CENTRAL LIMIT THEOREMS FOR PATTERNS IN MULTISET PERMUTATIONS AND SET PARTITIONS

BY VALENTIN FÉRAY

Institute of Mathematics, University of Zurich, valentin.feray@math.uzh.ch

We use the recently developed method of weighted dependency graphs to prove central limit theorems for the number of occurrences of any fixed pattern in multiset permutations and in set partitions. This generalizes results for patterns of size 2 in both settings, obtained by Canfield, Janson and Zeilberger and Chern, Diaconis, Kane and Rhoades, respectively.

1. Introduction.

1.1. *Background and informal presentation of the main results.* A natural parameter of interest in the framework of random combinatorial structures is the number of occurrences of a given substructure. When this substructure has a fixed size, we observe in many cases that this number is asymptotically normal. Such a central limit theorem (CLT) for substructures was first proved for random graphs [21, 29] and random words [13]. More recently, similar results for pattern occurrences in uniform random permutations were obtained: see the works of Fulman [14] (for inversions and descents), Goldstein [15], see in particular Example 3.2 (for consecutive patterns), Bóna [2] (for monotone patterns, both in the consecutive and classical settings), Janson, Nakamura and Zeilberger [24] (for general classical patterns) and Hofer [19] (for vincular patterns). We also refer to the work of the author [10] and Crane, DeSlavo and Elizalde [9] for results for Ewens distributed and Mallows distributed random permutations, respectively.

Most of these results are based on the theory of dependency graphs, used either in combination with cumulant estimates or Stein's method; see, for example, [19], Section 3, for an overview of these tools. One exception is the work of Janson, Nakamura and Zeilberger [24], which uses the theory of U-statistics [17, 22]. In all of these methods, a key feature is the independence of occurrences of the given fixed substructure in disjoint sets of positions.

In this paper, we investigate CLTs for substructures in two other families of combinatorial objects: multiset permutations and set partitions. For both objects, some notion of patterns have been studied in the literature; see [1] and [7], respectively. In both settings, a CLT is only known for the simplest kind of patterns: inversions in multiset permutations, where the central limit theorem was established by Canfield, Janson and Zeilberger [5] (see also [32]) and crossings in set partitions, from the work of Chern, Diaconis, Kane and Rhoades [8]. The methods used in these papers do not seem to be generalizable to longer patterns. Indeed they are based on the following facts, which only hold for inversions and crossings, respectively: the explicit generating functions of inversions in multiset permutations has a simple explicit product from; conditionally on the starting and ending points of blocks in set partitions, the number of crossings is a sum of independent random variables.

What makes patterns in multiset permutations and set partitions harder to study is that occurrences of a given pattern in disjoint sets of places are no longer independent events.

We overcome this difficulty by using the theory of *weighted dependency graphs*, recently developed by the author [11].

Our main results, Theorems 1.4 and 1.6, are *the asymptotic normality of the number of occurrences of any fixed patterns* in random multiset permutations and in random set partitions. For multiset permutations, we need a slight regularity assumption on the sequence of multisets that we consider. We refer to Sections 1.3 and 1.4 respectively, for precise definitions of the notions of patterns in both settings and for precise statements of the main results.

1.2. *Terminology and notation.* Before stating precisely our main theorem, we introduce some notation.

Sets and multisets. As usual, we write $[n]$ for the set $\{1, \ldots, n\}$. A multiset is an *unordered* collection of elements, with possible repetitions. Given a multiset B, we write $|B|$ and $\#B$ for its number of elements, and number of *distinct* elements, respectively.

Probability theory. Indicator functions will be denoted with the symbol $\mathbf{1}$, namely

$$\mathbf{1}[P] := \begin{cases} 1 & \text{if } P \text{ holds,} \\ 0 & \text{otherwise.} \end{cases}$$

Throughout the paper, we say that a sequence $(X_n)_{n \geq 1}$ of real-valued random variables is *asymptotically normal* if the following convergence in distribution holds:

$$(1) \qquad \frac{X_n - \mathbb{E} X_n}{\sqrt{\operatorname{Var} X_n}} \to_d Z \quad \text{as } n \to \infty,$$

where Z is a standard Gaussian variable.

Asymptotic notation. Besides, we use the standard $\mathcal{O}, \Theta, \tilde{\mathcal{O}}, \tilde{\Theta}$ symbols for asymptotic comparisons: we write $f(n) = \mathcal{O}(g(n))$ if there exists $C > 0$ such that $|f(n)| \leq C g(n)$, for all n sufficiently large. Furthermore, $f(n) = \Theta(g(n))$ stands for $f(n) = \mathcal{O}(g(n))$ and $g(n) = \mathcal{O}(f(n))$. Finally, in the set partition section, we use the tilde variants, for bounds up to logarithmic factors: more precisely $f(n) = \tilde{\mathcal{O}}(g(n))$ means that there exists $d \geq 0$ such that $f(n) = \mathcal{O}(g(n) \ln(n)^d)$; and we write $f(n) = \tilde{\Theta}(g(n))$ if $f(n) = \tilde{\mathcal{O}}(g(n))$ and $g(n) = \tilde{\mathcal{O}}(f(n))$. We try and make explicit on which parameters the above constant C (and exponent d) may depend or not. Nevertheless, as a rule of thumb, it depends on the pattern (denoted τ or $\mathcal{A}$) we consider and/or the order of the considered moment/cumulant, but neither on the size n of the objects (or the underlying multiset M for multiset permutations), nor on the positions $i_1, \ldots, i_\ell$ of a pattern occurrence.

1.3. *First main result: A CLT for patterns in multiset permutations.* Let M be a finite multiset of positive integers. Concretely, we can write $M = \{1^{a_1}, 2^{a_2}, \ldots\}$, where exponents are used to indicate multiplicities; let also $n = |M| = \sum_{j=1}^{\infty} a_j$. A *multiset permutation* (or permutation for short) of M is a word containing exactly a_j times the integer j (for each $j \geq 1$). Define S_M as the set of multiset permutations of M. Naturally, we have $|S_M| = \frac{n!}{a_1! a_2! \cdots}$.

EXAMPLE 1.1. The multiset $M = \{1^2, 2^2, 3\}$ has 30 permutations, one of which is $\sigma = 23112$.

We are interested in patterns in multiset permutations, following [1].

DEFINITION 1.2. Let τ be a permutation of size k. A multiset permutation σ has an occurrence of τ in position $(i_1, \ldots, i_\ell)$ $(i_1 < \cdots < i_\ell)$ if the subsequence $\sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_\ell}$ has distinct entries in the same relative order as τ; formally if

$$\sigma_{i_{\tau^{-1}(1)}} < \sigma_{i_{\tau^{-1}(2)}} < \cdots < \sigma_{i_{\tau^{-1}(\ell)}}.$$

For example, the (multiset) permutation $\sigma = 23112$ contains five occurrences of the pattern $\tau = 21$: in positions $(1, 3)$, $(1, 4)$, $(2, 3)$, $(2, 4)$ and $(2, 5)$.

We are interested in the random variable $\mathrm{Occ}^{\tau}_M := \mathrm{Occ}^{\tau}(\sigma)$, which gives the number of occurrences of τ in a uniform random element σ of S_M. Fixing τ and taking a sequence of multisets $M^{(m)}$, we get a sequence of random variables $\mathrm{Occ}^{\tau}_{M^{(m)}}$. Our main theorem is a central limit theorem for $\mathrm{Occ}^{\tau}_{M^{(m)}}$ under some regularity condition on the sequence $M^{(m)}$.

DEFINITION 1.3. Fix a positive integer ℓ. A sequence $M^{(m)}$ is called ℓ-regular, if there exists $K < 1$ and m_0, such that, for $m \geq m_0$, the sum of the ℓ largest multiplicities in $M^{(m)}$ is at most $K|M^{(m)}|$.

THEOREM 1.4. *Let τ be a pattern of size ℓ and $(M^{(m)})_{m \geq 1}$ be an ℓ-regular sequence of finite multisets. Then $\mathrm{Occ}^{\tau}_{M^{(m)}}$ is asymptotically normal.*

We also give estimates for the expectation and variance of $\mathrm{Occ}^{\tau}_{M^{(m)}}$: under the regularity hypothesis, the expectation is easily seen to be of order $\Theta(n^{\ell})$. Furthermore, it follows from our proof of Theorem 1.4 that the variance is $\mathcal{O}(n^{2\ell-1})$ and that this bound is tight up to subpolynomial factors; see Section 3.4.4.

REMARK 1.5. The above condition of ℓ-regularity is not optimal, as can already be observed for inversions ($\tau = 21$). In this case, our theorem gives the asymptotic normality under the condition that an asymptotically nonzero proportion of elements are different from the two most repeated parts. This asymptotic normality is in fact known to hold under the weaker condition that a nonbounded number of elements are different from the single most repeated part [5]. In general, our theorem could be improved if we could prove a lower bound for the variance with less restrictive conditions; see Question 3.7 below and the discussion after it.

1.4. *Second main result*: *A CLT for patterns in set-partitions*. A *set partition* π of $[n]$ is a set of nonempty pairwise disjoint *blocks* $B_1, B_2, \ldots, B_k$ with $\bigcup_{i=1}^{k} B_i = [n]$. We will denote the set of set partitions of $[n]$ by $\mathcal{P}([n])$. It is customary to represent graphically set partitions by a set of arcs that join every pair of consecutive elements in the same block; see an example in Figure 1. Formally, for $i, j \in [n]$, there is an *arc* from i to j in π if j is the smallest number that is greater than i and in the same block.

An *arc pattern* of length ℓ is a subset $\mathcal{A} \subseteq \{(i, j) \in [\ell] \times [\ell] : i < j\}$ such that for distinct elements (i_1, j_1) and (i_2, j_2) in $\mathcal{A}$, we have $i_1 \neq i_2$ *and* $j_1 \neq j_2$. Arc patterns of length ℓ are exactly the arc representations of set partitions of ℓ, but in the sequel, it is more natural to think of them as sets of arcs. We say that the pattern $\mathcal{A}$ occurs in a set partition π in positions $x_1 < x_2 < \cdots < x_{\ell}$ if for every $(i, j) \in \mathcal{A}$ there is an arc from x_i to x_j in π. The number of occurrences of $\mathcal{A}$ in a set partition π will be denoted $\mathrm{Occ}^{\mathcal{A}}(\pi)$. As an example, an occurrence of the arc pattern $\{(1, 3), (2, 4)\}$ is a crossing in a set partition π; the set partition of Figure 1 contains one such occurrence, in positions $\{1, 2, 3, 5\}$.

We are interested in the random variable $\mathrm{Occ}^{\mathcal{A}}_n := \mathrm{Occ}^{\mathcal{A}}(\pi)$, which gives the number of occurrences of a fixed arc pattern in a uniform random set partition π of $[n]$. The main result of this part is the following central limit theorem.
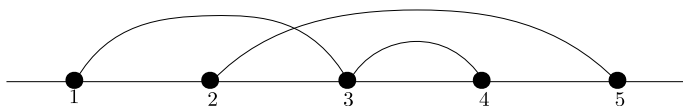

FIG. 1. *The arc representation of the set partition $\pi = \{\{1, 3, 4\}, \{2, 5\}\}$.*

THEOREM 1.6. *The number of occurrences $\mathrm{Occ}_n^{\mathcal{A}}$ of any fixed arc pattern $\mathcal{A}$ in a uniform random set partition of $[n]$ is asymptotically normal as $n \to \infty$. Moreover, asymptotically, we have $\mathbb{E}(\mathrm{Occ}_n^{\mathcal{A}}) = \tilde{\Theta}(n^{\ell-a})$ and $\mathrm{Var}(\mathrm{Occ}_n^{\mathcal{A}}) = \tilde{\Theta}(n^{2\ell-2a-1})$, where ℓ and a denote the length and the number of arcs of $\mathcal{A}$, respectively.*

REMARK 1.7. Arc patterns are a particular case of the statistics called *patterns* in [7]. The latter are more general in the sense that we can require in an occurrence that some x_i is the first (or the last) element in its block or that x_i and x_{i+1} are consecutive for some i. Then the authors of [7] consider sums of the kind $\sum Q(x_1, \ldots, x_\ell, n)$, where Q is a polynomial, and the sum runs over occurrences of a given pattern $\mathcal{A}$. This obviously generalizes the number of occurrences.

The asymptotic normality of such statistics could also be investigated through weighted dependency graphs. Indeed, we think that with little extra effort (but heavier notation), we could include in the weighted dependency graph of Section 4.1 some indicator variables F_i and L_i, indicating whether i is the first or the last element in its block, respectively. Nevertheless, at this level of generality, it seems hard to find good variance bounds and estimates for the parameters R and T_h of the weighted dependency graph.

1.5. *Discussion on the proofs.* As said above, the proof relies on the theory of weighted dependency graphs. There are two major difficulties in applying it.

- The first one is to prove that the relevant random variables admit a suited weighted dependency graphs; this consists in bounding their joint cumulants. Thanks to the general theory of weighted dependency graphs, this can be reduced to bound the joint cumulants of simple indicator random variables.

  In the case of multiset permutations, since the corresponding joint moments are explicit, this is relatively easy. In the case of set partitions, we use a construction of a uniform random set partition through a urn model with a random number M of urns, due to Stam [31]. We first bound the conditional cumulants with respect to M, and then use the law of total cumulance [3]. This needs deviation estimates on M and a quite delicate analysis (Section 5).

- The second difficulty is to find a lower bound on the variance. This is generally a delicate question, often left aside in the literature: in [6], the author writes "to show that the bound [on Kolmogorov's distance] is useful, we require a lower bound on σ^2. We prefer to think of that as a separate problem."

  In this paper, we provide such lower bounds on the variance for general patterns, both in multiset permutations and set partitions. In both cases, these bounds are based on the law of total variance (which is suited for finding lower bounds since it contains only nonnegative terms). For set partitions, we condition on the number M of urns in the urn model and it turns out that one of the terms in the law of total variance, namely the variance of the conditional expectation, is analyzable and already large enough to provide the needed lower bound (the analysis is however rather delicate, see Section 6). For multiset permutations, we condition on (one of) the position of the smallest element. This gives us a recursive inequality on the variance that can be analyzed to give the needed lower bound. This method has been recently introduced by Hofer [19], in the context of vincular patterns in (usual) permutations.

1.6. *Outline of the paper.* The paper is organized as follows. Section 2 introduces the necessary background on weighted dependency graphs, in particular, the normality criterion that we use. Sections 3 and 4 contain the proofs of our main results on multiset permutations and set partitions, respectively. These sections are independent from each other. Sections 5 and 6 are devoted to technical proofs of the set partition section.

2. Weighted dependency graphs.

2.1. *Motivation and definition.*   We first present the definition of usual dependency graphs and explain informally why we need a weighted version of it.

DEFINITION 2.1 (Janson [20]).   A graph G is a dependency graph for the family $\{Y_\alpha, \alpha \in A\}$ of random variables if the two following conditions are satisfied:

  (i) the vertex set of G is A;
  (ii) if A_1 and A_2 are subsets of A such that no edge connects a vertex of A_1 to one of A_2 in G, then $\{Y_\alpha, \alpha \in A_1\}$ and $\{Y_\alpha, \alpha \in A_2\}$ are independent.

Roughly, a dependency graph encodes the dependency relations between the variables Y_α: variables not related by edges must be independent.

Consider now a sequence $S_n = \sum_{\alpha \in A_n} Y_{\alpha,n}$ of sum of random variables and, for each n, a dependency graph G_n for the family $\{Y_{\alpha,n}, \alpha \in A_n\}$. If these dependency graphs are sparse enough, we might expect that S_n behaves as a sum of independent variables and is, under mild condition, asymptotically normal. A precise normality criterion which formalizes this intuition, has been given by Janson [20], Theorem 2; see also the work of Mikhailov [26], which has a higher range of applications.

As explained in the Introduction, dependency graphs are not suited to study patterns in multiset permutations or set partitions, since occurrences of a given pattern in disjoint sets of position are in general dependent events. Yet, such events are weakly correlated, and we will be able to use a weighted variant of dependency graphs that we now present. The rest of this section follows [11].

First, we need something that quantifies the dependency between random variables. To this purpose, we use mixed cumulants. Cumulants have a long history and many different names: they are referred to as Ursell functions (after Ursell [33]) or truncated/connected correlation functions in the physics literature, as *"déviation d'indépendence"* in a note of Schutzengberger [30], or as semi-invariants, for instance, in [25] and [20]. We refer to [23], Chapter 6, for a modern presentation of mixed cumulants and applications to random graph theory.

The (mixed) *cumulant* of a family of random variables $X_1, X_2, \ldots, X_r$ defined on the same probability space and having finite moments is defined as

$$(2) \qquad \kappa(X_1, X_2, \ldots, X_r) := \sum_{\pi \in \mathcal{P}([r])} \mu(\pi, [r]) \prod_{B \in \pi} \mathbb{E}\left(\prod_{i \in B} X_i\right),$$

where $\mathcal{P}([r])$ is the lattice of set partitions of $[r]$ and μ is its Möbius function. If $X_i = X$ for all i, we abbreviate $\kappa(X_1, X_2, \ldots, X_r)$ as $\kappa_r(X)$. Key properties of cumulants are the following:

  (i) If $\{X_1, X_2, \ldots, X_r\}$ can be written as a disjoint union of two independent nonempty sets of random variables, then $\kappa(X_1, X_2, \ldots, X_r) = 0$.
  (ii) A sequence Y_n of random variables converges in distribution to a standard normal variable as soon as $\kappa_r(Y_n) \to \mathbf{1}[r = 2]$ for all $r \geq 1$.

If $\{Y_\alpha, \alpha \in A\}$ is a family of random variables with dependency graph G, then property (i) above implies that, for indices $\alpha_1, \ldots, \alpha_r$ such that the induced graph $G[\alpha_1, \ldots, \alpha_r]$ is disconnected, we have

$$\kappa(Y_{\alpha_1}, \ldots, Y_{\alpha_r}) = 0.$$

For weighted dependency graphs, the idea behind the definition is that the smaller the edge weights in the induced subgraph $G[\alpha_1, \ldots, \alpha_r]$ are, the smaller the corresponding mixed cumulants should be.

In the sequel, a weighted graph is a graph with weights on its *edges*, belonging to $(0, 1]$. Nonedges can be interpreted as edges of weight 0, so that a weighted graph can be equivalently seen as an assignment of weights in $[0, 1]$ to the edges of the complete graph. All our definitions are compatible with this convention.

For a weighted graph H, we define $\mathcal{M}(H)$ to be the *maximal weight of a spanning tree* of H, the weight of a spanning tree being the product of the weights of its edges (if H is disconnected, there is no spanning tree and as a consequence of the above convention, we have $\mathcal{M}(G[B]) = 0$). The following definition was proposed in [11].

DEFINITION 2.2.   Let $C = (C_1, C_2, \ldots)$ be a sequence of positive real numbers. Let Ψ be a real-valued function on multisets of elements of A.

A weighted graph G is a (Ψ, C) weighted dependency graph for a family $\{Y_\alpha, \alpha \in A\}$ of random variables defined on the same probability space and having finite moments if, for any multiset $B = \{\alpha_1, \ldots, \alpha_r\}$ of elements of A, one has

$$(3) \qquad\qquad \left| \kappa(Y_\alpha; \alpha \in B) \right| \le C_r \Psi(B) \mathcal{M}(G[B]).$$

In examples of weighted dependency graphs, Ψ and C are simple or universal quantities, so that the meaningful term is $\mathcal{M}(G[B])$. Note that the smaller the weight on edges are, the smaller $\mathcal{M}(G[B])$ is, which is consistent with intuition.

2.2. *A criterion for asymptotic normality.*   Let G be a (Ψ, C) weighted dependency graph for a family of variables $\{Y_\alpha, \alpha \in A\}$. Let I and J be subsets of A. If I and J have an element in common, we set $W(I, J) = 1$. Otherwise, we define $W(I, J)$ as the maximal weight of an edge in G connecting an element of I to an element of J (if there is no such edge $W(I, J) = 0$, which is consistent with the fact that a nonedge can be replaced by an edge of weight 0).

Finally, we introduce the following parameters (h being a positive integer):

$$(4) \qquad\qquad R = \sum_{\alpha \in A} \Psi(\{\alpha\});$$

$$(5) \qquad\qquad T_h = \max_{\alpha_1, \ldots, \alpha_h \in A} \left[\sum_{\beta \in A} W(\{\beta\}, \{\alpha_1, \ldots, \alpha_h\}) \frac{\Psi(\{\alpha_1, \ldots, \alpha_h, \beta\})}{\Psi(\{\alpha_1, \ldots, \alpha_h\})} \right].$$

Admittedly, the definition of T_h is somewhat involved, but these parameters turn out to be easy to estimate in practice. In the particular case where Ψ is the constant function equal to 1, R is the number of variables, and each T_h is within a factor h of the maximal weighted degree of the graph; see [11], Remark 4.9.

Using these parameters, the following asymptotic normality criteria was given in [11], Theorem 4.11.

THEOREM 2.3.   *Suppose that, for each n, $\{Y_{n,i}, 1 \le i \le N_n\}$ is a family of random variables with finite moments defined on the same probability space. For each n, let Ψ_n a function on multisets of elements of $[N_n]$. We also fix a sequence $C = (C_r)_{r \ge 1}$, not depending on n.*

Assume that, for each n, one has a (Ψ_n, C) weighted dependency graph G_n for the family $\{Y_{n,i}, 1 \le i \le N_n\}$ and define the corresponding quantities $R_n, T_{1,n}, T_{2,n}, \ldots$, by equations (4) and (5).

Let $X_n = \sum_{i=1}^{N_n} Y_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(X_n)$.

Assume that there exist numbers γ_h and Q_n and an integer $s \geq 3$ such that

$$(6) \qquad\qquad T_{h,n} \leq \gamma_h Q_n;$$

$$(7) \qquad\qquad \left(\frac{R_n}{Q_n}\right)^{1/s} \frac{Q_n}{\sigma_n} \to 0 \quad as\ n \to \infty.$$

Then X_n is asymptotically normal.

Note that establishing (7) requires a lower bound on σ_n, which is often nontrivial to obtain. This theorem is proved by bounding cumulants of X_n. In particular, we get an upper bound on the variance.

PROPOSITION 2.4 ([11], Lemma 4.10). *We use the notation of Theorem* 2.3 (*but do not assume* (6) *and* (7)). *Then* $\mathrm{Var}(X_n) \leq 2C_2 R_n T_{1,n}$.

REMARK 2.5. The above normality criterion can be adapted to sequences of random variables with a $(\Psi_n, \boldsymbol{C}_n)$ weighted dependency graph G_n, where $\boldsymbol{C}_n$ depends on n. In such situation, we write $\boldsymbol{C}_n = (C_{r,n})_{r \geq 1}$. The condition (7) should then be replaced by the fact that, for some $s \geq 3$ and all $r \geq 1$, the quantity $(\frac{R_n}{Q_n})^{1/s} \frac{Q_n}{\sigma_n}$ tends to 0 faster than any power of $C_{r,n}$. The proof of this extension is a straightforward adaptation of the proof of [11], Theorem 4.11. We will use it in Section 4 below.

2.3. *Power of weighted dependency graphs.* An important property of weighted dependency graphs is the following stability property. Consider a family of random variables $\{Y_\alpha, \alpha \in A\}$ with a $(\Psi, \boldsymbol{C})$ weighted dependency graph G and fix some integer $d \geq 1$. We are interested in monomials $\boldsymbol{Y}_I := \prod_{\alpha_i \in I} Y_{\alpha_i}$ of degree at most d, that is, I is a multiset of elements of A of size at most d (counting repetitions), which we will denote as $I \in \mathrm{MSet}_{\leq d}(A)$. This new family of random variables $\{\boldsymbol{Y}_I, I \in \mathrm{MSet}_{\leq d}(A)\}$ admits a natural weighted dependency graph inherited from that of $\{Y_\alpha, \alpha \in A\}$.

To state this formally, we need to introduce some more notation. The dth power G^d of G is the weighted graph with vertex set $\mathrm{MSet}_{\leq d}(A)$ and having an edge of weight $W(I, J)$ between every pair of vertices (I, J) in $\mathrm{MSet}_{\leq d}(A)$ (here, we describe the weighted graph G^d as a complete graph with possibly zero edge-weights).

Finally, a function Ψ defined on multiset of A is naturally seen as a function on multiset of multisets of A by setting:

$$(8) \qquad\qquad \boldsymbol{\Psi}(\{I_1, \ldots, I_r\}) = \Psi(I_1 \uplus \cdots \uplus I_r).$$

PROPOSITION 2.6 ([11], Proposition 5.11). *Consider random variables* $\{Y_\alpha, \alpha \in A\}$ *with a* $(\Psi, \boldsymbol{C})$ *weighted dependency graph* G *and fix some integer* $d \geq 1$. *Then, with the above notation,* G^d *is a* $(\boldsymbol{\Psi}, \boldsymbol{D}_d)$ *weighted dependency graph for the family* $\{\boldsymbol{Y}_I, I \in \mathrm{MSet}_{\leq d}(A)\}$, *where the constants* $\boldsymbol{D}_d = (D_{d,r})_{r \geq 1}$ *depend only on* d, r *and* $\boldsymbol{C}$.

In applications, the above proposition is used as follows. We first find a weighted dependency graphs for some simple family of random variables: typically indicators of basic events, such as $\sigma(i) = j$ for (multiset) permutations, or the presence of an arc between given points i and j in set partitions. The above theorem gives us automatically a weighted dependency graph for more complicated random variables, such as indicators of having a fixed pattern in some given set of positions. Then the normality criterion (Theorem 2.3) gives a central limit theorem for the number of occurrences of this fixed pattern.

3. Permutation patterns in multiset permutations. In this section, we prove Theorem 1.4. We refer to Section 1.3 for notation.

3.1. *The weighted dependency graph.* We consider the following random variable X_i^j on S_M, defined by

$$X_i^j(\sigma) = \begin{cases} 1 & \text{if } \sigma_i = j; \\ 0 & \text{otherwise.} \end{cases}$$

Let us also denote $A^M := \{X_i^j : 1 \le i \le n, j \in M\}$. The purpose of this subsection is to prove the following.

THEOREM 3.1. *Consider the weighted complete graph G^M on vertex-set A^M with weights*

$$w(X_i^j, X_{i'}^{j'}) = \begin{cases} 1 & \text{if } i = i', \\ 1/a_j & \text{if } i \ne i' \text{ and } j = j', \\ 1/n & \text{otherwise.} \end{cases}$$

Then G^M is a $(\Psi, \mathbf{C})$ weighted dependency graph for the family A^M, where Ψ is the function on multisets B of elements in A^M defined by

$$\Psi(B) = \prod_{X_i^j \in B \text{ distinct}} \mathbb{E}[X_i^j] = \prod_{X_i^j \in B \text{ distinct}} \frac{a_j}{n}$$

and $\mathbf{C} = (C_1, C_2, \ldots)$ is a universal sequence of constants not depending on M.

PROOF. The proof is relatively easy using the tools given in [11], but these tools require to introduce some terminology/notation. As a start, recall that if G is a weighted graph and B a subset of its vertices, we denote $G[B]$ the subgraph induced by G in B (if B is a multiset, we see it as a set, by simply forgetting repetitions). Also, for a weighted graph H, $\mathcal{M}(H)$ is the maximal weight of a spanning tree of H.

We have to prove that for any multiset $B = \{X_{i_1}^{j_1}, \ldots, X_{i_r}^{j_r}\}$ of elements of A^M, one has

(9) $$\left| \kappa(X_{i_1}^{j_1}, \ldots, X_{i_r}^{j_r}) \right| \le C_r \Psi(B) \mathcal{M}(G^M[B]).$$

For a multiset B, we denote $B_1, B_2, \ldots$, the vertex-sets of the connected components of $G_1^M[B]$, the graph obtaining from $G^M[B]$ by keeping only edges of weight 1. Using [11], Proposition 5.2, it is equivalent to prove (9) or the following: for any multiset B of elements of A^M with $|B| = r$ we have

(10) $$\left| \kappa\left(\prod_{X_i^j \in B_1} X_i^j, \prod_{X_i^j \in B_2} X_i^j, \ldots \right) \right| \le D_r \Psi(B) \mathcal{M}(G^M[B]),$$

for some sequence D_r also independent of M.

By definition of G^M, vertices X_i^j and $X_{i'}^{j'}$ are connected in $G_1^M[B]$ if and only if $i = i'$ or both $j = j'$ and $a_j = 1$. In this case, $X_i^j X_{i'}^{j'} = 0$ a.s. unless $i = i'$ and $j = j'$. Of course, if one of the product on the left-hand side of (10) is a.s. 0, then the inequality is trivial. Thus it suffices to consider the case where each component B_k contains at most one distinct element X_i^j, say with multiplicity m. But since X_i^j is a Bernoulli random variable, we have $(X_i^j)^m = X_i^j$; the right-hand side of (10) is also insensitive to repetitions in B, so that we can

assume $m = 1$ (in each of the components B_k). In other words, we only need to prove (10) in the case the left-hand side is $|\kappa(X_{i_1}^{j_1}, \ldots, X_{i_r}^{j_r})|$, for *distinct* $i_1, \ldots, i_r$ (and such that repeated entries j in the list $j_1, \ldots, j_r$ fulfills $a_j > 1$, but we will not need this extra condition).

The proof of (9) in the all i distinct case is based on the formula for joint moments in this case: if C is such a subset of A^M, we have

$$(11) \qquad \mathbb{E}\left(\prod_{X_i^j \in C} X_i^j \right) = \frac{\binom{n - |C|}{a_1 - \#\{X_i^1 \in C\}, a_2 - \#\{X_i^2 \in C\}, \ldots}}{\binom{n}{a_1, a_2, \ldots}}.$$

Indeed, the numerator counts the number of multiset permutations σ of M with $\sigma(i) = j$ for all $X_i^j \in C$, while the denominator is the total number of multiset permutations of M. To get bounds on cumulants, we use this expression and the quasi-factorization technique, as developed in [11], Section 5.2.

Consider a family $\boldsymbol{u} = (u_\Delta)_{\Delta \subseteq [r]}$ of real numbers indexed by subsets of $[r]$ with $u_\varnothing \neq 0$. We furthermore assume that $u_\delta = 0$ implies that $u_\Delta = 0$ as well for all subsets Δ containing δ; we call this *the vanishing ideal condition*. In the following, all families under consideration fulfill the vanishing ideal condition. For such a $\boldsymbol{u}$, we define

$$\mathbb{P}_\Delta(\boldsymbol{u}) = \prod_{\delta \subseteq \Delta} (u_\delta)^{(-1)^{|\Delta| - |\delta|}}.$$

By convention, if the above fraction is $0/0$, we set $\mathbb{P}_\Delta(\boldsymbol{u}) = 0$. A simple Möbius inversion gives back

$$u_\Delta = \prod_{\delta \subseteq \Delta} P_\delta(\boldsymbol{u}).$$

Let H be a weighted graph on $[r]$. We say that $\boldsymbol{u}$ has the H-quasi factorization property if, for each $\Delta \subseteq [r]$ *of size at least* 2, we have

$$(12) \qquad P_\Delta(\boldsymbol{u}) = 1 + \mathcal{O}(\mathcal{M}(H[\Delta])).$$

The constants in the $\mathcal{O}$ symbol should depend only on r, in particular, in the following, they are independent of B and M.

Fix now a subset $B = \{X_{i_1}^{j_1}, \ldots, X_{i_r}^{j_r}\}$ of A^M with distinct $i_1, \ldots, i_r$. For $\Delta \subseteq [r]$, we define $\boldsymbol{u}_\Delta = \mathbb{E}[\prod_{t \in \Delta} X_{i_t}^{j_t}]$. (Note that the dependence in M and B is kept implicit here.) From equation (11), this is explicitly given as

$$\boldsymbol{u}_\Delta = \boldsymbol{u}_\Delta^{(1)} \boldsymbol{u}_\Delta^{(2)} \prod_{j \geq 1} \boldsymbol{u}_\Delta^{(3,j)},$$

with

$$\boldsymbol{u}_\Delta^{(1)} = \frac{1}{\binom{n}{a_1, a_2, \ldots}}; \qquad \boldsymbol{u}_\Delta^{(2)} = (n - |\Delta|)!;$$

$$\boldsymbol{u}_\Delta^{(3,j)} = \frac{1}{(a_j - \#\{t \in \Delta : j_t = j\})!}.$$

By convention, $\boldsymbol{u}_\Delta^{(3,j)} = 0$ if $a_j - \#\{t \in \Delta : j_t = j\} < 0$. Note that all of these families have the vanishing ideal property. We discuss the quasi-factorization property of each factor separately:

(i) The first factor $u_\Delta^{(1)}$ is independent of Δ and, therefore, $\boldsymbol{u}^{(1)}$ trivially satisfies the $H^{(1)}$-quasi factorization property, where $H^{(1)}$ is the graph on $[r]$ with no edges.

(ii) The family $\boldsymbol{u}^{(2)}$ defined by $u_\Delta^{(2)} = (n - |\Delta|)!$ satisfies the $H^{(2)}$-quasi factorization property, where $H^{(2)}$ is the complete graph on $[r]$ with weight $1/n$ on each edge [11], Proposition 5.10.

(iii) Fix $j \geq 1$ and consider the factor $u_\Delta^{(3,j)} = (a_j - \#\{i : X_i^j \in \Delta\})!^{-1}$. Let $H^{(3,j)}$ be the graph with vertex-set $[r]$, and an edge of weight $\frac{1}{a_j}$ between s and t if $j_s = j_t = j$ (in particular, if $j_s \neq j$, then s is isolated in $H^{(3,j)}$). We claim that $\boldsymbol{u}^{(3,j)}$ has the $H^{(3,j)}$ quasi factorization property. Indeed, if Δ contains an s such that $j_s \neq j$, then $u_{\delta \cup \{s\}}^{(3,j)} = u_\delta^{(3,j)}$ for $\delta \subseteq (\Delta \setminus \{s\})$ and this implies $P_\Delta(\boldsymbol{u}^{(3,j)}) = 1$ so that (12) holds trivially. On the other hand, if $\Delta \subseteq \{s : j_s = j\}$, then $u_\delta^{(3,j)} = (a_j - |\delta|)!^{-1}$ and the estimate (12) again follows from [11], Proposition 5.10.

Denote H the following graph with vertex-set $[r]$: the weight of the edge between s and t is the *maximum* of the weights of the corresponding edges in $H^{(1)}$, $H^{(2)}$ and in all the $H^{(3,j)}$. Note that H corresponds to $G^M[B]$, where G^M is defined in the statement of Theorem 3.1. Observe that, for any Δ, we have

$$P_\Delta(\boldsymbol{u}) = P_\Delta(\boldsymbol{u}^{(1)}) P_\Delta(\boldsymbol{u}^{(2)}) \prod_{j \geq 1} P_\Delta(\boldsymbol{u}^{(3,j)}),$$

where at most one factor in the infinite product is different from 1. Together with the above observations, this implies that $\boldsymbol{u}$ has the H quasi factorization property. In [11], it is proved that the H quasi-factorization property implies the so-called H small cumulant property,[1] that is, in particular the following inequality:

$$|\kappa_r(X_i^j, X_i^j \in B)| = \mathcal{O}\left(\mathcal{M}(H) \prod_{X_i^j \in B} \mathbb{E}(X_i^j)\right),$$

which is what we needed to prove. □

3.2. *A preliminary estimate.* Let $M = \{1^{a_1}, 2^{a_2}, \dots\}$ be a finite multiset of size $n = \sum_{i \geq 1} a_i$. We let $(b_j)_{j \geq 1}$ be the nonincreasing reordering of $(a_i)_{i \geq 1}$ and define, for $j \geq 1$,

$$(13) \qquad\qquad n_{(j)} = n - b_1 - \cdots - b_j.$$

Furthermore, for an integer $d \geq 1$, we denote by $e_d(M)$ the dth elementary symmetric function evaluated in the numbers $a_1, a_2, \dots$, that is,

$$e_d(M) := \sum_{j_1 < \cdots < j_d} a_{j_1} \cdots a_{j_d}.$$

These quantities turn out to be omnipresent when we evaluate the various parameters needed to prove our central limit theorem.

LEMMA 3.2.   *For any $d \geq 1$ and any multiset $M = \{1^{a_1}, 2^{a_2}, \dots\}$, we have*

$$\frac{1}{d!} n n_{(1)} \cdots n_{(d-1)} \leq e_d(M) \leq n n_{(1)} \cdots n_{(d-1)}.$$

In practice, the degree d will be fixed, while n tends to infinity, so that the above lemma gives us the exact order of magnitude of $e_d(M)$. For a $d - 1$-regular sequence of multiset partition, we clearly have $n_j = \Theta(n)$ for any $j \leq d - 1$, so that $e_d(M) = \Theta(n^d)$.

---

[1] See [11], Proposition 5.8; the implication is proved for families $\boldsymbol{u}$ with nonzero entries, but the proof extends readily to families with the vanishing ideal property.

PROOF.   We start with the upper bound. Assume, without loss of generality that the sequence $(a_i)_{i \geq 1}$ is nonincreasing, in which case $n_{(h)} = n - a_1 - \cdots - a_h$. Then we have

$$
\begin{aligned}
e_d(a_1, a_2, \ldots) &= \sum_{j_1 < \cdots < j_d} a_{j_1} \cdots a_{j_d} \\
&\leq \left(\sum_{j_1 \geq 1} a_{j_1} \right) \left(\sum_{j_2 \geq 2} a_{j_2} \right) \cdots \left(\sum_{j_d \geq d} a_{j_d} \right) \\
&= n n_{(1)} \cdots n_{(d-1)}.
\end{aligned}
$$

For the lower bound, we first observe that

$$
d! e_d(a_1, a_2, \ldots) = \sum_{j_1 \geq 1} a_{j_1} \left(\sum_{j_2 \neq j_1} a_{j_2} \left(\cdots \left(\sum_{j_d \neq j_1, \ldots, j_{d-1}} a_{j_d} \right) \right) \right).
$$

The inner sum over j_d can be bounded below as follows:

$$
\left(\sum_{j_d \neq j_1, \ldots, j_{d-1}} a_{j_d} \right) = n - a_{j_1} - \cdots - a_{j_{d-1}} \geq n_{(d-1)}.
$$

This bound does not depend on $j_1, \ldots, j_{d-1}$ and can therefore be factorized out of all sums. We then bound the sum over j_{d-1} by $n_{(d-2)}$, which can also be factorized out. Iterating this procedure, we get

$$
d! e_d(a_1, a_2, \ldots) \geq n_{(d-1)} n_{(d-2)} \cdots n_{(1)} n,
$$

as claimed.   □

3.3. *The central limit theorem.*   The goal of this section is to prove Theorem 1.4. Fix a pattern τ and a sequence of finite multisets, where $M^{(m)} = \{1^{a_1^{(m)}}, 2^{a_2^{(m)}}, \ldots\}$. Most of the time we will omit the superscript m and denote $n = |M^{(m)}|$. We first observe that the number Occ^τ of occurrences of a pattern τ in a uniform random element σ in S_M can be written as

(14)
$$
\mathrm{Occ}^\tau(\sigma) = \sum_{\substack{i_1 < \cdots < i_\ell \leq n \\ j_1 < \cdots < j_\ell}} X_{i_1}^{j_{\tau(1)}} \cdots X_{i_\ell}^{j_{\tau(\ell)}}.
$$

Combining Theorem 3.1 and Proposition 2.6, we know that the family

$$
A_\tau := \left\{ X_{i_1}^{j_{\tau(1)}} \cdots X_{i_\ell}^{j_{\tau(\ell)}}, i_1 < \cdots < i_\ell \leq n, j_1 < \cdots < j_\ell \right\}
$$

admits $(G^M)^\ell$ as $(\boldsymbol{D}, \boldsymbol{\Psi})$ dependency graph, where $\boldsymbol{D}$ and $\boldsymbol{\Psi}$ are as follows:

- With a multiset B of monomials in A_τ, the function $\boldsymbol{\Psi}$ associates $\prod_{X_i^j} \frac{a_j}{n}$, where the product runs over the distinct variables X_i^j appearing in some monomial in B;
- $\boldsymbol{D} = (D_1, D_2, \ldots)$ is a sequence of constants independent of the set partition M under consideration.

Our goal is to apply the normality criterion of Theorem 2.3 to this (sequence of) dependency graph(s). The first task is to estimate the parameters R and T_h. For R, this is immediate; indeed, we write

$$
R = \sum_{\substack{i_1 < \cdots < i_\ell \leq n, \\ j_1 < \cdots < j_\ell}} \prod_{t=1}^{\ell} \frac{a_{j_t}}{n} = \binom{n}{\ell} \frac{e_\ell(M)}{n^\ell} = \Theta(e_\ell(M)).
$$

We now consider T_h. By definition,

$$(15) \qquad T_h := \max_{\alpha_1,\ldots,\alpha_h \in A_\tau} \left[\sum_{\beta \in A_\tau} W(\{\beta\}, \{\alpha_1, \alpha_2, \ldots, \alpha_h\}) \frac{\Psi(\{\alpha_1, \alpha_2, \ldots, \alpha_h, \beta\})}{\Psi(\{\alpha_1, \alpha_2, \ldots, \alpha_h\})} \right].$$

We fix a set $S = \{\alpha_1, \alpha_2, \ldots, \alpha_l\} \subseteq A_\tau$ and write

$$F(\beta) := W(\{\beta\}, \{\alpha_1, \alpha_2, \ldots, \alpha_h\}) \frac{\Psi(\{\alpha_1, \alpha_2, \ldots, \alpha_h, \beta\})}{\Psi(\{\alpha_1, \alpha_2, \ldots, \alpha_h\})}.$$

We use the $\mathcal{O}$-notation with implicit constants depending on h. We split the sum over β into three parts:

(i) Consider first the monomial $\beta = X_{i_1}^{j_\tau(1)} \cdots X_{i_\ell}^{j_\tau(\ell)}$, which do not share any index with any of the α in S (i.e., neither an upper, nor a lower index). For such β, the W factor in $F(\beta)$ is $1/n$, while the quotient of Ψ is $\prod_t \frac{a_{j_t}}{n}$, which yields

$$F(\beta) = n^{-\ell-1} \prod_{t=1}^{\ell} a_{j_t}.$$

We should sum this over ordered ℓ-uplets $(i_1, \ldots, i_\ell)$ and $(j_1, \ldots, j_\ell)$ with values in $[n]$. The sum over $(j_1, \ldots, j_\ell)$ gives $n^{-\ell-1} e_\ell(M)$. This is independent of $i_1, \ldots, i_\ell$ so that summing over the $\mathcal{O}(n^\ell)$ possible ℓ-uplets $(i_1, \ldots, i_\ell)$ gives $\mathcal{O}(n^{-1} e_\ell(M))$.

Summing up, the total contribution of monomials β not sharing any index with any α in S to the sum in (15) is $\mathcal{O}(n^{-1} e_\ell(M))$.

(ii) We now consider monomials β which do not share a lower index with any $\alpha \in S$ but do share some upper index, say j_r, with some $\alpha \in S$. For such β, we have that $W(\{\beta\}, \{\alpha_1, \alpha_2, \ldots, \alpha_h\}) = \frac{1}{\alpha_{j_r}}$ and

$$F(\beta) = \frac{1}{a_{j_r}} \prod_{t=1}^{\ell} \frac{a_{j_t}}{n} = n^{-\ell} \prod_{t \neq r} a_{j_t}.$$

Again, we should sum over ordered ℓ-uplets $(i_1, \ldots, i_\ell)$ and $(j_1, \ldots, j_\ell)$ with values in $[n]$. But the number of possible values of j_r is finite since it must be chosen among the lower indices of $\alpha_1, \ldots, \alpha_\ell$. Up to a constant factor, we can therefore only consider the sum over $j_1, \ldots, j_{r-1}, j_{r+1}, \ldots, j_\ell$, which gives $\mathcal{O}(n^{-\ell} e_{\ell-1}(M))$. Finally, summing over the $\mathcal{O}(n^\ell)$ possible ℓ-uplets $(i_1, \ldots, i_\ell)$ yields $\mathcal{O}(e_{\ell-1}(M))$. This is the total contribution to the sum in (15) of monomial β in this second case.

(iii) The third and last case is that of monomials β sharing some lower index, say i_r with some $\alpha \in S$. In that case, the W factor in $F(\beta)$ is one. Define

$$T(\beta) = \{t \in [\ell] : X_{i_t}^{j_\tau(t)} \text{ is } not \text{ a factor of some } \alpha \text{ in } S\} \subseteq [\ell].$$

Then we have $F(\beta) = \prod_{t \in T(\beta)} \frac{a_{j_\tau(t)}}{n} = n^{-|T(\beta)|} \prod_{t \in T(\beta)} a_{j_\tau(t)}$. The number of possible values for the set $T(\beta)$ is finite, so that it is enough to bound the sum in (15) over β's with a given value of $T(\beta)$.

Given $T_0 \subseteq [\ell]$, a monomial β with $T(\beta) = T_0$ is described by the lists $(i_t)_{t \in T_0 \setminus [r]}$ and $(j_{\tau(t)})_{t \in T_0}$, and some additional finite choices (the values of i_t and $j_{\tau(t)}$, for $t \notin T_0$, as well as that of i_r if $r \in T_0$). Similarly, as above, summing $F(\beta)$ over $(j_{\tau(t)})_{t \in T_0}$ gives $n^{-|T_0|} e_{|T_0|}(M)$, while summing over $(i_t)_{t \in T_0}$ yields $n^{|T_0|-1}$ or $n^{|T_0|}$, depending on whether r is in T_0 or not. Therefore, the total contribution to the sum in (15) of monomials β in this third case with $T(\beta) = T_0$ is either $\mathcal{O}(n^{-1} e_d(M))$ for $d := |T_0| \leq \ell$ (in the case $r \in T_0$) or $\mathcal{O}(e_d(M))$ for $d := |T_0| \leq \ell - 1$ (in the case $r \notin T_0$).

From Lemma 3.2, we see that, as long as $n_{(\ell-1)}$ tends to infinity (so in particular for an ℓ-regular sequences of multiset partition $M^{(m)}$), the biggest of the above bounds is $\mathcal{O}(e_{\ell-1}(M))$. We therefore conclude that $T_h = \mathcal{O}(e_{\ell-1}(M))$.

Next, we need a lower bound on the variance. Here, we need our regularity assumption on the sequence of multiset partitions.

PROPOSITION 3.3. *Let τ be a fixed pattern of size ℓ and consider a ℓ-regular sequence $M^{(m)}$ of multisets. There exists a constant $K_1 > 0$ such that*

$$\mathrm{Var}\big(\mathrm{Occ}^{\tau}_{M^{(m)}}\big) \geq K_1 \big|M^{(m)}\big|^{2\ell - 3/2}.$$

The proof of this statement being rather technical, we postpone it to the end of the multiset partition part, that is, Section 3.4.

We can now prove Theorem 1.4, using the normality criterion given in Theorem 2.3. We assume that the sequence of multiset partitions $M^{(m)}$ is regular. This implies the estimate $e_{\ell-1}(M) = \Theta(n^{\ell-1})$, so that $T_h = \mathcal{O}(n^{\ell-1})$; therefore, we can take $Q_n = n^{\ell-1}$ in Theorem 1.4. Combining with Proposition 3.3, we find $\frac{Q_n}{\sigma_n} \leq \frac{1}{\sqrt{K_1}} \frac{n^{\ell-1}}{n^{\ell-3/4}} = \mathcal{O}(n^{-1/4})$. Furthermore, $R_n = \Theta(e_\ell(M)) = \mathcal{O}(n^\ell)$. Summing up, for $s = 5$, we have

$$\left(\frac{R_n}{Q_n}\right)^{1/s} \frac{Q_n}{\sigma_n} = \mathcal{O}(n^{1/5} n^{-1/4}) = \mathcal{O}(n^{-1/20}).$$

This tends to 0 as n tends to infinity, and (7) is satisfied. We conclude that $\mathrm{Occ}^{\tau}_{M^{(m)}}$ is asymptotically normal, as wanted.

3.4. *Variance estimate for pattern counts in multiset permutations.* As above, we fix a pattern τ of size ℓ. Throughout this section, we assume that $M^{(m)}$ is an ℓ-regular sequence of multisets. We will most of the time drop to superscript m and denote $n = |M^{(m)}|$. The goal of this section is to prove Proposition 3.3, that is, to bound $\mathrm{Var}(\mathrm{Occ}^{\tau}(\sigma))$ from below, where σ is a random uniform multiset permutation of M. We use the notation $v(M) = \mathrm{Var}(\mathrm{Occ}^{\tau}(\sigma))$, making the dependence in τ implicit.

3.4.1. *An initial bound and a recursive inequality on the variance.* Let $M = \{1^{a_1}, 2^{a_2}, \ldots\}$ be a multiset of size n and choose j_0 minimal such that $a_{j_0} > 0$. We denote M' the multiset obtained by removing a single copy of j_0 from M.

Then a random uniform multiset permutation σ of M can be constructed as follows:

- Choose P uniformly at random between 1 and n and set $\sigma_P = j_0$.
- Take a uniform random multiset permutation σ' of M', independent from P, and fill the other positions of σ in the same order as in σ'. Formally, we set, for $i \neq P$,

$$\sigma_i = \begin{cases} \sigma'_i & \text{if } i < P; \\ \sigma'_{i-1} & \text{if } i > P. \end{cases}$$

It is straightforward to check that, by construction, σ is a uniform random multiset permutation of M. Moreover, we can write $\mathrm{Occ}^{\tau}(\sigma) = B + C$, where:

- B is the number of occurrences of τ in σ, *not using* the position P—it is easy to see that $B = \mathrm{Occ}^{\tau}(\sigma')$;
- C is the number of occurrences of τ in σ, using the position P.

Using the law of total variance, we have

$$v(M) = \text{Var}(\text{Occ}^\tau(\sigma)) = \mathbb{E}\big[\text{Var}(\text{Occ}^\tau(\sigma)|P)\big] + \text{Var}\big[\mathbb{E}(\text{Occ}^\tau(\sigma)|P)\big].$$

Note that B is independent of P, so that $\mathbb{E}(B|P)$ is the *constant* random variable a.s. equal to $\mathbb{E}(B)$. This implies that

$$\text{Var}\big[\mathbb{E}(\text{Occ}^\tau(\sigma)|P)\big] = \text{Var}\big[\mathbb{E}(B) + \mathbb{E}(C|P)\big] = \text{Var}\big[\mathbb{E}(C|P)\big].$$

In particular, we get an initial bound

$$(16) \qquad\qquad v(M) \geq \text{Var}\big[\mathbb{E}(C|P)\big].$$

On the other hand, from Cauchy–Schwarz and Jensen inequalities, since B and P are independent, we have

$$\big|\mathbb{E}[\text{Cov}(B, C|P)]\big| \leq \mathbb{E}\big[\sqrt{\text{Var}(B)}\sqrt{\text{Var}(C|P)}\big] \leq \sqrt{\text{Var}(B)}\sqrt{\mathbb{E}[\text{Var}(C|P)]}.$$

Expanding $\text{Var}(\text{Occ}^\tau(\sigma)|P) = \text{Var}(B + C|P)$ by bilinearity, we find that

$$\mathbb{E}\big[\text{Var}(\text{Occ}^\tau(\sigma)|P)\big] = \text{Var}(B) + 2\mathbb{E}\big[\text{Cov}(B, C|P)\big] + \mathbb{E}\big[\text{Var}(C|P)\big]$$

$$\geq \text{Var}(B) - 2\sqrt{\text{Var}(B)}\sqrt{\mathbb{E}\big[\text{Var}(C|P)\big]}.$$

Note that $\text{Var}(B) = v(M')$ since $B = \text{Occ}^\tau(\sigma')$. Summing up, we get the following recursive inequality on $v(M)$:

$$(17) \qquad\qquad v(M) \geq v(M')\left(1 - 2\sqrt{\frac{\mathbb{E}[\text{Var}(C|P)]}{v(M')}}\right) + \text{Var}\big[\mathbb{E}(C|P)\big].$$

3.4.2. *Analyzing the initial bound* (16). Recall that C counts the number of occurrences of τ in σ that use the position P. Since the letter at position P is the smallest one in σ, it should correspond to 1 in the pattern τ. Formally, if r is the index such that $\tau(r) = 1$, an occurrence $(i_1, \ldots, i_\ell)$ of τ in σ using the position P must satisfy $i_r = P$. We therefore have, conditionally on P,

$$(18) \qquad\qquad C = \sum_{\substack{i_1 < \cdots < i_{r-1} < P \\ i_\ell > \cdots > i_{r+1} > P}} \left[\sum_{j_\ell > \cdots > j_2 > j_1 = j_0} X_{i_1}^{j_\tau(1)} \cdots X_{i_\ell}^{j_\tau(\ell)}\right].$$

Note that the factor $X_{i_r}^{j_\tau(r)} = X_P^{j_0} = 1$ because of the condition on the summation index. For any $i_1, \ldots, i_{r-1}, i_{r+1}, \ldots, i_\ell$, we have

$$(19) \qquad \begin{aligned} &\mathbb{E}\big[X_{i_1}^{j_\tau(1)} \cdots X_{i_{r-1}}^{j_\tau(r-1)} X_{i_{r+1}}^{j_\tau(r+1)} \cdots X_{i_\ell}^{j_\tau(\ell)}|P\big] \\ &= \mathbb{E}\big[(X')_{i_1}^{j_\tau(1)} \cdots (X')_{i_{r-1}}^{j_\tau(r-1)} (X')_{i_{r+1}-1}^{j_\tau(r+1)} \cdots (X')_{i_\ell-1}^{j_\tau(\ell)}\big], \end{aligned}$$

where the variables $(X')_i^j$ refer to the multiset permutation σ'. In particular, the right-hand side is a quantity $F(j_2, \ldots, j_\ell)$ that depends neither on P, nor $i_1, \ldots, i_\ell$ (the uniform random multiset permutation σ' is invariant by re-indexing). Since the number of choices for the indices $i_1, \ldots, i_{r-1}, i_{r+1}, \ldots, i_\ell$ is $\binom{P-1}{r-1}\binom{n-P}{\ell-r}$, we have

$$\mathbb{E}(C|P) = \binom{P-1}{r-1}\binom{n-P}{\ell-r}\left(\sum_{j_\ell > \cdots > j_2 > j_1 = j_0} F(j_2, \ldots, j_\ell)\right).$$

We take the variance of this function of the random variable P (the sum is independent of P, i.e., deterministic):

$$\text{Var}\big[\mathbb{E}(C|P)\big]$$

(20)
$$= \Bigg(\sum_{j_\ell > \cdots > j_2 > j_1 = j_0} F(j_2, \ldots, j_\ell) \Bigg)^2 \text{Var}\left(\binom{P-1}{r-1} \binom{n-P}{\ell-r} \right).$$

Since P is uniformly distributed in $\{1, \ldots, n\}$, we see easily that the variance of the product of binomials is of order $n^{2\ell-2}$. Moreover, equation (11) gives us (recall that the j_t are distinct here)

$$F(j_2, \ldots, j_\ell) = \frac{a_{j_2} \cdots a_{j_t}}{(n-1) \cdots (n - \ell + 1)} \geq n^{-\ell+1} \prod_{t=2}^{\ell} a_{j_t}.$$

Therefore, the sum in (20) is bigger than $e_{\ell-1}(M \setminus (j_0))n^{-\ell+1}$, where $M \setminus (j_0)$ is obtained form M by removing *all* copies of j_0. Note that, if a sequence $M^{(m)}$ of multiset partition is ℓ regular, then $e_{\ell-1}(M \setminus (j_0)) = \Theta(n^{\ell-1})$. We conclude that there exist $K_2, K_3 > 0$ such that

(21)
$$\text{Var}\big[\mathbb{E}(C|P)\big] \geq K_2 \big(e_{\ell-1}(M \setminus (j_0))\big)^2 \geq K_3 n^{2\ell-2}.$$

In particular, using (16), we have

(22)
$$v(M) \geq K_3 n^{2\ell-2}.$$

3.4.3. *Analyzing the recursive inequality.* The lower bound (22) is not sufficient to apply Theorem 2.3 directly. We shall use the recursive inequality equation (17) to improve it. To this end, we first need to analyze the term $\mathbb{E}[\text{Var}(C|P)]$.

LEMMA 3.4. *Let $M^{(m)}$ be an ℓ-regular sequence of multisets. There exist $K_4, K_5 > 0$ such that*

$$\mathbb{E}\big[\text{Var}(C|P)\big] \leq K_4 e_{\ell-1}(M \setminus (j_0)) e_{\ell-2}(M \setminus (j_0)) \leq K_5 n^{2\ell-3}.$$

PROOF. We start from equations (18) and (19). Conditionally on P, we have (note that the indices i_t for $t \geq r+1$ below are shifted by 1, compared to equations (18) and (19)):

$$C = \sum_{\substack{i_1 < \cdots < i_{r-1} < P \\ i_\ell > \cdots > i_{r+1} \geq P \\ j_\ell > \cdots > j_2 > j_1 = j_0}} (X')_{i_1}^{j_{\tau(1)}} \cdots (X')_{i_{r-1}}^{j_{\tau(r-1)}} (X')_{i_{r+1}}^{j_{\tau(r+1)}} \cdots (X')_{i_\ell}^{j_{\tau(\ell)}}.$$

In the sequel, we use $\boldsymbol{i}$ and $\boldsymbol{j}$ to represent lists $(i_1, \ldots, i_{r-1}, i_{r+1}, \ldots, i_\ell)$ and $(j_2, \ldots, j_\ell)$ as in the above summation index. We also write $(X')_{\boldsymbol{i}}^{\boldsymbol{j}}$ for the corresponding monomial. We have

$$\text{Var}(C|P) = \sum_{\boldsymbol{i}, \tilde{\boldsymbol{i}}, \boldsymbol{j}, \tilde{\boldsymbol{j}}} \text{Cov}\big((X')_{\boldsymbol{i}}^{\boldsymbol{j}}, (X')_{\tilde{\boldsymbol{i}}}^{\tilde{\boldsymbol{j}}}\big).$$

Since σ' is a uniform random multiset permutation of M', the variable $(X')_{\boldsymbol{i}}^{\boldsymbol{j}}$ admit a $(\boldsymbol{\Psi}, \boldsymbol{D})$-weighted dependency graph (by Theorem 1.4 and Proposition 2.6) and we have that, for some constant $K_4 > 0$

$$\big|\text{Cov}\big((X')_{\boldsymbol{i}}^{\boldsymbol{j}}, (X')_{\tilde{\boldsymbol{i}}}^{\tilde{\boldsymbol{j}}}\big)\big| \leq K_4 \boldsymbol{\Psi}\big((X')_{\boldsymbol{i}}^{\boldsymbol{j}}, (X')_{\tilde{\boldsymbol{i}}}^{\tilde{\boldsymbol{j}}}\big) W\big((X')_{\boldsymbol{i}}^{\boldsymbol{j}}, (X')_{\tilde{\boldsymbol{i}}}^{\tilde{\boldsymbol{j}}}\big).$$

With the same case distinction as in Section 3.3 on whether $(X')^j_i$ and $(X')^{\tilde{j}}_{\tilde{i}}$ share a lower index, an upper index or no index at all, we can prove that

$$\left|\mathrm{Var}(C|P)\right| \leq K_4 e_{\ell-1}(M \setminus (j_0)) e_{\ell-2}(M \setminus (j_0)).$$

The upper bound is independent of P, so that

$$\mathbb{E}\left[\mathrm{Var}(C|P)\right] \leq K_4 e_{\ell-1}(M \setminus (j_0)) e_{\ell-2}(M \setminus (j_0)).$$

For ℓ-regular sequences of multisets, the upper bound is smaller than $K_5 n^{2\ell-3}$ (for some $K_5 > 0$), concluding the proof of the lemma. $\square$

Plugging in equations (21) and (22) and Lemma 3.4 in equation (17), we get that, for some constant $K_6 > 0$,

$$(23) \qquad\qquad v(M) \geq v(M')\left(1 - K_6 n^{-1/2}\right) + K_3 n^{2\ell-2}.$$

We denote $M_{(i)} = ((M') \cdots)'$ the partition obtained by removing the i smallest parts of M (counting parts with repetitions). For $i \leq K_7 n^{1/2}$ (where K_7 is a positive constant that will be determined later), the sequence $M_{(i)}$ is still ℓ-regular. Therefore, we can apply (23) and we have for

$$v(M_{(i)}) \geq v(M_{(i+1)})\left(1 - K_6 (n-i)^{-1/2}\right) + K_3 (n-i)^{2\ell-2}.$$

We start from the initial inequality $v(M_{(K_7\sqrt{n})}) \geq K_3 (n - K_7\sqrt{n})^{2\ell-2}$ and iterate the above recursive inequality: we get

$$v(M) \geq \sum_{i=0}^{K_7\sqrt{n}-1} K_3 (n-i)^{2\ell-2}\left(1 - K_6(n-i)^{-1/2}\right) \cdots \left(1 - K_6 n^{-1/2}\right).$$

For K_7 sufficiently small, the product is bounded away from 0, so that each of the $\Theta(\sqrt{n})$ terms in the sum behaves as $\Theta(n^{2\ell-2})$. Therefore, there exists a constant $K_1 > 0$ such that

$$(24) \qquad\qquad v(M) \geq K_1 n^{2\ell-3/2},$$

which is exactly what we wanted to prove.

REMARK 3.5.   Plugging in the final inequality (24) in (22), we could improve (23) to

$$v(M) \geq v(M')\left(1 - K_6 n^{-3/4}\right) + K_3 n^{2\ell-2}.$$

Then, arguing as above with $i \leq K_8 n^{3/4}$, we have $v(M) \geq K_9 n^{2\ell-5/4}$ (for some $K_9 > 0$). Iterating this argument shows that for any $\varepsilon > 0$, there exists $K_{10}(\varepsilon) > 0$ such that $v(M) \geq K_{10}(\varepsilon) n^{2\ell-1-\varepsilon}$. However, the bound $v(M) \geq K_1 n^{2\ell-3/2}$ given above is sufficient to apply our normality criterion.

3.4.4. *Comparison with an upper bound.* Proposition 2.4 and the estimates $R = \mathcal{O}(e_\ell(M))$, $T_1 = \mathcal{O}(e_{\ell-1}(M))$ from Section 3.3 yield the following upper bound on the variance, which is valid without the regularity hypothesis.

PROPOSITION 3.6.   *There exists a constant $K_{11} > 0$, such that for each multiset permutation $M = \{1^{a_1}, 2^{a_2}, \dots\}$, we have*

$$\mathrm{Var}(\mathrm{Occ}^\tau) \leq K_{11} e_\ell(M) e_{\ell-1}(M).$$

We note in the case $\tau = 21$, that is, when we are interested in inversions in random multiset permutations, this upper bound is tight (up to a multiplicative constant): see [5], equation (1.10) or Lemma 3.1. It is natural to ask whether this is also the case for longer patterns.

QUESTION 3.7. *Fix a pattern τ. Does there exist a constant $K_{12} > 0$ (depending on τ), such that for each multiset permutation $M = \{1^{a_1}, 2^{a_2}, \ldots\}$, we have*

$$(25) \qquad \mathrm{Var}(\mathrm{Occ}^\tau) \geq K_{12} e_\ell(M) e_{\ell-1}(M)?$$

An affirmative answer to this question would imply that the central limit in Theorem 1.4 holds under the less restrictive condition that the sequence of multiset permutations satisfy $n_{(\ell-1)} \to \infty$ ($n_{(j)}$ is defined in (13)). This condition is easily seen to be necessary for having a central limit theorem (see [5], Section 5, for the case of inversions).

For a regular sequence $M = M^{(m)}$ of multisets, we have the estimate $e_\ell(M) e_{\ell-1}(M) = \Theta(n^{2\ell-1})$, and we can prove $\mathrm{Var}(\mathrm{Occ}^\tau) \geq K_{10}(\varepsilon) n^{2\ell-1-\varepsilon}$ for arbitrarily small $\varepsilon > 0$; see Remark 3.5. Therefore, in this case, the suggested lower bound (25) holds, at least up to subpolynomial factors.

4. Arc patterns in set partitions.

In this section, we prove Theorem 1.6. We refer to Section 1.4 for notation.

4.1. *Stam's urn model and the weighted dependency graph.* Stam's urn model [31] gives a simple way to uniformly sample from $\mathcal{P}([n])$. It works as follows: the first step consists in picking the number of urns M according to the distribution

$$(26) \qquad \mathbb{P}(M = m) = \frac{1}{e B_n} \frac{m^n}{m!},$$

where $B_n = |\mathcal{P}([n])|$ is the nth Bell number. This is indeed a probability measure, thanks to Dobiński's formula; see [28] for an insightful proof.

In the second step, we drop each number $i \in [n]$ into one of M urns with uniform probability $1/M$. We denote by π the random set partition of $[n]$ in which i and j are in the same block if and only if they were dropped into the same urn. This construction of a random set partition will be referred to as Stam's urn model.

PROPOSITION 4.1 ([31]). *The random set partition π constructed by Stam's urn model is uniformly distributed on $\mathcal{P}([n])$. Moreover, the number of empty urns in the process has law* Poisson(1) *and is independent from π.*

For $1 \leq i < j \leq n$, we consider the random variable

$$X_{ij} = \mathbf{1}[\text{there is an arc from } i \text{ to } j \text{ in } \pi].$$

THEOREM 4.2. *Consider the weighted graph G^n on vertex-set A^n with weights*

$$w(X_{ij}, X_{i'j'}) = \begin{cases} 1 & \text{if } i = i' \text{ or } j = j'; \\ 1/n & \text{otherwise.} \end{cases}$$

Then, for each $n \geq 1$, the graph G^n is a (Ψ, C_n) weighted dependency graph for the family A^n, where Ψ is the function on multisets of elements of A^n defined by

$$\Psi(B) = n^{-\#(B)}$$

and $(C_n)_{n \geq 1} = (C_{r,n})_{r \geq 1, n \geq 1}$ is a doubly indexed sequence of coefficients such that for each r, we have $C_{r,n} = \tilde{\mathcal{O}}(1)$ as n tends to infinity.

The proof is postponed to Section 5, but let us say a few words on its structure. We should bound joint cumulants $\kappa(X_{ij}, X_{ij} \in B)$ for multiset B of elements A^n; with the same kind of observations as in the proof of Theorem 3.1, we can assume w.l.o.g. that B is a (repetition-free) subset of A^n. The difficulty here is that the joint moments of X_{ij} do not have a simple form as in equation (11). However, conditionally on M, joint moments do have a nice multiplicative expression. Indeed, a random set partition π generated by Stam's urn model contains all the arcs in some set B if and only if:

(i) for every arc (i, j) in B, we drop its endpoint j in the same urn as its starting point i;

(ii) for every arc (i, j) in B, none of the integers between i and j is dropped in the same urn as i; equivalently, if g is not the endpoint of an arc of B, it should be dropped in a different urn from the starting points of those arcs in B that go over g.

Since all balls are dropped uniformly independently, this happens with probability

$$(27) \qquad \mathbb{E}\left(\prod_{X_{ij} \in B} X_{ij} | M \right) = \frac{1}{M^a} \prod_{g:g \notin \{j, X_{ij} \in B\}} \frac{M - a(g)}{M},$$

where $a(g)$ is the number of arcs in B going over g. Using the multiplicative form of equation (27) and general results from the theory of weighted dependency graphs, we can show that the *conditional* joint cumulants $\kappa(X_{ij} \in B | M)$ are small.

To go back to unconditional cumulants, we use Brillinger's *law of total cumulance* [3], which we now recall. If $X_1, \ldots, X_r$ and Y are random variables with finite moments defined on the same probability space, then

$$(28) \qquad \kappa(X_1, \ldots, X_r) = \sum_{\rho} \kappa\left(\kappa(X_i, i \in B | Y), B \in \rho \right),$$

where the sum runs over all set partitions ρ of $[r]$ and the B's are the blocks of ρ.

Note that the inner conditional cumulants are functions of the random variable Y, that is, in our setting, of M. We therefore need estimates for cumulants of particular functions of M, which we derive in Section 5.2 as a preliminary to the proof of Theorem 4.2.

In the rest of this section, we admit Theorem 4.2 and show how to deduce Theorem 1.6 from it.

4.2. *A lower bound for the variance.* As in the statement of Theorem 1.6, we fix an arc pattern $\mathcal{A}$ of length ℓ with a arcs and denote $\mathrm{Occ}_n^{\mathcal{A}}$ its number of occurrences in a uniform random set partition of size n. We would like a lower bound on its variance.

To this end, we use the law of total variance, conditioning on the value of the number M of urns in Stam's construction:

$$\mathrm{Var}(\mathrm{Occ}_n^{\mathcal{A}}) = \mathbb{E}\left[\mathrm{Var}(\mathrm{Occ}_n^{\mathcal{A}} | M) \right] + \mathrm{Var}\left[\mathbb{E}(\mathrm{Occ}_n^{\mathcal{A}} | M) \right].$$

Both terms are nonnegative; it turns out that the second one is relatively easy to evaluate and gives us a good enough lower bound (in fact, this lower bound is optimal up to logarithmic factors, as explained at the end of Section 4.3).

LEMMA 4.3. $\mathrm{Var}(\mathrm{Occ}_n^{\mathcal{A}}) \geq \mathrm{Var}[\mathbb{E}(\mathrm{Occ}_n^{\mathcal{A}} | M)] = \tilde{\Theta}(n^{2\ell - 2a - 1}).$

PROOF. We start by introducing notation. For i in $\{2, \ldots, \ell\}$, we denote by a_i the number of arcs above the segment $[i - 1, i]$ in $\mathcal{A}$. For example, if $\mathcal{A}$ is the arc pattern of Figure 1, then $a_2 = a_5 = 1$ and $a_3 = a_4 = 2$. By convention, we set $a_1 = 0$.

Recall from equation (27) above that

$$(29) \qquad \mathbb{P}\big((x_1, \ldots, x_\ell) \text{ occurrence of } \mathcal{A} | M\big) = \frac{1}{M^a} \prod_{g: g \notin \{x_j, (i,j) \in \mathcal{A}\}} \frac{M - a(g)}{M}.$$

The product can be split in two parts.

- First, consider factors indexed by g in $\{x_1, \ldots, x_\ell\} \setminus \{x_j, (i,j) \in \mathcal{A}\}$, that is, by elements of the pattern that are not the ending point of an arch. There are a fixed number of such g, and the corresponding numbers $a(g)$ do not depend on $x_1, \ldots, x_\ell$ nor on n. Thus the product of such factors is a Laurent polynomial of degree 0 in M, say $R_0(M)$, independent of $x_1, \ldots, x_j$ (and of n);
- Second, we consider factors corresponding to $g \notin \{x_1, \ldots, x_\ell\}$: for each i in $[\ell]$, we get $x_i - x_{i-1} - 1$ factors equal to $\frac{M - a_i}{M}$ (by convention $x_0 = 0$).

Hence (29) rewrites as

$$\mathbb{P}\big((x_1, \ldots, x_\ell) \text{ occurrence of } \mathcal{A} | M\big) = \frac{R_0(M)}{M^a} \prod_{i \le \ell} \left(\frac{M - a_i}{M}\right)^{x_i - x_{i-1} - 1}.$$

Summing this conditional probability over $x_1 < x_2 < \cdots < x_\ell \le n$, we get the conditional expectation of $\mathrm{Occ}_n^{\mathcal{A}}$. For convenience, we rather write this formula with summation indices $y_i = x_i - x_{i-1}$ (with the convention $x_0 = 0$, i.e., $y_1 = x_1$):

$$\mathbb{E}\big[\mathrm{Occ}_n^{\mathcal{A}} | M\big] = \frac{R_0(M)}{M^a} \sum_{\substack{y_1, \ldots, y_\ell \ge 1 \\ y_1 + \cdots + y_\ell \le n}} \prod_{i \le \ell} \left(\frac{M - a_i}{M}\right)^{y_i - 1}.$$

We denote t the number of i such that $a_i \ne 0$. Since the above expression is symmetric in the a_i, we may assume as well that $a_1, \ldots, a_t$ are nonzero, while $a_{t+1} = \cdots = a_\ell = 0$. The summand in the above display does not depend on $(y_i, i > t)$, so that we can write

$$(30) \qquad \mathbb{E}\big[\mathrm{Occ}_n^{\mathcal{A}} | M\big] = \frac{R_0(M)}{M^a} \sum_{\substack{y_1, \ldots, y_t \ge 1 \\ y_1 + \cdots + y_t \le n}} \binom{n - y_1 - \cdots - y_t}{\ell - t} \prod_{i \le t} \left(1 - \frac{a_i}{M}\right)^{y_i - 1}.$$

The next step is to see how the sum behaves when $n \to \infty$, and M is closed to its expectation

$$m_n := \mathbb{E}(M) \sim n / \ln(n).$$

We will see later that M concentrates around m_n; see beginning of Section 5.2. Informally, the dominant term is obtained as follows:

(i) replace the binomial coefficient by its dominant part, which is

$$\frac{1}{(\ell - t)!} (n - y_1 - \cdots - y_t)^{\ell - t};$$

(ii) forget the condition $y_1 + \cdots + y_t \le n$ in the sum;
(iii) use the approximation

$$\sum_{y \ge 1} y^j \left(1 - \frac{a}{M}\right)^{y - 1} \approx \sum_{y \ge 1} (y + j - 1) \cdots y \left(1 - \frac{a}{M}\right)^{y - 1}$$

$$= j! \left(\frac{1}{1 - (1 - \frac{a}{M})}\right)^{j+1} = j! \left(\frac{M}{a}\right)^{j+1}.$$

(For the middle equality, note that both sides are the jth derivative of $1/(1-x)$ evaluated in $x = 1 - \frac{a}{M}$.) With this heuristic, we get

$$\sum_{\substack{y_1,\ldots,y_t \geq 1 \\ y_1 + \cdots + y_t \leq n}} \binom{n - y_1 - \cdots - y_t}{\ell - t} \prod_{i \leq t} \left(1 - \frac{a_i}{M}\right)^{y_i - 1}$$

$$\approx \frac{1}{(\ell - t)!} \sum_{\substack{j_0,\ldots,j_t \geq 0 \\ j_0 + \cdots + j_t = \ell - t}} \binom{\ell - t}{j_0, \ldots, j_t} n^{j_0} \prod_{i \leq t} \left[\sum_{y_i \geq 1} (-y_i)^{j_i} \left(1 - \frac{a_i}{M}\right)^{y_i - 1}\right]$$

$$\approx \sum_{\substack{j_0,\ldots,j_t \geq 0 \\ j_0 + \cdots + j_t = \ell - t}} (-1)^{j_1 + \cdots + j_t} \frac{n^{j_0} M^{j_1 + \cdots + j_t + t}}{j_0! a_1^{j_1 + 1} \cdots a_t^{j_t + 1}}.$$

The summand corresponding to $j_0 = \ell - t$, $j_1 = \cdots = j_t = 0$ dominates the above sum. Inserting this estimate back into equation (30), we deduce heuristically the following: uniformly on M in $[m_n - n^{3/4}, m_n + n^{3/4}]$, one has

$$(31) \qquad\qquad \mathbb{E}[\mathrm{Occ}_n^{\mathcal{A}} | M] = \frac{n^{\ell - t} M^{t - a}}{(\ell - t)!}(1 + o(1)).$$

Taking the variance of this function of M, we expect to get (see Section 5.2 for a justification of this heuristic):

$$(32) \qquad \mathrm{Var}(\mathbb{E}[\mathrm{Occ}_n^{\mathcal{A}} | M]) = \mathbb{E}[\mathbb{E}[\mathrm{Occ}_n^{\mathcal{A}} | M]^2]\tilde{\Theta}(n^{-1}) = \tilde{\Theta}(n^{2\ell - 2a - 1}).$$

In Section 6, we prove formally equations (31) and (32). In particular, this completes the proof of the lemma. $\square$

4.3. *The central limit theorem.*   In this section, we prove Theorem 1.6. The quantity $\mathrm{Occ}_n^{\mathcal{A}}$ of interest can be written in terms of the X_{ij} as follows:

$$\mathrm{Occ}_n^{\mathcal{A}} = \sum_{s_1 < s_2 < \cdots < s_\ell} \prod_{(i,j) \in \mathcal{A}} X_{s_i s_j}.$$

From Theorem 4.2 and Proposition 2.6, the family of monomials $\prod_{(i,j) \in \mathcal{A}} X_{s_i s_j}$ has a $(\boldsymbol{\Psi}_n, \boldsymbol{D}_n)$-weighted dependency graph, where:

- $\boldsymbol{\Psi}_n$ is the function on multisets of monomials $\prod_{(i,j) \in \mathcal{A}} X_{s_i s_j}$, which is $n^{-\#\mathrm{arcs}}$, where #arcs denotes the total number of distinct arcs $X_{s_i s_j}$ appearing in some monomial in the multiset;
- $(\boldsymbol{D}_n)_{n \geq 1} = (D_{r,n})_{r \geq 1, n \geq 1}$ is a doubly indexed sequence of numbers such that, for each $r \geq 1$, we have $D_{r,n} = \tilde{\mathcal{O}}(1)$.

We will prove the central limit theorem for $\mathrm{Occ}_n^{\mathcal{A}}$ by applying Theorem 2.3. For this, we need to evaluate the quantities R_n and $T_{h,n}$ defined in equations (4) and (5). We have

$$R_n := \sum_{1 \leq s_1 < \cdots < s_\ell \leq n} \Psi\left(\left\{\prod_{(i,j) \in \mathcal{A}} X_{s_i s_j}\right\}\right) = \Theta(n^{\ell - a}).$$

(Recall that a is the number of arcs in $\mathcal{A}$.) We now consider $T_{h,n}$. In the following, we use $\alpha^{(k)}$ or β to denote some ℓ-uple of positive integers $s_1 < \cdots < s_\ell \leq n$ and $\Pi(\alpha^{(k)})$ or $\Pi(\beta)$ for the corresponding monomial $\prod_{(i,j) \in \mathcal{A}} X_{s_i s_j}$. By definition,

$$T_{h,n} = \max_{\alpha^{(1)},\ldots,\alpha^{(h)}}\left[\sum_\beta F(\beta)\right],$$

where

$$F(\beta) = W\big(\{\Pi(\beta)\}, \{\Pi(\alpha^{(1)}), \dots, \Pi(\alpha^{(h)})\}\big) \frac{\Psi(\{\Pi(\beta), \Pi(\alpha^{(1)}), \dots, \Pi(\alpha^{(h)})\})}{\Psi(\{\Pi(\alpha^{(1)}), \dots, \Pi(\alpha^{(h)})\})}.$$

To evaluate this quantity, we fix $\alpha^{(1)}, \dots, \alpha^{(h)}$. All constants in $\mathcal{O}$ terms below depend implicitly on h, but not on $\alpha^{(1)}, \dots, \alpha^{(h)}$.

- Consider first the terms where β is disjoint from $\alpha^{(1)} \cup \dots \cup \alpha^{(h)}$. For such β, the weight in $F(\beta)$ is $1/n$, while the quotient of Ψ is n^{-a}. Thus, we have $F(\beta) = n^{-1-a}$. Since there are $\mathcal{O}(n^{\ell})$ such terms, their total contribution is $\mathcal{O}(n^{\ell-a-1})$.
- Consider now terms where β intersects $\alpha^{(1)} \cup \dots \cup \alpha^{(h)}$, but there is no factor in common between $\Pi(\beta)$ and any $\Pi(\alpha^{(k)})$. There are only $\mathcal{O}(n^{\ell-1})$ such terms. In this case, we bound the weight in $F(\beta)$ by 1, and the quotient of Ψ's is still n^{-a}. The total contribution of such terms is therefore also $\mathcal{O}(n^{\ell-a-1})$.
- We finally consider terms, where some factors of $\Pi(\beta)$, say g of them, are already present in some $\Pi(\alpha^{(k)})$ (possibly different factors are in different $\Pi(\alpha^{(k)})$). This forces $\beta \cap (\alpha^{(1)} \cup \dots \cup \alpha^{(h)})$ to be of size at least $g+1$, so that the number of such β is $\mathcal{O}(n^{\ell-g-1})$. Again we bound the weight in $F(\beta)$ by 1, but now the quotient of Ψ's is n^{-a+g}. The total contribution of such terms is therefore also $\mathcal{O}(n^{\ell-a-1})$ as well.

From this discussion, for any fixed $h \geq 1$, we have $T_{h,n} = \mathcal{O}(n^{\ell-a-1})$. We can therefore set $Q_n = n^{\ell-a-1}$ in Theorem 2.3.

From Lemma 4.3, we know that

$$\sigma_n := \sqrt{\mathrm{Var}(\mathrm{Occ}_n^{\mathcal{A}})} \geq \tilde{\Theta}(n^{\ell-a-1/2}).$$

Therefore, for $s = 3$, we have

$$\left(\frac{R_n}{Q_n}\right)^{1/s} \frac{Q_n}{\sigma_n} = \tilde{\Theta}(n^{-1/6})$$

and thus, this quantity tends to 0 faster than any power of $D_{r,n}$ (for any fixed $r \geq 1$). From Theorem 2.3 and Remark 2.5, we conclude that $\mathrm{Occ}_n^{\mathcal{A}}$ is asymptotically normal.

We still need to justify the expectation and variance estimates. From equation (31) above and using that $\mathbb{E}(M^{t-a}) = \tilde{\Theta}(n^{t-a})$ (see equation (37) below), we have $\mathbb{E}(\mathrm{Occ}_n^{\mathcal{A}}) = \tilde{\Theta}(n^{\ell-a})$. For the variance, a lower bound is obtained in Lemma 4.3. A matching upper bound comes from Proposition 2.4, using the estimates $R_n = \Theta(n^{\ell-a})$ and $T_{1,n} = \mathcal{O}(n^{\ell-a-1})$ given above.

5. The weighted dependency graphs for set partitions.
The goal of this section is to prove Theorem 4.2, that is, that a given weighted graph is a weighted dependency graphs for the presence of arcs in set partitions.

5.1. *Two general simple estimate for cumulants.*
We start by two easy bounds on cumulants. Denote $B_r = \sum_{\pi \in \mathcal{P}([r])} |\mu(\pi, [r])|$, which is a universal constant depending only on r. We use the standard notation for the r-norm $\|X\|_r := \mathbb{E}[|X|^r]^{1/r}$.

LEMMA 5.1. *For any random variables $X_1, \dots, X_r$ with finite moments defined on the same probability space, we have*

(33) $$\left|\kappa(X_1, \dots, X_r)\right| \leq B_r \prod_{i=1}^{r} \|X_i\|_r.$$

PROOF.   This is an immediate consequence of the moment-cumulant formula (2), combined with Hölder inequality and the monotonicity of r-norms ($\|X\|_s \le \|X\|_r$ is $s \le r$). $\square$

For an event A, we denote A^c its complement.

LEMMA 5.2.   *Let* $X_1, \ldots, X_r$ *be random variables on the same probability space. Then,* *for any event* A, *we have*

$$(34) \qquad \left| \mathbb{E}\big(\mathbf{1}[A] X_1 \cdots X_r\big) - \mathbb{E}(X_1 \cdots X_r) \right| \le \mathbb{P}\big(A^c\big)^{1/(r+1)} \prod_{i=1}^{r} \|X_i\|_{r+1};$$

$$\left| \kappa\big(X_1 \mathbf{1}[A], \ldots, X_r \mathbf{1}[A]\big) - \kappa\big(X_1, \ldots, X_r\big) \right|$$
$$(35)$$
$$\le r B_r \mathbb{P}\big(A^c\big)^{1/(r+1)} \prod_{i=1}^{r} \|X_i\|_{r+1}.$$

PROOF.   The first formula is a trivial consequence of Hölder inequality:

$$\mathbb{E}\big(\mathbf{1}[A] X_1 \cdots X_r\big) - \mathbb{E}(X_1 \cdots X_r) = \mathbb{E}\big(\mathbf{1}[A^c] X_1 \cdots X_r\big)$$

$$\le \big\|\mathbf{1}[A^c]\big\|_{r+1} \prod_{i=1}^{r} \|X_i\|_{r+1}.$$

For the second, we use the moment-cumulant formula (2) and write

$$\kappa\big(X_1 \mathbf{1}[A], \ldots, X_r \mathbf{1}[A]\big) - \kappa\big(X_1, \ldots, X_r\big)$$
$$(36)$$
$$= \sum_{\pi \in \mathcal{P}([r])} \mu(\pi, [r]) \prod_{B \in \pi} \left[\mathbb{E}\bigg(\mathbf{1}[A] \prod_{i \in B} X_i\bigg) - \mathbb{E}\bigg(\prod_{i \in B} X_i\bigg) \right].$$

For a block B of size s we have, since r-norms are increasing in r,

$$\left| \mathbb{E}\bigg(\mathbf{1}[A] \prod_{i \in B} X_i\bigg) \right| \le \left| \mathbb{E}\bigg(\prod_{i \in B} |X_i|\bigg) \right| \le \prod_{i \in B} \|X_i\|_s \le \prod_{i \in B} \|X_i\|_{r+1};$$

and

$$\left| \left[\mathbb{E}\bigg(\mathbf{1}[A] \prod_{i \in B} X_i\bigg) - \mathbb{E}\bigg(\prod_{i \in B} X_i\bigg) \right] \right| \le \mathbb{P}\big(A^c\big)^{1/(s+1)} \prod_{i=1}^{r} \|X_i\|_{s+1}$$

$$\le \mathbb{P}\big(A^c\big)^{1/(r+1)} \prod_{i \in B} \|X_i\|_{r+1}.$$

Combining this with the classical inequality,

$$|a_1 \cdots a_t - b_1 \cdots b_t| \le \sum_{i=1}^{t} |b_1| \cdots |b_{i-1}| |a_i - b_i| |a_{i+1}| \cdots |a_t|$$

implies that, for any set partition π, we have

$$\left| \prod_{B \in \pi} \left[\mathbb{E}\bigg(\mathbf{1}[A] \prod_{i \in B} X_i\bigg) - \mathbb{E}\bigg(\prod_{i \in B} X_i\bigg) \right] \right| \le |\pi| \mathbb{P}\big(A^c\big)^{1/(r+1)} \prod_{i=1}^{r} \|X_i\|_{r+1}.$$

Since we always have $|\pi| \le r$, plugging this inequality back into equation (36) proves equation (35). $\square$

5.2. *Cumulants of rational functions in M.* In this section, we bound (joint) cumulants of various functions of M. The important recurrent feature is that all cumulants have a smaller order of magnitude than what we could naively expect, the difference being a factor n^{1-r} (for joint cumulants of order r, up to logarithmic factors). As usual, constants in $\mathcal{O}$ and $\tilde{\mathcal{O}}$ symbols do depend on the order r of the cumulant under consideration.

Recall that the distribution of M is given in (26). In particular, M depends on n, even if this is implicit in the notation. For a fixed integer r (either positive or negative), one has

$$(37) \qquad \mathbb{E}[M^r] = \frac{B_{n+r}}{B_n} = \tilde{\Theta}(n^r);$$

the first equality is indeed a direct consequence of the formula for the distribution of M (equation (26)), while the second equality follows from asymptotic results for Bell numbers; see, for example, [4], equation (4). This implies $\|M\|_r = \mathcal{O}(n)$. Regarding cumulants, the following estimates for $r \leq 3$ were given in [12], Section 8.3:

$$(38) \qquad m_n := \mathbb{E}[M] = \frac{n}{\ln n}(1 + o(1)),$$

$$(39) \qquad \sigma_n^2 := \mathrm{Var}(M) = \frac{n}{(\ln n)^2}(1 + o(1)),$$

$$(40) \qquad \kappa_3(M) = \frac{2n}{(\ln n)^2}(1 + o(1)).$$

Having such asymptotic equivalent for all cumulants seem hard, but we can easily get a $\mathcal{O}(n)$ bound, which will be sufficient for us.

LEMMA 5.3. *We have* $|\kappa_r(M)| = \mathcal{O}(n) = m_n^r \tilde{\mathcal{O}}(n^{1-r})$.

PROOF. Note that the second equality follows from (38). We focus therefore on the first one.

From Proposition 4.1, it follows that M the same law as $X_n + P$, where X_n is the number of parts in a uniform random set partition of $[n]$, and is independent from the Poisson(1) random variable P. Besides, we know from Harper [16], page 413, that X_n can be written as a sum of n independent Bernoulli variables $X_n = \sum_{i=1}^n B_{i,n}$; the parameters of these Bernoulli variables are here irrelevant. Summing up, we get for M the following useful representation:

$$(41) \qquad M \overset{\text{law}}{=} P + \sum_{i=1}^n B_{i,n}.$$

Using the additivity of cumulants on independent random variables, we have

$$\kappa_r(M) = \kappa_r(P) + \sum_{i=1}^n \kappa_r(B_{i,n}).$$

Since $\kappa_r(B_{i,n})$ is bounded by a constant D_r, independently on the parameter of $B_{i,n}$, the lemma is proved. □

This lemma has the following easy consequence. Consider the following normalized version of M:

$$Z_n := \frac{M - m_n}{\sigma_n},$$

then its cumulants behave as follows: $\kappa_1(Z_n) = 0$ and, for $r \geq 2$,

$$(42) \qquad |\kappa_r(Z_n)| = \sigma_n^{-r}\kappa_r(M) = \tilde{\mathcal{O}}(n^{1-r/2}).$$

For $r \geq 3$, the upper bound tends to 0, so that Z_n converges in distribution and in moments to a standard normal variable. (The convergence in distribution is stated in [8], Theorem 2.1, see also Harper [16].)

We now give bounds for joint cumulants of powers of M.

COROLLARY 5.4. *For any integers* $i_1, \ldots, i_r \geq 1$, *we have*

$$\kappa(M^{i_1}, \ldots, M^{i_r}) = \mathcal{O}(n^{i_1 + \cdots + i_r - r + 1}) = m_n^{i_1 + \cdots + i_r} \tilde{\mathcal{O}}(n^{1-r}),$$

where the constants in $\mathcal{O}$ *and* $\tilde{\mathcal{O}}$ *symbols depend on* $i_1, \ldots, i_r$. *Consequently, we have*

$$(43) \qquad \kappa(Z_n^{i_1}, \ldots, Z_n^{i_r}) = \left(\frac{m_n}{\sigma_n}\right)^{i_1 + \cdots + i_r} \tilde{\mathcal{O}}(n^{-r+1}).$$

PROOF. The Leonov–Shiryaev formula for cumulants of products of random variables [25] gives

$$\kappa(M^{i_1}, \ldots, M^{i_r}) = \sum_{\pi} \prod_{B \in \pi} \kappa_{|B|}(M),$$

where the sum runs over all set partitions π of $[i_1] \uplus \cdots \uplus [i_r]$ such that

$$(44) \qquad \pi \vee \{[i_1], \ldots, [i_r]\} = \{[i_1] \uplus \cdots \uplus [i_r]\}.$$

(Here, $\vee$ is the joint operation on the set partition lattice, ordered by refinement.) Using Lemma 5.3, we have

$$\kappa(M^{i_1}, \ldots, M^{i_r}) = \mathcal{O}\left(\max_{\pi} n^{\#(\pi)}\right),$$

where $\#(\pi)$ is the number of parts of π and the maximum is taken on set partitions π satisfying (44). A simple combinatorial argument (left to the reader) shows that (44) implies $\#(\pi) \leq i_1 + \cdots + i_r + 1 - r$, proving the corollary. $\square$

Finally, we consider joint cumulants of $1/M$ and Z_n/M, which will be useful in the next section. To this end, we need some concentration inequality for M. Let us introduce the following event:

$$A_n = \{m_n - n^{3/4} \leq M \leq m_n + n^{3/4}\}.$$

Note that $n^{3/4}$ is larger than the standard deviation $\sigma_n = \tilde{\mathcal{O}}(n^{1/2})$ of M, so that we expect A_n to hold with large probability. Indeed, the following holds.

LEMMA 5.5. $\mathbb{P}[A_n^c]$ *tends to* 0 *faster than any rational function of* n.

PROOF. We use the representation (41) of M as a sum of independent variables:

$$M \stackrel{\text{law}}{=} P + \sum_{i=1}^{n} B_{i,n},$$

where P follows a Poisson law of parameter 1 and the $B_{i,n}$ are independent Bernoulli variables, whose parameters are not relevant. A standard tail estimate for Poisson distribution (see, e.g., [27], page 97) gives

$$\mathbb{P}\left[|P - \mathbb{E}(P)| \geq \frac{1}{2}n^{3/4}\right] \leq \mathbb{P}\left[P \geq \frac{1}{2}n^{3/4}\right] \leq e^{-1}\left(\frac{2e}{n^{3/4}}\right)^{n^{3/4}},$$

while Hoeffding's inequality [18], Theorem 1, equation (2.3), tells us that

$$\mathbb{P}\left[\left|\sum_{i=1}^{n}(B_{i,n} - \mathbb{E}(B_{i,n}))\right| \geq \frac{1}{2}n^{3/4}\right] \leq 2\exp\left(-\frac{1}{2}n^{1/2}\right).$$

Combining both inequalities tells us yields, for n big enough,

$$\mathbb{P}[|M - m_n| \geq n^{3/4}] \leq 3\exp\left(-\frac{1}{2}n^{1/2}\right). \qquad \square$$

From Lemma 5.2, adding/removing $\mathbf{1}[A_n]$ from joint cumulants/joint moments of powers of Z_n, M and $1/M$ change the resulting value by an error that is smaller than any rational function of n (indeed, for each fixed $r \geq 1$, the r-norms of Z_n, M and $1/M$ are bounded by $\tilde{\mathcal{O}}(\sqrt{n})$, $\mathcal{O}(n)$ and 1, resp.). We will denote by $oe(n)$ such error terms and use this fact repeatedly below.

We can now turn back to bounds on cumulants.

PROPOSITION 5.6. *For nonnegative integers r_1, r_2 with $r_1 + r_2 \geq 1$, we have*

$$\kappa(\underbrace{M^{-1}, \ldots, M^{-1}}_{r_1 \text{ times}}, \underbrace{Z_n/M, \ldots, Z_n/M}_{r_2 \text{ times}}) = m_n^{-r_1}\sigma_n^{-r_2}\tilde{\mathcal{O}}(n^{1-r_1-r_2}).$$

PROOF. Fix some integer $k \geq 0$ (that will be specified later). Elementary analysis asserts that, for any $x \geq -1/2$,

$$(1+x)^{-1} = \sum_{i=0}^{k-1}(-x)^i + \varepsilon_k(x),$$

where $|\varepsilon_k(x)| \leq 2|x|^k$. When A_n holds, for n large enough, we can use this expansion for $x = \frac{M-m_n}{m_n} = \frac{\sigma_n}{m_n}Z_n$ and write

$$\mathbf{1}[A_n]\frac{m_n}{M} = \mathbf{1}[A_n]\left(1 + \frac{\sigma_n}{m_n}Z_n\right)^{-1}$$

$$= \mathbf{1}[A_n]\left[\sum_{i=0}^{k-1}\left(-\frac{\sigma_n}{m_n}Z_n\right)^i + \varepsilon_k\left(\frac{\sigma_n}{m_n}Z_n\right)\right].$$

Multiplying this by $\frac{\sigma_n}{m_n}Z_n$ and setting $\varepsilon_k'(x) = x\varepsilon_{k-1}(x)$ (so that we have the bound $|\varepsilon_k'(x)| \leq 2|x|^k$, as for $\varepsilon_k(x)$), we get (after a shift of index)

$$\mathbf{1}[A_n]\sigma_n\frac{Z_n}{M} = \mathbf{1}[A_n]\left[-\sum_{i=1}^{k-1}\left(-\frac{\sigma_n}{m_n}Z_n\right)^i + \varepsilon_k'\left(\frac{\sigma_n}{m_n}Z_n\right)\right].$$

We now consider the cumulants

$$(45) \qquad k_{r_1,r_2} := \kappa\left(\underbrace{\mathbf{1}[A_n]\frac{m_n}{M}, \ldots, \mathbf{1}[A_n]\frac{m_n}{M}}_{r_1 \text{ times}}, \underbrace{\mathbf{1}[A_n]\sigma_n\frac{Z_n}{M}, \ldots, \mathbf{1}[A_n]\sigma_n\frac{Z_n}{M}}_{r_2 \text{ times}}\right).$$

One the one hand, since A_n^c has exponentially small probability, we can forget the indicators $\mathbf{1}[A_n]$ in the definition of k_{r_1,r_2}, up to an error term $oe(n)$. Therefore, we have

$$(46) \qquad k_{r_1,r_2} = m_n^{r_1}\sigma_n^{r_2}\kappa(\underbrace{M^{-1}, \ldots, M^{-1}}_{r_1 \text{ times}}, \underbrace{Z_n/M, \ldots, Z_n/M}_{r_2 \text{ times}}) + oe(n).$$

On the other hand, we can expand (45) by multilinearity and bound separately each summand (the number of summands is independent of n). We set $r = r_1 + r_2$ and distinguish two types of summands:

- First consider, for $0 \leq i_1, \ldots, i_r \leq k - 1$, the summand

$$\pm \kappa \left[\mathbf{1}[A_n] \left(-\frac{\sigma_n}{m_n} Z_n \right)^{i_1}, \ldots, \mathbf{1}[A_n] \left(-\frac{\sigma_n}{m_n} Z_n \right)^{i_r} \right],$$

which does not involve any of the functions ε_k. Removing the indicators $\mathbf{1}[A_n]$ yields an error of order $oe(n)$. Then, using (43), we see that such terms are of order $\tilde{\mathcal{O}}(n^{1-r})$.

- We now consider summands that involves some function ε_k or ε'_k. Again the indicator functions can be forgotten up to an error of order $oe(n)$. For such terms, we use Lemma 5.1 and, therefore, we only have to bound expressions of the form

$$\mathbb{E}\left[\left| \frac{\sigma_n}{m_n} Z_n \right|^{r i_1} \right]^{1/r} \cdots \mathbb{E}\left[\left| -\frac{\sigma_n}{m_n} Z_n \right|^{r i_t} \right]^{1/r}$$

$$\cdots \mathbb{E}\left[\left| \varepsilon_k \left(\frac{\sigma_n}{m_n} Z_n \right) \right|^r \right]^{\frac{s_1}{r}} \mathbb{E}\left[\left| \varepsilon'_k \left(\frac{\sigma_n}{m_n} Z_n \right) \right|^r \right]^{\frac{s_2}{r}}$$

for triples $(t, s_1, s_2) \neq (r, 0, 0)$ of sum r and integers $i_1, \ldots, i_t \geq 0$. Denoting $A = k(s_1 + s_2) + i_1 + \cdots + i_t$ and using the bound for ε_k and ε'_k, the above expression is smaller than

$$2^{s_1+s_2} \left(\frac{\sigma_n}{m_n} \right)^A \mathbb{E}[|Z_n|^{r i_1}]^{1/r} \cdots \mathbb{E}[|Z_n|^{r i_t}]^{1/r} \mathbb{E}[|Z_n|^{kr}]^{\frac{s_1+s_2}{r}}.$$

Since cumulants of Z_n converges, so does its moments and its absolute moments, and we get a bound in $\tilde{\mathcal{O}}(n^{-A/2})$. Take $k = 2(r-1)$. Since $s_1 + s_2 \geq 1$, we have $A \geq k$ and we get that all summands are $\tilde{\mathcal{O}}(n^{1-r})$.

Finally, we conclude that $k_{r_1,r_2} = \tilde{\mathcal{O}}(n^{1-r})$. Together with equation (46), this concludes the proof. $\square$

COROLLARY 5.7.   *For any nonnegative integers* $i_1, \ldots, i_r$ *and* $j_1, \ldots, j_r$, *we have*

$$\kappa\left[\left(\frac{1}{M} \right)^{i_1} \left(\frac{Z_n}{M} \right)^{j_1}, \ldots, \left(\frac{1}{M} \right)^{i_r} \left(\frac{Z_n}{M} \right)^{j_r} \right] = m_n^{-i_1-\cdots-i_r} \sigma_n^{-j_1-\cdots-j_r} \tilde{\mathcal{O}}(n^{1-r}),$$

where the constant in the $\tilde{\mathcal{O}}$ *symbols depend on* $i_1, \ldots, i_{r_1}, j_1, \ldots, j_{r_2}$.

PROOF.   Similar to the proof of Corollary 5.4. $\square$

We consider polynomials P in the variables $(\mathbf{x} = \{x_1, \ldots, x_p\}, y, z)$ and define the following gradations. If P is a *monomial* in these variables, we denote by $\deg_{\mathbf{x}}(P)$ its total degree in $x_1, \ldots, x_p$; by $\deg_y(P)$ its degree in y; and by $\deg_{\mathbf{x}, y^{-1}}(P)$ the difference $\deg_{\mathbf{x}}(P) - \deg_y(P)$. As usual, for any of the three above notions of degree, the degree of a polynomial P is the maximal degree of a monomial with a nonzero coefficient in P.

We claim that Corollary 5.7 implies by linearity the following bound. Let $(P_i)_{1 \leq i \leq s}$ be polynomials in $(\mathbf{x} = \{x_1, \ldots, x_p\}, y, z)$ and d_i be integers with $\deg_{\mathbf{x}, y^{-1}}(P_i) \leq d_i$. Then, uniformly for all values $\mathbf{a} = (a_1, \ldots, a_p)$ and $\mathbf{b} = (b_1, \ldots, b_s)$ in $[0, n]$, we have

$$(47) \qquad \kappa\left(P_i \left(\mathbf{a}, M^{-1}, \frac{b_i \sigma_n Z_n}{M \cdot m_n} \right) : 1 \leq i \leq s \right) = n^{\sum_{i \leq s} d_i} \tilde{\mathcal{O}}(n^{1-s}).$$

Indeed it is enough to check this for monomials $P_i = \mathbf{x}^{\mathbf{e}_i} y^{f_i} z^{g_i}$ (for i in $[s]$), the general case following by linearity. For monomials we have, using Corollary 5.7

$$\left| \kappa\left(P_i \left(\mathbf{a}, M^{-1}, \frac{b_i \sigma_n Z_n}{M \cdot m_n} \right) : 1 \leq i \leq s \right) \right|$$

$$\leq \prod_{i \leq s} \left(\mathbf{a}^{\mathbf{e}_i} \left(\frac{b_i \sigma_n}{m_n} \right)^{g_i} \right) \cdot \left| \kappa\left(M^{-f_i} (Z_n/M)^{g_i} : i \leq s \right) \right|$$

$$\leq n^{\sum_{i \leq s} \deg_x(P_i)} \left(\frac{1}{\sqrt{n}}\right)^{g_1 + \cdots + g_s} m_n^{-f_1 - \cdots - f_s} \sigma_n^{-g_1 - \cdots - g_s} \tilde{\mathcal{O}}(n^{1-s})$$

$$\leq n^{\sum_{i \leq s} (\deg_x(P_i) - f_i)} \tilde{\mathcal{O}}(n^{1-s}),$$

concluding the proof of (47).

5.3. *Proof of Theorem* 4.2. By definition of weighted dependency graphs, we have to prove that for any multiset $B = \{X_{i_1 j_1}, \ldots, X_{i_r j_r}\}$ of elements of A^n, one has

(48) $$\left|\kappa(X_{i_1 j_1}, \ldots, X_{i_r j_r})\right| \leq C_{r,n} \Psi(B) \mathcal{M}(G^n[B]).$$

Recall that $\mathcal{M}(G^n[B])$ is by definition the maximal weight of a spanning tree of the graph induced by G^n on B.

As in the proof of Theorem 3.1, [11], Proposition 5.2, allows us to reduce the proof of (48) to the case where $G^n[B]$ has no edges of weight 1, that is, where $i_1, \ldots, i_r$ and $j_1, \ldots, j_r$ are two lists of distinct integers (but we may have $i_t = j_s$ for some s, t in $[r]$). Indeed, this reduction was based on the following facts, which hold true here as well:

- variables linked by an edge of weight 1 are incompatible (i.e., their product is a.s. 0);
- our random variables take values in $\{0, 1\}$;
- $\Psi(B)$ is insensitive to repetitions.

In the case where $i_1, \ldots, i_r$ and $j_1, \ldots, j_r$ are two lists of distinct integers, the bound to be proven is

(49) $$\left|\kappa(X_{i_1 j_1}, \ldots, X_{i_r j_r})\right| \leq C_{r,n} n^{1-2r} = \tilde{\mathcal{O}}(n^{1-2r}),$$

where the constant in $\tilde{\mathcal{O}}$ symbol depend on r but neither on n nor on the indices $i_1, \ldots, i_r, j_1, \ldots, j_r$.

Fix $r \geq 2$. Take $i_1, i_2, \ldots, i_r$ distinct and $j_1, j_2, \ldots, j_r$ distinct with $i_s < j_s$ for all s. We identify B to the set of arcs $\{(i_1, j_1), (i_2, j_2), \ldots, (i_r, j_r)\}$ and will use letters C and D for subsets of B.

We first condition on the number of urns M and get bound for the conditional cumulants. Consider a set partition π generated through the urn model with M urns. As explained in (27) above, we have

(50) $$\mathbb{E}\left(\prod_{(i,j) \in D} X_{ij} \Big| M\right) = \prod_{g \in [n]} p_g(D),$$

where $p_g(D)$ is the probability of g being dropped in a correct urn, that is,

$$p_g(D) = \begin{cases} \dfrac{1}{M} & \text{if } g = j_t \text{ for some } (i_t, j_t) \text{ in D;} \\ \dfrac{M - a_g(D)}{M} & \text{otherwise.} \end{cases}$$

Here, $a_g(D) := \{(i, j) \in D : i < g < j\}$ is the number of arcs in D above g.

As in the proof of Theorem 3.1, we will use the small-cumulant/quasi-factorization equivalence. For a subset C of B, we define

$$Q_C(M) := \prod_{D \subseteq C} \mathbb{E}\left(\prod_{(i,j) \in D} X_{ij} \Big| M\right)^{(-1)^{|D|}}.$$

For each $D \subseteq C$, we can use (50) for the corresponding joint moment and we get

$$Q_C(M) = \prod_{g \in [n]} \prod_{D \subseteq C} p_g(D)^{(-1)^{|D|}}.$$

The integers g indexing the product can be divided in three categories:

(i) There exists an arc (i_0, j_0) in C that is not above g; more precisely such that $g \leq i_0$ or $g > j_0$, the large and strict inequalities being important. Then

$$\prod_{D \subseteq C} p_g(D)^{(-1)^{|D|}} = 1,$$

since the map $D \leftrightarrow D \bigtriangleup \{(i, j)\}$ ($\bigtriangleup$ is the symmetric difference) gives a fixed-point free sign-reversing involution of the factors.

(ii) If all arcs (i, j) in C are above g (i.e., $i < g < j$ for all (i, j) in C), then for all $D \subseteq C$, we have $a_g(D) = |D|$. Therefore,

$$\prod_{D \subseteq C} p_g(D)^{(-1)^{|D|}} = \prod_{D \subseteq C} \left(\frac{M - |D|}{M} \right)^{(-1)^{|D|}} = \prod_{D \subseteq C} (M - |D|)^{(-1)^{|D|}}.$$

The right-hand side clearly depends only on M and of the size $|C|$ of C. By [11], Lemma A.1, it writes as $1 + R_{|C|}(M)$ for some rational function $R_{|C|}$ of degree at most $-|C|$.

(iii) The remaining case is when both $\min(j_1, \ldots, j_s) > \max(i_1, \ldots, i_s)$ and $g = \min(j_1, \ldots, j_s)$ hold. To simplify notation, assume $g = j_1$. Then $p_g(D) = \frac{1}{M}$ whenever $(i_1, j_1) \in D$. The corresponding factors cancel out and we get, with similar arguments as above for the other factors,

$$\prod_{D \subseteq C} p_g(D)^{(-1)^{|D|}} = \prod_{D \subseteq C \setminus \{(i_1, j_1)\}} (M - |D|)^{(-1)^{|D|}} = 1 + R_{|C|-1}(M).$$

In conclusion, we have

$$Q_C(M) = \left(1 + R_{|C|}(M)\right)^{\ell_C} \left(1 + R_{|C|-1}(M)\right)^{\mathbf{1}[\ell_C > 0]},$$

where ℓ_C is the number of $g \in [n]$ that are below all arcs, that is,

$$\ell_C := \max\left[(\min(j_1, \ldots, j_s) - \max(i_1, \ldots, i_s)), 0 \right].$$

Note that we always have $\ell_C \leq n$.

We now perform an asymptotic expansion of Q_C, using $*$ for coefficients which depend only on $|C|$ and do not need to be made explicit: if $\ell_C > 0$, we have

$$
\begin{aligned}
\left(1 + R_{|C|}(M)\right)^{\ell_C} \\
&= \exp\left[\ell_C \log(1 + R_{|C|}(M)) \right] \\
&= \exp\left[\ell_C \log(1 + *M^{-|C|} + *M^{-|C|+1} + \cdots + \mathcal{O}(M^{-r})) \right] \\
&= \exp\left[*\ell_C M^{-|C|} + *\ell_C M^{-|C|+1} + \cdots + \mathcal{O}(\ell_C M^{-r})) \right] \\
&= 1 + *\ell_C M^{-|C|} + \cdots + \mathcal{O}(\ell_C M^{-r}) + *\ell_C^2 M^{-2|C|} + \cdots + \mathcal{O}(\ell_C^r M^{-r|C|}).
\end{aligned}
$$

For $|C| \geq 2$, under the assumption that $M \geq \frac{1}{2}m_n$, both error terms are smaller than $\tilde{\mathcal{O}}(n^{1-r})$ (we use the universal bound $\ell_C \leq n$). The main term is a polynomial in ℓ_C and M^{-1}. After subtracting 1, its total degree in ℓ_C and M is at most $-|C| + 1$. Note that both the constants in the error terms and this polynomial only depend on the size of $|C|$. The same applies to the other factor $1 + R_{|C|-1}(M)$ of Q_C. We therefore conclude that, for all C such that $\ell_C > 0$,

$$(51) \qquad\qquad Q_C(M) = 1 + P_{|C|}(\ell_C, M^{-1}) + \tilde{\mathcal{O}}_{|C|}(n^{1-r}),$$

where $P_{|C|}$ is a polynomial in x and y such that $\deg_{x, y^{-1}}(P_{|C|}) \leq -|C| + 1$ and the error term $\tilde{\mathcal{O}}_{|C|}(n^{1-r})$ is uniform for all C of a given size and all $M \geq \frac{1}{2}m_n$. For $\ell_C = 0$, we

have $Q_C(M) = 1$, so that (51) holds trivially in this case as well (though, with different polynomials $P_{|C|}$ and error terms $\tilde{\mathcal{O}}_{|C|}(n^{1-r})$).

We shall also need an expansion of the conditional expectation of single variable X_{ij}:

$$
\begin{aligned}
\mathbb{E}[X_{ij}|M] &= M^{-1}(1 - M^{-1})^{j-i-1} \\
&= M^{-1}\exp[-(j-i)M^{-1}]\exp[M^{-1} + *(j-i-1)M^{-2} + \cdots \\
&\quad + \mathcal{O}((j-i)M^{-r})].
\end{aligned}
\tag{52}
$$

The second exponential writes as $\tilde{P}_1(j-i, M^{-1}) + \tilde{\mathcal{O}}(n^{1-r})$ for some polynomial $\tilde{P}_1$ with $\deg_{\boldsymbol{x}, y^{-1}}(\tilde{P}_1) \leq 0$. The exponent of the first exponential however is typically large, forcing us to use the concentration property of M and exhibit the deterministic dominant term. We have

$$
\exp[-(j-i)M^{-1}] = \exp[-(j-i)m_n^{-1}]\exp\left[\frac{(j-i)\sigma_n Z_n}{M \cdot m_n}\right],
$$

where we recall that $\sigma_n Z_n = M - m_n$ by definition. Assume now that we have $|M - m_n| \leq n^{3/4}$, which implies $Z_n = \tilde{\mathcal{O}}(n^{1/4})$. Then the argument of the second exponential is $\tilde{\mathcal{O}}(n^{-1/4})$ and we can perform a series expansion, uniformly on (i, j) and M:

$$
\exp[-(j-i)M^{-1}] = \exp[-(j-i)m_n^{-1}]\left(\hat{P}_1\left(\frac{(j-i)\sigma_n Z_n}{M \cdot m_n}\right) + \tilde{\mathcal{O}}(n^{1-r})\right),
\tag{53}
$$

for some polynomial $\hat{P}_1(z)$. Combining equations (52) and (53), we get

$$
\begin{aligned}
\mathbb{E}[X_{ij}|M] &= M^{-1}\exp[-(j-i)m_n^{-1}] \\
&\quad \times \left[P_1\left(j-i, M^{-1}, \frac{(j-i)\sigma_n Z_n}{M \cdot m_n}\right) + \tilde{\mathcal{O}}(n^{1-r})\right],
\end{aligned}
\tag{54}
$$

where P_1 is a polynomial with $\deg_{\boldsymbol{x}, y^{-1}}(P_1) \leq 0$ and the error term is uniform on (i, j) and M in the interval $[m_n - n^{3/4}; m_n + n^{3/4}]$.

We now come back to the conditional cumulant

$$
\kappa_C(M) := \kappa(X_{ij} : (i, j) \in C|M).
$$

As observed in [11], Proof of Proposition 5.8, it can be written in terms of conditional expectations and quotients Q as follows:

$$
\kappa_C(M) = \left(\prod_{(i,j)\in C}\mathbb{E}[X_{ij}|M]\right) \cdot \sum_{\{C_1,\ldots,C_m\}}(Q_{C_1} - 1)\cdots(Q_{C_m} - 1),
$$

where the sum is over all sets of (distinct) subsets $C_1, \ldots, C_m \subseteq C$ such that $|C_i| \geq 2$ and the hypergraph with edges $C_1, \ldots, C_m$ is connected. Using the asymptotic expansions (51) and (54), we get

$$
\begin{aligned}
\kappa_C(M) &= M^{-|C|}\prod_{(i,j)\in C}\exp[-(j-i)m_n^{-1}] \\
&\quad \times \left[P_C\left(\boldsymbol{\ell}_C, M^{-1}, \frac{(j-i)\sigma_n Z_n}{M \cdot m_n}\right) + \tilde{\mathcal{O}}(n^{1-r})\right],
\end{aligned}
\tag{55}
$$

where $\boldsymbol{\ell}_C$ represent the vector $(\ell_D)_{D\subseteq C}$ and P_C is a polynomial in the variables ($\boldsymbol{x} = (x_D)_{D\subseteq C}, y, z$). Furthermore, the error term $\tilde{\mathcal{O}}(n^{1-r})$ is uniform for M in the interval $[m_n - n^{3/4}; m_n + n^{3/4}]$. Both the polynomials P_C and the constants in the error terms depend only on C through its size and which of ℓ_D ($D \subseteq C$) are nonzero. Therefore, the error

term can thus be chosen uniformly on C with $|C| \leq r$. Moreover, the total degree of P_C in x and y^{-1} is at most

$$\max_{\{C_1,...,C_m\}} (-|C_1| + 1) + \cdots + (-|C_m| + 1),$$

where the maximum is taken over sets of subsets of C such that the hypergraph with edges $C_1, \ldots, C_m$ is connected. It is a simple combinatorial exercise to see that this condition implies

$$(-|C_1| + 1) + \cdots + (-|C_m| + 1) \leq -|C| + 1,$$

so that $\deg_{x,y^{-1}}(P_C) \leq -|C| + 1$.

We now use the law of total cumulance (28). Recall that, from Lemma 5.5, the event $A_n = \{|M - m_n| \leq n^{3/4}\}$ has probability $1 - oe(n)$. We have

$$\kappa(X_{ij} : (i, j) \in B) = \sum_{\pi \in \mathcal{P}(B)} \kappa(\kappa_C(M) : C \in \pi)$$

$$= \sum_{\pi \in \mathcal{P}(B)} \kappa(\mathbf{1}[A_n]\kappa_C(M) : C \in \pi) + oe(n).$$

When A_n is satisfied, the conditional cumulants $\kappa_C(M)$ can be evaluated by (55) and we get

$$\kappa(X_{ij} : (i, j) \in B) + oe(n)$$

$$= \left(\prod_{(i,j) \in B} \exp[-(j-i)m_n^{-1}] \right) \cdot$$

$$\times \sum_{\pi \in \mathcal{P}(B)} \kappa \left(\mathbf{1}[A_n] M^{-|C|} \left[P_C \left(\ell_C, M^{-1}, \frac{(j-i)\sigma_n Z_n}{M \cdot m_n} \right) + \tilde{\mathcal{O}}(n^{1-r}) \right] : C \in \pi \right).$$

We now bound the right-hand side. The product of exponential factors is simply bounded by 1. Each summand of the sum over set partitions is bounded as follows. We expand by multilinearity the joint cumulant.

- One of the terms is

$$\kappa \left(\mathbf{1}[A_n] M^{-|C|} \left[P_C \left(\ell_C, M^{-1}, \frac{(j-i)\sigma_n Z_n}{M \cdot m_n} \right) \right] : C \in \pi \right).$$

  Using equation (47) and the fact $\deg_{x,y^{-1}}(P_C) \leq -|C| + 1$, we get that this term is $\tilde{\mathcal{O}}(n^e)$, where

$$e \leq 1 - |\pi| + \sum_{C \in \pi} (-2|C| + 1) = 1 - 2r.$$

- The other terms are joint cumulants of random variables, at least one of which is $\tilde{\mathcal{O}}(n^{1-r})$ uniformly on the event A_n. Since all P_C have nonpositive degree in x and y^{-1}, they are bounded on the event A_n by $\tilde{\mathcal{O}}(1)$. Moreover, on A_n, we have $M^{-|C|} = \tilde{\mathcal{O}}(n^{-|C|})$. We therefore use the easy bounds (33) and (35) for cumulants and get that all these terms are

$$\tilde{\mathcal{O}}(n^{1-r-\sum_{C \in \pi} |C|}) = \tilde{\mathcal{O}}(n^{1-2r}).$$

To conclude, we get that

$$\kappa(X_{ij} : (i, j) \in B) = \tilde{\mathcal{O}}(n^{1-2r}).$$

Note that the constants in the $\tilde{\mathcal{O}}$ symbols only depend on which ℓ_D are nonzero, for subsets D of B, and hence can be chosen uniformly for all sets B of size r (assuming, as we have always done so far, that $i_1, \ldots, i_r$, resp., $j_1, \ldots, j_r$, are distinct). We have thus proved (48) for lists $i_1, \ldots, i_r$ and $j_1, \ldots, j_r$ with distinct entries. As argued at the beginning of the proof, this completes the proof of Theorem 4.2.

6. Technical statements for the variance estimate. The goal of this section is to prove the estimates on $\mathbb{E}[\mathrm{Occ}_n^{\mathcal{A}} | M]$ and its variance given in equations (31) and (32). According to (30), we need to understand sums of the following kind (where we take $a_1, \ldots, a_t$ to be positive integers and $\ell \geq t$):

$$F_{a_1,\ldots,a_t;\ell}(n, M) = \sum_{\substack{y_1,\ldots,y_t \geq 1 \\ y_1+\cdots+y_t \leq n}} \binom{n - y_1 - \cdots - y_t}{\ell - t} \prod_{i \leq t}\left(1 - \frac{a_i}{M}\right)^{y_i - 1}.$$

Note that they satisfy the following recursive formula (setting $y = y_t$):

$$(56) \qquad F_{a_1,\ldots,a_t;\ell}(n, M) = \sum_{1 \leq y \leq n} F_{a_1,\ldots,a_{t-1};\ell-1}(n - y, M)\left(1 - \frac{a_t}{M}\right)^{y-1}.$$

The general strategy is the following: using this recursive formula, we will prove that such functions belong to some specific graded space V, analyze their highest degree terms and show that the expectation and the variance of an element of V can be bounded above given its degree.

We let V to be the $\mathbb{Q}(M)[n]$-span of $(1 - b/M)^n$, where b runs over the set of nonnegative integers (it is easy to see that the functions $(1 - b/M)^n$ are linearly independent over $\mathbb{Q}[n](M)$).

We now define a gradation deg on V. Both variables n and M are of degree 1. Moreover, the degree of a term $(1 - b/M)^n$ is chosen to be $-b$. With this convention, we can check that an element of V of degree at most d behaves as $\tilde{\mathcal{O}}(n^d)$, as n and M to infinity with $|M - m_n| \leq n^{3/4}$.

LEMMA 6.1. *For all positive integers $a_1, \ldots, a_t$ and $\ell \geq t$, the quantity $F_{a_1,\ldots,a_t;\ell}(n, M)$ is in V. Moreover,*

$$F_{a_1,\ldots,a_t;\ell}(n, M) = \sum_{\substack{j_0,\ldots,j_t \geq 0 \\ j_0+\cdots+j_t = \ell-t}} \frac{(-1)^{j_1+\cdots+j_t}}{j_0! a_1^{j_1+1} \cdots a_t^{j_t+1}} n^{j_0}(M)^{j_1+\cdots+j_t+t} + \mathrm{Err}(n, M),$$

where the error term $\mathrm{Err}(n, M)$ has degree at most $\ell - 1$.

PROOF. For the sake of this proof, we introduce a second gradation $\deg'$ on V, for which n and M still have degree 1 but any $(1 - b/M)^n$ has degree 0. Obviously for any f in V, we have $\deg(f) \leq \deg'(f)$. To distinguish both gradations, we will refer below to deg as the *standard degree*, while $\deg'$ is called the *modified degree*.

We introduce some additional notation: let

$$I_a[G(n, M)] = \sum_{1 \leq y \leq n} G(n - y, M)\left(1 - \frac{a}{M}\right)^{y-1}$$

be the $\mathbb{Q}(M)$-linear operator appearing in equation (56). Finally, NoRn is the $\mathbb{Q}(M)[n]$-linear projection from V to V which sends the basis element 1 to 1 and $(1 - b/M)^n$ to 0 for $b > 0$.

The proof of the lemma involves three claims.

CLAIM 1. *The functional I_a maps V to V and increases the modified degree by at most 1.*

PROOF. By $\mathbb{Q}(M)$-linearity, it suffices to consider $I_a[G(n, M)]$, where $G(n, M) = P(n)(1 - b/M)^n$ has modified degree d (i.e., $\deg(P)d$). By definition and elementary manipulations,

$$(57) \qquad I_a[P(n)(1 - b/M)^n] = (1 - b/M)^{n-1}\sum_{1 \leq y \leq n} P(n - y)\left(\frac{1 - a/M}{1 - b/M}\right)^{y-1}.$$

If $a = b$, then the above sum is a polynomial of degree $\deg(P) + 1 = d + 1$ in n and therefore $I_a[P(n)(1 - b/M)^n]$ is an element of P of modified degree $d + 1$.

Consequently, we focus on the case $a \neq b$. The quantity $P(n - y)$ is a polynomial in n and y and can be decomposed as

$$(58) \qquad P(n - y) = \sum_{j \leq \deg(P)} \alpha_j \tilde{P}_j(n) y(y + 1) \cdots (y + j - 1),$$

for some constants α_j and polynomials $\tilde{P}_j$ satisfying $\deg(\tilde{P}_j) + j \leq \deg(P)$. This implies

$$I_a[P(n)(1 - b/M)^n] = (1 - b/M)^{n-1} \sum_{j \leq \deg(P)} \alpha_j \tilde{P}_j(n) S_j,$$

where

$$S_j := \sum_{1 \leq y \leq n} y(y + 1) \cdots (y + j - 1) \left(\frac{1 - a/M}{1 - b/M} \right)^{y-1}$$

$$(59) \qquad = \frac{d^j}{dx^j} \frac{1 - x^{n+j}}{1 - x} \Big|_{x = \frac{1-a/M}{1-b/M}}$$

$$= \left(\frac{j!}{(1-x)^{j+1}} - \sum_{h=0}^{j} \binom{j}{h} h! \frac{(n+j)!}{(n+h)!} \frac{x^{n+h}}{(1-x)^{h+1}} \right) \Big|_{x = \frac{1-a/M}{1-b/M}}.$$

We set $R(M) := \frac{1-a/M}{1-b/M}$. The above formula shows that each S_j is a $\mathbb{Q}(M)[n]$ linear combination of 1 and $R(M)^n$, so that $(1 - b/M)^{n-1} S_j$ is a $\mathbb{Q}(M)[n]$ linear combination of $(1 - b/M)^n$ and $(1 - b/M)^n R(M)^n = (1 - a/M)^n$. This proves that $I_a[P(n)(1 - b/M)^n]$ is in V.

Let us analyze its modified degree: $R(M)$ has degree 0, while $\frac{1}{1-R(M)}$ has degree 1 in M. Since both $(1 - b/M)^n$ and $(1 - b/M)^n R(M)^n = (1 - a/M)^n$ have modified degree 0, we see that the modified degree of $(1 - b/M)^n S_j$ is at most $j + 1$. From the inequality $\deg(\tilde{P}_j) + j \leq \deg(P)$, we conclude that $I_a[P(n)(1 - b/M)^n]$ has modified degree at most $\deg(P) + 1 = d + 1$. This completes the proof of Claim 1. □

CLAIM 2.　*For $a > 0$, we have* $\mathrm{NoRn} \circ I_a = \mathrm{NoRn} \circ I_a \circ \mathrm{NoRn}$.

PROOF.　Again, by $\mathbb{Q}(M)$-linearity, it is again to check the equality applied to $G(n, M) = P(n)(1 - b/M)^n$. If $b = 0$, this is trivial since, in this case, $\mathrm{NoRn}(G(n, M)) = G(n, M)$. For $b > 0$, $\mathrm{NoRn}(G(n, M)) = 0$ and we should check that $\mathrm{NoRn} \circ I_a(G(n, M)) = 0$.

This is a consequence of equation (59) and the discussion after it; indeed, for $b > 0$, the quantity $I_a[P(n)(1 - b/M)^n]$ is a $\mathbb{Q}(M)[n]$ linear combination of $(1 - b/M)^n$ and $(1 - a/M)^n$. Since a and b are both positive, these two basis elements are sent to 0 by NoRn proving Claim 2. □

CLAIM 3.　*Let $G(n, M)$ be a homogeneous element of degree d in $\mathbb{Q}(M)[n]$ (i.e., without terms of the form $(1 - b/M)^n$). Then the top homogeneous component of $\mathrm{NoRn} \circ I_a[G(n, M)]$ is $\Phi_a[G(n, M)]$, where Φ_a is the $\mathbb{Q}(M)$ linear map defined by*

$$\Phi_a \left(\frac{n^d}{d!} \right) = \sum_{j=0}^{d} (-1)^j \frac{n^{d-j}}{(d-j)!} \frac{M^{j+1}}{a^{j+1}}.$$

PROOF.   We focus on the case $G(n, M) = P(n) = n^d$, the general case following by $\mathbb{Q}(M)$-linearity. In this case, the decomposition (58) writes

$$P(n - y) = (n - y)^d$$

$$= \sum_{j=0}^{d} \binom{d}{j} (-1)^j n^{d-j} y^j$$

$$= \sum_{j=0}^{d} \binom{d}{j} (-1)^j n^{d-j} y(y + 1) \cdots (y + j - 1) + \text{smaller degree terms.}$$

Consequently,

$$I_a[n^d] = \sum_{j=0}^{d} \binom{d}{j} (-1)^j n^{d-j} S_j + \text{smaller modified degree terms,}$$

where S_j is given by equation (59), setting $b = 0$. From equation (59), we observe that

$$\text{NoRn}(S_j) = \frac{j!}{(1 - x)^{j+1}} \Big|_{x=1-a/M} = j! \frac{M^{j+1}}{a^{j+1}}.$$

Therefore, we have

$$\text{NoRn} \circ I_a[n^d] = \sum_{j=0}^{d} \binom{d}{j} (-1)^j n^{d-j} j! \frac{M^{j+1}}{a^{j+1}},$$

concluding the proof of Claim 3. $\square$

We come back to the proof of the lemma. Unwrapping the recursion in equation (56), we have

$$F_{a_1,\ldots,a_t;\ell}(n, M) = I_{a_1} \circ I_{a_2} \cdots \circ I_{a_t} \left[\binom{n}{\ell - t} \right].$$

From Claim 1, it immediately follows that $F_{a_1,\ldots,a_t;\ell}(n, M)$ is in V and has modified degree at most ℓ.

We are interested in its homogeneous component of standard degree ℓ. Since we know that its modified degree is at most ℓ, this top homogeneous part cannot contain any term $(1 - b/M)^n$ with $b > 0$; indeed such term have a bigger modified degree as standard degree. We can therefore consider

$$\text{NoRn}\big(F_{a_1,\ldots,a_t;\ell}(n, M)\big) = \text{NoRn} \circ I_{a_1} \circ I_{a_2} \cdots \circ I_{a_t} \left[\binom{n}{\ell - t} \right]$$

$$= \text{NoRn} \circ I_{a_1} \circ \text{NoRn} \circ I_{a_2} \cdots \text{NoRn} \circ I_{a_t} \left[\binom{n}{\ell - t} \right],$$

where the second equality follows from Claim 2.

Since $\binom{n}{\ell-t}$ has degree $\ell - t$ and that each one of the t operators $\text{NoRn} \circ I_{a_s}$ increases the degree by at most 1, we can only keep the highest degree component all along the computation. Therefore, using Claim 3, the highest degree component of $F_{a_1,\ldots,a_t;\ell}(n, M)$ is

$$\Phi_{a_1} \circ \Phi_{a_2} \cdots \circ \Phi_{a_t} \left[\frac{n^{\ell-t}}{(\ell - t)!} \right] = \sum_{\substack{j_0,\ldots,j_t \geq 0 \\ j_0 + \cdots + j_t = \ell - t}} \frac{(-1)^{j_1 + \cdots + j_t}}{j_0! a_1^{j_1+1} \cdots a_t^{j_t+1}} n^{j_0} M^{j_1 + \cdots + j_t + t}.$$

This completes the proof of the lemma. $\square$

In particular, the above lemma implies that, $\mathrm{Err}(n, M) = \tilde{\mathcal{O}}(n^{\ell-1})$ as N tends to infinity, uniformly on all M satisfying $|M - m_n| \leq n^{3/4}$. This proves (31).

We are now interested in controlling the variance of $\mathbb{E}[\mathrm{Occ}_n^{\mathcal{A}}|M]$. We start by a general result for the variance of some element in V.

LEMMA 6.2.   *Let $f(n, M)$ be a function in V of degree d. Assume that $f(n, M)$ is bounded by a polynomial function of n and M. Then*

$$\mathrm{Var}(f(n, M)) = \tilde{\mathcal{O}}(n^{2d-1}).$$

PROOF.   Note that any polynomial function of n and M is bounded in 2-norm by a polynomial of n (recall that $\|M\|_r = \mathcal{O}(n)$ for any $r \geq 1$). Moreover, from Lemma 5.5, the event $A_n = \{|M - m_n| \leq n^{3/4}\}$ has probability $1 - oe(n)$. Therefore,

$$\mathrm{Var}(f(n, M)) = \mathrm{Var}(f(n, M)\mathbf{1}_{A_n}) + oe(n)$$

and we aim to prove that

$$\mathrm{Var}(f(n, M)\mathbf{1}_{A_n}) = \tilde{\mathcal{O}}(n^{2d-1}).$$

By $\mathbb{Q}[n]$-bilinearity and the Cauchy–Schwarz inequality, it is enough to consider the case where $f(n, M) = R(M)(1 - b/M)^n$, for some $b \in \mathbb{Z}_{\geq 0}$. Performing a Taylor series expansion of the logarithm, we write, for some integer $A > 0$,

$$(1 - b/M)^n = \exp[n \log(1 - b/M)] = \exp\left[\frac{-bn}{M} + n P_1(M^{-1}) + \mathcal{O}(M^{-A})\right],$$

where P_1 is a polynomial in M^{-1} with no constant and linear terms. Continuing the computation, we have

$$(1 - b/M)^n = \exp\left[\frac{-bn}{m_n}\right] \exp\left[\frac{-bn(M - m_n)}{Mm_n}\right] \exp[n P_1(M^{-1})](1 + \mathcal{O}(M^{-A})).$$

When $|M - m_n| \leq n^{3/4}$, we have $\frac{-bn(M-m_n)}{Mm_n} = \tilde{\mathcal{O}}(n^{-1/4})$ and $n P_1(M^{-1}) = \tilde{\mathcal{O}}(n^{-1})$. Doing some series expansion of the second and third exponential factors above, we get

$$\frac{(1 - b/M)^n}{\exp[\frac{-bn}{m_n}]} = P(n, M, m_n) + \tilde{\mathcal{O}}(n^{-A}),$$

where $P(n, M, m_n)$ is a Laurent polynomial of total degree 0. We now multiply by $R(M)$ and perform a series expansion of $R(M)$ on the right-hand side: this gives

$$\frac{R(M)(1 - b/M)^n}{\exp[\frac{-bn}{m_n}]} = \tilde{P}(n, M, m_n) + \tilde{\mathcal{O}}(n^{d_R - A}),$$

where $\tilde{P}(n, M, m_n)$ is a Laurent polynomial of total degree $d_R := \deg(R)$. Since, for all integers r, we have $\mathrm{Var}(M^r) = n^{2r}\tilde{\mathcal{O}}(n^{-1})$ (see Corollaries 5.4 and 5.7), it is clear that all Laurent monomials degree d in n, M and m_n have variance $\tilde{\mathcal{O}}(n^{2d-1})$. Hence using the Cauchy–Schwarz inequality and choosing $A = 1$, we have

$$\mathrm{Var}\left(\frac{R(M)(1 - b/M)^n}{\exp[\frac{-bn}{m_n}]}\mathbf{1}_{A_n}\right) = \mathrm{Var}(\tilde{P}(n, M, m_n) + \tilde{\mathcal{O}}(n^{d_R - A}))$$

$$= \tilde{\mathcal{O}}(n^{2d_R - 1}).$$

Since the denominator is deterministic and has order n^{-b}, this implies

$$\mathrm{Var}(R(M)(1 - b/M)^n\mathbf{1}_{A_n}) = \tilde{\mathcal{O}}(n^{2(d_R - b)-1}),$$

as wanted. $\square$

COROLLARY 6.3. *If $R_0(M)$ denotes the Laurent polynomial appearing in* (30), *we have*

$$\text{Var}\left(\frac{R_0(M)}{M^a}\text{Err}(n, M)\right) = \tilde{\mathcal{O}}(n^{2\ell-2a-3}).$$

PROOF. We know that $\text{Err}(n, M)$ is an element of V of degree at most $\ell - 1$, which implies that $\frac{R_0(M)}{M^a}\text{Err}(n, M)$ is also an element of V, of degree $\ell - a - 1$. Moreover, since $\mathbb{E}(\text{Occ}_n^{\mathcal{A}}|M)$ is bounded by a polynomial function in n and M (namely $n^{|\mathcal{A}|}$), the quantity $\frac{R_0(M)}{M^a}\text{Err}(n, M)$ also has this property. We can therefore apply Lemma 6.2 and conclude. $\square$

Now consider Laurent polynomials $P(n, M)$ in the variables n and M. We order monomials $n^x M^y$, with the lexicographic order on $(x + y, x)$; this is consistent with the natural asymptotic ordering when A_n holds.

LEMMA 6.4. *Assume that*

$$(60) \qquad\qquad P(n, M) = n^{x_0} M^{y_0} + \textit{smaller degree terms}$$

with $y_0 \neq 0$ (in particular, M does appear in the dominant monomial). Then

$$\text{Var}(P(n, M)) = \text{Var}(n^{x_0} M^{y_0})(1 + o(1)) = \tilde{\Theta}(n^{2x_0+2y_0-1}).$$

PROOF. As above, we write $M = m_n + \sigma_n Z_n$, with Z_n asymptotically normal. For any x_1, y_1, x_2, y_2 with $y_1 \neq y_2$, we have

$$\text{Cov}(n^{x_1} M^{y_1}, n^{x_2} M^{y_2})$$

$$= n^{x_1+x_2} \sum_{k_1,k_2 \geq 0} \binom{y_1}{k_1}\binom{y_2}{k_2} m_n^{y_1+y_2-k_1-k_2}\sigma_n^{k_1+k_2} \text{Cov}(Z_n^{k_1}, Z_n^{k_2}),$$

where the sum might be finite or infinite depending on the signs of y_1 and y_2. The summand with the largest asymptotic behaviour correspond to $k_1 = k_2 = 1$ (if $k_1 = 0$ or $k_2 = 0$, the corresponding summand is 0). Therefore, when the sum is finite (i.e., when $y_1, y_2 > 0$), we have

$$\text{Cov}(n^{x_1} M^{y_1}, n^{x_2} M^{y_2}) = n^{x_1+x_2} m_n^{y_1+y_2-2}\sigma_n^2(1 + o(1)).$$

The same can be proved when y_1 and/or y_2 is negative, using the technique of Proposition 5.6; details are left to the reader.

Writing $P(n, M)$ as a sum of monomials and expanding $\text{Var}(P(n, M))$ concludes the proof. $\square$

We now give a similar result when $P(n, M)$ is perturbed by an error term with sufficiently small variance.

LEMMA 6.5. *We take a Laurent polynomial $P(n, M)$ as in* (60) *and let $E(n, M)$ be a function with* $\text{Var}(E(n, M)) = \tilde{\mathcal{O}}(n^{2x_0+2y_0-3})$. *Then*

$$\text{Var}(P(n, M) + E(n, M)) = \text{Var}(P(n, M))(1 + o(1)) = \tilde{\Theta}(n^{2x_0+2y_0-1}).$$

PROOF. This is a trivial application of the Cauchy–Schwarz inequality after expanding the variance. $\square$

Looking at equation (30), Lemma 6.1, and Corollary 6.3, we see that $\mathbb{E}(\mathrm{Occ}_n^{\mathcal{A}}\,|M)$ is of the form $P(n, M) + E(n, M)$, where $\mathrm{Var}(E(n, M)) = \tilde{\mathcal{O}}(n^{2\ell-2a-3})$ and $P(n, M)$ is a Laurent polynomial with dominant term $\frac{1}{(\ell-t)!}n^{\ell-t}M^{t-a}$. If $t \neq a$, Lemma 6.5 directly applies and (32) is proved. If $t = a$, we simply apply Lemma 6.5 to $\mathbb{E}(\mathrm{Occ}_n^{\mathcal{A}}\,|M) - \frac{1}{(\ell-t)!}n^{\ell-t}$: this difference has the same variance as $\mathbb{E}(\mathrm{Occ}_n^{\mathcal{A}}\,|M)$ (we removed a deterministic quantity) and is of the desired form: its dominant term is $\Theta(n^{\ell-t-1}M^{-1})$. This concludes the proof of (32).

REFERENCES

[1] ALBERT, M. H., ALDRED, R. E. L., ATKINSON, M. D., HANDLEY, C. and HOLTON, D. (2001). Permutations of a multiset avoiding permutations of length 3. *European J. Combin.* **22** 1021–1031. MR1861045 https://doi.org/10.1006/eujc.2001.0538

[2] BÓNA, M. (2010). On three different notions of monotone subsequences. In *Permutation Patterns. London Mathematical Society Lecture Note Series* **376** 89–114. Cambridge Univ. Press, Cambridge. MR2732825 https://doi.org/10.1017/CBO9780511902499.005

[3] BRILLINGER, D. (1969). The calculation of cumulants via conditioning. *Ann. Inst. Statist. Math.* **21** 215–218.

[4] CANFIELD, E. R. (1995). Engel's inequality for Bell numbers. *J. Combin. Theory Ser. A* **72** 184–187. MR1354972 https://doi.org/10.1016/0097-3165(95)90033-0

[5] CANFIELD, E. R., JANSON, S. and ZEILBERGER, D. (2011). The Mahonian probability distribution on words is asymptotically normal. *Adv. in Appl. Math.* **46** 109–124. MR2794017 https://doi.org/10.1016/j.aam.2009.10.001

[6] CHATTERJEE, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610. MR2435859 https://doi.org/10.1214/07-AOP370

[7] CHERN, B., DIACONIS, P., KANE, D. M. and RHOADES, R. C. (2014). Closed expressions for averages of set partition statistics. *Res. Math. Sci.* **1** Art. 2, 32. MR3338726 https://doi.org/10.1186/2197-9847-1-2

[8] CHERN, B., DIACONIS, P., KANE, D. M. and RHOADES, R. C. (2015). Central limit theorems for some set partition statistics. *Adv. in Appl. Math.* **70** 92–105. MR3388867 https://doi.org/10.1016/j.aam.2015.06.008

[9] CRANE, H., DESALVO, S. and ELIZALDE, S. (2018). The probability of avoiding consecutive patterns in the Mallows distribution. *Random Structures Algorithms* **53** 417–447. MR3854041 https://doi.org/10.1002/rsa.20776

[10] FÉRAY, V. (2013). Asymptotic behavior of some statistics in Ewens random permutations. *Electron. J. Probab.* **18** no. 76, 32. MR3091722 https://doi.org/10.1214/EJP.v18-2496

[11] FÉRAY, V. (2018). Weighted dependency graphs. *Electron. J. Probab.* **23** Paper No. 93, 65. MR3858921 https://doi.org/10.1214/18-EJP222

[12] FÉRAY, V., MÉLIOT, P.-L. and NIKEGHBALI, A. (2016). *Mod-φ Convergence: Normality Zones and Precise Deviations. SpringerBriefs in Probability and Mathematical Statistics.* Springer, Cham. MR3585777 https://doi.org/10.1007/978-3-319-46822-8

[13] FLAJOLET, P., SZPANKOWSKI, W. and VALLÉE, B. (2006). Hidden word statistics. *J. ACM* **53** 147–183. MR2212002 https://doi.org/10.1145/1120582.1120586

[14] FULMAN, J. (2004). Stein's method and non-reversible Markov chains. In *Stein's Method: Expository Lectures and Applications. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **46** 69–77. IMS, Beachwood, OH. MR2118603

[15] GOLDSTEIN, L. (2005). Berry–Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42** 661–683. MR2157512 https://doi.org/10.1239/jap/1127322019

[16] HARPER, L. H. (1967). Stirling behavior is asymptotically normal. *Ann. Math. Stat.* **38** 410–414. MR0211432 https://doi.org/10.1214/aoms/1177698956

[17] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196

[18] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[19] HOFER, L. (2017). A central limit theorem for vincular permutation patterns. *Discrete Math. Theor. Comput. Sci.* **19** Paper No. 9, 26. MR3783365

[20] JANSON, S. (1988). Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs. *Ann. Probab.* **16** 305–312. MR0920273

[21] JANSON, S. (1994). Orthogonal decompositions and functional limit theorems for random graph statistics. *Mem. Amer. Math. Soc.* **111** vi+78. MR1219708 https://doi.org/10.1090/memo/0534

[22] JANSON, S. (1997). *Gaussian Hilbert Spaces. Cambridge Tracts in Mathematics* **129**. Cambridge Univ. Press, Cambridge. MR1474726 https://doi.org/10.1017/CBO9780511526169

[23] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley Interscience, New York. MR1782847 https://doi.org/10.1002/9781118032718

[24] JANSON, S., NAKAMURA, B. and ZEILBERGER, D. (2015). On the asymptotic statistics of the number of occurrences of multiple permutation patterns. *J. Comb.* **6** 117–143. MR3338847 https://doi.org/10.4310/JOC.2015.v6.n1.a8

[25] LEONOV, V. P. and SIRJAEV, A. N. (1959). On a method of calculation of semi-invariants. *Theory Probab. Appl.* **4** 319–329. MR0123345 https://doi.org/10.1137/1104031

[26] MIKHAĬLOV, V. G. (1991). On a theorem of Janson. *Theory Probab. Appl.* **36** 173–176. MR1109028 https://doi.org/10.1137/1136018

[27] MITZENMACHER, M. and UPFAL, E. (2005). *Probability and Computing*: *Randomized Algorithms and Probabilistic Analysis*. Cambridge Univ. Press, Cambridge. MR2144605 https://doi.org/10.1017/CBO9780511813603

[28] ROTA, G.-C. (1964). The number of partitions of a set. *Amer. Math. Monthly* **71** 498–504. MR0161805 https://doi.org/10.2307/2312585

[29] RUCIŃSKI, A. (1988). When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** 1–10. MR0940863 https://doi.org/10.1007/BF00718031

[30] SCHUTZENBERGER, M.-P. (1947). Sur certains paramètres caractéristiques des systèmes d'événements compatibles et dépendants et leur application au calcul des cumulants de la répétition. *C. R. Math. Acad. Sci. Paris* **225** 277–278. MR0021668

[31] STAM, A. J. (1983). Generation of a random partition of a finite set by an urn model. *J. Combin. Theory Ser. A* **35** 231–240. MR0712107 https://doi.org/10.1016/0097-3165(83)90009-2

[32] THIEL, M. (2016). The inversion number and the major index are asymptotically jointly normally distributed on words. *Combin. Probab. Comput.* **25** 470–483. MR3482667 https://doi.org/10.1017/S0963548315000346

[33] URSELL, H. D. (1927). The evaluation of Gibbs' phase-integral for imperfect gases. *Math. Proc. Cambridge Philos. Soc.* **23** 685–697.