Be}(\frac12)$, and let $\mathcal{W} := 10$. A simple calculation yields $f_1(x) = \frac12(x - \frac12) = -f_2(x)$, and $\sigma_{11} = \sigma_{22} = -\sigma_{12} = 1/16$; thus Corollary 3.5 yields, see (3.16),

$$\frac{N_n - n/4}{\sqrt{n}} \xrightarrow{\mathrm{d}} N\Big(0, \frac{1}{48}\Big). \tag{5.3}$$

Furthermore, calculations as in Example 5.1 show that the functional limit (5.1) holds in this case too, with $\sigma = 1/4$.

Example 5.3 (Patterns in permutations)**.** Let $\pi = \pi_1 \cdots \pi_n$ be a uniformly random permutation of length n, and let the *pattern* $\sigma = \sigma_1 \cdots \sigma_m$ be a fixed permutation of length m. The *number of occurences* of σ in π, denoted by $N_n = N_\sigma(\pi)$ is the number of substrings (see Example 5.2) of π that have the same relative order as σ.

We can generate the random permutation π by taking i.i.d. random variables $X_1, \dots, X_n \sim U(0,1)$, and then replacing these numbers by their ranks. Then N_n is the U-statistic with $d = m$ given by the function

$$f(x_1, \dots, x_m) = \mathbf{1}\{x_1 \cdots x_m \text{ have the same relative order as } \sigma_1 \cdots \sigma_m\}. \tag{5.4}$$

Corollary 3.5 shows that N_n is asymptotically normal as $n \to \infty$. For details, including explicit variance calculations, see [23]; see also the earlier proof of asymptotic normality by Bóna [3, 4].

For example, taking $\sigma = 21$, N_n is the number of inversions in π, and we obtain by simple calculations the well-known result, see e.g. [8, Section X.6],

$$\frac{N_n - n^2/4}{n^{3/2}} \overset{\mathrm{d}}{\longrightarrow} N\Bigl(0, \frac{1}{36}\Bigr). \tag{5.5}$$

Example 5.4 (Restricted permutations I)**.** Fix a set T of permutations, and consider only permutations π of length n that *avoid* T, in the sense that there is no occurence of any $\tau \in T$ in π. Let π be uniformly random from this set, for a given n.

Several cases are studied in [22], and some of them yield asymmetric U-statistics, sometimes stopped or conditioned as in Theorem 3.11 or 3.13. We sketch two examples here and in the next example, and refer to [22] for details and further similar examples.

A permutation π avoids $\{231, 312\}$ if and only if π is an increasing sequence of *blocks* that all are decreasing; in other words,

$$\pi = (L_1, \dots, 1, L_1 + L_2, \dots, L_1 + 1, L_1 + L_2 + L_3, \dots, L_1 + L_2 + 1, \dots), \tag{5.6}$$

see [31, Proposition 12]. Let the (random) number of blocks be $B \geqslant 1$ and the block lengths $L_1, \dots, L_B$; thus $L_i \geqslant 1$ and $L_1 + \cdots + L_B = n$. Then, any such sequence $L_1, \dots, L_B$ is possible, and it determines π uniquely. Hence, taking $f(L) := 1$ and thus $U_n = S_n = \sum_1^n L_i$, it is easily seen that $(L_1, \dots, L_B)$ has the same distribution as the first $N_-(n)$ elements of an i.i.d. sequence $(L_k)_k$ of geometric random variables $L_k \sim \mathrm{Ge}(1/2)$, conditioned on $U_{N_-(n)} = n$.

Let σ be a fixed permutation that avoids $\{231, 312\}$, with block lengths $\ell_1, \dots, \ell_b$. Then the number $N_{\sigma,n} = N_\sigma(\pi)$ of occurrences of σ in π is given by a U-statistic, with $d = b$, based on the sequence of variables $L_1, \dots, L_B$ and the function

$$\tilde{f}(x_1, \dots, x_b) := \prod_{j-1}^{b} \binom{x_i}{\ell_i}. \tag{5.7}$$

Theorem 3.13(iii) applies and shows asymptotic normality in the form

$$\frac{N_{\sigma,n} - n^b/b!}{n^{b-1/2}} \overset{\mathrm{d}}{\longrightarrow} N\bigl(0, \gamma^2\bigr), \tag{5.8}$$

for some $\gamma^2 > 0$ depending on σ.

For example, taking $\sigma = 21$, so $N_{21,n}$ is the number of inversions in π, $b = 1$ and, by a calculation, $\gamma^2 = 6$; hence

$$\frac{N_{21,n} - n}{n^{1/2}} \overset{\mathrm{d}}{\longrightarrow} N\bigl(0, 6\bigr). \tag{5.9}$$

We here applied the conditional result in Theorem 3.13. Alternatively (since a geometric distribution has no memory), we may avoid the conditioning above and instead truncate the last element L_B such that the sum becomes exactly n; using a simple approximation argument, we can then apply the unconditional Theorem 3.11.

Example 5.5 (Restricted permutations II)**.** Continuing Example 5.4, now let π be a uniformly random permutation of a given length n such that π avoids $\{231, 312, 321\}$.

A permutation π avoids $\{231, 312, 321\}$ if and only if π is of the form (5.6) and furthermore every block length $L_i \leqslant 2$, see [31, Proposition 15^*]. Taking again $f(L) := 1$, it is easily seen that $(L_1, \dots, L_B)$ has the same distribution as the first $N_-(n)$ elements of an i.i.d. sequence $(L'_k)_1^\infty$, conditioned on $U_{N_-(n)} = n$, where we now let

$$\mathbb{P}(L'_i = 1) = p, \qquad \mathbb{P}(L'_i = 2) = p^2, \tag{5.10}$$

where $p + p^2 = 1$ and thus p is the golden ratio

$$p := \frac{\sqrt{5} - 1}{2}. \tag{5.11}$$

Let σ be a fixed permutation that avoids $\{231, 312, 321\}$, with block lengths $\ell_1, \dots, \ell_b \in \{1, 2\}$. Then the number $N_{\sigma,n} = N_\sigma(\pi)$ of occurrences of σ in π is given by a U-statistic based on $L_1, \dots, L_B$, with $d = b$ and the function $\tilde{f}$ in (5.7). Theorem 3.13(iii) applies and shows asymptotic normality in the form

$$\frac{N_{\sigma,n} - \mu n^b / b!}{n^{b-1/2}} \xrightarrow{\mathrm{d}} N\bigl(0, \gamma^2\bigr), \tag{5.12}$$

for some $\mu > 0$ and $\gamma^2 > 0$ depending on σ.

For example, taking $\sigma = 21$, so $N_{21,n}$ is the number of inversions in π, $b = 1$ and, by calculations, see [22], $\mu = (3 - \sqrt{5})/2$ and $\gamma^2 = 5^{-3/2}$; hence

$$\frac{N_{21,n} - \frac{3-\sqrt{5}}{2} n}{n^{1/2}} \xrightarrow{\mathrm{d}} N\bigl(0, 5^{-3/2}\bigr). \tag{5.13}$$

6 Further comments and open problems

Remark 6.1. In Theorems 3.16 and 3.17, we assume (for simplicity) existence of all moments for f and $\tilde{f}$, and conclude convergence of all moments in (3.19)–(3.22). If we only want to conclude convergence of a specific moment, e.g. convergence of second moments in (3.20) or (3.22), the proofs above show that it suffices to assume existence of some specific moment for f and $\tilde{f}$. However, we do not know the best possible moment conditions for this, and we leave it as an open problem to find optimal conditions. (The proofs above are not optimized; furthermore, the methods used there are not necessarily optimal.) In particular, we do not know whether convergence of first and second moments always holds in (3.20) and (3.22) without further moment assumptions. (For some results when $d = \tilde{d} = 1$, see [19] and [12, Chapter 3].)

Remark 6.2. In the case when f is bounded, subgaussian estimates for large deviations of the left-hand side of (3.13) are shown in [18] and [21]. This and the definitions (3.17)–(3.18) lead to large deviation estimates for $N_\pm$, and, provided also $\tilde{f}$ is bounded, then further to large deviation estimates for the left-hand side in (3.22). We leave the details to the reader.

Remark 6.3. As said in the introduction, the results above are of most interest in the non-degenerate case, where $\Sigma = (\sigma_{ij})$ defined by (3.2) is non-zero. In the degenerate case, when all $\sigma_{ij} = 0$, or equivalently, $f_i(X) = 0$ a.s. for every i, the results still hold but then the limits in e.g. Theorem 3.2 are degenerate, see also (3.15). A typical degenerate example is the anti-symmetric $f(X_1, X_2) = \sin(X_1 - X_2)$, with X uniformly distributed on $[0, 2\pi)$ (best regarded as the unit circle), where $f_1 = f_2 = 0$.

In the degenerate case, one can instead normalize using a smaller power of n than in Theorem 3.2 and obtain non-degenerate limits; this is well-known in the symmetric case, see e.g. [9], [11], [29], [6], [20, Chapter 11] for univariate results and [27], [15], [5], [7], [28], [20, Remark 11.11] for functional limits. This extends to the asymmetric case; univariate results are given in [20, Chapter 11.2] with the possibility of functional limits briefly mentioned in [20, Remark 11.25], and the case $d = 2$ and f antisymmetric was studied in [24] (functional limits for both the degenerate and non-degenerate cases), see Example 5.1. We do not consider such refined results for the degenerate case in the present paper.

Remark 6.4. For multi-sample U-statistics, i.e., variables of the form

$$U_{n_1,\dots,n_\ell} := \sum f\bigl(X^{(1)}_{i_{1,1}},\dots,X^{(1)}_{i_{1,d(1)}},\dots,X^{(\ell)}_{i_{\ell,1}},\dots,X^{(\ell)}_{i_{\ell,d(\ell)}}\bigr), \tag{6.1}$$

summing over $1 \leqslant i_{j,1} < \dots < i_{j,d(j)} \leqslant n_j$ for every $j = 1,\dots,\ell$, a multi-dimensional functional limit theorem has been given by Sen [30] in the symmetric case (i.e., with f symmetric in each of the ℓ sets of variables); see also e.g. [27], [15], [7]. We expect that this too can be extended to the asymmetric case, but we leave this to the interested reader.

Remark 6.5. There is a standard trick to convert an asymmetric U-statistic to a symmetric one, see e.g. [20]. Let $Y_i \sim U(0,1)$ be i.i.d. random variables, independent of $(X_j)_1^\infty$, let $Z_i := (X_i, Y_i) \in \tilde{S} := S \times \mathbb{R}$, and define $F : \tilde{S}^n \to \mathbb{R}$ by

$$F\bigl((x_1,y_1),\dots,(x_d,y_d)\bigr) := f(x_1,\dots,x_d)\mathbf{1}\{y_1 < \dots < y_d\} \tag{6.2}$$

and its symmetrized version

$$F^*(z_1,\dots,z_d) := \sum_{\sigma\in S_d} F\bigl(z_{\sigma(1)},\dots,z_{\sigma(d)}\bigr), \tag{6.3}$$

summing over the $d!$ permutations of $\{1,\dots,d\}$. Then, letting $\sum^*$ denote the sum over distinct indices,

$$\begin{aligned} U_n(f) &\overset{\mathrm{d}}{=} \sideset{}{^*}\sum_{\substack{i_1,\dots,i_d\leqslant n\\ Y_{i_1}<\dots<Y_{i_d}}} f\bigl(X_{i_1},\dots,X_{i_d}\bigr) = \sideset{}{^*}\sum_{i_1,\dots,i_d\leqslant n} F\bigl((X_{i_1},Y_{i_1}),\dots,(X_{i_d},Y_{i_d})\bigr) \\ &= \sum_{1\leqslant i_1<\dots<i_d\leqslant n} F^*\bigl(Z_{i_1},\dots,Z_{i_d}\bigr) = U_n(F^*). \end{aligned} \tag{6.4}$$

This trick often makes it possible to transfer results for symmetric U-statistics to the general, asymmetric case. However, this trick works only for a single n, and we do not know of any similar trick that can handle the process $(U_n)_{n=0}^\infty$. Hence this method does not seem useful for the results above.

Remark 6.6. In the symmetric case, it is easily seen that $U_n/\binom{n}{d}$, $n \geqslant d$, is a reverse martingale, which for example yields a simple proof of the law of large numbers; see [1] and e.g. [13, Chapter 10.16.2]. This does not hold in general; thus we used above (in the proof of Lemma 4.4) instead forward martingales similarly to [17].

References

[1] Robert H. Berk, Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37** (1966), 51–58. MR-0189176

[2] Patrick Billingsley, *Convergence of Probability Measures*. Wiley, New York, 1968. MR-0233396

[3] Miklós Bóna, The copies of any permutation pattern are asymptotically normal. Preprint, 2007. arXiv:0712.2792

[4] Miklós Bóna, On three different notions of monotone subsequences. *Permutation Patterns*, 89–114, London Math. Soc. Lecture Note Ser., 376, Cambridge Univ. Press, Cambridge, 2010. MR-2732825

[5] Herold Dehling, Manfred Denker & Walter Philipp, Invariance principles for von Mises and U-statistics. *Z. Wahrsch. Verw. Gebiete* **67** (1984), no. 2, 139–167. MR-0758070

[6] Víctor H. de la Peña and Evarist Giné, *Decoupling*. Springer-Verlag, New York, 1999. MR-1666908

[7] M. Denker, Ch. Grillenberger & G. Keller, A note on invariance principles for v. Mises' statistics. *Metrika* **32** (1985), no. 3-4, 197–214. MR-0824454

[8] William Feller, *An Introduction to Probability Theory and Its Application*, volume I, third edition, Wiley, New York, 1968. MR-0228020

[9] A. A. Filippova, The theorem of von Mises on limiting behaviour of functionals of empirical distribution functions and its statistical applications. (Russian.) *Teor. Verojatnost. i Primenen.* **7** (1962), 26–60. MR-0150870

[10] Philippe Flajolet, Wojciech Szpankowski and Brigitte Vallée. Hidden word statistics. *J. ACM* **53** (2006), no. 1, 147–183. MR-2212002

[11] Gavin G. Gregory, Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5** (1977), no. 1, 110–123. MR-0433669

[12] Allan Gut, *Stopped Random Walks* 2nd ed. Springer, New York, 2009. MR-2489436

[13] Allan Gut, *Probability: A Graduate Course*, 2nd ed. Springer, New York, 2013. MR-2977961

[14] Allan Gut & Svante Janson, Converse results for existence of moments and uniform integrability for stopped random walks. *Ann. Probab.* **14** (1986), 1296–1317. MR-0866351

[15] Peter Hall, On the invariance principle for U-statistics. *Stochastic Process. Appl.* **9** (1979), no. 2, 163–174. MR-0548836

[16] Wassily Hoeffding, A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* **19** (1948), 293–325. MR-0026294

[17] Wassily Hoeffding, The strong law of large numbers for U-statistics. Institute of Statistics, Univ. of North Carolina, Mimeograph series 302 (1961). https://repository.lib.ncsu.edu/handle/1840.4/2128

[18] Wassily Hoeffding, Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** (1963), 13–30. MR-0144363

[19] Svante Janson, Moments for first passage and last exit times, the minimum, and related quantities for random walks with positive drift. *Adv. Appl. Probab.* **18** (1986), 865–879. MR-0867090

[20] Svante Janson, *Gaussian Hilbert Spaces*, Cambridge Univ. Press, Cambridge, UK, 1997. MR-1474726

[21] Svante Janson, Large deviations for sums of partly dependent random variables. *Random Structures Algorithms* **24** (2004), no. 3, 234–248. MR-2068873

[22] Svante Janson, Patterns in random permutations avoiding some sets of multiple patterns. Preprint, 2018. arXiv:1804.06071

[23] Svante Janson, Brian Nakamura & Doron Zeilberger, On the asymptotic statistics of the number of occurrences of multiple permutation patterns. *Journal of Combinatorics* **6** (2015), no. 1-2, 117–143. MR-3338847

[24] Svante Janson & Michael J. Wichura, Invariance principles for stochastic area and related stochastic integrals. *Stochastic Process. Appl.* **16** (1984), no. 1, 71–84. MR-0723644

[25] Olav Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002. MR-1876169

[26] R. G. Miller, Jr. & Pranab Kumar Sen, Weak convergence of U-statistics and von Mises' differentiable statistical functions. *Ann. Math. Statist.* **43** (1972), 31–41. MR-0300321

[27] Georg Neuhaus, Functional limit theorems for U-statistics in the degenerate case. *J. Multivariate Anal.* **7** (1977), no. 3, 424–439. MR-0455084

[28] A. F. Ronzhin, A functional limit theorem for homogeneous U-statistics with degenerate kernel. (Russian) *Teor. Veroyatnost. i Primenen.* **30** (1985), no. 4, 759–762. English transl.: *Theory Probab. Appl.* 30 (1985), no. 4, 806–809. MR-0816288

[29] H. Rubin & R.A. Vitale, Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** (1980), 165–170. MR-0557561

[30] Pranab Kumar Sen, Weak convergence of generalized U-statistics *Ann. Probability* **2** (1974), 90–102. MR-0402844

[31] Rodica Simion and Frank W. Schmidt, Restricted permutations. *European J. Combin.* **6** (1985), no. 4, 383–406. MR-0829358

[32] Raymond N. Sproule, Asymptotic properties of U-statistics. *Trans. Amer. Math. Soc.* **199** (1974), 55–64. MR-0350826