A Central Limit Theorem for Vincular Permutation Patterns

Lisa Hofer

Institute of Mathematics, University of Zurich, Switzerland

We study the number of occurrences of any fixed vincular permutation pattern. We show that this statistics on uniform random permutations is asymptotically normal and describe the speed of convergence. To prove this central limit theorem, we use the method of dependency graphs. The main difficulty is then to estimate the variance of our statistics. We need a lower bound on the variance, for which we introduce a recursive technique based on the law of total variance.

Keywords: permutation patterns, central limit theorem, dependency graphs, vincular patterns, variance estimate

1 Introduction

Permutation patterns are well studied objects in combinatorics and, more recently, also in probability theory. In combinatorics, most of the research is focused on pattern avoidance, *i.e.* finding the number of permutations with no occurrences of a given pattern, see Bóna (2012); Kitaev (2011). Another problem is the study of statistics and their distribution. Those objects are studied combinatorially, for instance with multivariate generating functions as in Chatterjee and Diaconis (2017); Crane et al. (2016), but also in probability theory. One statistics of interest is the number of occurrences of a given pattern in a random permutation, where various distributions on permutations are considered. Often, people study the case of uniform permutations, as in Bóna (2010); Fulman (2004); Janson et al. (2015). Among patterns, descents and inversions are the most well-known, see Fulman (2004).

The problem we consider in this article is the following: for a fixed pattern π, what is the asymptotic behaviour of the number of occurrences of π in a uniform random permutation σ_n of size n going to infinity? We consider this problem when π is a so-called vincular pattern.

To explain the precise meaning of this problem, we begin by describing different types of permutation patterns. When we study patterns, permutations are represented in one-line form, *i.e.* as a sequence. A permutation of size n is a reordering of the monotone sequence $12 \ldots n$. The study of patterns is the study of subsequences and their order. For example, consider the permutation

$$\sigma = \mathbf{58}213476$$

and its subsequence 524. The unique permutation whose elements are listed in the same relative order is 312. We say that the subsequence 524 is an occurrence of the *classical* pattern 312. Occurrences of classical patterns can be any subsequences. Additional constraints on the subsequences lead to different types of patterns: *tight*, *very tight*, *vincular* and *bivincular* patterns (terminology from Bóna (2012); Kitaev (2011)). To count as an occurrence of such a type of pattern, a subsequence must fulfil the constraints listed below:

- Tight: all entries have adjacent positions.

- Very tight: all entries have adjacent positions and values.

- Vincular: some entries have adjacent positions or in other words appear in blocks.

- Bivincular: some entries have adjacent positions and some (maybe different) entries have adjacent values.

In the literature, patterns are sometimes called consecutive instead of tight and generalized or dashed instead of vincular. Note that vincular patterns generalize both classical and tight patterns.

As an example of a vincular pattern, consider $3\underline{12}$ where the underlined symbols indicate that the last two entries are required to be in adjacent positions. In the permutation σ above, the subsequence 524 is therefore not an occurrence of this vincular pattern, but the following are: 513, 534, 813, 834, 847. The number of occurrences of $3\underline{12}$ in σ is 5. We say the pattern $3\underline{12}$ has two *blocks*, the first is one isolated entry and the second consists in two adjacent entries. Formally, we write a vincular pattern as a pair (π, A), where π is a permutation giving the order constraint and A gives the required adjacencies.

Instead of counting occurrences of a pattern in a deterministic permutation σ as above, we look at σ_n which is a uniform permutation of size n. Considering a random variable counting the number of occurrences of a fixed pattern of size k in σ_n, we ask how it behaves asymptotically, as n goes to infinity. Since there are different types of patterns, there is actually a whole family of such problems.

The following answers to these problems are known.

- In Fulman (2004), J. Fulman proves asymptotic normality for inversions and descents (classical and tight patterns of size $k = 2$). In addition, he provides a rate of convergence.

- M. Bóna establishes in Bóna (2010) asymptotic normality for classical and tight patterns which are monotone (increasing or decreasing).

- In Janson et al. (2015), asymptotic normality is shown for all classical patterns. In fact, the authors also establish the joint convergence.

- In (Crane et al., 2016, Section 8), asymptotic normality and a rate of convergence are provided for tight patterns of size $k \geq 3$ in random permutations distributed with the so-called Mallows measure of parameter q (for $q = 1$, this measure specializes to the uniform distribution which is of interest in this article). These results are obtained under the assumption that the highest degree component of the variance does not vanish. Proving this kind of variance estimate is often difficult, as we will discuss later in the introduction.

- L. Goldstein provides in (Goldstein, 2005, Example 3.2) a rate of convergence for tight patterns in case of the uniform measure. As in Crane et al. (2016), this rate depends on the variance. In contrast to Crane et al. (2016), the method used to obtain this rate applies to a larger family of statistics.

- Very tight patterns behave differently: for $k = 2$, they are asymptotically Poisson distributed, see Corteel et al. (2006); Kaplansky (1945). It is easy to see that for $k > 2$, the probability to find such a pattern tends to zero (see (Corteel et al., 2006, p. 3–4)).

- In Crane and DeSalvo (2017), Poisson limit theorems are also obtained for tight patterns in Mallows permutations of size n, if the parameter of the Mallows distribution $q(n)$ is a function of n of a specific form or if q is fixed but the size of the pattern tends to infinity. This setting is quite orthogonal to the one of other papers (including this one).

In the present article, we generalize the result of asymptotic normality to vincular patterns and we also describe the speed of convergence. Our main result, proved in Section 3, is the following.

Let Z denote a standard normal random variable and d_K denote the Kolmogorov distance which is the maximum distance between distribution functions. Denoting $\overline{X^{(\pi,A)}(\sigma_n)}$ the renormalized (mean 0, variance 1) random variable which counts the number of occurrences of (π, A) in σ_n, it holds that for some positive constant C,

$$d_K\left(\overline{X^{(\pi,A)}(\sigma_n)}, Z\right) \leq Cn^{-1/2}.$$

This implies immediately that

$$\overline{X^{(\pi,A)}(\sigma_n)} \xrightarrow{d} Z,$$

where $\xrightarrow{d}$ denotes convergence in distribution, hence proving asymptotic normality. In addition, the bound on d_K quantifies the accuracy of the approximation of $\overline{X^{(\pi,A)}(\sigma_n)}$ by Z. Note that this result encompasses the results from Bóna (2010); Crane et al. (2016); Fulman (2004); Goldstein (2005); Janson et al. (2015) previously mentioned except for the joint convergence in Janson et al. (2015) and the case of a general parameter $q \neq 1$ in Crane et al. (2016).

Let us now discuss the method of proof. In the literature, the following methods have been used for normal approximation:

- *U-statistics* in Janson et al. (2015). However, the number of occurrences of vincular patterns is not a U-statistics (unlike for classical patterns).

- *Exchangeable Stein pairs* in Fulman (2004) for patterns of size $k = 2$. Here, we did not succeed in finding such a pair for patterns of any size.

- *Size-bias couplings* in Goldstein (2005) for tight patterns. We are not aware of such a coupling for vincular patterns.

- *Dependency graphs* in Bóna (2010) and Crane et al. (2016). Such a graph captures the dependencies in a family of random variables. This is useful for our problem since $X^{(\pi,A)}(\sigma_n)$ can be decomposed as a sum of partially dependent random variables (see Equation (2), p. 7).

The last three methods are based on *Stein's method* (except for dependency graphs in Bóna (2010)). This method is used to prove convergence in distribution as well as to describe the approximation error, see Section 2.7 for more details.

In Section 3, we present two approaches to bound the Kolmogorov distance both based on dependency graphs: one using the Stein machinery following Chen and Röllin (2010); Ross (2011) and one using the moment method following Féray et al. (2016); Janson (1988); Saulis and Statulevičius (1991). While their application is easy, there is one difficulty: we need a lower bound on the variance of $X^{(\pi,A)}(\sigma_n)$ to prove that $d_K(\overline{X^{(\pi,A)}(\sigma_n)}, Z)$ goes to 0.

The method to find that lower bound is discussed in Section 4.2. First, we show that $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ is a polynomial in n. Denoting j the number of blocks of the vincular pattern (π, A), the polynomiality implies that

$$\mathrm{Var}(X^{(\pi,A)}(\sigma_n)) = Cn^{2j-1} + \mathcal{O}(n^{2j-2}), \text{ with } C \geq 0.$$

If we can show that C is different from 0, then $\frac{1}{2}Cn^{2j-1}$ is a sharp lower bound (for n big enough). The most natural approach to prove that C is larger than 0 is to find a formula for C by expressing the variance in terms of covariances (see Equation (8), p. 16). This is Bóna's approach in Bóna (2010). Such a formula for C is a *signed* sum of binomials, which in our case is hard to examine. Instead, we introduce a new technique: a recurrence based on the law of total variance. It provides a lower bound for $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ of the form $C'n^{2j-3/2}$. Thanks to the polynomiality, this is enough to prove that C is larger than 0 (see Section 4.2).

We conclude this introduction with further directions of research.

- In this article, we establish the asymptotic normality for vincular patterns. Moreover, the asymptotic behaviour of very tight patterns is characterized in Corteel et al. (2006): as Poisson distributed ($k = 2$) or rare ($k > 2$). Both these types are contained in the larger class of bivincular patterns. One could try to classify such patterns in terms of their limiting distribution.

  To study the asymptotically normal case, we suggest using so-called *interaction graphs* or *weighted dependency graphs* introduced in Chatterjee and Diaconis (2017); Féray (2016). In Chatterjee and Diaconis (2017), interaction graphs are used for a similar statistics, "number of descents plus number of descents in the inverse" which also has constraints in positions and values. The classical tool of dependency graphs does not apply anymore since constraints along these two directions imply that the random variables in a natural sum decomposition are all pairwise dependent (unlike in Equation (2), p. 7). Using the mentioned extensions of dependency graphs avoids this problem, but the main difficulty will again be estimating the variance.

- In Janson et al. (2015), joint convergence is established for classical patterns and (Janson et al., 2015, Theorem 4.5) describes how dependent the single limit random variables are (for patterns of the same size). It would be interesting to study these questions for vincular patterns. Moreover, one could try to examine the speed of convergence in the multivariate case.

- The optimality of the bound on d_K obtained in this article could be investigated. We believe it is optimal since no better bounds are obtained in similar problems (*e.g.* subgraph counts in random graphs), but we do not have concrete mathematical evidence for it.

The outline of this article is as follows. In Section 2, we give all the necessary background and set up the notation. In Section 3, we show two approaches to prove the main result. Both approaches rely on a lower bound on the variance, which is established in Section 4.

2   Background and notation

Throughout this article, we write a permutation σ in one-line notation, *i.e.* $\sigma = \sigma_1 \sigma_2 \ldots \sigma_n$. The length of the sequence $\sigma_1 \sigma_2 \ldots \sigma_n$ is the *size* of σ, denoted by $|\sigma|$. The set of all permutations of size n is denoted by S_n. We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$ and $\binom{[n]}{k}$ for the set of all subsets of $[n]$ which are of size k.

2.1   Vincular patterns

We refer to Bóna (2012) and Kitaev (2011), which discuss the notion of permutation patterns. In Bóna (2012), the reader can find information about classical patterns while Kitaev (2011) discusses patterns in more generality, *e.g.* also the case of vincular patterns. A classical pattern is defined as follows.

Definition 2.1. *Let $\sigma \in S_n$ and $\pi \in S_k$. An* occurrence *of the* classical pattern π *in σ is a subsequence $\sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_k}$ of length k of σ such that:*

- $\pi_r < \pi_s \iff \sigma_{i_r} < \sigma_{i_s}$, *for all $r, s \in [k]$.*

We say that σ has an occurrence of the classical pattern π at positions $i_1, i_2, \ldots, i_k$.

Example 2.2. *Let $\sigma = 2374561$. The subsequence $\sigma_2 \sigma_5 \sigma_7 = 351$ is an occurrence of the classical pattern $\pi = 231$.*

In contrast to classical patterns, vincular patterns have additional constraints on the subsequences that are allowed to be counted as an occurrence of the pattern. Certain parts of the pattern are required to be adjacent or, in other words, to appear in blocks.

Definition 2.3. *Let $\sigma \in S_n$, $\pi \in S_k$ and let $A \subseteq [k-1]$. An* occurrence *of the* vincular pattern (π, A) *in σ is a subsequence $\sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_k}$ of length k of σ such that:*

- $\sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_k}$ *is an occurrence of π in the classical sense,*

- $i_{a+1} = i_a + 1$, *for all $a \in A$.*

We call A the set of adjacencies.

Example 2.4. *Let $\sigma = 2374561$. The subsequence $\sigma_2 \sigma_3 \sigma_7 = 371$ is an occurrence of the vincular pattern $(\pi, A) = (231, \{1\})$. Note that the subsequence from the previous example, $\sigma_2 \sigma_5 \sigma_7 = 351$, is not an occurrence of (π, A) since 3 and 5 are not adjacent in σ.*

Remark 2.5. *In (Kitaev, 2011, Definition 1.3.1), the definition of vincular patterns is even more general, allowing also constraints on the beginning and on the end of a pattern occurrence. However, such constraints are rarely considered in the literature, and not included in the present work. The central limit theorem would not hold if we constraint the first entry of the pattern occurrence to be at the beginning, as can be seen in the pattern 12 with 1 forced to be the first entry of the permutation.*

In this article, we will work with the above definition but sometimes, it is more convenient to see adjacencies as *blocks*. By block, we mean a maximal subsequence of the pattern whose entries are required to be adjacent. An equivalent way to encode the adjacency information of a vincular pattern is to give a list of block sizes. For example, the vincular pattern $(231, \{1\})$ would be written as $(231, (2, 1))$, where the list $(2, 1)$ describes a first block of size 2 followed by a block of size 1. This idea appears also in (Kitaev, 2011, Definition 7.1.2), where the list of block sizes is called the *type* of a vincular pattern.

Now, note that the block sizes add up to the size of the pattern, leading to the notion of *composition*. The following definition can be found in (Flajolet and Sedgewick, 2009, p. 39).

Definition 2.6. *A composition of an integer n is a sequence $(x_1, x_2, \ldots, x_\ell)$ of integers, for some ℓ, such that $n = x_1 + x_2 + \cdots + x_\ell$ and $x_i \geq 1$ for all i.*

For example, $(2, 1)$ is a composition of 3. How one can go from one encoding of vincular patterns (by adjacencies or block sizes) to the other is explained by a bijection between subsets of $[k-1]$ and compositions of k, where $k = |\pi|$. This bijection associates for instance the composition $(4, 2, 2, 1)$ of size 9 to $\{1, 2, 3, 5, 7\} \subseteq [8]$. The formal construction of the bijection is given below. It will help understanding the rephrasing of the adjacency condition.

For $A \subseteq [k-1]$, consider $j = k - |A|$ and $\{c_1, c_2, \ldots, c_j\} = [k] \setminus A$ with $c_1 < c_2 < \cdots < c_j = k$, and construct iteratively:

$$\begin{cases} b_1 & = c_1, \\ b_2 & = c_2 - c_1, \\ & \vdots \\ b_j & = c_j - c_{j-1}. \end{cases} \tag{1}$$

Then, $(b_1, b_2, \ldots, b_j)$ is a composition of k, since $\sum_{i=1}^{j} b_i = c_j = k$. On the other hand, if $(b_1, b_2, \ldots, b_j)$ is a composition of k, then the inverse construction is:

$$A = [k] \setminus \left\{ \sum_{i=1, \ldots, \ell} b_i \ \middle|\ \ell \in [j] \right\},$$

which is a subset of $[k-1]$.

So, indeed, a vincular pattern (π, A) can be equivalently defined as $(\pi, (b_1, b_2, \ldots, b_j))$, where the adjacency condition is rephrased by:

- $i_{a+1} = i_a + 1$, for all $a \in [k] \setminus \left\{ \sum_{i=1,\ldots,\ell} b_i \ \middle|\ \ell \in [j] \right\}$,

or, in other words, the first b_1 elements should be adjacent, the next b_2 also, and so on. We say the vincular pattern $(\pi, (b_1, b_2, \ldots, b_j))$ has j blocks, numbered from left to right, where $b_1, b_2, \ldots, b_j$ are

the respective block sizes. When we speak about a vincular pattern (π, A) with j blocks, it is the above bijection which is underlying.

In the literature, vincular patterns are commonly represented as permutations where some adjacent parts may be underlined, see (Kitaev, 2011, Definition 1.3.1). What is underlined are the non-trivial blocks of the pattern, *i.e.* blocks of size at least 2. This representation is visual and we will use it when we work with concrete examples. For example, we would write $\underline{23}1$ for the vincular pattern $(231, (2, 1))$, or equivalently for $(231, \{1\})$.

2.2 Vincular pattern statistics on uniform permutations

Let σ_n be a uniform random permutation of size n and let (π, A) be a vincular pattern of size k. The vincular pattern statistics for the pattern (π, A) on σ_n is a random variable counting the number of occurrences of (π, A) in σ_n. We denote it by $X^{(\pi, A)}(\sigma_n)$.

Since it is a counting statistics, $X^{(\pi, A)}(\sigma_n)$ can be naturally decomposed as a sum of indicator random variables. First, we introduce a notation to collect all sets of positions that are admissible for occurrences of (π, A) in a permutation of size n:

$$\mathcal{I}(n, k, A) := \Big\{ \{i_1, i_2, \ldots, i_k\} \in \binom{[n]}{k} \ \Big| \ i_{a+1} = i_a + 1, \text{ for all } a \in A, \text{ where } i_1 < i_2 < \cdots < i_k \Big\}.$$

Now, the sum decomposition of $X^{(\pi, A)}(\sigma_n)$ is the following:

$$X^{(\pi, A)}(\sigma_n) = \sum_{I \in \mathcal{I}(n, k, A)} X_I^{\pi}(\sigma_n), \tag{2}$$

where $X_I^{\pi}(\sigma_n)$ is 1 if σ_n has an occurrence of the (classical) pattern π at positions given by I and it is 0 otherwise.

In Bóna (2010), the reader can find this sum decomposition for other types of patterns. The difference lies in the positions over which the summation runs. In our case, the total amount of admissible positions is counted as follows.

Lemma 2.7. *For $n \geq k - j$, where $j = k - |A|$, it holds that:*

$$|\mathcal{I}(n, k, A)| = \binom{n - k + j}{j}.$$

Proof: As for patterns, a set of positions $\{i_1, i_2, \ldots, i_k\} \in \mathcal{I}(n, k, A)$, with $i_1 < i_2 < \cdots < i_k$, can be split into j blocks using the adjacency information of A. These blocks are ordered and of prescribed size. So, essentially, the set $\{i_1, i_2, \ldots, i_k\}$ is determined by the set $\{i_1, i_{c_1+1}, i_{c_2+1}, \ldots, i_{c_{j-1}+1}\}$, containing only the first position for each block (see Eq. (1) for the description of c_i). The trick is to count such sets, but we have to be careful that between the first entries of the blocks there is enough space for the whole blocks. To overcome this problem, we shift everything according to the block sizes, as it is illustrated in Fig. 1. More precisely, we associate to $\{i_1, i_2, \ldots, i_k\}$ a set of positions in $\binom{[n-k+j]}{j}$ where each element i_{c_t+1} is shifted to $i_{c_t+1} - (c_t - t)$. For example, to the set $\{3, 4, 5, 8, 9, 10, 11\} \in \mathcal{I}(11, 7, \{1, 2, 4, 6\})$ corresponding to 3 ordered blocks of sizes 3, 2 and 2, we associate the set $\{3, 6, 7\} \in \binom{[7]}{3}$, as shown in
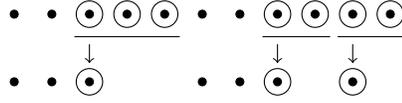
Figure 1: Illustration of the bijection.

Fig. 1. It is easy to see that this construction describes a bijection between $\mathcal{I}(n, k, A)$ and $\binom{[n-k+j]}{j}$, so that we have:

$$|\mathcal{I}(n, k, A)| = \left| \binom{[n-k+j]}{j} \right| = \binom{n-k+j}{j}. \qquad \qquad \Box$$

Remark 2.8. *The restriction $n \geq k - j$ is necessary, since otherwise the binomial coefficient is not defined, but for $n < k$, we simply have $|\mathcal{I}(n, k, A)| = 0$.*

2.3 Representation of uniform permutations

We start with the definition of *reduction*, which is given in (Janson et al., 2015, p. 1).

Definition 2.9. *Let $x_1 x_2 \ldots x_n$ be a sequence of n distinct real numbers. The* reduction *of $x_1 x_2 \ldots x_n$, which is denoted by $\mathrm{red}(x_1 x_2 \ldots x_n)$, is the unique permutation $\sigma \in S_n$ such that order relations are preserved, i.e. $\sigma_i < \sigma_j$ if and only if $x_i < x_j$ for all $i, j \in [n]$.*

As pointed out in (Janson et al., 2015, proof of Theorem 4.1), it is a standard trick to construct a uniform permutation of size n as the reduction of n independent and identically distributed (i.i.d.) random variables which are uniform in the interval $[0, 1]$.

Lemma 2.10. *Let $U_1, \ldots, U_n$ be i.i.d. random variables, uniform in $[0, 1]$. Then, $\mathrm{red}(U_1 \ldots U_n)$ is a uniform permutation of size n.*

Since we could not find a reference where Lemma 2.10 is proved, we provide a short proof here.

Proof: First, the random variables $U_1, \ldots, U_n$ are almost surely all distinct. Now, let $\sigma \in S_n$ be arbitrary. Because the random variables U_i are i.i.d., we have for any $\pi \in S_n$,

$$\begin{aligned} \mathbb{P}\big(\mathrm{red}(U_1 \ldots U_n) = \sigma \big) &= \mathbb{P}\big(\mathrm{red}(U_{\pi_1} \ldots U_{\pi_n}) = \sigma \big) \\ &= \mathbb{P}\big(\mathrm{red}(U_1 \ldots U_n)\pi = \sigma \big) \\ &= \mathbb{P}\big(\mathrm{red}(U_1 \ldots U_n) = \sigma\pi^{-1} \big), \end{aligned}$$

showing that $\mathrm{red}(U_1 \ldots U_n)$ takes all values in S_n with the same probability. So, $\mathrm{red}(U_1 \ldots U_n)$ is a uniform permutation in S_n. $\qquad \Box$

This representation of a uniform permutation is particularly adapted for our purpose since it relies on *independent* random variables.

2.4 Dependency graphs

The dependencies, and not only the pairwise dependencies, within a family of random variables can be represented in a graph as follows, see Baldi and Rinott (1989); Bóna (2010); Féray et al. (2016); Rinott (1994).

Definition 2.11. *A graph G with vertex set V is called a* dependency graph *for a family of random variables $\{X_v\}_{v \in V}$ if it satisfies the following property:*

> *If V_1 and V_2 are disjoint subsets of V which are not connected by an edge in G, then the sets of random variables $\{X_v\}_{v \in V_1}$ and $\{X_v\}_{v \in V_2}$ are independent.*

A family of random variables may have several dependency graphs, capturing sometimes more, sometimes less dependency information. The least information is contained in the complete graph, which is always a dependency graph. As pointed out in Féray et al. (2016), dependency graphs are often used to work with sums of *partly* dependent random variables. For example, $X^{(\pi,A)}(\sigma_n)$ decomposed as in Eq. (2) falls in this category. In Section 3, we will see two methods to show asymptotic normality of $X^{(\pi,A)}(\sigma_n)$ using the concept of dependency graphs.

2.5 Law of total variance

The following decomposition formula for the variance can be found in (Weiss, 2005, p. 385–386).

Proposition 2.12. *For two random variables Y and J, defined on the same probability space, the following holds:*

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|J)] + \mathrm{Var}(\mathbb{E}[Y|J]).$$

This is called the *law of total variance*. It relates the variance of a random variable to its conditional variance and its conditional expectation. The proof uses the *tower property*: $\mathbb{E}[\mathbb{E}[Y|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[Y|\mathcal{H}]$ with $\mathcal{H} = \{\emptyset, \Omega\}$ and $\mathcal{G} = \sigma(J)$, the smallest σ-algebra such that J is measurable.

It is a natural question if a statement similar to Proposition 2.12 holds for the conditional variance. The equivalent result is called the *law of total conditional variance* and its proof uses the tower property with $\mathcal{H} = \sigma(J_1)$ and $\mathcal{G} = \sigma(J_1, J_2)$.

Proposition 2.13. *For three random variables Y, J_1 and J_2, defined on the same probability space, the following holds:*

$$\mathrm{Var}(Y|J_1) = \mathbb{E}\big[\mathrm{Var}(Y|J_1, J_2)\big|J_1\big] + \mathrm{Var}\big(\mathbb{E}[Y|J_1, J_2]\big|J_1\big).$$

A similar expression can be found for $\mathrm{Var}(Y|J_1, J_2)$, the inner conditional variance term in the above equation, and also for more conditioning random variables. Starting from Proposition 2.12 and iteratively using these expressions for the inner conditional variance terms, we can obtain an expression for the variance involving several random variables as conditions.

Proposition 2.14. *For some random variables Y and $J_1, J_2, \ldots, J_m$, all defined on the same probability space, the following holds:*

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|J_1, \ldots, J_m)] + \sum_{i=0,\ldots,m-1} \mathbb{E}\big[\mathrm{Var}\big(\mathbb{E}[Y|J_1, \ldots, J_{i+1}]\big|J_1, \ldots, J_i\big)\big].$$

Observe that the summand for $i = 0$ simplifies to $\mathrm{Var}(\mathbb{E}[Y|J_1])$.

The only reference we could find for these formulas is (Bowsher and Swain, 2012, Eqs. 13 and 15), but it is likely that they appeared independently in other sources. This general decomposition formula for the variance will be very helpful in Section 4.2, where we need to find a lower bound for the variance of $X^{(\pi,A)}(\sigma_n)$.

2.6 Asymptotic notation

Since we will speak about the asymptotic behaviour of functions, we recall some standard notation.

Definition 2.15. *Let* $f, g : \mathbb{N} \to \mathbb{R}$ *be two functions. We say that:*

- *f is $\mathcal{O}(g)$, written $f = \mathcal{O}(g)$, if: $\exists\, C > 0, n_0$ such that $\forall\, n \geq n_0,\ |f(n)| \leq C|g(n)|$,*

- *f is of* order *g, written $f \asymp g$, if: $f = \mathcal{O}(g)$ and $g = \mathcal{O}(f)$,*

- *f is* asymptotically equivalent *to g, written $f \sim g$, if: $\displaystyle\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.*

Remark 2.16.   *1. If there exists C such that $f \sim Cn^k$, then $f \asymp n^k$.*

2. Let p be a polynomial in n. If $p \asymp n^k$, then there exists C such that $p \sim Cn^k$.

2.7 Stein's method for central limit theorems

Stein's method is a technique invented by Charles Stein to bound the distance between two probability distributions. It is used to prove central limit theorems as well as approximation by the Poisson, exponential and other distributions. The survey article Ross (2011) gives an overview of applications to different distributions and discusses methods to further analyse the bounds. We recall that applications of Stein's method to permutation patterns can be found in Crane and DeSalvo (2017); Crane et al. (2016); Fulman (2004); Goldstein (2005). To illustrate the main concept of Stein's method for central limit theorems, we follow (Ross, 2011, p. 6–9).

First, recall that the *Kolmogorov distance* is a probability metric defined as follows, see (Chen and Röllin, 2010, p. 9), (Ross, 2011, p. 5).

Definition 2.17. *For two $\mathbb{R}$-valued random variables X and Y, let F_X and F_Y be their distribution functions. The Kolmogorov distance between X and Y is defined as:*

$$d_K(X, Y) = \sup_{t \in \mathbb{R}} |F_X(t) - F_Y(t)|.$$

In particular, since convergence of the distribution functions in all continuity points implies convergence in distribution (denoted $\xrightarrow{d}$), we have the following.

Lemma 2.18. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}$-valued random variables and let Y be some $\mathbb{R}$-valued random variable. Then:*

$$d_K(X_n, Y) \underset{n\to\infty}{\to} 0 \implies X_n \xrightarrow{d} Y.$$

In general, Stein's method proves convergence in distribution using a functional equation. The next lemma can be used to prove a central limit theorem (CLT), see (Ross, 2011, Lemma 2.1).

Lemma 2.19 (Stein's Lemma). *Define the functional operator $\mathcal{A}$ by*

$$\mathcal{A}f(x) = f'(x) - xf(x).$$

Then, for a real-valued random variable X, the following are equivalent.

1. *X has the standard normal distribution.*

2. *For all absolutely continuous functions $f : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[|f'(X)|] < \infty$, it holds that $\mathbb{E}[\mathcal{A}f(X)] = 0$.*

The first direction (1) $\implies$ (2) is simply integration by parts. More interesting is the second direction. Let Z be a standard normal random variable, F_Z its distribution function and $\mathbb{I}_a$ the indicator function which is 1 if and only if condition a holds. In (Ross, 2011, Lemma 2.2) it is shown that for any $t \in \mathbb{R}$, there exists a unique bounded solution f_t of the differential equation

$$f'_t(x) - xf_t(x) = \mathbb{I}_{x \le t} - F_Z(t). \tag{3}$$

Taking X to be any real-valued random variable, we obtain the equation

$$|\mathbb{E}[f'_t(X) - Xf_t(X)]| = |F_X(t) - F_Z(t)|. \tag{4}$$

Hence, the maximum distance between the distribution functions of X and Z is given by

$$d_K(X, Z) = \sup_{t \in \mathbb{R}} |F_X(t) - F_Z(t)| = \sup_{t \in \mathbb{R}} |\mathbb{E}[f'_t(X) - Xf_t(X)]|.$$

By Lemma 2.18, to prove the second direction (2) $\implies$ (1), it is enough that (2) holds for all f_t with $t \in \mathbb{R}$ which are solutions to Eq. (3).

To prove a CLT, we prove that $(X_n)_{n \in \mathbb{N}}$ almost satisfies the functional equation in Stein's Lemma. We use Eq. (4) to bound $d_K(X_n, Z)$ which also quantifies the rate of convergence. In practice, Stein's method is useful because there exist various techniques to estimate the quantity $\mathbb{E}[f'_t(X) - Xf_t(X)]$: *dependency graphs*, *exchangeable pairs*, *zero- and size-bias couplings*. Theorem 3.5 used in Approach I relies on the dependency graph method which is often useful if the random variable X is a sum of partially dependent random variables like $X^{(\pi,A)}(\sigma_n)$. However, Stein's method may also be used for random variables with different structures.

We conclude this short summary of Stein's method with two remarks.

Remark 2.20. *1. The* Stein-Chen method *is a version developed for Poisson approximation, see Barbour et al. (1992); Crane and DeSalvo (2017).*

2. *It is sometimes easier to study other probability metrics than the Kolmogorov distance. For normal approximation, often the Wasserstein distance is studied, see Rinott (1994). In Crane and DeSalvo (2017), the total variation distance is studied for Poisson approximation.*

3 Central limit theorem

In this section, we prove a central limit theorem (CLT) for our random variable $X^{(\pi,A)}(\sigma_n)$, the vincular pattern statistics on uniform permutations. Let (π, A) be a fixed vincular pattern and assume $n \geq |\pi| \geq 2$ (the other cases are trivial). We normalize $X^{(\pi,A)}(\sigma_n)$:

$$\overline{X^{(\pi,A)}(\sigma_n)} = \frac{X^{(\pi,A)}(\sigma_n) - e_n}{\sqrt{v_n}},$$

where $v_n = \mathrm{Var}(X^{(\pi,A)}(\sigma_n))$, $e_n = \mathbb{E}[X^{(\pi,A)}(\sigma_n)]$. The following theorem is the main result of this article, where $\mathcal{N}(0,1)$ denotes the standard normal distribution.

Theorem 3.1. *Let $Z \sim \mathcal{N}(0,1)$, let (π, A) be a fixed vincular pattern and for any n, let σ_n be uniform in S_n. Then, there exists $C > 0, n_0$ such that for $n \geq n_0$:*

$$d_K\left(\overline{X^{(\pi,A)}(\sigma_n)}, Z\right) \leq C n^{-1/2}.$$

Consequently, it holds that:

$$\overline{X^{(\pi,A)}(\sigma_n)} \xrightarrow{d} Z.$$

Remark 3.2. *The second claim of Theorem 3.1 follows from the first claim and Lemma 2.18.*

Note that Theorem 3.1 is not only a CLT result. It also contains information about the speed of convergence, measured in the metric d_K. We present two different approaches to prove Theorem 3.1. One works with Stein's method and the other one with cumulants, but both have in common that they use dependency graphs. Before we start the two different proofs, we give a dependency graph for our problem.

3.1 Dependency graph for the $X^{\pi}_I(\sigma_n)$'s

We need the following observation about the dependencies between the $X^{\pi}_I(\sigma_n)$'s from the sum decomposition of $X^{(\pi,A)}(\sigma_n)$ (see Eq. (2), p. 7). In (Crane and DeSalvo, 2017, Lemma 5.3), this is called the property of *dissociation*.

Lemma 3.3. *Let (π, A) be a vincular pattern with $|\pi| = k$ and let $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{I}(n, k, A)$. If $\mathcal{F}_1$ and $\mathcal{F}_2$ are such that $\bigcup_{I \in \mathcal{F}_1} I$ and $\bigcup_{I \in \mathcal{F}_2} I$ are disjoint, then the corresponding families of indicators, $\{X^{\pi}_I(\sigma_n)\}_{I \in \mathcal{F}_1}$ and $\{X^{\pi}_I(\sigma_n)\}_{I \in \mathcal{F}_2}$, are independent.*

Proof: Let $U_1, \ldots, U_n$ be independent and uniform in $[0, 1]$. By Lemma 2.10, we can represent a uniform permutation as $\mathrm{red}(U_1 \ldots U_n)$. Since pattern occurrence in non-intersecting subsequences depends on disjoint subsets of the set $\{U_1, \ldots, U_n\}$, the independence of these subsets proves the independence of the corresponding families of indicators. □

With the help of Lemma 3.3, we can now construct a dependency graph for the family of indicator random variables $\{X^{\pi}_I(\sigma_n)\}_{I \in \mathcal{I}(n,k,A)}$, where $k = |\pi|$. We define its vertex set V_n and its edge set E_n as follows:

$$\begin{cases} V_n = \mathcal{I}(n, k, A), \\ E_n = \{\{I_1, I_2\} \subseteq \mathcal{I}(n, k, A) \mid I_1 \neq I_2, \ I_1 \cap I_2 \neq \emptyset\}. \end{cases} \tag{5}$$

If $\mathcal{F}_1$ and $\mathcal{F}_2$ are disjoint subsets of $\mathcal{I}(n, k, A)$ that are not connected by an edge in the graph, then by construction, Lemma 3.3 applies to $\mathcal{F}_1$ and $\mathcal{F}_2$, ensuring that $\{X_I^\pi(\sigma_n)\}_{I \in \mathcal{F}_1}$ and $\{X_I^\pi(\sigma_n)\}_{I \in \mathcal{F}_2}$ are independent. Hence, the dependency graph condition is fulfilled.

An important parameter is the maximal degree of the dependency graph.

Lemma 3.4. *For a fixed vincular pattern* (π, A) *with* j *blocks, let* $D - 1$ *be the maximal degree of the dependency graph given by* (5). *Then:*

$$D \asymp n^{j-1}.$$

Proof: Let $k = |\pi|$. In the proof of Lemma 2.7, we have counted $|\mathcal{I}(n, k, A)|$. It is of order n^j. Now, for any fixed vertex $H \in \mathcal{I}(n, k, A)$, consider the quantity $|\{I \in \mathcal{I}(n, k, A)| \ I \cap H \neq \emptyset\}|$. It can be bounded from above and from below as follows:

$$\max_{h \in H} |\{I \in \mathcal{I}(n, k, A)| \ h \in I\}| \leq |\{I \in \mathcal{I}(n, k, A)| \ I \cap H \neq \emptyset\}|$$

$$\leq \sum_{h \in H} |\{I \in \mathcal{I}(n, k, A)| \ h \in I\}|,$$

with

$$|\{I \in \mathcal{I}(n, k, A)| \ h \in I\}| = \sum_{\ell \in [j]} |\{I \in \mathcal{I}(n, k, A)| \ h \text{ is in the } \ell\text{-th block of } I\}|.$$

The quantity $|\{I \in \mathcal{I}(n, k, A)| \ h \text{ is in the } \ell\text{-th block of } I\}|$ is counted similarly to $|\mathcal{I}(n, k, A)|$. The constraint "h is in the ℓ-th block of I" means that $i_{c_\ell + 1}$ may admit only a *finite* set of values (see proof of Lemma 2.7, setting $c_0 = 0$). Compared to $|\mathcal{I}(n, k, A)|$, not j but only $j - 1$ blocks are free which decreases the order from n^j to n^{j-1}. Since the number of terms in the two sums is independent of n, we have:

$$D = \max_{H \in \mathcal{I}(n, k, A)} |\{I \in \mathcal{I}(n, k, A)| \ I \neq H, \ I \cap H \neq \emptyset\}| + 1 \asymp n^{j-1}. \qquad \square$$

The dependency graph we just constructed will be used in the next two sections to prove Theorem 3.1.

3.2   Approach I: Dependency graphs and Stein's method

The following theorem can be obtained from results of Chen and Röllin (2010) and Ross (2011).

Theorem 3.5. *Let* $Z \sim \mathcal{N}(0, 1)$. *Let* G *be a dependency graph for* $\{X_i\}_{i=1}^N$ *and* $D - 1$ *be the maximal degree of* G. *Assume there is a constant* $B > 0$ *such that* $|X_i - \mathbb{E}[X_i]| \leq B$ *for all* i. *Then, for* $W = \frac{1}{\sigma} \sum_{i=1}^N (X_i - \mathbb{E}[X_i])$, *where* σ^2 *is the variance of the sum, it holds that:*

$$d_K(W, Z) \leq \frac{8B^2 D^{3/2} N^{1/2}}{\sigma^2} + \frac{8B^3 D^2 N}{\sigma^3}.$$

In particular, if $\sigma^2 \asymp B^2 DN$ *(or* $\sigma^2 \geq CB^2 DN$ *for some constant* $C > 0$*), then:*

$$d_K(W, Z) = \mathcal{O}\left(\sqrt{\frac{D}{N}}\right).$$

Proof: Without loss of generality, we assume that $\mathbb{E}[X_i] = 0$ for all i. Let

$$A_i = \{i\} \cup \{1 \leq j \leq N | \text{ vertices } i \text{ and } j \text{ are connected in } G\}.$$

Clearly, $|A_i| \leq D$ for all i. Using (Chen and Röllin, 2010, Construction 2B), we obtain a Stein coupling for W so that with (Chen and Röllin, 2010, Corollary 2.6) (for $\alpha = \frac{NB}{\sigma}$, $\beta = \frac{DB}{\sigma}$), we have:

$$d_K(W, Z) \leq \frac{2}{\sigma^2} \sqrt{\mathrm{Var}\left(\sum_{i=1}^{n} \sum_{j \in A_i} X_i X_j \right)} + \frac{8B^3 D^2 N}{\sigma^3}. \tag{6}$$

Under the assumption of a dependency graph, from the end of the proof of (Ross, 2011, Theorem 3.5), we have that:

$$\mathrm{Var}\left(\sum_{i=1}^{N} \sum_{j \in A_i} X_i X_j \right) \leq 13 D^3 \sum_{i=1}^{N} \mathbb{E}[X_i^4] \leq 13 B^4 D^3 N. \tag{7}$$

The final bound follows from Eqs. (6) and (7). $\qquad\square$

We apply Theorem 3.5 to our problem.

Proof of Theorem 3.1 (variant I): Let (π, A) have j blocks and let $k = |\pi|$. Consider the family of random variables $\{X_I^\pi(\sigma_n)\}_{I \in \mathcal{I}(n,k,A)}$ and the dependency graph constructed for it in (5). Denote by N the size of the family and denote by $D - 1$ the maximal degree of the dependency graph. Set $v_n = \mathrm{Var}(X^{(\pi,A)}(\sigma_n))$. From Lemma 2.7 and Lemma 3.4, we have:

$$N \asymp n^j, \quad D \asymp n^{j-1}.$$

We will see in Theorem 4.1 that we have:

$$v_n \asymp n^{2j-1}.$$

The proof being technical, it is postponed to Section 4. Clearly, it holds that $v_n \asymp DN$. Moreover, for all $I \in \mathcal{I}(n, k, A)$, we have:

$$|X_I^\pi(\sigma_n) - \mathbb{E}[X_I^\pi(\sigma_n)]| \leq 1.$$

Using Theorem 3.5 with $B = 1$, we obtain:

$$d_K\left(\overline{X^{(\pi,A)}(\sigma_n)}, Z \right) = \mathcal{O}\left(\sqrt{\frac{n^{j-1}}{n^j}} \right) = \mathcal{O}(n^{-1/2}). \qquad\square$$

Remark 3.6. *Almost the same bound as in Theorem 3.5 (giving also a $\mathcal{O}(n^{-1/2})$ in our case) has been obtained by Y. Rinott (Rinott, 1994, Theorem 2.2).*

Remark 3.7. *Using (Ross, 2011, Theorem 3.5), the same bound as in Theorem 3.1 can be obtained for the Wasserstein distance d_W, which is defined in (Ross, 2011, p. 5). The bound on d_W would then give a bound on d_K, however not an equally good one as Theorem 3.1 ($n^{-1/4}$ instead of $n^{-1/2}$).*

3.3 Approach II: Dependency graphs and cumulants

For any random variable X, denote by $\kappa^{(r)}(X)$ its r-th cumulant. As in (Saulis and Statulevičius, 1991, p. 16), we say that X satisfies condition $(S_{\gamma,\Delta})$ for some $\gamma \geq 0$, $\Delta > 0$ if:

$$|\kappa^{(r)}(X)| \leq \frac{(r!)^{1+\gamma}}{\Delta^{r-2}}, \text{ for all } r \geq 3.$$

The following result can be found in (Saulis and Statulevičius, 1991, Corollary 2.1).

Theorem 3.8. *Let* $Z \sim \mathcal{N}(0,1)$*. For any random variable* X *satisfying condition* $(S_{\gamma,\Delta})$*, it holds that:*

$$d_K(X, Z) \leq \frac{108}{\left(\Delta \frac{\sqrt{2}}{6}\right)^{\frac{1}{1+2\gamma}}}.$$

To prove that condition $(S_{\gamma,\Delta})$ is satisfied by our random variable $X^{(\pi,A)}(\sigma_n)$, we use the following result from (Féray et al., 2016, p. 71) giving a bound on cumulants of sums of random variables. A slightly weaker version has been established by S. Janson (Janson, 1988, Lemma 4), see also (Féray et al., 2016, p. 71).

Theorem 3.9. *Let* $\{X_v\}_{v \in V}$ *be a family of random variables with dependency graph* G*. Denote by* N *the number of vertices of* G *and by* $D-1$ *the maximal degree of* G*. Assume that the* X_v*'s are uniformly bounded by a constant* B*. Then, if* $X = \sum_{v \in V} X_v$*, for any integer* $r \geq 1$*, one has:*

$$|\kappa^{(r)}(X)| \leq 2^{r-1} r^{r-2} N D^{r-1} B^r.$$

With the help of Theorem 3.9, we apply Theorem 3.8 to our problem.

Proof of Theorem 3.1 (variant II): Let (π, A) have j blocks and let $k = |\pi|$. Consider the normalized indicator random variables $X_I^\pi(\sigma_n)$:

$$\overline{X_I^\pi(\sigma_n)} = \frac{X_I^\pi(\sigma_n) - e_I}{\sqrt{v_n}},$$

where $v_n = \text{Var}(X^{(\pi,A)}(\sigma_n))$, $e_I = \mathbb{E}[X_I^\pi(\sigma_n)]$. Clearly, $\overline{X^{(\pi,A)}(\sigma_n)}$ is the sum of the $\overline{X_I^\pi(\sigma_n)}$'s. It is easy to see that the dependency graph constructed in (5) is also a dependency graph for the family $\{\overline{X_I^\pi(\sigma_n)}\}_{I \in \mathcal{I}(n,k,A)}$. Denote by N its number of vertices and denote by $D-1$ its maximal degree. By Lemma 2.7 and Lemma 3.4, there exist $C_1, C_2 > 0$ such that:

$$N \leq C_1 n^j, \quad D \leq C_2 n^{j-1}.$$

By Theorem 4.1 (whose proof is postponed), we have $v_n \asymp n^{2j-1}$. Since $|X_I^\pi(\sigma_n) - e_I| \leq 1$, there exists $C_3 > 0$ such that for all $I \in \mathcal{I}(n, k, A)$:

$$|\overline{X_I^\pi(\sigma_n)}| \leq C_3 n^{1/2-j}.$$

We now use Theorem 3.9 to estimate the cumulants of $\overline{X^{(\pi,A)}(\sigma_n)}$. We use the following simple inequality for factorials: $r^r \leq r!e^r$, valid for $r \geq 1$. For any $r \geq 3$, we obtain that there exists $C > 0$ such that:

$$\left| \kappa^{(r)}\left(\overline{X^{(\pi,A)}(\sigma_n)} \right) \right| \leq 2^{r-1} r^{r-2} (C_1 n^j)(C_2 n^{j-1})^{r-1}(C_3 n^{1/2-j})^r$$

$$\leq C^r r! n^{-(r-2)/2}$$

$$\leq \frac{r!}{\Delta^{r-2}},$$

where $\Delta = C^{-3} n^{1/2}$. Here we used that $r/(r-2) \leq 3$ for all $r \geq 3$. Since condition $(S_{\gamma,\Delta})$ is satisfied with Δ and $\gamma = 0$, by Theorem 3.8, we obtain:

$$d_K\left(\overline{X^{(\pi,A)}(\sigma_n)}, Z \right) = \mathcal{O}(n^{-1/2}). \qquad \square$$

4 Variance estimate

The main result of this section is the following theorem about the asymptotic behaviour of the variance of $X^{(\pi,A)}(\sigma_n)$ (we still assume $|\pi| \geq 2$). This result has been used in both proofs of Theorem 3.1 given above.

Theorem 4.1. *For a vincular pattern (π, A) with j blocks and for σ_n uniform in S_n, there exists $C > 0$ such that:*

$$\mathrm{Var}(X^{(\pi,A)}(\sigma_n)) \sim Cn^{2j-1}.$$

The proof consists in two steps. First, we show in Section 4.1 that $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ is a polynomial in n of degree at most $2j-1$. This immediately implies that Cn^{2j-1} is an upper bound for the variance (see Corollary 4.4). The second step is to find a lower bound of the same form (see Proposition 4.7). The lower bound is more important for the CLT result, but it does not follow from the polynomiality. To find it, we present in Section 4.2 a proof technique building a recurrence from the law of total variance.

4.1 Polynomiality and upper bound

Using the sum decomposition of $X^{(\pi,A)}(\sigma_n)$ (see Eq. (2), p. 7) and Lemma 3.3, we have:

$$\mathrm{Var}(X^{(\pi,A)}(\sigma_n)) = \sum_{\substack{I,J \in \mathcal{I}(n,k,A): \\ I \cap J \neq \emptyset}} \mathrm{Cov}(X_I^{\pi}(\sigma_n), X_J^{\pi}(\sigma_n)). \qquad (8)$$

By Lemma 3.3, the covariances are 0 for any $I, J \in \mathcal{I}(n, k, A)$ that do not intersect, explaining the summation index in the above formula.

We use this expression to prove that $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ is a polynomial in n.

Lemma 4.2. *Let (π, A) be a fixed vincular pattern of size k with j blocks and, for $n \geq 1$, let σ_n be uniform in S_n. Then, for $n \geq 2(k-j)$, $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ is a polynomial in n whose degree is at most $2j-1$.*

Proof: The proof idea is to split the sum in Eq. (8) according to I, J that have the same covariances, and then to count the number of pairs (I, J) in each of these covariance-groups. We will see that the cardinalities of all those groups are polynomials in n and that the number of groups and the covariance values do not depend on n. Then, this implies that also $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ is a polynomial in n.

Let $I, J \in \mathcal{I}(n, k, A)$, $I = \{i_1, i_2, \ldots, i_k\}$ with $i_1 < i_2 < \cdots < i_k$ and $J = \{j_1, j_2, \ldots, j_k\}$ with $j_1 < j_2 < \cdots < j_k$. To split the sum, we consider $I \cup J = \{u_1, u_2, \ldots, u_t\}$ with $u_1 < u_2 < \cdots < u_t$ and $k \le t \le 2k$. Note that $t = 2k$ if and only if $I \cap J = \emptyset$. Depending on the intersections between I and J, $I \cup J$ looks different, see Fig. 2. Each u_s in $I \cup J$ comes either from I, or from J, or from $I \cap J$ here
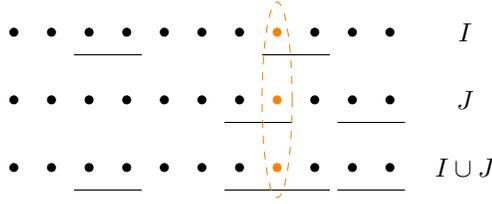
Figure 2: Block structure of $I \cup J$.

shown in orange. The following function $f_{I,J} : [t] \to \{I, J, (I, J)\}$ gives to each position s the origin of u_s:

$$
f_{I,J}(s) = \begin{cases} I & \text{if } u_s \in I \text{ but } u_s \notin J \\ J & \text{if } u_s \in J \text{ but } u_s \notin I \\ (I, J) & \text{if } u_s \in I \cap J \end{cases} .
$$

Since σ_n is uniform, $\mathrm{Cov}(X_I^\pi(\sigma_n), X_J^\pi(\sigma_n))$ depends on I, J only through the positions of the intersections of I and J. In the example of Fig. 2, I and J intersect at one position only: position 3 in I and 2 in J. In particular, when two functions $f_{I,J}$ and $f_{I',J'}$ are the same, then the covariances $\mathrm{Cov}(X_I^\pi(\sigma_n), X_J^\pi(\sigma_n))$ and $\mathrm{Cov}(X_{I'}^\pi(\sigma_n), X_{J'}^\pi(\sigma_n))$ are the same. We split the sum:

$$
\sum_{\substack{I,J \in \mathcal{I}(n,k,A): \\ I \cap J \ne \emptyset}} \mathrm{Cov}(X_I^\pi(\sigma_n), X_J^\pi(\sigma_n)) = \sum_{t=k}^{2k-1} \sum_{f:[t] \to \{I,J,(I,J)\}} \sum_{\substack{I,J \in \mathcal{I}(n,k,A): \\ |I \cup J| = t, \ f_{I,J} = f}} \mathrm{Cov}(X_I^\pi(\sigma_n), X_J^\pi(\sigma_n)).
$$

Each pair (t, f) defines a covariance-group. The number of such pairs and the covariance associated to a pair (t, f) are independent of n. To count the size of the groups, we consider the blocks of $I \cup J$. Similarly to patterns, I and J come in blocks. The inherited block structure of $I \cup J$ has three different types of blocks: blocks from I, blocks from J and *merged* blocks, coming both from I and J. In Fig. 2, the block of size 3 in $I \cup J$ is a merged block. Instead of blocks, it is equivalent to describe the adjacencies of $I \cup J$. As the blocks, they are inherited from I, or from J, or both from I and J. Neighbouring (as in Fig. 2) or shared adjacencies from I and J take care of the block merges. If $u_s \in I$, then define $r_I(s)$ as the index ℓ such that $u_s = i_\ell$ and similarly for r_J. For a function $f : [t] \to \{I, J, (I, J)\}$, we define the set

$A_f \subseteq [t-1]$ by:

$$s \in A_f \iff \begin{cases} f(s) \in \{I, (I,J)\} \text{ and } r_I(s) \in A, \\ \text{or } f(s) \in \{J, (I,J)\} \text{ and } r_J(s) \in A. \end{cases}$$

If $I, J \in \mathcal{I}(n, k, A)$ with $|I \cup J| = t$ and $f_{I,J} = f$, then A_f is the set of adjacencies of $I \cup J$. More precisely, then $I \cup J \in \mathcal{I}(n, t, A_f)$. But not every f can occur this way since it has to respect the block structure of I and J given by A. For example, $f(s) = I$ or $f(s) = (I,J)$ with $r_I(s) \in A$ but $f(s+1) = J$ for some s, is not valid. For a fixed pair (t, f), one of the two following cases occurs:

- There are no $I, J \in \mathcal{I}(n, k, A)$ with $f_{I,J} = f$.

- The map $\begin{cases} \{(I,J) \in \mathcal{I}(n,k,A)^2 \mid f_{I,J} = f\} & \to \mathcal{I}(n,t,A_f), \\ (I,J) & \mapsto I \cup J, \end{cases}$ is a bijection.
  Indeed, if $K \in \mathcal{I}(n, t, A_f)$, then (I, J) can be reconstructed from $K = I \cup J$ and $f = f_{I,J}$. By construction of A_f, I and J will be in $\mathcal{I}(n, k, A)$.

So, the cardinalities of the covariance-groups are either 0 or given by the cardinality of the corresponding $\mathcal{I}(n, t, A_f)$. By Lemma 2.7, for $n \geq t - (t - |A_f|) = |A_f|$, we have:

$$|\mathcal{I}(n, t, A_f)| = \binom{n - t + (t - |A_f|)}{t - |A_f|} = \binom{n - |A_f|}{t - |A_f|}.$$

This is a polynomial in n, since t and $|A_f|$ do not depend on n. The maximal value $|A_f|$ can take is $2(k - j)$. So, for $n \geq 2(k - j)$, $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ is the sum of polynomials in n, which is again a polynomial in n. Its maximal degree is the maximal value for $t - |A_f|$ (the number of blocks of $I \cup J$) which is $2j - 1$. $\square$

Remark 4.3. *This polynomiality result can be used to compute expressions of the variance for small patterns by polynomial interpolation (whence the desire to be precise on the range of values of n for which the polynomiality holds).*

From Lemma 4.2, we directly obtain an upper bound for $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$.

Corollary 4.4. *Let (π, A) be a fixed vincular pattern of size k with j blocks and let σ_n be uniform in S_n. Then, there exists $C > 0, n_0$ such that for $n \geq n_0$:*

$$\mathrm{Var}(X^{(\pi,A)}(\sigma_n)) \leq Cn^{2j-1}.$$

4.2   Lower bound

Our proof technique for finding a sharp lower bound on $\mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ uses a recurrence that we obtain from the law of total variance. Working directly with the variance decomposition (see Eq. (8)) would be more difficult since covariances can be negative whereas the law of total variance involves only non-negative terms. We first discuss what conditioning we want to use in the law of total variance. Then, we show how to obtain the recurrence relation. And finally, we deduce a recursive estimation from which we then derive the lower bound.

For the rest of this section, let (π, A) be a fixed vincular pattern of size k with j blocks. Let $U_1, U_2, \ldots$ be independent and uniform in $[0, 1]$. For any n, set $\sigma_n = \mathrm{red}(U_1 \ldots U_n)$. By Lemma 2.10, σ_n is uniform in S_n. Moreover, for any n, we define $v_n = \mathrm{Var}(X^{(\pi, A)}(\sigma_n))$. For simplicity, we also set $Y = X^{(\pi, A)}(\sigma_n)$.

We apply the general law of total variance (see Proposition 2.14) on Y where we shall condition on the last few entries of σ_n. The number of these entries is the size of the last block of the pattern (π, A), denoted b_j. More precisely, we condition on $U_n, U_{n-1}, \ldots, U_{n-b_j+1}$. We obtain the following expression for v_n:

$$\begin{aligned} v_n = \mathrm{Var}(Y) &= \mathbb{E}[\mathrm{Var}(Y|U_n, \ldots, U_{n-b_j+1})] \\ &+ \sum_{i=0,\ldots,b_j-1} \mathbb{E}\big[\,\mathrm{Var}\,\big(\mathbb{E}[Y|U_n, \ldots, U_{n-i}]\big|U_n, \ldots, U_{n-i+1}\big)\big]. \end{aligned} \tag{9}$$

We now turn to the recurrence where we will see why this conditioning is a good choice. We split Y in two parts as follows: $Y = B + C$ with

$$B = \sum_{I \in \mathcal{I}(n,k,A):\, n \notin I} X_I^\pi(\sigma_n), \quad C = \sum_{I \in \mathcal{I}(n,k,A):\, n \in I} X_I^\pi(\sigma_n).$$

Observe that $B = X^{(\pi, A)}(\sigma_{n-1})$, since σ_n is uniform in S_n and σ_{n-1} is uniform in S_{n-1}. Applying Proposition 2.14 on B, with the conditions $U_{n-1}, \ldots, U_{n-b_j+1}$, we have:

$$\begin{aligned} v_{n-1} = \mathrm{Var}(B) &= \mathbb{E}[\mathrm{Var}(B|U_{n-1}, \ldots, U_{n-b_j+1})] \\ &+ \sum_{i=0,\ldots,b_j-2} \mathbb{E}\big[\,\mathrm{Var}\,\big(\mathbb{E}[B|U_{n-1}, \ldots, U_{n-i-1}]\big|U_{n-1}, \ldots, U_{n-i}\big)\big]. \end{aligned} \tag{10}$$

In the expression for v_n, Eq. (9), we want to recover v_{n-1} from $\mathbb{E}[\mathrm{Var}(Y|U_n, \ldots, U_{n-b_j+1})]$. Since $Y = B + C$ and since B is independent of U_n, we have:

$$\begin{aligned} \mathbb{E}[\mathrm{Var}(Y|U_n, \ldots, U_{n-b_j+1})] &= \mathbb{E}[\mathrm{Var}(B|U_{n-1}, \ldots, U_{n-b_j+1})] \\ &+ \mathbb{E}[\mathrm{Var}(C|U_n, \ldots, U_{n-b_j+1})] \\ &+ 2\mathbb{E}[\mathrm{Cov}(B, C|U_n, \ldots, U_{n-b_j+1})]. \end{aligned} \tag{11}$$

Because $\mathbb{E}[\mathrm{Var}(B|U_{n-1}, \ldots, U_{n-b_j+1})]$ also appears in the expression for v_{n-1}, using Eqs. (9), (10) and (11), we obtain the following recurrence relation:

$$\begin{aligned} v_n - v_{n-1} =\; &2\mathbb{E}[\mathrm{Cov}(B, C|U_n, \ldots, U_{n-b_j+1})] \\ &+ \mathbb{E}[\mathrm{Var}(C|U_n, \ldots, U_{n-b_j+1})] \\ &+ \Big(\sum_{i=0,\ldots,b_j-2} \mathbb{E}\big[\,\mathrm{Var}\,\big(\mathbb{E}[Y|U_n, \ldots, U_{n-i}]\big|U_n, \ldots, U_{n-i+1}\big)\big] \\ &\quad - \sum_{i=0,\ldots,b_j-2} \mathbb{E}\big[\,\mathrm{Var}\,\big(\mathbb{E}[B|U_{n-1}, \ldots, U_{n-i-1}]\big|U_{n-1}, \ldots, U_{n-i}\big)\big]\Big) \\ &+ \mathbb{E}\big[\,\mathrm{Var}\,\big(\mathbb{E}[Y|U_n, \ldots, U_{n-b_j+1}]\big|U_n, \ldots, U_{n-b_j+2}\big)\big]. \end{aligned} \tag{12}$$

The right-hand side of Eq. (12) is grouped in four terms. To find a lower bound with the help of this recurrence relation, we examine all these terms. We will need the following result which can be proved very similarly to Lemma 3.3.

Lemma 4.5. *Let (π, A) be a vincular pattern of size k with j blocks and a last block of size b_j. Let $U_1, \ldots, U_n$ be independent and uniform in $[0, 1]$ and let $\sigma_n = \mathrm{red}(U_1 \ldots U_n)$. Then, for $I, J \in \mathcal{I}(n, k, A)$ with $I \cap J \subseteq \{n - b_j + 1, \ldots, n\}$, conditionally on $U_{n-b_j+1}, \ldots, U_n$, the random variables $X_I^{(\pi, A)}(\sigma_n)$ and $X_J^{(\pi, A)}(\sigma_n)$ are independent.*

Lemma 4.5 is the reason why we want to condition on the last b_j entries of σ_n. Moreover, we need the following definition in order to work with sorted sequences.

Definition 4.6. *Let $x_1 x_2 \ldots x_n$ be a sequence of n distinct real numbers. Then, $\mathrm{sort}(x_1 \ldots x_n)$ is the* sorted *sequence which contains $x_1, x_2, \ldots, x_n$ but in increasing order.*

We now examine separately the four terms of the recurrence relation (see Eq. (12)). We want to find lower bounds for each of them. The following computations hold for n large enough.

First term   First, we have:

$$\mathrm{Var}(C|U_n, \ldots, U_{n-b_j+1}) = \sum_{I, J \in \mathcal{I}(n,k,A): \, n \in I \cap J} \mathrm{Cov}(X_I^\pi(\sigma_n), X_J^\pi(\sigma_n)|U_n, \ldots, U_{n-b_j+1}).$$

The constraint $n \in I \cap J$ means that I and J intersect at least in the whole last block. By Lemma 4.5, for the above covariances to be non-zero, I and J must intersect at least in one more block. Similar arguments as in the proof of Lemma 3.4 show that the number of non-zero covariances is $\mathcal{O}(n^{2j-3})$. Since the covariances are bounded by 1, this implies:

$$\mathrm{Var}(C|U_n, \ldots, U_{n-b_j+1}) = \mathcal{O}(n^{2j-3}).$$

Note that the constant in the $\mathcal{O}$-term does not depend on $U_n, \ldots, U_{n-b_j+1}$, which is important when we take the expectation. Then, by the Cauchy-Schwarz inequality, it holds that:

$$|\mathrm{Cov}(B, C|U_n, \ldots, U_{n-b_j+1})| \le \mathrm{Var}(B|U_n, \ldots, U_{n-b_j+1})^{1/2} \mathcal{O}(n^{j-3/2}).$$

And by the Jensen inequality:

$$\mathbb{E}[|\mathrm{Cov}(B, C|U_n, \ldots, U_{n-b_j+1})|] \le \mathbb{E}[\mathrm{Var}(B|U_n, \ldots, U_{n-b_j+1})]^{1/2} \mathcal{O}(n^{j-3/2}).$$

Equation (10) implies that $\mathbb{E}[\mathrm{Var}(B|U_n, \ldots, U_{n-b_j+1})] \le v_{n-1}$, so that:

$$\mathbb{E}[|\mathrm{Cov}(B, C|U_n, \ldots, U_{n-b_j+1})|] \le v_{n-1}^{1/2} \mathcal{O}(n^{j-3/2}).$$

Finally, there exists $C_1 > 0$ such that:

$$2\mathbb{E}[\mathrm{Cov}(B, C|U_n, \ldots, U_{n-b_j+1})] \ge -C_1 n^{j-3/2} v_{n-1}^{1/2}. \tag{13}$$

Second term   We will simply use the trivial inequality

$$\mathbb{E}[\mathrm{Var}(C|U_n, \ldots, U_{n-b_j+1})] \ge 0. \tag{14}$$

Third term For $0 \leq m \leq b_j - 1$, define

$$A_m = \sum_{\substack{I \in \mathcal{I}(n,k,A): \\ n,\ldots,n-m+1 \notin I, n-m \in I}} X_I^\pi(\sigma_n),$$

$$A_{b_j} = \sum_{\substack{I \in \mathcal{I}(n,k,A): \\ n,\ldots,n-b_j+1 \notin I}} X_I^\pi(\sigma_n).$$

Clearly, $Y = A_0 + A_1 + \cdots + A_{b_j}$, while $C = A_0$ and $B = A_1 + A_2 + \cdots + A_{b_j}$. Note that for $0 \leq m \leq b_j$, A_m is independent of $U_n, \ldots, U_{n-m+1}$. For any $0 \leq i \leq b_j - 1$, we have:

$$\mathbb{E}[Y|U_n, \ldots, U_{n-i}] = \sum_{m=0}^{i} \mathbb{E}[A_m|U_{n-m}, \ldots, U_{n-i}] + \sum_{m=i+1}^{b_j} \mathbb{E}[A_m].$$

Taking the variance, the second sum will not contribute since it is deterministic. For the first sum, *i.e.* the case $0 \leq m \leq i$, we compute $\mathbb{E}[A_m|U_{n-m}, \ldots, U_{n-i}]$. Before giving the general formula, we consider a simple example.

We explain how to obtain a formula for the random variable $\mathbb{E}[A_0|U_n, U_{n-1}]$. Consider the pattern $\pi = 54\underline{231}$. Then, A_0 is the sum of the indicators $X_I^\pi(\sigma_n)$ for $I \in \mathcal{I}(n, 5, \{3, 4\})$ such that $n \in I$. This implies automatically that $n - 2, n - 1 \in I$ due to the adjacencies of the given pattern. Assume $I = \{i_5, i_4, n-2, n-1, n\}$ with $i_5 < i_4 < n - 2$. Then, we have:

$$\mathbb{P}\big(X_I^\pi(\sigma_n) = 1 \big| U_n, U_{n-1}\big) = \mathbb{P}\big(\text{red}(U_{i_5} U_{i_4} U_{n-2} U_{n-1} U_n) = 54231 \big| U_n, U_{n-1}\big)$$
$$= \mathbb{P}(U_{n-1} > U_{n-2} > U_n|U_n, U_{n-1})\mathbb{P}(U_{i_5} > U_{i_4} > U_{n-1}|U_{n-1})\mathbb{I}_{U_{n-1}>U_n}$$
$$= (U_{n-1} - U_n)\frac{(1 - U_{n-1})^2}{2!}\mathbb{I}_{U_{n-1}>U_n},$$

where $\mathbb{I}_x$ is the indicator function which is 1 if and only if condition x holds. We used that the U_i's are independent and uniform in $[0, 1]$. Since there are $\binom{n-3}{2}$ choices for i_5 and i_4, we have:

$$\mathbb{E}[A_0|U_n, U_{n-1}] = \sum_{I \in \mathcal{I}(n,5,\{3,4\}): \, n \in I} \mathbb{P}\big(X_I^\pi(\sigma_n) = 1 \big| U_n, U_{n-1}\big)$$
$$= \binom{n-3}{2}\left((U_{n-1} - U_n)\frac{(1 - U_{n-1})^2}{2!}\mathbb{I}_{U_{n-1}>U_n}\right).$$

This is the explicit formula for the random variable $\mathbb{E}[A_0|U_n, U_{n-1}]$.

For the general formula, let $a_1 \ldots a_{1+i-m} = sort(\pi_{k-(i-m)} \ldots \pi_k)$ (fixed, since π is) and define $\ell_1, \ldots, \ell_{1+i-m}$ such that $U_{\ell_1} \ldots U_{\ell_{1+i-m}} = sort(U_{n-i} \ldots U_{n-m})$ (not fixed). Then:

$$\mathbb{E}[A_m|U_{n-m}, \ldots, U_{n-i}] = B_m(n) \times S(U_{n-m}, \ldots, U_{n-i}),$$

where

$$B_m(n) = \binom{n - m - k + j - 1}{j - 1},$$

$$S(U_{n-m}, \ldots, U_{n-i}) = \frac{U_{\ell_1}^{a_1-1}(U_{\ell_2} - U_{\ell_1})^{a_2-a_1-1} \cdots (1 - U_{\ell_{1+i-m}})^{k-a_{1+i-m}}}{(a_1 - 1)!(a_2 - a_1 - 1)! \cdots (k - a_{1+i-m})!} \mathbb{I}_{\substack{\mathrm{red}(U_{n-i}\ldots U_{n-m}) \\ =\mathrm{red}(\pi_{k-(i-m)}\ldots\pi_k)}}.$$

The binomial coefficient counts the number of $I \in \mathcal{I}(n, k, A)$ fulfilling the constraints for A_m: this set is in bijection with $\mathcal{I}(n - b_j - m, k - b_j, A \cap [k - b_j])$ so that its cardinality is given by Lemma 2.7. For each such I, the probability that the corresponding values in σ_n are in the good order is given by the fraction. The indicator takes care of the order of the given U_i's.

We need these computations to compare the expectations (of Y and of B) appearing in the third term in Eq. (12). For Y, we have:

$$\mathbb{E}\big[\operatorname{Var}\big(\mathbb{E}[Y|U_n, \ldots, U_{n-i}]\big|U_n, \ldots, U_{n-i+1}\big)\big]$$

$$= \mathbb{E}\bigg[\operatorname{Var}\bigg(\sum_{m=0}^{i} B_m(n)S(U_{n-m}, \ldots, U_{n-i})\bigg|U_n, \ldots, U_{n-i+1}\bigg)\bigg] \tag{15}$$

$$= n^{2j-2}\mathbb{E}\bigg[\operatorname{Var}\bigg(\sum_{m=0}^{i} \frac{1}{(j-1)!}S(U_{n-m}, \ldots, U_{n-i})\bigg|U_n, \ldots, U_{n-i+1}\bigg)\bigg] + \mathcal{O}(n^{2j-3}).$$

And for B, we obtain:

$$\mathbb{E}\big[\operatorname{Var}\big(\mathbb{E}[B|U_{n-1}, \ldots, U_{n-i-1}]\big|U_{n-1}, \ldots, U_{n-i}\big)\big]$$

$$= \mathbb{E}\bigg[\operatorname{Var}\bigg(\sum_{m=1}^{i+1} B_m(n)S(U_{n-m}, \ldots, U_{n-i-1})\bigg|U_{n-1}, \ldots, U_{n-i}\bigg)\bigg]$$

$$= n^{2j-2}\mathbb{E}\bigg[\operatorname{Var}\bigg(\sum_{m=1}^{i+1} \frac{1}{(j-1)!}S(U_{n-m}, \ldots, U_{n-i-1})\bigg|U_{n-1}, \ldots, U_{n-i}\bigg)\bigg] + \mathcal{O}(n^{2j-3})$$

$$= n^{2j-2}\mathbb{E}\bigg[\operatorname{Var}\bigg(\sum_{m=0}^{i} \frac{1}{(j-1)!}S(U_{n-m}, \ldots, U_{n-i})\bigg|U_n, \ldots, U_{n-i+1}\bigg)\bigg] + \mathcal{O}(n^{2j-3}),$$

where the last step uses that the U_i's are independent and identically distributed (i.i.d.) so that we can replace $(U_{n-1}, \ldots, U_{n-i-1})$ by $(U_n, \ldots, U_{n-i})$. In particular:

$$\sum_{i=0,\ldots,b_j-2} \mathbb{E}\big[\operatorname{Var}\big(\mathbb{E}[Y|U_n, \ldots, U_{n-i}]\big|U_n, \ldots, U_{n-i+1}\big)\big]$$

$$- \sum_{i=0,\ldots,b_j-2} \mathbb{E}\big[\operatorname{Var}\big(\mathbb{E}[B|U_{n-1}, \ldots, U_{n-i-1}]\big|U_{n-1}, \ldots, U_{n-i}\big)\big] \tag{16}$$

$$= \mathcal{O}(n^{2j-3}).$$

Fourth term From Eq. (15), we have:

$$\mathbb{E}\big[\, \mathrm{Var}\,\big(\mathbb{E}[Y|U_n,\ldots,U_{n-b_j+1}]\big|U_n,\ldots,U_{n-b_j+2}\big)\big] = n^{2j-2}\mathbb{E}[f(U_n,\ldots,U_{n-b_j+2})] + \mathcal{O}(n^{2j-3}),$$

where

$$f(U_n,\ldots,U_{n-b_j+2}) = \mathrm{Var}\bigg(\sum_{m=0}^{b_j-1}\frac{1}{(j-1)!}S(U_{n-m},\ldots,U_{n-b_j+1})\bigg|U_n,\ldots,U_{n-b_j+2}\bigg).$$

Note that with positive probability, $\mathrm{red}(U_{n-b_j+2}\ldots U_n) = \mathrm{red}(\pi_{k-b_j+2}\ldots\pi_k)$. On this event, as a function of U_{n-b_j+1}, the random variable $\sum_{m=0,\ldots,b_j-1}S(U_{n-m},\ldots,U_{n-b_j+1})$ is not a.s. constant (in particular, $\mathbb{I}_{\mathrm{red}(U_{n-b_j+1}\ldots U_n)=\mathrm{red}(\pi_{k-b_j+1}\ldots\pi_k)}$ takes value 1 or 0 with positive probability each). So, with positive probability, $f(U_n,\ldots,U_{n-b_j+2})$ is non-zero which implies $\mathbb{E}[f(U_n,\ldots,U_{n-b_j+2})] > 0$. Furthermore, since the U_i's are i.i.d., $\mathbb{E}[f(U_n,\ldots,U_{n-b_j+2})]$ does not depend on n. Hence, there exists $C_2 > 0$ such that:

$$\mathbb{E}\big[\, \mathrm{Var}\,\big(\mathbb{E}[Y|U_n,\ldots,U_{n-b_j+1}]\big|U_n,\ldots,U_{n-b_j+2}\big)\big] \geq C_2 n^{2j-2}. \tag{17}$$

Conclusion Putting Eqs. (13), (14), (16) and (17) together, we obtain for some $C_3 > 0$:

$$\begin{aligned} v_n &\geq v_{n-1}\big(1 - C_1 n^{j-3/2}v_{n-1}^{-1/2}\big) + C_2 n^{2j-2} + \mathcal{O}(n^{2j-3}) \\ &\geq v_{n-1}\big(1 - C_1 n^{j-3/2}v_{n-1}^{-1/2}\big) + C_3 n^{2j-2}. \end{aligned} \tag{18}$$

Using first Eq. (9) and then Eq. (15) (where $i = 0$ gives $\mathbb{E}[\mathrm{Var}(\mathbb{E}[Y|U_n])] = \mathrm{Var}(\mathbb{E}[Y|U_n])$), we directly obtain:

$$v_n \geq \mathrm{Var}(\mathbb{E}[Y|U_n]) = n^{2j-2}\,\mathrm{Var}\Big(\frac{1}{(j-1)!}S(U_n)\Big) + \mathcal{O}(n^{2j-3}),$$

where

$$S(U_n) = \frac{U_n^{\pi_k-1}(1-U_n)^{k-\pi_k}}{(\pi_k-1)!(k-\pi_k)!}.$$

Because we assume $k \geq 2$, $S(U_n)$ is not a.s. constant. Moreover, its variance does not depend on n, so that there exists $C_4 > 0$ such that:

$$v_n \geq C_4 n^{2j-2}. \tag{19}$$

Using Eq. (19) for v_{n-1}, we can refine Eq. (18) for some $C_5 > 0$:

$$v_n \geq v_{n-1}\big(1 - C_5 n^{-1/2}\big) + C_3 n^{2j-2}. \tag{20}$$

Equation (20) is the recursive estimation that we now use to obtain a lower bound for the variance $v_n = \mathrm{Var}(X^{(\pi,A)}(\sigma_n))$ which is sharper than the one given by Eq. (19).

Proposition 4.7. *Let (π, A) be a fixed vincular pattern of size k with j blocks and let σ_n be uniform in S_n. Then, there exists $C > 0, n_0$ such that for $n \geq n_0$:*

$$\mathrm{Var}(X^{(\pi,A)}(\sigma_n)) \geq Cn^{2j-1}.$$

Proof: For any n, let $v_n = \mathrm{Var}(X^{(\pi,A)}(\sigma_n))$. Let n_1 be large enough such that $1 - \frac{C_5}{\sqrt{n_1}} \geq 0$ and such that Eq. (20) holds for $n \geq n_1$. By Eq. (20), we recursively obtain:

$$v_n \geq \sum_{\ell=n_1}^{n} C_3 \ell^{2j-2} \prod_{m=\ell+1}^{n} \left(1 - \frac{C_5}{\sqrt{m}}\right).$$

Since $1 - x \geq e^{-2x} \geq 1 - 2x$ for $x \leq 1$, we have for ℓ big enough:

$$\prod_{m=\ell+1}^{n} \left(1 - \frac{C_5}{\sqrt{m}}\right) \geq \prod_{m=\ell+1}^{n} e^{-2\frac{C_5}{\sqrt{m}}} = e^{-2\sum_{m=\ell+1}^{n} \frac{C_5}{\sqrt{m}}} \geq 1 - 2\sum_{m=\ell+1}^{n} \frac{C_5}{\sqrt{m}}.$$

Hence, there exists $C_6 > 0$ such that for n big enough:

$$
\begin{aligned}
v_n &\geq \sum_{\ell=n-C_6\sqrt{n}}^{n} C_3 \ell^{2j-2} \left(1 - 2C_5 \sum_{m=\ell+1}^{n} \frac{1}{\sqrt{m}}\right) \\
&\geq \sum_{\ell=n-C_6\sqrt{n}}^{n} C_3 \ell^{2j-2} \left(1 - 2C_5 \frac{C_6\sqrt{n}}{\sqrt{n - C_6\sqrt{n}}}\right) \\
&\geq \sum_{\ell=n-C_6\sqrt{n}}^{n} C_3 \ell^{2j-2} (1 - 4C_5 C_6) \\
&\geq C_3 (1 - 4C_5 C_6) C_6 \sqrt{n} \left(\frac{n}{2}\right)^{2j-2} \\
&\geq C_7 n^{2j-3/2},
\end{aligned}
$$

for some $C_7 > 0$. Because v_n is a polynomial in n for $n \geq 2(k - j)$ (see Lemma 4.2), there exists $C > 0$ and $n_0 > 0$ such that for $n \geq n_0$, $v_n \geq Cn^{2j-1}$. $\qquad\square$

Acknowledgements

References

P. Baldi and Y. Rinott. On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.*, 17(4):1646–1650, 1989.

A. D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*. Oxford University Press, 1992.

M. Bóna. On three different notions of monotone subsequences. In *Permutation patterns*, volume 376 of *London Math. Soc. Lecture Note Series*, pages 89–114. Cambridge University Press, 2010.

M. Bóna. *Combinatorics of permutations (2nd edition)*. Chapman-Hall and CRC Press, 2012.

C. G. Bowsher and P. S. Swain. Identifying sources of variation and the flow of information in biochemical networks. *Proc. Natl. Acad. Sci. USA*, 109(20):E1320–E1328, 2012.

S. Chatterjee and P. Diaconis. A central limit theorem for a new statistic on permutations. *Indian J. Pure Appl. Math.*, 48(4):561–573, 2017.

L. H. Y. Chen and A. Röllin. Stein couplings for normal approximation. arXiv preprint, `arXiv:1003.6039`, 2010.

S. Corteel, G. Louchard, and R. Pemantle. Common intervals in permutations. *Discrete Math. Theor. Comput. Sci.*, 8(1):189–216, 2006.

H. Crane and S. DeSalvo. Pattern avoidance for random permutations. To appear in *Permutation Patterns 2016*, special issue of *Discrete Math. Theor. Comput. Sci.*, `arXiv:1509.07941v3`, 2017.

H. Crane, S. DeSalvo, and S. Elizalde. The probability of avoiding consecutive patterns in the Mallows distribution. To appear in *Random Struct. Algor.*, `arXiv:1609.01370`, 2016.

V. Féray. Weighted dependency graphs. arXiv preprint, `arXiv:1605.03836`, 2016.

V. Féray, P.-L. Méliot, and A. Nikeghbali. *Mod-ϕ Convergence: Normality Zones and Precise Deviations*. Springer Briefs in Probability and Statistics, 2016.

P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, 2009.

J. Fulman. Stein's method and non-reversible Markov chains. In *Stein's Method: Expository Lectures and Applications*, volume 46 of *Lecture Notes–Monograph Series*, pages 66–74. Institute of Mathematical Statistics, 2004.

L. Goldstein. Berry Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.*, 42(3):661–683, 2005.

S. Janson. Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. *Ann. Probab.*, 16(1):305–312, 1988.

S. Janson, B. Nakamura, and D. Zeilberger. On the asymptotic statistics of the number of occurrences of multiple permutation patterns. *J. Comb.*, 6:117–143, 2015.

I. Kaplansky. The asymptotic distributions of runs of consecutive elements. *Ann. Math. Statist.*, 16(2): 200–203, 1945.

S. Kitaev. *Patterns in permutations and words*. Springer, 2011.

Y. Rinott. On normal approximation rates for certain sums of dependent random variables. *J. Comp. Appl. Math.*, 55(2):135–143, 1994.

N. Ross. Fundamentals of Stein's method. *Probab. Surveys*, 8:210–293, 2011.

L. Saulis and V. A. Statulevičius. *Limit Theorems for Large Deviations*. Kluwer Academic Publications, 1991.

N. A. Weiss. *A course in probability*. Addison–Wesley, 2005.