The probability of non-existence of small substructures via clusters and cumulants

Rui-Ray Zhang

(legal name: Rui Zhang)

Bachelor of Physics, Sichuan University, 2016

Master of Computer Science, University of Chinese Academy of Sciences, 2019

A thesis submitted for the degree of

Doctor of Philosophy

at Monash University in 2023

School of Mathematics, Monash University

Melbourne, Australia

2023

Copyright notice

Abstract

The central focus of this thesis is to obtain the asymptotics of the probability of non-existence of small substructures in random objects. It contains several parts that all fit into one bigger picture of using clusters and cumulants to characterise dependencies in probabilistic combinatorics.

In Chapter 2, we express the probability that a binomial random hypergraph contains no copy of some given small hypergraphs in terms of clusters, by showing a cumulant series obtained by Mousset, Noever, Panagiotou and Samotij in 2020, approximating the same probability, is equivalent to a truncated cluster expansion series. In addition, we use the tree-graph approach to bound the cumulants of graph-dependent variables, rederiving a bound by Féray, Méliot and Nikeghbali in 2016, and give an asymptotic normality criterion that generalizes the one by Janson in 1988. Moreover, we also obtain the asymptotic distribution of maxima of graph-dependent sequences by using convergent cluster expansions.

In Chapter 3, we obtain the limiting distribution of maxima of various extension counts (codegrees, clique-extensions, common neighbours) in random hypergraphs by studying the probability of the non-occurrence of exceedances, that is, variables whose values are greater than certain specified threshold. Under certain weakly dependent conditions, we show that the distribution of the maximum entry of a random vector and the distribution of the maximum of their independent entries are asymptotically equivalent.

In Chapter 4, our focus is the Eulerian orientations, which are orientations of all edges of a graph such that every vertex is balanced (that is, its in-degree being equal to its out-degree). Studying the probability of non-existence of unbalanced vertices for random orientations is essentially equivalent to the enumeration of Eulerian orientations. We give accurate asymptotic enumerations of Eulerian orientations of graphs, regular tournaments, Eulerian digraphs, and Eulerian oriented graphs extending McKay's results in 1990. This is by using the saddle point method applied to a certain high dimensional integral, and truncated cumulant series. We derive accurate approximations of moment-generating function of higher-order Lipschitz functions, which can be of independent interest. Our formula yields estimates of the Eulerian orientations of square lattices, triangular lattices, cubic lattices, and hypercube, etc. Our values turn out to be close to the 'naive' estimates by Pauling in 1935, and the only two known exact values for ice-type models by Lieb in 1967, and by Baxter in 1969, respectively.

In Chapter 5, we use the perturbation method to study the probability of the non-existence of small subhypergraphs in random hypergraphs $H_r(n,p)$, which is also obtained by Mousset, Noever, Panagiotou and Samotij using an alternative method, as describe in Chapter 2. Our results have an advantage of extending easily to the case of $H_r(n,m)$ that is more complicated and not well studied before. As a special case, we give the asymptotic probability of a random hypergraph being linear. This extends results by McKay and Tian in 2021. Additionally, by keeping track of the numbers of clusters, we obtain approximations of the conditional probabilities of avoiding certain sets of clusters given the counts for smaller ones and the non-existence of even larger ones for $H_r(n,p)$ and $H_r(n,m)$, which can be of independent interest.

In all the above studies, we utilize clusters of combinatorial structures and cumulants of certain suitably defined dependent variables to capture the high-order terms in random combinatorial problems.

Keywords: Random graphs, linear hypergraphs, directed graphs, cumulants, cluster expansion, spanning tree, independence polynomial, chromatic polynomial, asymptotic normality, m-dependence, extreme value theory, extremal index, Gumbel distribution, mixing coefficient, Eulerian orientations, regular tournaments, Eulerian digraphs, Eulerian oriented graphs, graph Laplacian matrix, ice-type model.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Rui Zhang
1st Sep 2023 at Brisbane
revised 10th Dec 2023 at Barcelona

Publications during enrolment

7. Cumulant expansion for counting Eulerian orientations (arXiv, submitted).
   with Mikhail Isaev and Brendan McKay.

6. Generalization bounds for learning under graph-dependence: A survey. (arXiv, submitted)
   with Massih-Reza Amini.

5. Extremal independence in discrete random systems.
   (arXiv, accepted in *Annales de l'Institut Henri Poincaré B: Probabilités et Statistiques*)
   with Mikhail Isaev, Igor Rodionov, Maksim Zhukovskii.

   Published:

4. Asymptotic linearity of binomial random hypergraphs via cluster expansion under graph-dependence.
   Advances in Applied Mathematics 2022. (journal version, arXiv)

3. When Janson meets McDiarmid: Concentration under graph-dependence.
   Statistics & Probability Letters 2022. (journal version, arXiv)

2. Extreme value theory for triangular arrays of dependent random variables.
   with Mikhail Isaev, Igor Rodionov, Maksim Zhukovskii.
   Russian Mathematical Surveys 2020. (journal version). Uspekhi Matematicheskikh Nauk 2020. (Russian journal version)

1. McDiarmid-type inequalities for graph-dependent variables and stability bounds.
   with Xingwu Liu, Yuyi Wang, Liwei Wang.
   Advances in Neural Information Processing Systems 32 (NeurIPS 2019) (conference version, arXiv).

Any fool can know. The point is to understand.

—Albert Einstein

What I cannot create, I do not understand.

—Richard Feynman

Acknowledgements

I would like to thank my supervisor Nick Wormald for the invaluable guidance and extreme patience over the years. Many thanks to him for being extremely rigorous on every aspect of mathematics and non-mathematics, providing sanity checks on various types of mathematics, and especially for countless criticism of the writing that would be very useful for future career. Additionally, thanks for other help that I received very quickly whenever I need over the years, including various financial support, including 6 months' stipends from March 2023. Thanks to Hania for her hospitality during many visits.

I would like to sincerely thank my second supervisor Mikhail Isaev for working together when he is available, during which we are able to progress at a fast speed and obtain many results.

I would like to thank Brendan McKay for his significant help, especially his simulations, computations, and programming in various projects, which provide the best computational results known on the planet.

I would like to thank my other collaborators (chronological order): Maksim Zhukovskii, Igor Rodionov, Massih-Reza Amini, Gábor Lugosi, Andrea Collevecchio, Guillem Perarnau, Xavier Pérez-Giménez, etc. For each of them, there exists something that I learned from. Special thanks to Gábor Lugosi for offering a postdoc position.

In addition, I would like to thank (chronological order): Will Perkins, Ian Wanless, David Wood, Tim Garoni, Wojciech Samotij, Catherine Greenhill, Andrzej Ruciński, Thomas Łuczak, Bruce Reed, Marco Scarsini, and many others for helpful discussions and meetings.

I would like to thank many journal editors, and anonymous referees for comments, suggestions, and criticism, which led to many modifications, including improvements, of various writings.

Special thanks to John Chan and many others for very efficient handling and timely reminders of the administrative issues of various kinds over the years.

Thanks to Monash office mates Michael and Angus for having non-academic chats on a wide range of topics.

Special thanks to Lily for accompanying, and to my families for help in various forms.

Special thanks to two thesis examiners for pointing out typographical errors, grammar mistakes, and inaccurate descriptions, and providing very helpful comments, suggestions, and new insightful ideas!

Contributions to chapters

- A part of Chapter 2 is based on *Asymptotic linearity of binomial random hypergraphs via cluster expansion under graph-dependence* [104] published in Advances in Applied Mathematics 2022.
  - Brendan McKay provided significant help, including checking numerical values via simulations, providing his computational results for comparison, and pointing out calculation errors.
  - Nick Wormald commented on the writing.
  - An anonymous referee pointed out some typographical errors and inaccurate descriptions, and provided comments and suggestions.
  - It is benefited from discussions with Will Perkins on cluster expansion via Zoom during Covid in 2021.
- Chapter 3 is based on papers: *Extreme value theory for triangular arrays of dependent random variables* [48] published in Russian Mathematical Surveys 2020, and *Extremal independence in discrete random systems* accepted in Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques [49]; both are with Mikhail Isaev, Igor Rodionov and Maksim Zhukovskii.
  - The Gaussian part is by Igor Rodionov, and is therefore not included. Other parts are with Mikhail Isaev and Maksim Zhukovskii.
  - An anonymous referee pointed out some typographical errors and inaccurate descriptions, and provided comments and suggestions.
  - Anonymous referees from *Combinatorics, Probability and Computing* also provided helpful comments and suggestions.
- Chapter 4 is based on drafts *Cumulant expansion for Eulerian orientation count* [47] and *Residual entropy and Eulerian orientations of random graphs with given degrees* (under preparation) with Mikhail Isaev and Brendan McKay.
  - Mikhail Isaev and Brendan McKay provided the method, commented on the writing, and worked together on many proofs, edits, calculations, and computations.
- Chapter 5 is based on a draft *The non-existence of small subhypergraphs in moderately sparse random hypergraphs via the perturbation method* with Nick Wormald.
  - Nick Wormald provided the method, worked together on many proofs, and commented *a lot* on the writing.
- The content and ideas in Section 6.1, and future work (fw5) are suggested by Will Perkins.

Contents

Chapter 1

Introduction and preliminaries

This chapter is to give an overview of the content of this thesis, and to set the context for the remainder of this thesis. Section 1.2 contains basic definitions and notations.

1.1 Overview of chapters

Here we describe the contents of each chapter, starting with a illustration by Figure 1.1.

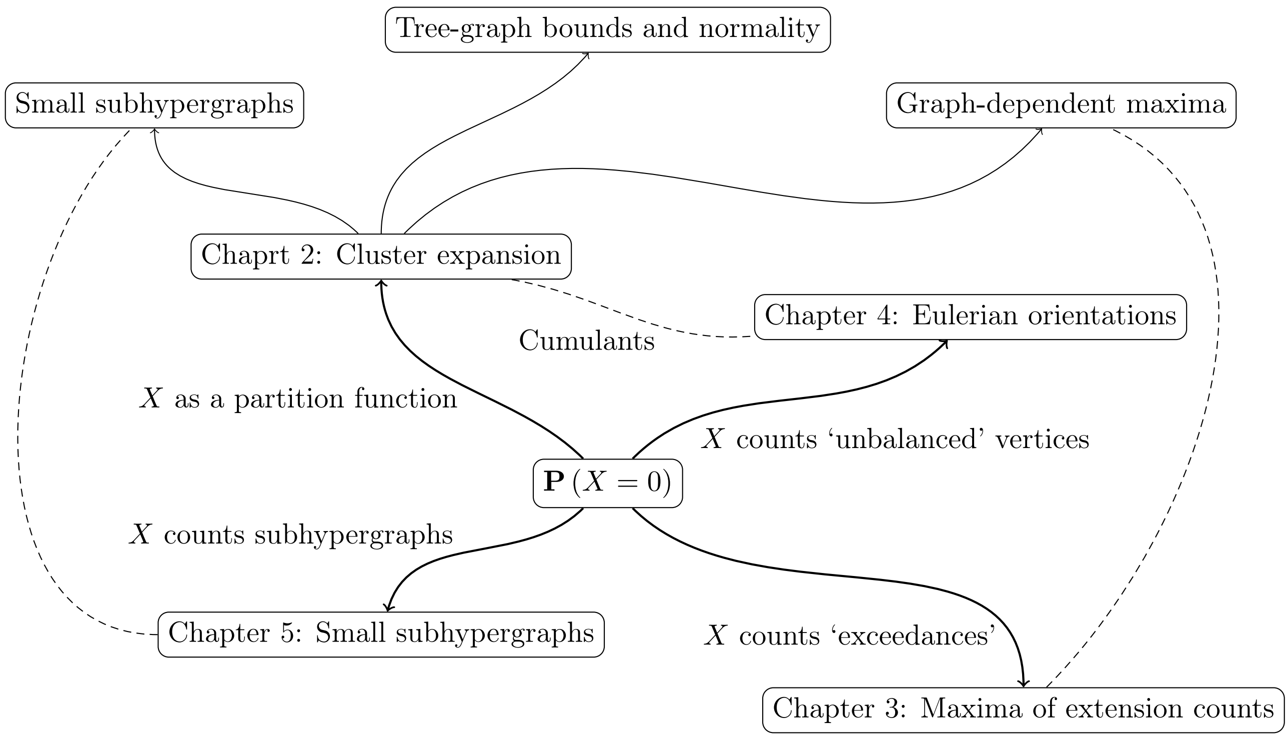


Figure 1.1: A mixed graph illustrating relations among chapters, where the mixed graph is a graph with edges and arcs, and random variable X counts small substructures in random objects.

- Chapter 2 contains three applications of cluster expansion.

  Cluster expansion is a powerful tool in the rigorous study of statistical mechanics. It was pioneered by Mayer in the 1930s and remains widely used nowadays, see, for example, [30, Chapter 5]. The cluster expansion expresses the logarithm of a certain partition function as an infinite summation over clusters.

- We express the probability that a binomial random hypergraph contains no copy of some given small hypergraphs in terms of clusters, by showing a cumulant series obtained by Mousset, Noever, Panagiotou and Samotij [77] approximating the same probability is equivalent to a truncated cluster expansion series. We use the formal cluster expansion after writing the probability of interest as a partition function. As a special case, we extend the result on the asymptotic probability of a random binomial hypergraph $H_r(n, p)$ being linear obtained by McKay and Tian in [74], with explicit computation for $r = 3$ and $p = o(n^{-7/5})$.
- We use the tree-graph approach for the cluster expansion to bound cumulant of graph-dependent variables. This provides an alternative proof of the bound on cumulant by Féray, Méliot and Nikeghbali in [26], and further leads to an asymptotic normality criterion that generalises Janson's [54].
- We obtain the asymptotic distribution of maxima under graph-dependence via cluster expansion, which gives accurate asymptotic distribution of the maxima of m-dependent random variables. This extends the seminal work of Newell's [79] that originates the study of clustering of exceedances in extreme value theory. Our method is by using Koteckỳ-Preiss criterion [61] to obtain the absolute convergence of the cluster expansion. The new asymptotic formula also provides new insights into the extremal index.

- Chapter 3 studies the limiting distribution of the maximum of weakly dependent variables.

  Under certain weakly dependent conditions that is a special notion of φ-mixing, we show that the distribution of the maximum entry of a random vector and the distribution of the maximum of their independent entries are asymptotically equivalent.

  Our result on extremal independence relies on new lower and upper bounds for the probability of the non-existence, that were inspired by Lovász local lemma [21] and Dubickas' bound [20].

  As applications, we obtain the distribution of various extremal characteristics of random discrete structures such as the maximum codegree in binomial random hypergraphs, and the maximum number of cliques sharing a given vertex in binomial random graphs, etc. We show that their limiting distributions are all standard Gumbel, extending the results for the maximum degree of binomial random graphs by Bollobás [10].

- Chapter 4 uses cumulants to give accurate asymptotic enumeration results.

  We study Eulerian orientations, which are orientations of all edges of a graph such that every vertex is balanced with its in-degree being equal to the out-degree. Studying the probability of non-existence of unbalanced vertices for random orientations is essentially equivalent to the enumeration of Eulerian orientations. We give accurate asymptotic enumerations of Eulerian orientations of graphs, regular tournaments, Eulerian digraphs, and Eulerian oriented graphs extending the results by McKay [73]. This is by using the saddle point method applied to a certain high dimensional integral, and truncated cumulant series. We use the saddle point method applied to some high-dimensional integrals and truncated cumulant series. In particular, we derive accurate approximations of the moment-generating function of higher-order Lipschitz functions, which can be of independent interest.

We also use our series to estimate the Eulerian orientations of square lattices, triangular lattices, cubic lattices, hypercubes, etc. Our values turn out to be close to the estimates by Pauling [81] that is simply ignoring the dependence among vertices, and the only known exact values by Lieb [66] and Baxter [7] for ice-type models.

- Chapter 5 studies the non-existence of small subhypergraphs in random hypergraphs via the perturbation method.

  We extend the perturbation method introduced by Nick Wormald [103], and its generalization with Stark [98], to obtain the asymptotics of the probability of the non-existence of small subhypergraphs in random hypergraphs $H_r(n,p)$ and $H_r(n,m)$ for moderately large p and m. The case of $H_r(n,p)$ is also obtained by Mousset, Noever, Panagiotou and Samotij using an alternative method. Our results have a advantage of extending easily to the case of $H_r(n,m)$ that is more complicated and not well-studied before besides [103, 98].

  As a direct corollary, we derive the asymptotic probability of a random hypergraph being linear. In the case of fixed r, this relaxes the restriction on p for the asymptotic probability obtained by McKay and Tian [74]. For $H_r(n,p)$, when $r = 3$ and $p = o(n^{-7/5})$, the asymptotic probability matches the one obtained in Chapter 2.

  Similar to the graph case in [98], by keeping track of the numbers of clusters, we obtain stronger results, giving an approximation of the conditional probabilities of avoiding certain sets of clusters of hypergraphs given the counts for smaller ones and the non-existence of even larger ones, which is of independent interest.

- Finally, we conclude with a discussion of a plan for future work in the last chapter.

1.2 Basic definitions and notations

To make the descriptions and concepts clear and specific, we give basic definitions and the framework. First, we introduce the basic definitions and notations that will be used in this thesis.

For all positive integer $n \geqslant 1$, let $[n]$ denote the integer set $\{1, 2, \ldots, n\}$. Let $[n]_t := n(n-1)\cdots(n-t+1)$ denote the t-th falling factorial for every non-negative integer t. For any set V and $1 \leqslant r \leqslant |V|$, let $\binom{V}{r}$ denote the set containing all r-elements subset of V.

1.2.1 Graph-theoretical notation

A graph is a pair $G = (V(G), E(G))$, where $V(G)$ is a set whose elements are called vertices, and $E(G) \subseteq \binom{V(G)}{2}$ is a set of paired vertices, whose elements are called edges.

Definition 1.1. Given an undirected graph $G = (V(G), E(G))$.

(d1) A connected component of G is a maximal set of vertices such that every pair of vertices is connected by a path. The number of connected components of G is denoted by $c(G)$.

(d2) The set of *polymers* $\mathcal{C}(G)$ of G is the family of vertex sets of all connected induced subgraphs of G, namely,

$$\mathcal{C}(G) = \{C \subseteq V(G) : c(G[C]) = 1\},$$

where $G[C]$ denotes the subgraph of G induced by the vertex set C.

For any two distinct polymers $C_i, C_j \in \mathcal{C}(G)$, we write $C_i \sim C_j$ if $C_i \cup C_j \in \mathcal{C}(G)$; otherwise, $C_i \nsim C_j$. Equivalently, $C_i \nsim C_j$ if $d_G(C_i, C_j) > 1$ and otherwise, $C_i \sim C_j$. Note that if we have that $C_i \sim C_j$ and $C_i \cap C_j = \emptyset$, then C_i and C_j are adjacent in G, that is, there exists an edge in $E(G)$, with one endpoint in C_i and the other in C_j.

(d3) The size of a polymer, denoted by $|C|$, is the number of vertices in it. We will use

$$\mathcal{C}_k(G) := \{C \in \mathcal{C}(G) : |C| = k\}, \quad \text{and} \quad \mathcal{C}_{\leqslant k}(G) := \bigcup_{i \in [k]} \mathcal{C}_i(G),$$

to denote the set of all polymers of size k, and at most k respectively.

(d4) For every non-empty ordered multiset of polymers $(C_1, \ldots, C_n) \in \mathcal{C}(G)^n$, let $\mathbb{G}(C_1, \ldots, C_n) = \mathbb{G}_G(C_1, \ldots, C_n)$ be the graph on $[n]$ with $\{i, j\} \in E(\mathbb{G})$ if $C_i \sim C_j$ in G.

For instance, fix a polymer $C \in \mathcal{C}(G)$. For a multiset of n copies of C, we have $\mathbb{G}(C, \ldots, C) = K_n$, where K_n denotes the complete graph on $[n]$.

If we have that $\{C_1, \ldots, C_n\}$ is a partition of the vertex set of graph G, then graph $\mathbb{G}(C_1, \ldots, C_n)$ corresponds to the so-called *quotient graph* of G with respect to $\{C_1, \ldots, C_n\}$.

In graph theory, a *vertex identification* (also called vertex contraction) is to contract a pair of vertices u and v of a graph and produces a graph in which the two vertices u and v are replaced with a single vertex t such that t is adjacent to the union of the vertices to which u and v were originally adjacent. In other words, it is by identifying the vertices in each part, deleting loops, and replacing parallel edges with a single edge. Note that in vertex contraction, it does not matter if u and v are connected by an edge; if they are, the edge is simply removed upon contraction, this special case of vertex identification is also called *edge contraction*.

(d5) A *cluster* γ is a non-empty ordered multiset of polymers $(C_1, \ldots, C_{|\gamma|})$ such that $\mathbb{G}(\gamma) = \mathbb{G}(C_1, \ldots, C_{|\gamma|})$ is connected. The size of a cluster γ, denoted by $|\gamma|$, is the number of polymers in it, and the number of vertices of a cluster γ, denoted by $\|\gamma\|$, is the sum of sizes of polymers it contains, that is, $\|\gamma\| = \sum_{C \in \gamma} |C|$. The set of all clusters of G is denoted by $\Gamma(G)$.

Note that the cluster is redefined and used in Chapter 5, to mean some connected structure that is similar to the polymers defined here, but not exactly.

1.2.2 Asymptotic notation

All asymptotics in this thesis are with respect to $n \to \infty$. We use the following asymptotic notations. Let $g(n) > 0$ for all large enough integer n.

- $f(n) = o(g(n))$ if for every $\varepsilon > 0$ there exists N_ε such that

$$|f(n)| \leqslant \varepsilon g(n)$$

for $n \geqslant N_\varepsilon$; in other words, $f(n)/g(n) \to 0$.

- $f(n) = O(g(n))$ if there exist constants C and N such that

$$|f(n)| \leqslant Cg(n)$$

for $n \geqslant N$; in other words, $f(n)/g(n)$ is bounded.

- $f(n) = \Omega(g(n))$ if there exist constants $C > 0$ and N such that

$$f(n) \geqslant Cg(n)$$

for $n \geqslant N$.

- $f(n) = \omega(g(n))$ if for all $C > 0$ there exists a constant N_C such that

$$f(n) > Cg(n)$$

for $n \geqslant N$.

- $f(n) = \Theta(g(n))$ if there exist constants $C_1, C_2 > 0$ and N such that

$$C_1 g(n) \leqslant f(n) \leqslant C_2 g(n)$$

for $n \geqslant N$; in other words, $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

- $f(n) \ll g(n)$ if $f(n) \geqslant 0$ and $f(n) = o(g(n))$.
- $f(n) \sim g(n)$ if $f(n) = (1 + o(1))g(n)$; in other words, $f(n)/g(n) \to 1$.

1.2.3 Random (hyper)graphs

Here we introduce the random hypergraph models that we study. Define the binomial random r-uniform hypergraph $H_r(n, p)$ to be the r-uniform hypergraph (r-graph for short) on the vertex set $[n]$ such that each r-element subset (r-set for short) is an edge independently with probability p. We use $H_r(n, m)$ to denote the random r-graphs on n vertices obtained by choosing uniformly at random from the $\binom{\binom{n}{r}}{m}$ r-graphs having m hyperedges. When $r = 2$, we have $H_r(n, p)$ and $H_r(n, m)$ become the Erdős-Rényi random graphs $\mathcal{G}(n, p)$ and $\mathcal{G}(n, m)$ respectively.

1.2.4 Dependency graphs

Dependency graphs can be used to characterize the dependencies among variables. They have been widely used in probability and statistics to establish normal or Poisson approximation via Stein's method, cumulants, etc. (see, for example, [54, 55]). They are also heavily used in probabilistic combinatorics, such as Lovász local lemma [21], Janson's inequality [57], concentration inequalities [105], etc.

Given a graph $G = (V, E)$, we say that random variables $\{X_i\}_{i \in V}$ are G-*dependent* if for any disjoint $S, T \subset V$ such that $d_G(S, T) > 1$, random variables $\{X_i\}_{i \in S}$ and $\{X_j\}_{j \in T}$ are independent. In particular, random variables $\{X_i\}_{i \in C_1}$ and $\{X_j\}_{j \in C_2}$ are independent for any two distinct polymers C_1 and C_2 of the graph G such that $C_1 \not\sim C_2$.

Definition 1.2 (Dependency graphs)**.** An undirected graph G is called a dependency graph of random variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ if

(b1) $V(G) = [n]$.

(b2) For all disjoint $I, J \subset [n]$, if I, J are not adjacent in G, then $\{X_i\}_{i\in I}$ and $\{X_j\}_{j\in J}$ are independent.

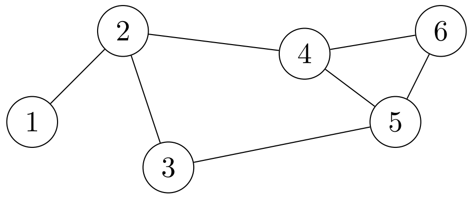


Figure 1.2: A dependency graph G for random variables $\{X_i\}_{i\in[6]}$. Random variables $\{X_1, X_2\}$ and $\{X_5, X_6\}$ are independent, since disjoint vertex sets $\{1, 2\}$ and $\{5, 6\}$ are not adjacent in G.

The above definition of dependency graphs is a strong version; there are ones with weaker assumptions, such as the one used in Lovász local lemma. Note that the dependency graph for a set of random variables may not be necessarily unique and the sparser ones are the more interesting ones. Let K_n denote the complete graph on n vertices, that is, all pairs of vertices are connected by an edge. Since no two disjoint $S, T \subset [n]$ are non-adjacent in K_n, then the trivial dependency graph K_n is a valid dependency graph for any set of variables $\{X_i\}_{i\in[n]}$.

Given G-dependent random variables $\{X_i\}_{i\in V(G)}$, for every set of vertices $S \subseteq V(G)$, the *joint moment* of random variables $\{X_i\}_{i\in S}$ is defined by

$$\mu(S) = \mathbf{E}\left[\prod_{i\in S} X_i\right], \tag{1.2.1}$$

with $\mu(\emptyset) := 1$. We will sometimes simplify the notation and use the $\mu(i)$ to denote $\mu(\{i\})$, and similarly, $\mu(i, j)$ to denote $\mu(\{i, j\})$.

Let $\{C_i\}_{i\in[n]}$ be a set of pairwise non-adjacent disjoint polymers of G, in other words, for all distinct $i, j \in [n]$, we have $C_i \not\sim C_j$, or equivalently, $d_G(C_i, C_j) > 1$. Then one important factorisation property for G-dependent variables, following from the definition of dependency graph, is that

$$\mu\left(\bigcup_{i\in[n]} C_i\right) = \prod_{i\in[n]} \mu(C_i). \tag{1.2.2}$$

1.2.5 Cumulants

Then *joint cumulant* (or *mixed cumulant*) is a fundamental tool in probability theory. Let $X_1, \ldots, X_r$ be random variables with finite moments. The joint cumulant is defined by

$$\kappa(X_1, \ldots, X_r) = [t_1 \ldots t_r] \log\left(\mathbf{E}\left[\exp\left(\sum_{i=1}^{r} t_i X_i\right)\right]\right), \tag{1.2.3}$$

where $[t_1 \ldots t_r]$ stands for the coefficient of $t_1 \ldots t_r$ in the series expansion.

Given G-dependent random variables $\{X_v\}_{v\in V(G)}$, for every set of vertices $S \subseteq V(G)$, we have an equivalent combinatorial definition of the joint cumulant of random variables $\{X_i\}_{i\in S}$ by

$$\kappa(S) = \sum_{\pi\in\Pi(S)} (-1)^{|\pi|-1}(|\pi|-1)! \prod_{P\in\pi} \mu\left(P\right), \tag{1.2.4}$$

where $\Pi(S)$ denotes the set of all partitions of S.

The joint cumulant $\kappa(S)$ can be regarded as a measure of the mutual dependencies of the variables in S. An important property of the joint cumulant $\kappa(S)$ is that if S can be partitioned into two subsets S_1 and S_2 such that the variables in S_1 are independent of the variables in S_2, then $\kappa(S) = 0$. In other words, if $S \notin \mathcal{C}(G)$, then $\kappa(S) = 0$ (see, for example, [93]). This reveals the natural connections between cumulants and clusters. More details will be discussed in Section 2.4 using spanning trees.

1.3 Classical results through the lens of clusters illustrated with triangles in $\mathcal{G}(n, p)$

Here we present some classical probabilistic results with a unified formulation in terms of clusters. We illustrate and compare the classical results by giving the distribution of the triangles in random binomial graphs $\mathcal{G}(n, p)$. To determine the probability that $\mathcal{G}(n, p)$ does not contain a copy of some given 'forbidden' graph is a fundamental question in the random graph theory since the seminal paper of Erdős and Rényi [23]. To avoid triviality, we assume that $0 < p < 1$.

For each $\{i, j, k\} \in \binom{[n]}{3}$, let X_{ijk} be the indicator for the occurrence of triangle with vertex set $\{i, j, k\}$ in $\mathcal{G}(n, p)$. Then a dependency graph for random indicators $\{X_{ijk}\}_{\{i,j,k\}\in\binom{[n]}{3}}$ can be defined by

$$G := \left(\binom{[n]}{3}, \left\{\{\mathbf{v}_1, \mathbf{v}_2\} \in \binom{\binom{[n]}{3}}{2} : |\mathbf{v}_1 \cap \mathbf{v}_2| = 2\right\}\right).$$

Let $X := \sum_{\{i,j,k\}\in\binom{[n]}{3}} X_{ijk}$. Then X counts the copies of triangles in $\mathcal{G}(n, p)$.

A simple lower bound on the probability of non-existence of the triangles in $\mathcal{G}(n, p)$ is by the probability of getting an empty graph, that is,

$$\mathbf{P}\left(X = 0\right) \geqslant \mathbf{P}\left(E(\mathcal{G}(n, p)) = \emptyset\right) = (1-p)^{\binom{n}{2}} = \exp\left(-\frac{1}{2}[n]_2 p + O\left(n^2p^2\right)\right) = e^{-\Theta\left(n^2p\right)}. \tag{1.3.1}$$

1.3.1 Non-existence probability: Suen's inequality

Suen [99] obtained upper and lower bounds on $\mathbf{P}\left(X = 0\right)$, where X is a sum of G-dependent random indicators. Later, Janson [56] improved Suen's inequality and obtained strengthened inequalities that become the most commonly used and cited version.

Theorem 1.3 (Suen's inequality). *Let $\{X_v\}_{v\in V(G)}$ be G-dependent random indicators and $X = \sum_{v\in V(G)} X_v$. Let*

$$\Delta^\star := \sum_{\{i,j\}\in\mathcal{C}_2(G)} \frac{\mu(i,j)}{\widetilde{\mu}(i,j)} \quad \textit{and} \quad \Delta^\star_0 := \sum_{\{i,j\}\in\mathcal{C}_2(G)} \frac{\mu(i)\mu(j)}{\widetilde{\mu}(i,j)} \tag{1.3.2}$$

where $\mathcal{C}_k(G)$ denotes the set of all polymers of size k, and

$$\widetilde{\mu}(i,j) := (1-\mu(i))(1-\mu(j)) \prod_{k:\{i,j,k\}\in\mathcal{C}_3(G)} (1-\mu(k)).$$

Then we have

$$1-\Delta_0^\star \exp(\Delta^\star) \leqslant \frac{\mathbf{P}(X=0)}{\prod_{i\in V(G)}(1-\mu(i))} \leqslant \exp(\Delta^\star).$$

For triangles in $\mathcal{G}(n,p)$, we have, from (1.3.2), that

$$\Delta^\star = \frac{[n]_4 p^5}{4(1-p^3)^{5n-18}} \quad \text{and} \quad \Delta_0^\star = \frac{[n]_4 p^6}{4(1-p^3)^{5n-18}}.$$

Therefore, we have

$$\mathbf{P}(X=0) \leqslant (1-p^3)^{\binom{n}{3}} \exp\left(\frac{[n]_4 p^5}{4(1-p^3)^{5n-18}}\right) = \exp\left(-\frac{1}{6}[n]_3 p^3 + \frac{1}{4}[n]_4 p^5 + O\left(n^3p^6 + n^4p^8\right)\right).$$

The lower bound is useful only when $\Delta_0^\star \leqslant 1$ and $\Delta^\star = o(1)$, by noting that $\Delta^\star > 0$. If we have $n^4p^5 = o(1)$, then

$$\begin{aligned}\mathbf{P}(X=0) &\geqslant (1-p^3)^{\binom{n}{3}} \left(1 - \frac{[n]_4 p^6}{4(1-p^3)^{5n-18}} \exp\left(\frac{[n]_4 p^5}{4(1-p^3)^{5n-18}}\right)\right) \\ &= \exp\left(-\frac{1}{6}[n]_3 p^3 + O\left(n^3p^6 + n^4p^6\right)\right).\end{aligned}$$

1.3.2 Non-existence probability under correlation: Harris-FKG, Janson's inequalities

Let $\Omega = \{0,1\}^n$ and define a partial ordering of the elements in Ω by

$$(x_1,\dots,x_n) \preceq (y_1,\dots,y_n)$$

if and only if $x_i \leqslant y_i$ for all $i \in [n]$. We say that an event $A \subseteq \Omega$ is increasing if $x \in A$ and $x \preceq y$ implies that $y \in A$. We say that an event $A \subseteq \Omega$ is decreasing if $x \in A$ and $y \preceq x$ implies that $y \in A$. Harris [39] obtained that events are positively correlated if they are both increasing or both decreasing.

Theorem 1.4 (Harris' inequality)**.** *Let Ω be a finite set and let X and Y be random variables defined on a product probability space over $\{0,1\}^\Omega$. If X and Y are both non-decreasing (or non-increasing), then*

$$\mathbf{E}[XY] \geqslant \mathbf{E}[X]\,\mathbf{E}[Y].$$

If X is non-decreasing and Y is non-increasing, then

$$\mathbf{E}[XY] \leqslant \mathbf{E}[X]\,\mathbf{E}[Y].$$

A more generalized version is the FKG inequality, attributed to Fortuin, Kasteleyn and Ginibre [29], which will be discussed in the next subsection.

Let R be a random subset of Ω by independently choosing each $r \in \Omega$ with certain probability. Let $\{A_i\}_{i\in I}$ be some subsets of Ω, where I is a finite index set. Let X_i be the indicator of event $A_i \subseteq R$ and

$X=\sum_{i\in I}X_i$. Then X counts the number of $A_i\subseteq R$. The upper bound on $\mathbf{P}(X=0)$ in this binomial random subset is obtained by Janson, Łuczak and Ruciński [57] under the name of *Janson's inequality.*

Theorem 1.5 (Janson's inequality)**.** *Let $\{X_v\}_{v\in V(G)}$ be G-dependent random indicators for the occurrences of sets and $X=\sum_{v\in V(G)}X_v$. Then we have*

$$\mathbf{P}(X=0)\leqslant\min\left(\exp\left(-\mathbf{E}[X]+\Delta\right),\exp\left(-\frac{\mathbf{E}[X]^2}{\mathbf{E}[X]+2\Delta}\right)\right),\tag{1.3.3}$$

where

$$\Delta:=\sum_{\{i,j\}\in\mathcal{C}_2(G)}\mu(i,j)$$

is a summation over unordered dependent pairs (polymers of size 2*).*

Now we can use FKG inequality and Janson's inequality to obtain asymptotics. For $x\in[0,1-\varepsilon]$, we have $\log(1-x)\geqslant -x-x^2/\varepsilon$, and therefore if $\mathbf{E}[X_i]\leqslant 1-\varepsilon$ for some constant $\varepsilon>0$, then

$$\mathbf{P}(X=0)\geqslant\prod_{i\in V(G)}(1-\mathbf{E}[X_i])\geqslant\exp\left(-\mathbf{E}[X]-\frac{1}{\varepsilon}\sum_{i\in V(G)}\mathbf{E}[X_i]^2\right).\tag{1.3.4}$$

Combining this lower bound with the upper bound given by Janson's inequality, we have

$$\log\mathbf{P}(X=0)=-\mathbf{E}[X]+O\left(\sum_{i\in V(G)}\mathbf{E}[X_i]^2+\sum_{\{i,j\}\in\mathcal{C}_2(G)}\mu(i,j)\right).\tag{1.3.5}$$

For triangles in $\mathcal{G}(n,p)$, if $p\leqslant 1-\varepsilon$ for some $\varepsilon>0$, then we have that

$$\log\mathbf{P}(X=0)=-\frac{1}{6}[n]_3p^3+O\left(n^3p^6+n^4p^5\right).\tag{1.3.6}$$

Moreover, using Janson's inequality, we get

$$\mathbf{P}(X=0)=\begin{cases}e^{-\Theta\left(n^2p\right)} & \text{if } p=\omega(n^{-1/2}),\\ e^{-\Theta\left(n^3p^3\right)} & \text{if } p=O(n^{-1/2}).\end{cases}$$

This is by noting that if $p=\omega(n^{-1/2})$, we have from (1.3.3), that

$$\mathbf{P}(X=0)\leqslant\exp\left(-\frac{\left(\frac{1}{6}[n]_3p^3\right)^2}{\frac{1}{6}[n]_3p^3+\frac{1}{2}[n]_4p^5}\right)=e^{-\Theta\left(n^2p\right)},$$

which matches the empty graph lower bound (1.3.1) asymptotically.

1.3.3 Mousset-Noever-Panagiotou-Samotij cumulant series for binomial subsets

The FKG lattice condition [29, Eq. (2.1)] is also called the log-supermodularity condition. It is the condition under which the Fortuin–Kasteleyn–Ginibre (FKG) correlation inequality holds, that is, for all $U, V \subseteq V(G)$,

$$\mu(U)\mu(V) \leqslant \mu(U \cup V)\, \mu(U \cap V)\,, \tag{1.3.7}$$

where $\mu(\cdot)$ was defined by (1.2.1).

Given a graph G, let $N_G^+(v)$ denote the inclusive neighbours of vertex v in graph G, that is,

$$N_G^+(v) := \{u \in V(G) : \{v, u\} \in E(G)\} \cup \{v\}.$$

For any set $U \subseteq V(G)$, let

$$N_G^+(U) := \bigcup_{u \in U} N_G^+(u).$$

Let $i > 0$ be an integer, and define $\kappa_i(D)$ to be the sum of joint cumulants over polymers of size i in the dependency graph D, that is,

$$\kappa_i(D) = \sum_{C \in \mathcal{C}_i(D)} \kappa(C).$$

Exploiting the property that the indicators of the appearance of subsets in a binomial random set satisfy the the FKG lattice condition, Mousset, Noever, Panagiotou and Samotij [77, Theorem 11] obtain an approximation of $\log \mathbf{P}\,(X = 0)$ using a truncated series with terms up to a constant order.

Theorem 1.6. *Let $\varepsilon > 0$ and k be a positive integer. Let $\{X_v\}_{v \in V(G)}$ be indicators of the appearance of subsets in a binomial random set. If*

$$\min_{U \subseteq V(G) : |U| \in [k+1]} \mathbf{P} \left(\sum_{i \in N^+(U)} X_i = 0 \right) \geqslant \varepsilon, \tag{1.3.8}$$

then there exists a constant $K = K(\varepsilon, k)$ such that

$$\left| \log \mathbf{P}\,(X = 0) - \sum_{i \in [k]} (-1)^i \kappa_i(D) \right| \leqslant K \left(\sum_{C \in \mathcal{C}_{\leqslant K}(G)} \mu(C) \max_{v \in C} \mu\,(v) + \sum_{C \in (\mathcal{C}_K(G) \setminus \mathcal{C}_k(G))} \mu(C) \right). \tag{1.3.9}$$

For triangles in $\mathcal{G}(n, p)$, the following result is first obtained by Stark and Wormald [98, Theorem 1.2] and also by Mousset, Noever, Panagiotou and Samotij [77, Corollary 15] as a more or less direct consequence of Theorem 1.6.

Theorem 1.7. *Let X count the number of triangles in $\mathcal{G}(n, p)$. If $p = o\left(n^{-7/11}\right)$, then*

$$\mathbf{P}\,(X = 0) = \exp\left(-\tfrac{1}{6} n^3 p^3 + \tfrac{1}{4} n^4 p^5 - \tfrac{7}{12} n^5 p^7 + \tfrac{1}{2} n^2 p^3 - \tfrac{3}{8} n^4 p^6 + \tfrac{27}{16} n^6 p^9 + o\,(1) \right). \tag{1.3.10}$$

1.3.4 Lovász local lemma and Shearer's lemma

Rederiving the results by Dobrushin [17] and Shearer [89], Scott and Sokal [88] point out the connections between the repulsive hard-core gas in statistical mechanics and the Lovász local lemma, and discovered that the assumption in the Lovász local lemma provides a sufficient condition for the absolute convergence of the cluster expansion of the partition function of hard-core models.

Specifically, given a graph G and a vector $\mathbf{x} = (x_v)_{v \in V(G)}$, the partition function of the hard-core model (also the independence polynomial) on G is defined by

$$\mathbf{I}_G(\mathbf{x}) := \sum_{U \in \mathcal{I}(G)} \prod_{i \in U} x_i, \tag{1.3.11}$$

where $\mathcal{I}(G)$ denotes the set of all independent sets for every graph G.

Let $R(G)$ be the convergence region of the cluster expansion of the hard-core model on graph G, that is,

$$R(G) := \left\{ \mathbf{p} \in (0,1)^{V(G)} : \mathbf{I}_G(-\mathbf{p}\mathbf{1}_S) \geqslant 0 \text{ for all } S \subseteq V(G) \right\}.$$

The following is Shearer's lemma [89, Theorem 1] formulated in [88, Theorem 4.1].

Theorem 1.8 (Shearer's lemma)**.** *Let G be a graph and $\{X_v\}_{v \in V(G)}$ be random indicators. Suppose that $(p_i)_{i \in V(G)}$ are real numbers in $[0,1]$ such that, for each v and each $U \subseteq V(G) \setminus N_G^+(v)$, we have*

$$\mathbf{P}\left(X_v = 1 \,\middle|\, \sum_{u \in U} X_u = 0 \right) \leqslant p_v \,. \tag{1.3.12}$$

If $\mathbf{p} \in R(G)$, then

$$\mathbf{P}\left(\sum_{v \in V(G)} X_v = 0 \right) \geqslant \mathbf{I}_G(-\mathbf{p}) > 0 \tag{1.3.13}$$

and more generally

$$\mathbf{P}\left(\sum_{v \in V} X_v = 0 \,\middle|\, \sum_{u \in U} X_u = 0 \right) \geqslant \frac{\mathbf{I}_G(-\mathbf{p}\,\mathbf{1}_{V \cup U})}{\mathbf{I}_G(-\mathbf{p}\,\mathbf{1}_U)} > 0 \tag{1.3.14}$$

for any subsets $V, U \subseteq V(G)$.

Moreover, this lower bound is best possible in the sense that there exists a probability space on which there can be constructed a family of G-dependent events $(B_i)_{i \in V(G)}$ with probabilities $\mathbf{P}(B_i) = p_i$, such that

$$\mathbf{P}\left(\sum_{v \in V(G)} X_v = 0 \right) = \mathbf{I}_G(-\mathbf{p}).$$

A commonly used corollary of Shearer's Lemma is for the symmetric case, in which all events are given the same probability bound.

Corollary 1.9 ([89, 88])**.** *Let $\{X_v\}_{v \in V(G)}$ be random indicators such that for each $v \in V(G)$, variable*

X_v is independent of all but Δ other variables ($\Delta \geqslant 2$), and

$$\mathbf{P}\left(X_v = 1\right) \leqslant \frac{(\Delta-1)^{\Delta-1}}{\Delta^{\Delta}} := p_{Shearer}(\Delta). \tag{1.3.15}$$

Then

$$\mathbf{P}\left(\sum_{v \in V(G)} X_v = 0\right) > 0.$$

Another easily testable sufficient condition for $\mathbf{P}\left(X = 0\right) > 0$ appears in the Lovász local lemma.

Theorem 1.10 (Lopsided Lovász local lemma [21, 22]). *Let G be a graph and $\{X_v\}_{v \in V(G)}$ be random indicators. Suppose that $(p_i)_{i \in V(G)}$ are real numbers in $[0, 1]$ such that, for each v and each $U \subseteq V(G) \setminus N_G^+(v)$, we have*

$$\mathbf{P}\left(X_v = 1 \,\middle|\, \sum_{u \in U} X_u = 0\right) \leqslant x_v \prod_{u : \{v,u\} \in \mathcal{C}_2(G)} \left(1 - x_u\right). \tag{1.3.16}$$

Then

$$\mathbf{P}\left(X = 0\right) \geqslant \prod_{v \in V(G)} \left(1 - x_v\right) > 0.$$

Moreover, let Δ be the maximum degree of G, and $p_m := \max_{i \in V(G)} p_i$. If

$$e(\Delta + 1)p_m \leqslant 1, \tag{1.3.17}$$

then $\mathbf{P}\left(X = 0\right) > 0$.

The graph in the Lopsided Lovász local lemma is called a negative dependency graph. We will introduce a similar but more general dependency graph using the notion of φ-mixing coefficient, and discuss this in Chapter 3.

Scott and Sokal [88] discovered that the assumption (1.3.16) in the Lovász local lemma is a sufficient condition for the convergence of the cluster expansion, combining Shearer's lemma (1.3.13), simply implies that $\mathbf{P}\left(X = 0\right) > 0$.

For triangles in $\mathcal{G}(n, p)$, by noting that $\Delta = 3(n - 3)$, the Lovász local lemma gives that if

$$p \leqslant \left(\frac{1}{e(\Delta+1)}\right)^{1/3} \leqslant \left(\frac{1}{e(3n-8)}\right)^{1/3},$$

then

$$\mathbf{P}\left(X = 0\right) \geqslant \left(1 - \frac{1}{\Delta+1}\right)^{\binom{n}{3}} = \exp\left(-\frac{[n]_3}{6(3n-8)} - \frac{[n]_3}{12(3n-8)^2} - \frac{[n]_3}{18(3n-8)^3} + O\left(n^{-1}\right)\right) = e^{-\Theta\left(n^2\right)}.$$

However, this is not very useful as the empty graph lower bound (1.3.1) is better.

1.3.5 Chen-Stein method under local dependence

Poisson approximation via Chen-Stein method is for sums of "locally" dependent variables [4, 5], which are essentially the graph-dependent variables. Let $d_{\mathrm{tv}}(\cdot,\cdot)$ denote the total variation distance between two probability measures.

Theorem 1.11 ([4, 5]). *Let $\{X_v\}_{v\in V(G)}$ be G-dependent random indicators and $X=\sum_{v\in V(G)} X_v$. Let $Z\sim \mathrm{Poi}\left(\mathbf{E}\left[X\right]\right)$. Then*

$$\begin{aligned} d_{\mathrm{tv}}(X,Z) &\leqslant \min(1,\mathbf{E}\left[X\right]^{-1})\left(\sum_{i\in V(G)}\mathbf{E}\left[X_i\right]^2+2\sum_{\{i,j\}\in\mathcal{C}_2(G)}\left(\mu\left(i\right)\mu\left(j\right)+\mu\left(ij\right)\right)\right)\\ &=\min\left(\widetilde{L}_{G,1}^{-1}\widetilde{L}_{G,2},\widetilde{L}_{G,2}\right),\end{aligned}$$

where

$$\widetilde{L}_{G,k}:=\sum_{\gamma\in\Gamma(G):\|\gamma\|=k}\prod_{C\in\gamma}\mu\left(C\right).$$

It turns out that the dependency graph can be defined with greater flexibility. The G-dependence assumption can be relaxed, and this will give rise to an additional error term that measures the dependence (see [4, 5]). We will discuss this in detail in Chapter 3.

For triangles in $\mathcal{G}(n,p)$, we have

$$\begin{aligned} d_{\mathrm{tv}}\left(X,\mathrm{Poi}\left(\tfrac{1}{6}[n]_3p^3\right)\right) &\leqslant \min\left(\frac{1}{6}[n]_3p^6+\frac{1}{2}[n]_4p^5+\frac{1}{2}[n]_4p^6,\frac{\frac{1}{6}[n]_3p^6+\frac{1}{2}[n]_4p^5+\frac{1}{2}[n]_4p^6}{\frac{1}{6}[n]_3p^3}\right)\\ &\leqslant \min\left(\frac{1}{6}[n]_3p^6+\frac{1}{2}[n]_4p^5,p^3+6np^2\right). \end{aligned}\tag{1.3.18}$$

This bound is useful only when np^2 is small. If $np^2=o(1)$, then we have, from (1.3.18), that $d_{\mathrm{tv}}\left(X,\mathrm{Poi}\left(\frac{1}{6}[n]_3p^3\right)\right)=o(1)$, as the upper bound is by taking the minimum, and therefore,

$$\left|\mathbf{P}\left(X=0\right)-\exp\left(-\tfrac{1}{6}[n]_3p^3\right)\right|=o(1).$$

Note that Chen-Stein method is not strong enough to give accurate asymptotics unless $np=O\left(1\right)$, under which we have

$$\mathbf{P}\left(X=0\right)=\exp\left(-\tfrac{1}{6}[n]_3p^3+o(1)\right),$$

which is implied by Harris-Janson bound (1.3.6). For the case that p satisfies $p=o(n^{-1/2})$ and $p=\omega(n^{-1})$, Theorem 1.6 gives more accurate asymptotics.

Chapter 2

Cluster expansion, cumulants, non-existence probability, and asymptotic normality

2.1 Introduction

In this chapter, we consider the following three applications of cluster expansion.

- To obtain the probability of the non-existence of small hypergraphs in random hypergraphs.

  We express the probability that a binomial random hypergraph contains no copy of some given small hypergraphs in terms of clusters, by showing a cumulant series obtained by Mousset, Noever, Panagiotou and Samotij [77] approximating the same probability is equivalent to a truncated cluster expansion series. We use the formal cluster expansion after writing the probability of interest as a partition function. To illustrate the formula, we extend the result of the asymptotic probability that a random binomial hypergraph $H_r(n, p)$ is linear obtained by McKay and Tian in [74], for fixed r to the wider range of p, by a computation for $r = 3$ and $p = o(n^{-7/5})$.

- To use tree-graph bounds to estimate the cumulant, and using it to give asymptotic normality criteria.

  To establish the absolute convergence of the cluster expansion series, one difficulty is to estimate a summation over connected graphs of arbitrary sizes in the Ursell function (2.2.2). It turns out that the summation over connected graphs can be reduced to a summation over spanning trees, which can be easier to deal with. This leads to the approach using tree-graph bounds, which is utilized to bound the cumulants of graph-dependent variables, giving the Féray-Méliot-Nikeghbali bound in [26]. This further yields an asymptotic normality criterion that generalizes Janson's [54].

- To obtain the limiting distribution of the maximum of graph-dependent sequences.

  We use the Kotecký-Preiss criterion [61] to establish the absolute convergence of the cluster expansion under a locally dependent assumption. This gives the asymptotic distribution of maxima for m-dependent random variables. This extends the seminal work of Newell's [79] that originates the study of clustering of exceedances in extreme value theory. The new asymptotic formula also provides new insights into the extremal index.

2.2 Cluster expansion and the probability of non-existence

Here we introduce the standard cluster expansion setting, which is formulated in a way that is convenient for our applications. Recall the dependency graph defined in Definition 1.2, the cluster expansion method can be naturally combined with dependency graphs. Let $\{X_v\}_{v\in V(G)}$ be G-dependent random indicators and

$$X = \sum_{v\in V(G)} X_v.$$

In our application, each indicator indicates the occurrence of some combinatorial structure, and dependencies among indicators is characterized by a dependency graph.

By writing the probability of the non-existence of some combinatorial structure $\mathbf{P}\left(X=0\right)$ as a partition function, the cluster expansion then gives the formal expansion formula as a sum over clusters. The asymptotic value of $\mathbf{P}\left(X=0\right)$ can be estimated by truncating this infinite cluster expansion series. This method is inspired by [88], in which they also treat $\mathbf{P}\left(X=0\right)$ as a partition function and investigate the connections between cluster expansion and the Lovász local lemma, giving a lower bound for $\mathbf{P}\left(X=0\right)$.

Let the set of all connected spanning subgraphs of G be

$$\mathrm{CSpan}(G) := \{(V(G), E) : E \subseteq E(G), c((V(G), E)) = 1\},$$

where c denotes the number of connected components. Then for every graph $H \in \mathrm{CSpan}(G)$, we have $V(H) = V(G)$, $E(H) \subseteq E(G)$, and $c(H) = 1$. The standard cluster expansion gives the formal cluster expansion

$$\log \mathbf{P}\left(X=0\right) = \sum_{\gamma\in\Gamma(G)} \frac{\phi(\gamma)}{|\gamma|!}(-1)^{\|\gamma\|} \prod_{C\in\gamma} \mu\left(C\right), \tag{2.2.1}$$

with *Ursell function* $\phi : \Gamma(G) \to \mathbb{R}$ defined by

$$\phi(\gamma) = \sum_{H\in\mathrm{CSpan}(\mathbb{G}(\gamma))} (-1)^{e_H}, \tag{2.2.2}$$

where e_H denotes the number of edges of the graph H, and $\mathbb{G}$ is defined in Section 1.2.1. Note that if the cluster γ contains one single polymer $C \in \mathcal{C}(G)$, then $\phi(\gamma) = 1$, since $\mathbb{G}(C) = K_1$.

For completeness, we include a simple derivation of (2.2.1), following the routine cluster expansion derivation procedure (see, for example, [88, Section 2.2] or [30, Proposition 5.3.]). First, the inclusion-exclusion formula gives

$$\mathbf{P}\left(X=0\right) = \sum_{S\subseteq V(G)} (-1)^{|S|}\mu\left(S\right). \tag{2.2.3}$$

Let $\mathcal{G}_c$ be a graph on vertex set $\mathcal{C}(G)$ such that for all distinct $C_i, C_j \in \mathcal{C}(G)$, if $C_i \sim C_j$, then $\{C_i, C_j\} \in E(\mathcal{G}_c)$. Next, we utilize the factorisation property as shown in (1.2.2) to prove that the right-hand side of (2.2.3) can be written as some partition function of the hard-core model, more specifically, as a summation over independent sets of graph $\mathcal{G}_c$,

$$\mathbf{P}\left(X=0\right)=\sum_{U\in\mathcal{I}(\mathcal{G}_c)}\prod_{C\in U}(-1)^{|C|}\mu\left(C\right). \tag{2.2.4}$$

Recall that $\mathcal{I}(G)$ denotes the set of all independent sets for every graph G.

For every $S\subseteq V(G)$ such that $S\in\mathcal{C}(G)$, we have $\{S\}\in\mathcal{I}(\mathcal{G}_c)$. For every $S\subseteq V(G)$ such that $S\notin\mathcal{C}(G)$, we have that S induces a union of pairwise non-adjacent maximal connected subgraphs, that is, there exists a unique set of polymers $U\in\mathcal{I}(\mathcal{G}_c)$ such that $S=\cup_{C\in U}C$, and $C_i\not\sim C_j$ for all pairs of distinct $C_i,C_j\in U$. The factorisation property (1.2.2) then gives

$$(-1)^{|S|}\mu\left(S\right)=\prod_{C\in S}(-1)^{|C|}\mu(C). \tag{2.2.5}$$

Conversely, for every $U\in\mathcal{I}(\mathcal{G}_c)$, we have $\cup_{C\in U}C\subseteq V(G)$, thus U determines S uniquely; combining with (2.2.5), it follows that (2.2.3) and (2.2.4) are equivalent.

Now we derive the formal cluster expansion. Recall that $\binom{S}{i}$ denotes the family of i-sets of S. From (2.2.4), we have

$$\begin{aligned}\mathbf{P}\left(X=0\right)&=\sum_{U\subseteq\mathcal{C}(G)}\prod_{C\in U}(-1)^{|C|}\mu(C)\prod_{\{C_i,C_j\}\in\binom{U}{2}}\mathbf{1}_{\{C_i\not\sim C_j\}}\\&=1+\sum_{n\geqslant1}\frac{1}{n!}\sum_{(C_1,\dots,C_n)\in\mathcal{C}(G)^n}\prod_{i\in[n]}(-1)^{|C_i|}\mu\left(C_i\right)\prod_{1\leqslant i<j\leqslant n}\mathbf{1}_{\{C_i\not\sim C_j\}}.\end{aligned}$$

Note that a simple expansion yields

$$\prod_{1\leqslant i<j\leqslant n}\mathbf{1}_{\{C_i\not\sim C_j\}}=\prod_{1\leqslant i<j\leqslant n}\left(1-\mathbf{1}_{\{C_i\sim C_j\}}\right)=\sum_{H\in\mathfrak{G}_n}(-1)^{e_H}\prod_{\{i,j\}\in E(H)}\mathbf{1}_{\{C_i\sim C_j\}},$$

where $\mathfrak{G}_n$ denotes the set of all graphs on n vertices. Then formally, we obtain

$$\mathbf{P}\left(X=0\right)=1+\sum_{n\geqslant1}\frac{1}{n!}\sum_{H\in\mathfrak{G}_n}W(H),$$

where

$$W(H)=\sum_{(C_1,\dots,C_{v_H})\in\mathcal{C}(G)^{v_H}}(-1)^{e_H}\prod_{\{i,j\}\in E(H)}\mathbf{1}_{\{C_i\sim C_j\}}\prod_{k\in[v_H]}(-1)^{|C_k|}\mu\left(C_k\right),$$

and $W(H)$ satisfies

(a1) $W(H)=W(H')$ whenever H and H' are isomorphic $H\cong H'$, that is, differ only by vertices relabelling;

(a2) $W(H)=W(H_1)W(H_2)$ whenever H is isomorphic to the disjoint union of H_1 and H_2.

Let $\mathfrak{C}_n$ be the set of all connected graphs on n vertices. Via the exponential formula [96, Corollary

5.1.6], we reduce the sum over the set of all graphs to the set of all connected graphs

$$\log \mathbf{P}\left(X=0\right)=\sum_{n\geqslant 1}\frac{1}{n!}\sum_{H\in\mathfrak{C}_n}W(H)=\sum_{\gamma\in\Gamma(G)}\frac{1}{|\gamma|!}\sum_{H\in\mathrm{CSpan}(\mathbb{G}(\gamma))}(-1)^{e_H+\|\gamma\|}\prod_{C\in\gamma}\mu\left(C\right),$$

where $\Gamma(G)$ denotes the set of all clusters of G. A similar derivation of the cluster expansion utilizing the exponential formula also appears in [30, Proposition 5.3]. Then (2.2.1) follows.

Remark 2.1. *(r1) The cluster expansion is essentially the multivariate Taylor series for $\log \mathbf{I}_H(\boldsymbol{p})$ in variables $\{p_v\}_{v\in V(H)}$ around $\mathbf{0}$. Let $\boldsymbol{\mu}=((-1)^{|C|}\mu(C))_{C\in\mathcal{C}(G)}$. Then (2.2.4) can be regarded as the partition function $\mathbf{I}_{\mathcal{G}_c}(\boldsymbol{\mu})$ of the hard-core model on $\mathcal{G}_c$.*

(r2) For independent indicators $\{X_i\}_{i\in[n]}$, if $0\leqslant \mathbf{E}\left[X_i\right]<1$ for all $i\in[n]$, then we have the Taylor series of logarithmic function

$$\log \mathbf{P}\left(X=0\right)=\sum_{i\in[n]}\log\left(1-\mathbf{E}\left[X_i\right]\right)=-\sum_{i\in[n]}\sum_{j\geqslant 1}\frac{1}{j}\mathbf{E}\left[X_i\right]^j. \tag{2.2.6}$$

The empty graph $\overline{K}_n:=([n],\emptyset)$ is a valid dependency graph for this independent case. Since the polymers of $\overline{K}_n$ are all of size one containing a single vertex, and the clusters of $\overline{K}_n$ are all multisets containing multiple copies of the same vertex, then for independent indicators $\{X_v\}_{v\in[n]}$, expansion in (2.2.1) becomes

$$\log \mathbf{P}\left(X=0\right)=\sum_{\gamma\in\Gamma[\overline{K}_n]}\frac{\phi(\gamma)}{|\gamma|!}(-1)^{\|\gamma\|}\prod_{C\in\gamma}\mu\left(C\right)=\sum_{i\in[n]}\sum_{j\geqslant 1}\frac{1}{j!}\sum_{H\in CSpan(K_j)}(-1)^{e_H}(-1)^j\mathbf{E}\left[X_i\right]^j. \tag{2.2.7}$$

Comparing (2.2.6) and (2.2.7), it follows that

$$\sum_{H\in CSpan(K_n)}(-1)^{e_H}=(-1)^{n-1}(n-1)!, \tag{2.2.8}$$

which is well-known, see, for example, [88, Eq. (2.13)] or [97, Eq. (3.37)].

Remark 2.2. *The Mousset-Noever-Panagiotou-Samotij series (1.3.9) in Theorem 1.6 utilizes only clusters of disjoint polymers in the summation on the left, with clusters of overlapping polymers all absorbed in the error term on the right-hand side of (1.3.9), as the cumulants involve partitions that contain pairwise disjoint elements. It is worth mentioning that, by a careful inspection of its proof, the first error term comes from the contribution of clusters with intersecting polymers, and it is upper bounded by applying the lattice condition (1.3.7).*

2.3 Non-existence of small subhypergraphs

Given a family $\mathcal{F}$ of r-graphs, we study the probability that the binomial random hypergraph $H_r(n,p)$ is $\mathcal{F}$-*free*, that is, it simultaneously avoids all copies of all r-graphs in $\mathcal{F}$. Note that removing isomorphic duplicates from $\mathcal{F}$ does not affect the probability that we are interested in, we assume that the r-graphs in $\mathcal{F}$ are pairwise non-isomorphic. We also assume that no hypergraphs in $\mathcal{F}$ have isolated vertices.

2.3.1 The Mousset-Noever-Panagiotou-Samotij series and cluster expansion

Here we formulate the Mousset-Noever-Panagiotou-Samotij series in terms of a cluster expansion series. The complete r-graph on n vertices, denoted by $K_{n,r}$, is the hypergraph consisting of n vertices and all possible edges of size r, that is,

$$K_{n,r} = ([n], \{S \subseteq [n] : |S| = r\}). \tag{2.3.1}$$

For every $F \in \mathcal{F}$, let A^F be the set of all subgraphs of $K_{n,r}$ that are isomorphic to F. There are $[n]_{v_F}/|\mathrm{aut}(F)|$ such subgraphs, where $|\mathrm{aut}(F)|$ denotes the number of automorphisms of the hypergraph F. Let $A^{\mathcal{F}} = \cup_{F\in\mathcal{F}} A^F$. Then random variable

$$X_{\mathcal{F}} := \sum_{F\in A^{\mathcal{F}}} \mathbf{1}_{\{F\subset H_r(n,p)\}}$$

counts all copies of all forbidden r-graphs of $\mathcal{F}$ occurring in $H_r(n,p)$.

Next we define a dependency graph D with vertex set $A^{\mathcal{F}}$ such that for two distinct subgraphs $F_1, F_2 \in A^{\mathcal{F}}$, we have edge $(F_1, F_2) \in E(D)$ if and only if two subgraphs share edges, specifically,

$$D = \left(A^{\mathcal{F}}, \left\{\{F_1, F_2\} \in \binom{A^{\mathcal{F}}}{2} : E(F_1) \cap E(F_2) \neq \emptyset\right\}\right). \tag{2.3.2}$$

It is obvious that graph D is a dependency graph for random indicators $\{\mathbf{1}_{\{F\subset H_r(n,p)\}}\}_{F\in A^{\mathcal{F}}}$.

Using the above dependency graph for random indicators of the forbidden structures, we obtain the the probability that a binomial random r-uniform hypergraph is $\mathcal{F}$-free. In this setting, a polymer is a set of forbidden subgraphs whose induced subgraph in D is connected.

The set of all clusters of G with pairwise disjoint polymers is denoted by

$$\Gamma_\emptyset(G) = \{\gamma \in \Gamma(G) : C_i \cap C_j = \emptyset \text{ for any distinct } C_i, C_j \in \gamma\}.$$

Note that each element in $\Gamma_\emptyset(G)$ is a cluster whose elements form a partition of a polymer, since for every $\gamma \in \Gamma_\emptyset(G)$, polymers $\{C : C \in \gamma\}$ are disjoint and their union $\cup_{C\in\gamma} C \in \mathcal{C}(G)$. For every integer $k > 0$, denote the k-th term of the cluster expansion and the k-th truncated expansion with *disjoint* polymers as

$$L^{\emptyset}_{G,k} := \sum_{\gamma\in\Gamma_\emptyset(G):\|\gamma\|=k} \frac{\phi(\gamma)}{|\gamma|!}(-1)^{\|\gamma\|} \prod_{C\in\gamma} \mu(C) \quad \text{and} \quad T^{\emptyset}_{G,k} := \sum_{i\in[k]} L^{\emptyset}_{G,i}. \tag{2.3.3}$$

The density of a graph G is defined by $d(G) = e_G/v_G$, where v_G and e_G are the numbers of vertices and edges of G respectively. Another commonly used (see, for example, [86, 77]) density measure $m_\star(G)$ is defined by

$$m_\star(G) = \min_{H\subseteq G, e_H\geqslant 1} \frac{e_G - e_H}{v_G - v_H}. \tag{2.3.4}$$

Now we are ready to approximate $\mathbf{P}(X_{\mathcal{F}} = 0)$ using a truncated cluster expansion by reformulating the Mousset-Noever-Panagiotou-Samotij series in [77, Corollary 12].

Theorem 2.3. *Let $\mathcal{F}$ be a finite family of r-graphs and $p = p(n) \in (0,1)$ satisfy*

$$np^{m_\star(\mathcal{F})} = o\left(1\right) \qquad \text{and} \qquad np^{2d(\mathcal{F})} = o\left(1\right), \tag{2.3.5}$$

where

$$m_\star(\mathcal{F}) = \min_{G\in\mathcal{F}} m_\star(G) \qquad \text{and} \qquad d(\mathcal{F}) = \min_{G\in\mathcal{F}} d(G).$$

Then, for every integer $k > 0$, we have

$$\mathbf{P}\left(X_{\mathcal{F}} = 0\right) = \exp\left(T^{\emptyset}_{D,k} + O\left(\Delta_{k+1}(D)\right) + o\left(1\right)\right). \tag{2.3.6}$$

Moreover, if $np^{m_\star(\mathcal{F})} = n^{-\varepsilon}$ for some $\varepsilon > 0$, then there exists an integer $k = k(\varepsilon, \mathcal{F}) > 0$ such that $\Delta_{k+1}(D) = o\left(1\right)$.

The above theorem is a reformulation of [77, Corollary 12] in view of the following lemma.

Lemma 2.4. *Let $\{X_v\}_{v\in V(G)}$ be G-dependent random indicators and $k > 0$ be an integer. Then*

$$T^{\emptyset}_{G,k} = \sum_{C\in\mathcal{C}_{\leqslant k}(G)} (-1)^{|C|}\kappa(C). \tag{2.3.7}$$

We introduce an auxiliary lemma for its proof.

Lemma 2.5. *For any connected graph H, we have*

$$\sum_{\pi\in\Pi(V(H))}\sum_{G\in CSpan(K_{|\pi|})} (-1)^{e_G} \prod_{P\in\pi} \mathbf{1}_{\{P\in\mathcal{I}(H)\}} = \sum_{G\in CSpan(H)} (-1)^{e_G}. \tag{2.3.8}$$

To prove Lemma 2.5, we first introduce the chromatic polynomial. Given a graph H and a positive integer λ, a (proper) λ-colouring of H is a map $\Phi : V(H) \to [\lambda]$ such that $\Phi(u) \neq \Phi(v)$ for all $\{u,v\} \in E(H)$. The *chromatic polynomial* $P_H(\lambda)$ of H is the number of λ-colourings of H.

Given a graph H and a positive integer k, a partition containing k subsets $\{V_1, \ldots, V_k\}$ of $V(H)$ is called a k-independent partition of H if for every $i \in [k]$, we have $V_i \neq \emptyset$ and $V_i \in \mathcal{I}(H)$. Let $\alpha(H,k)$ count the k-independent partition of H. Then we have the chromatic polynomial in factorial form

$$P_H(\lambda) = \sum_{k=1}^{v_H} \alpha(H,k)[\lambda]_k, \tag{2.3.9}$$

(see, for example, [18, Theorem 1.4.1]). An equivalent formula for $P_H(\lambda)$ written as a polynomial in λ, known as the Whitney-Tutte-Fortuin-Kasteleyn representation (see, for example, [69, Eq. (A.11)] or [92, Eq. (1.2)]) is

$$P_H(\lambda) = \sum_{E\subseteq E(H)} (-1)^{|E|}\lambda^{c(E)}, \tag{2.3.10}$$

where $c(E) = c(V(H), E)$ counts the number of the connected components of subgraph $(V(H), E)$ for every edge set $E \subseteq E(H)$.

Proof of Lemma 2.5. By inspecting (2.3.10) and (2.2.2), one observes that the Ursell function is the linear term of the chromatic polynomial (this is also a well-known fact, see, for example, [1]). Then we have the right-hand side of (2.3.8)

$$\sum_{G\in \mathrm{CSpan}(H)}(-1)^{e_G}=\frac{\mathrm{d}P_H(\lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=0}=\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\sum_{k=1}^{v_H}\alpha(H,k)[\lambda]_k\right)\Big|_{\lambda=0}=\sum_{k=1}^{v_H}\alpha(H,k)(-1)^{k-1}(k-1)!. \quad (2.3.11)$$

Using the combinatorial identity obtained before in (2.2.8), the left hand side of (2.3.8) can be rewritten as

$$\sum_{\pi\in\Pi(V(H))}\sum_{G\in \mathrm{CSpan}(K_{|\pi|})}(-1)^{e_G}\prod_{P\in\pi}\mathbf{1}_{\{P\in\mathcal{I}(H)\}}=\sum_{\pi\in\Pi(V(H))}(-1)^{|\pi|-1}(|\pi|-1)!\prod_{P\in\pi}\mathbf{1}_{\{P\in\mathcal{I}(H)\}}. \quad (2.3.12)$$

Notice that the right-hand side of (2.3.12) is a sum of $|\pi|$-independent partition for any $\pi\in\Pi(V(H))$. Thus we have

$$\begin{aligned}\sum_{\pi\in\Pi(V(H))}(-1)^{|\pi|-1}(|\pi|-1)!\prod_{P\in\pi}\mathbf{1}_{\{P\in\mathcal{I}(H)\}}&=\sum_{k=1}^{v_H}\sum_{\pi\in\Pi(V(H)):|\pi|=k}(-1)^{k-1}(k-1)!\prod_{P\in\pi}\mathbf{1}_{\{P\in\mathcal{I}(H)\}}\\&=\sum_{k=1}^{v_H}\alpha(H,k)(-1)^{k-1}(k-1)!. \quad (2.3.13)\end{aligned}$$

Then combining (2.3.11) and (2.3.13), we complete the proof. □

Proof of Lemma 2.4. From the cluster expansion, we have

$$\begin{aligned}T^{\emptyset}_{G,k}=\sum_{i\in[k]}L^{\emptyset}_{G,i}&=\sum_{\gamma\in\Gamma_{\emptyset}(G):\|\gamma\|\in[k]}\frac{\phi(\gamma)}{|\gamma|!}\prod_{C\in\gamma}(-1)^{|C|}\mu(C)\\&=\sum_{\substack{(C_1,\ldots,C_n)\in\Gamma_{\emptyset}(G)\\ \sum_{i\in[n]}|C_i|\in[k]}}\frac{1}{n!}\sum_{H\in \mathrm{CSpan}(\mathbb{G}(C_1,\ldots,C_n))}(-1)^{e_H}\prod_{i\in[n]}(-1)^{|C_i|}\mu(C_i)\\&=\sum_{\substack{\{C_1,\ldots,C_n\}\in\Gamma_{\emptyset}(G)\\ \sum_{i\in[n]}|C_i|\in[k]}}\sum_{H\in \mathrm{CSpan}(\mathbb{G}(C_1,\ldots,C_n))}(-1)^{e_H}\prod_{i\in[n]}(-1)^{|C_i|}\mu(C_i), \quad (2.3.14)\end{aligned}$$

where the first summation in the last line is an abuse of notation, and denotes the summation over (unordered) sets of polymers. From the definition of joint cumulants, we get

$$\begin{aligned}\sum_{C\in\mathcal{C}_{\leqslant k}(G)}(-1)^{|C|}\kappa(C)&=\sum_{C\in\mathcal{C}_{\leqslant k}(G)}(-1)^{|C|}\sum_{\pi\in\Pi(C)}(-1)^{|\pi|-1}(|\pi|-1)!\prod_{P\in\pi}\mu(P)\\&=\sum_{C\in\mathcal{C}_{\leqslant k}(G)}\sum_{\pi\in\Pi(C)}(-1)^{|\pi|-1}(|\pi|-1)!\prod_{P\in\pi}(-1)^{|P|}\mu(P).\end{aligned}$$

Combining with identity in (2.2.8), it follows that

$$\sum_{C\in\mathcal{C}_{\leqslant k}(G)}(-1)^{|C|}\kappa(C)=\sum_{C\in\mathcal{C}_{\leqslant k}(G)}\sum_{\pi\in\Pi(C)}\sum_{H\in \mathrm{CSpan}(K_{|\pi|})}(-1)^{e_H}\prod_{P\in\pi}(-1)^{|P|}\mu(P).$$

Fix an arbitrary polymer $C\in\mathcal{C}_{\leqslant k}(G)$, for every partition $\pi=\{P_1,\ldots,P_m\}\in\Pi(C)$, by factorizing into pairwise non-adjacent maximal connected subgraphs with vertex sets $C_1, \dots, C_n$ respectively, there exists a unique finest partition consisting of only polymers $\pi' = \pi'(\pi) = \{C_1, \dots, C_n\} \in \Pi(C)$ such that

(p1) $n \geqslant m$,

(p2) for all $i \in [n]$, we have $C_i \in \mathcal{C}(G)$,

(p3) $(C_1, \dots, C_n) \in \Gamma_{\emptyset}(G)$, and

(p4) $\prod_{P\in\pi} \mu(P) = \prod_{C\in\pi'} \mu(C)$.

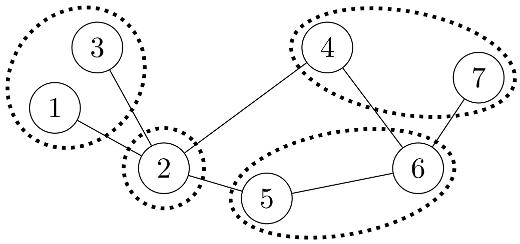


Figure 2.1: A polymer of size seven with a partition $\{\{1,3\},\{2\},\{4,7\},\{5,6\}\}$ and the corresponding polymer partition $\{\{1\},\{3\},\{2\},\{4\},\{7\},\{5,6\}\}$ such that $\mu(1,3)\,\mu(2)\,\mu(4,7)\,\mu(5,6) = \mu(1)\,\mu(2)\,\mu(3)\,\mu(4)\,\mu(5,6)\,\mu(7)$.

Then, we have

$$\begin{aligned}(-1)^{|C|}\kappa(C) &= \sum_{\pi\in\Pi(C)} \sum_{H\in\mathrm{CSpan}(K_{|\pi|})} (-1)^{e_H} \prod_{C\in\pi'(\pi)} (-1)^{|C|}\mu(C) \\ &= \sum_{\substack{\pi\in\Pi(C)\\ \pi'(\pi)=\{C_1,\dots,C_n\}}} \sum_{H\in\mathrm{CSpan}(K_{|\pi|})} (-1)^{e_H} \prod_{i\in[n]} (-1)^{|C_i|}\mu(C_i)\,.\end{aligned}$$

Since $\Gamma_{\emptyset}(G)$ is the set of all clusters of G with pairwise disjoint polymers, we then rearrange the partitions according to their corresponding polymer partitions and have that

$$\{\pi\in\Pi(C) : C\in\mathcal{C}_{\leqslant k}(G)\} = \left\{\pi\in\Pi' : \{C_1,\dots,C_n\}\in\Gamma_{\emptyset}(G) : \sum_{i\in[n]} |C_i| \in [k]\right\},$$

where

$$\Pi' := \left\{\pi\in\Pi\left(\cup_{i\in[n]}C_i\right) : \pi'(\pi) = (C_1,\dots,C_n)\right\} \tag{2.3.15}$$

denotes the set of partitions of $\cup_{i\in[n]}C_i$ for a given set of polymers $\{C_1,\dots,C_n\}\in\Gamma_{\emptyset}(G)$. Hence

$$\sum_{C\in\mathcal{C}_{\leqslant k}(G)} (-1)^{|C|}\kappa(C) = \sum_{\substack{\{C_1,\dots,C_n\}\in\Gamma_{\emptyset}(G)\\ \sum_{i\in[n]}|C_i|\in[k]}} \sum_{\pi\in\Pi'} \sum_{H\in\mathrm{CSpan}(K_{|\pi|})} (-1)^{e_H} \prod_{i\in[n]} (-1)^{|C_i|}\mu(C_i)\,. \tag{2.3.16}$$

Note that we have

$$\prod_{P\in\pi} (-1)^{|P|}\mu(P) = \prod_{i\in[n]} (-1)^{|C_i|}\mu(C_i)$$

if and only if every element of the partition $\pi \in \Pi'$ is an independent set of $\mathbb{G}(C_1, \dots, C_n)$. Then by comparing (2.3.14) and (2.3.16), it suffices to show that for all $(C_1, \dots, C_n) \in \Gamma_{\emptyset}(G)$,

$$\sum_{\pi \in \Pi'} \sum_{G \in \mathrm{CSpan}(K_{|\pi|})} (-1)^{e_G} \prod_{P \in \pi} \mathbf{1}_{\{P \in \mathcal{I}(\mathbb{G}(C_1,\dots,C_n))\}} = \sum_{H \in \mathrm{CSpan}(\mathbb{G}(C_1,\dots,C_n))} (-1)^{e_H},$$

which follows from Lemma 2.5. $\square$

2.3.2 Computation of the asymptotic probability that $\mathcal{G}(n, p)$ is triangle-free

The goal of this section is to compute the terms in Theorem 1.7 explicitly using the series in Theorem 2.3 in terms of clusters.

(c1) Clusters γ such that $|\gamma| = 1$.

In this case, we have a single polymer in the cluster, and the contributing polymers are listed as follows.

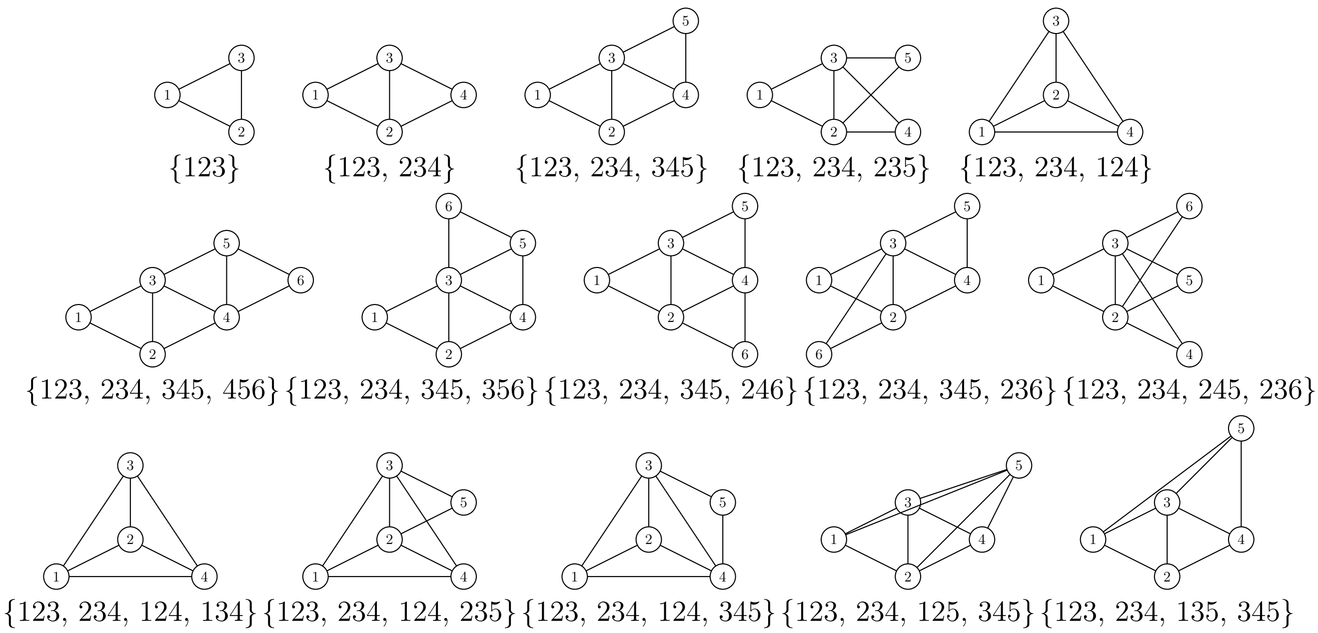


Figure 2.2: The set below each diagram indicates the triangles that correspond to dependent indicators.

Therefore we have

$$\sum_{C \in \mathcal{C}_1(G)} (-1)^{|C|} \mu(C) = -\frac{[n]_3 p^3}{6} + \frac{[n]_4 p^5}{4} - \left(\frac{[n]_5 p^7}{2} + \frac{[n]_5 p^7}{12} + \frac{[n]_4 p^6}{6} \right)$$
$$+ \frac{[n]_6 p^9}{2} + \frac{[n]_6 p^9}{2} + \frac{[n]_6 p^9}{6} + \frac{[n]_6 p^9}{2} + \frac{[n]_6 p^9}{48} + \frac{[n]_4 p^6}{24} + O\left(n^5 p^8\right).$$

(c2) Clusters γ such that $|\gamma| = 2$.

In this case, we have two polymers. In order to form clusters, two polymers share at least one edge, and we have the following contributing clusters for $p = o\left(n^{-7/5}\right)$.

For any two edge-sharing polymers $(C_i, C_j) \in \mathcal{C}(G)^2$, we have $\mathbb{G}(C_i, C_j) = K_2$, thus $\phi(C_i, C_j) =$

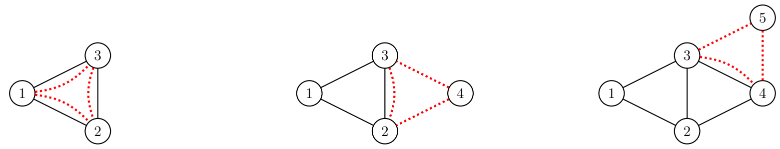


Figure 2.3: The set below each diagram indicates a cluster containing two polymers. Note that the first type corresponds to a multiset, which is not included in (2.3.3), and its contribution is absorbed in the first error term on the right-hand side of (1.3.9).

-1. Then the contribution of pairs of clusters is

$$\frac{1}{2}\sum_{(C_i,C_j)\in\mathcal{C}(G)^2}(-1)^{1+|C_i|+|C_j|}\mu\left(C_i\right)\mu\left(C_j\right)=-\frac{[n]_3p^6}{6}-\frac{[n]_4p^6}{4}+O\left(n^5p^8\right)=-\frac{[n]_4p^6}{4}+o\left(1\right).$$

It can be shown that all other configurations contribute negligibly as in [77, Corollary 15]. Adding up all the above contributions gives the final asymptotic formula (1.3.10) for $p=o\left(n^{-7/11}\right)$.

2.3.3 Linearity of binomial random hypergraphs

In this section, we obtain the asymptotic probability of a random hypergraph being linear. Linear hypergraphs have been well studied in many contexts (sometimes under the name 'simple hypergraphs'). A hypergraph is linear if every pair of hyperedges intersects in at most one vertex. We accordingly define a set $\mathcal{F}$ of 'forbidden' hypergraphs containing all r-graphs having two distinct hyperedges e_1 and e_2 and vertex set $e_1\cup e_2$, such that $2\leqslant|e_1\cap e_2|<r$. Then the probability that a random hypergraph is linear equals the probability of avoiding all copies of all 'forbidden' hypergraphs in $\mathcal{F}$.

Here we study the probability of a random hypergraphs $H_r(n,p)$ being linear, and improve the following result by McKay and Tian [74] by giving more accurate asymptotics of the probability. Let $\mathcal{L}_r(n)$ be the set of all linear r-uniform hypergraphs with n vertices.

Theorem 2.6 ([74, Theorem 1.2])**.** *Let $r=r(n)\geqslant 3$. If $p\binom{n}{r}=O\left(r^{-2}n\right)$, then*

$$\mathbf{P}\left(H_r(n,p)\in\mathcal{L}_r(n)\right)=\exp\left(-\frac{[r]_2^2}{4n^2}\binom{n}{r}^2p^2+O\left(\frac{r^6}{n^3}\binom{n}{r}^2p^2\right)\right).$$

If $r^{-2}n\leqslant p\binom{n}{r}=o\left(r^{-3}n^{3/2}\right)$, then

$$\mathbf{P}\left(H_r(n,p)\in\mathcal{L}_r(n)\right)=\exp\left(-\frac{[r]_2^2}{4n^2}\binom{n}{r}^2p^2+\frac{(3r-5)[r]_2^3}{6n^4}\binom{n}{r}^3p^3+O\left(\frac{\log^3(r^{-2}n)}{\sqrt{\binom{n}{r}p}}+\frac{r^6}{n^3}\binom{n}{r}^2p^2\right)\right).$$

For random 3-uniform hypergraphs, the above theorem gives that if $p=o\left(n^{-3/2}\right)$, then

$$\mathbf{P}\left(H_3(n,p)\in\mathcal{L}_3(n)\right)=\exp\left(-\frac{1}{4}n^4p^2+\frac{2}{3}n^5p^3+o\left(1\right)\right).\qquad(2.3.17)$$

The probability of a random hypergraph being linear is equal to the probability of the non-existence of hyperedge pairs intersecting in more than one vertex. Recall from above that the set of 'forbidden' hypergraphs $\mathcal{F}$ contains all r-graphs $(e_1\cup e_2,\{e_1,e_2\})$ on vertex set $e_1\cup e_2$ such that $2\leqslant|e_1\cap e_2|<r$,

that is,

$$\mathcal{F} := \bigcup_{2\leqslant t\leqslant r-1} \Big\{(e_1\cup e_2, \{e_1,e_2\}) : |e_1| = |e_2| = r, |e_1\cap e_2| = t\Big\}. \tag{2.3.18}$$

First, recall the definitions of $L^{\emptyset}_{G,k}$ and $T^{\emptyset}_{G,k}$ in (2.3.3).

Theorem 2.7. *Let $r = r(n) \geqslant 3$. If $p = o\left(n^{2-r}\right)$, then for every integer $k > 0$,*

$$\mathbf{P}\left(H_r(n,p) \in \mathcal{L}_r(n)\right) = \exp\Big(T^{\emptyset}_{D,k} + O\left(\Delta_{k+1}(D)\right) + o\left(1\right)\Big), \tag{2.3.19}$$

where D is the dependency graph for the indicators of forbidden r-graphs defined by (2.3.2), and $\Delta_i(D)$ denotes the sum of joint moments over polymers of size i in graph D, that is,

$$\Delta_i(D) = \sum_{C\in\mathcal{C}_i(D)} \mu(C).$$

Moreover, for any $\varepsilon > 0$, if $p = o\left(n^{2-r-\varepsilon}\right)$, then there exists an integer $k = k(\varepsilon) > 0$ such that

$$\mathbf{P}\left(H_r(n,p) \in \mathcal{L}_r(n)\right) = \exp\Big(T^{\emptyset}_{D,k} + o\left(1\right)\Big). \tag{2.3.20}$$

Theorem 2.7 gives the asymptotics of the probability of a random hypergraph being linear. We next consider a specific example, by restricting to the 3-uniform hypergraphs case, and computing only the first few terms of the series explicitly for illustration purposes. This extends the formula for the asymptotic probability of linearity for $H_3(n,p)$ given by McKay and Tian (2.3.17).

Corollary 2.8. *If $p = o\left(n^{-7/5}\right)$, then*

$$\mathbf{P}\left(H_3(n,p) \in \mathcal{L}_3(n)\right) = \exp\Big(-\frac{1}{4}n^4p^2 + \frac{2}{3}n^5p^3 - \frac{55}{24}n^6p^4 + \frac{3}{2}n^3p^2 + o\left(1\right)\Big). \tag{2.3.21}$$

Proof of Theorem 2.7. Recall that the set of forbidden r-graphs is defined by $\mathcal{F}$ in (2.3.18). Since each r-set of $[n]$ is an edge independently with probability p in $H_r(n,p)$, we have for distinct subgraphs $F_1, F_2 \in A^{\mathcal{F}}$, indicators $\mathbf{1}_{\{F_1\subset H_r(n,p)\}}$ and $\mathbf{1}_{\{F_2\subset H_r(n,p)\}}$ are dependent if $E(F_1)\cap E(F_2) \neq \emptyset$. Additionally, the graph D defined by (2.3.2) is a dependency graph for random indicators $\mathbf{1}_{\{F\subset H_r(n,p)\}}$ for $F \in A^{\mathcal{F}}$.

Next, we verify the assumptions in (2.3.5). Since

$$m_\star(\mathcal{F}) = \min_{G\in\mathcal{F}} \min_{H\subseteq G, e_H\geqslant 1} \frac{e_G - e_H}{v_G - v_H} = \frac{1}{\max_{G\in\mathcal{F}} \max_{H\subseteq G, e_H\geqslant 1}(v_G - v_H)} = \frac{1}{r-2}$$

and

$$d(\mathcal{F}) = \min_{G\in\mathcal{F}} d(G) = \min_{G\in\mathcal{F}} \frac{e_G}{v_G} = \frac{2}{\max_{G\in\mathcal{F}} v_G} = \frac{1}{r-1},$$

then $2d(\mathcal{F}) \geqslant m_\star(\mathcal{F})$ for all $r \geqslant 3$. Therefore, we have that Theorem 2.7 follows from Theorem 2.3. ☐

2.3.4 Computation of the asymptotic probability that $H_3(n,p)$ is linear

The goal of this section is to compute the terms in Theorem 2.7 explicitly to prove Corollary 2.8.

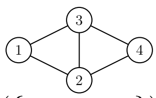
Proof of Corollary 2.8. For 3-uniform hypergraphs, the forbidden hypergraph is on four vertices with a pair of 3-sets sharing two vertices, we call it a *link*. Then for random indicators of links, we construct the dependency graph D following (2.3.2), such that two links are adjacent if and only if they share one hyperedge. A polymer $C \in \mathcal{C}(D)$ of size k is a set of links $\{F_1, \ldots, F_k\}$ whose induced subgraph in D is connected.

We first enumerate all contributing non-isomorphic types of clusters, and compute value $\phi(\gamma)(-1)^{\|\gamma\|} \prod_{C\in\gamma} \mu(C)/|\gamma|!$ for each cluster type γ. Then we multiply each value with the size of the respective isomorphism class. More precisely, noting a cluster is a set of link sets, an isomorphism between two clusters γ_1, γ_2 is a bijection between their vertices (the union of vertices in all links): $\cup_{C\in\gamma_1} \cup_{F\in C} V(F) \to \cup_{C\in\gamma_2} \cup_{F\in C} V(F)$, which induces a bijection from the hyperedges of γ_1 to the hyperedges of γ_2, and a bijection from the polymers of γ_1 to the polymers of γ_2. An automorphism of a cluster is an isomorphism to itself. For each cluster $\gamma \in \Gamma(D)$, we consider the distinct copies of γ in the complete r-graph on n vertices by choosing all the vertices in $\cup_{C\in\gamma_1} \cup_{F\in C} V(F)$ from $[n]$ (ordered selections without repetition), and every element of $\Gamma(D)$ isomorphic to γ is counted once for every automorphism of γ.

Now, we compute the terms in (2.3.20) for $p = o\left(n^{-7/5}\right)$ explicitly.

(c1) Clusters γ such that $\|\gamma\| = 1$.

There is only one cluster type, a single forbidden link, namely, a hypergraph with two hyperedges intersecting in two vertices.



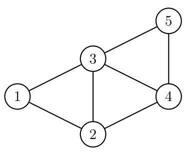
$(\{123{+}234\})$

Thus, we have that

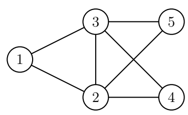
$$L^{\emptyset}_{D,1} = -\sum_{C\in\mathcal{C}(D):|C|=1} \mu(C) = -\frac{[n]_4 p^2}{4} = -\frac{1}{4}n^4p^2 + \frac{3}{2}n^3p^2 + o(1).$$

(c2) Clusters γ such that $\|\gamma\| = 2$.

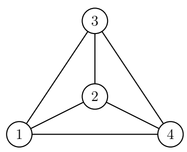
There are two cluster types: one polymer of size two, and two polymers of size one, namely, one polymer consisting of two edge-sharing forbidden links, or two edge-sharing polymers with each being a single forbidden link. Also note that for any two (not necessarily distinct) polymers $(C_i, C_j) \in \mathcal{C}(D)^2$ such that $C_i \sim C_j$, we have that $\mathbb{G}(C_i, C_j) = K_2$, thus $\phi(C_i, C_j) = -1$.



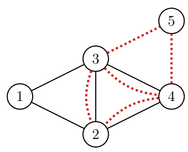
$(\{123{+}234,\ 234{+}345\})$



$(\{123{+}234,\ 123{+}235\})$



$(\{123{+}234,\ 123{+}124\})$



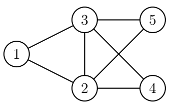
$(\{123+234\}, \{234+345\})$

Therefore, we get

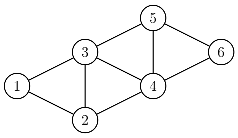
$$
\begin{aligned}
L^{\emptyset}_{D,2} &= \sum_{\gamma\in\Gamma_{\emptyset}(D):\|\gamma\|=2} \frac{\phi(\gamma)}{|\gamma|!}(-1)^{\|\gamma\|}\prod_{C\in\gamma}\mu\left(C\right) \\
&= \sum_{C\in\mathcal{C}_2(D)}(-1)^{|C|}\mu\left(C\right) + \sum_{\substack{(C_1,C_2)\in\Gamma_{\emptyset}(D)\\ |C_1|=|C_2|=1}} -\frac{1}{2}(-1)^2\mu\left(C_1\right)\mu\left(C_2\right) \\
&= \frac{[n]_5p^3}{2} + \frac{[n]_5p^3}{4} + \frac{[n]_4p^3}{2} - \frac{[n]_5p^4}{4} = \frac{3}{4}n^5p^3 + o\left(1\right).
\end{aligned}
$$

(c3) Clusters γ such that $\|\gamma\| = 3$.

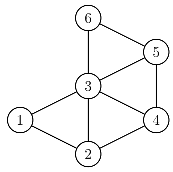
We only focus on one cluster type: one polymer of size three, namely, one polymer consisting of three edge-sharing forbidden links, since if the cluster is formed by more then one polymer, then it must be extended from clusters γ such that $\|\gamma\| = 2$ and more than one polymer, which are already asymptotically negligible.



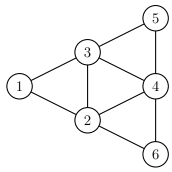
$(\{123+234,$
$123+235,$
$234+235\})$



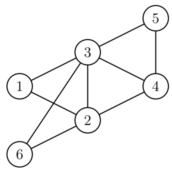
$(\{123+234,$
$234+345,$
$345+456\})$



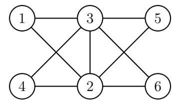
$(\{123+234,$
$234+345,$
$345+356\})$



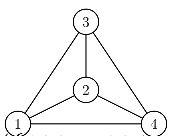
$(\{123+234,$
$234+345,$
$234+246\})$



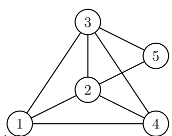
$(\{123+234,$
$234+345,$
$234+236\}),$
$(\{123+234,$
$234+345,$
$123+236\})$



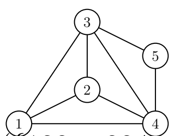
$(\{123+234,$
$123+235,$
$123+236\}),$
$(\{123+234,$
$234+235,$
$235+236\})$



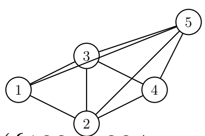
$(\{123+234,$
$123+124,$
$234+124\})$



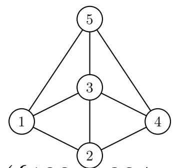
$(\{123+234,$
$123+124,$
$123+235\}),$
$(\{123+234,$
$123+124,$
$234+235\})$



$(\{123+234,$
$123+124,$
$234+345\})$



$(\{123+234,$
$123+125,$
$234+345\})$



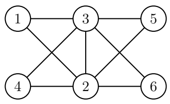
$(\{123+234,$
$123+135,$
$234+345\})$

Hence, we have that

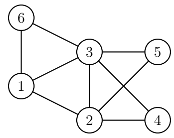
$$
\begin{aligned}
L_{D,3}^{\emptyset} &= \sum_{\gamma\in\Gamma_{\emptyset}(D):\|\gamma\|=3} \frac{\phi(\gamma)}{|\gamma|!}(-1)^{\|\gamma\|}\prod_{C\in\gamma}\mu\left(C\right)\\
&= \sum_{C\in\mathcal{C}_3(D)} (-1)^{|C|}\mu\left(C\right) + O\left(n^4p^3\right) + O\left(n^5p^4\right)\\
&= -\frac{[n]_5p^3}{2\times 3!} - \frac{[n]_6p^4}{2} - \frac{[n]_6p^4}{2} - \frac{[n]_6p^4}{3!} - \frac{[n]_6p^4}{2} - [n]_6p^4 - \frac{[n]_6p^4}{2\times 3!} - \frac{[n]_6p^4}{2\times 2} + o(1),
\end{aligned}
$$

where the last row of the types of polymers are of contribution $O\left(n^4p^3\right) = o(1)$.

(c4) Clusters γ such that $\|\gamma\| = 4$.
As before, we only focus on one cluster type: one polymer of size four, since clusters with more than one polymer contribute negligibly.



$(\{123+234, 123+235, 123+236, 234+235\})$,
$(\{123+234, 234+235, 235+236, 123+236\})$

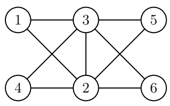


$(\{123+234, 123+235,$
$234+235, 123+136\})$

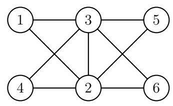
We then have

$$
L_{D,4}^{\emptyset} = \sum_{\gamma\in\Gamma_{\emptyset}(D):\|\gamma\|=4} \frac{\phi(\gamma)}{|\gamma|!}(-1)^{\|\gamma\|}\prod_{C\in\gamma}\mu\left(C\right) = \frac{[n]_6p^4}{2\times 2} + \frac{[n]_6p^4}{2\times 8} + \frac{[n]_6p^4}{2} + o(1).
$$

(c5) Clusters γ such that $\|\gamma\| \in \{5,6\}$.



$(\{123+234, 123+235,$
$123+236, 234+235, 234+236\})$



$(\{123+234, 123+235, 123+236,$
$234+235, 234+236, 235+236\})$

Then we have

$$
L_{D,5}^{\emptyset} + L_{D,6}^{\emptyset} = \sum_{\gamma\in\Gamma_{\emptyset}(D):\|\gamma\|\in\{5,6\}} \frac{\phi(\gamma)}{|\gamma|!}(-1)^{\|\gamma\|}\prod_{C\in\gamma}\mu\left(C\right) = -\frac{[n]_6p^4}{2\times 2\times 2} + \frac{[n]_6p^4}{2\times 4!} + o(1).
$$

Since there is a finite number of types of polymers with size seven, we thus have $\Delta_7(D) = o\left(1\right)$. Hence, we ignore the remaining terms by (2.3.19). Adding up the contributing terms for $p = o\left(n^{-7/5}\right)$ gives the asymptotic probability of $H_3(n,p)$ being linear in Corollary 5.3. $\square$

2.4 Tree-graph bounds on cumulants and asymptotic normality

In this section, we use the tree-graph approach to bound the cumulants, and to obtain an asymptotic normality criterion.

2.4.1 Asymptotic normality via cumulants

Cumulants can be utilized the obtain the asymptotic normality. It suffices to simply show that there are only two cumulant terms that do not vanish asymptotically using the following result by Janson [54].

Theorem 2.9 ([54]). *Let $X_1, X_2, \ldots$ be a sequence of random variables such that, as $n \to \infty$,*

$$\begin{aligned} &\kappa_1(X_n) = \mathbf{E}[X_n] \to \mu, \\ &\kappa_2(X_n) = \mathrm{var}(X_n) \to \sigma^2, \\ &\kappa_j(X_n) \to 0 \end{aligned}$$

for every $j \geqslant 3$, where $-\infty < \mu < \infty$ and $\sigma^2 \geqslant 0$. Then as $n \to \infty$, we have

$$X_n \to \mathcal{N}(\mu, \sigma^2).$$

Moreover, all moments of X_n converge to the corresponding moments of $\mathcal{N}(\mu, \sigma^2)$.

Note that there is no probability distribution that has only non-zero low-order cumulants (orders 3 to $m-1$ for $m > 3$), that is, no probability distribution satisfies $\kappa_1 \neq 0, \ldots, \kappa_{m-1} \neq 0$ and $\kappa_j = 0$ for $j \geqslant m$. In other words, the cumulant generating function cannot be a finite polynomial of degree greater than 2.

2.4.2 Tree-graph bounds and Penrose identity

The difficulty with the convergence of cluster expansions is to estimate a summation over connected graphs of arbitrary sizes. It turns out that the sum over connected graphs can be reduced to a sum over spanning trees that can be smaller. This approach has been introduced by Penrose [82] and are widely used.

Various partition schemes are often used to obtain estimates of the summation of connected graphs in terms of the summation of trees. While his original argument involved a particular partition scheme, it works equally well for any other choice, as emphasized by Scott and Sokal [88]. Next, we give a brief description of the approach based on [88, Section 2.2].

Fix a graph G, and recall that $\mathrm{CSpan}(G)$ denotes the set of all connected spanning subgraphs of G. Let $\mathcal{ST}(G)$ denote the family of spanning trees of G. Note that $\mathcal{ST}(G) \subseteq \mathrm{CSpan}(G)$. The set $\mathrm{CSpan}(G)$ admits a partial ordering by edge inclusion, specifically,

$$G \prec \widetilde{G} \iff E(G) \subset E(\widetilde{G}) .$$

If $G \prec \widetilde{G}$, we define a set of connected spanning subgraphs of G by

$$[G, \widetilde{G}] := \left\{ \widehat{G} \in \mathrm{CSpan}(G) : G \prec \widehat{G} \prec \widetilde{G} \right\}.$$

We call a *partition scheme* for the family $\mathrm{CSpan}(G)$ to any map $R : \mathcal{ST}(G) \to \mathrm{CSpan}(G)$ such that

(i) $E\big(R(T)\big) \supset E(T)$, and

(ii) $\mathrm{CSpan}(G)$ is a disjoint union of the sets $[T, R(T)]$ for $T \in \mathcal{ST}(G)$, that is,

$$\mathrm{CSpan}(G) = \bigcup_{T \in \mathcal{ST}(G)} [T, R(T)].$$

A number of such partition schemes are now available (see references in [88, Section 2.2]). The one proposed by Penrose is constructed in the following way: we fix an ordering $v_0, v_1, \ldots, v_n$ of the vertices of G, and for each spanning tree $T \in \mathcal{ST}(G)$, choose v_0 as its root, let $d(i)$ be the tree distance of the vertex v_i to v_0 in tree T.

Penrose scheme associates to T the graph $R_{\mathrm{Pen}}(T)$ formed by adding to T all edges $\{v_i, v_j\} \in E(G) \setminus E(T)$ such that either:

(p1) between vertices of the same generation: $d(i) = d(j)$, or

(p2) connecting to predecessors with smaller index: $d(i) = d(j) - 1$ and $i < j$.

For a partition scheme R, we denote the set of trees by

$$\mathcal{T}_R(G) := \Big\{ T \in \mathcal{ST}(G) : R(T) = T \Big\}.$$

The following is a generalized Penrose identity that is well-known, see, for example, [88, Proposition 2.3] or [27, Proposition 5].

Lemma 2.10. *For any partition scheme R, we have*

$$\sum_{H \in CSpan(G)} (-1)^{|E(H)|} \;=\; (-1)^{|V(G)|-1} \big|\mathcal{T}_R(G)\big| \,.$$

Proof. For any numbers x_e for $e \in E(G)$, we have

$$\begin{aligned} \sum_{G \in \mathrm{CSpan}(G)} \prod_{e \in E(G)} x_e \;&=\; \sum_{T \in \mathcal{ST}(G)} \prod_{e \in E(T)} x_e \sum_{\mathcal{F} \subset E(R(T)) \setminus E(T)} \prod_{e \in \mathcal{F}} x_e \\ &=\; \sum_{T \in \mathcal{ST}(G)} \prod_{e \in E(T)} x_e \prod_{e \in E(R(T)) \setminus E(T)} (1 + x_e) \,, \end{aligned}$$

where the first equality is due to property (ii) of partition schemes. If $x_e = -1$, the last factor kills the contributions of all trees with $E(R(T)) \setminus E(T) \neq \emptyset$. Furthermore, for any tree T, we have $e_T = v_T - 1$. $\square$

Recall the Ursell function $\phi : \Gamma(G) \to \mathbb{R}$ is defined by

$$\phi(\gamma) = \sum_{H \in \mathrm{CSpan}(\mathbb{G}(\gamma))} (-1)^{e_H}.$$

Then the above Penrose identity gives a simple upper bound on the Ursell function using the number of spanning trees. This is the well-known "tree-graph bound" in the statistical physics community.

Corollary 2.11.

$$\phi(\gamma) = (-1)^{|v_{\mathbb{G}(\gamma)}|-1} \left|\mathcal{T}_R(\mathbb{G}(\gamma))\right| \leqslant \left|\mathcal{ST}(\mathbb{G}(\gamma))\right|.$$

2.4.3 Féray-Méliot-Nikeghbali bound on cumulants and asymptotic normality

To derive asymptotic normality, we need to bound cumulants. The following bound on cumulants of a summation of random variables is by Janson [54, Lemma 4].

Theorem 2.12. *Let $\{X_i\}_{i\in[n]}$ be G-dependent random variables that are uniformly bounded by M. Then for any integer $r \geqslant 1$, there exists some constant C_r such that*

$$\left|\kappa_r\left(\sum_{v\in V(G)} X_v\right)\right| \leqslant C_r n(\Delta+1)^{r-1}M^r.$$

In most applications for counting substructures in random objects, each variable X_v is an indicator variable, so that the bounded assumption is not too restrictive. This theorem is often used to prove some central limit theorem. Döring and Eichelsbacher [19] have analyzed Janson's original proof and claim that the above theorem holds with

$$C_r = (2e)^r(r!)^3.$$

Then they use this new bound to obtain moderate deviation results.

Later these bounds on cumulants are improved by Féray, Méliot and Nikeghbali in [26]. Using the new bounds they obtained, they prove precise large or moderate deviations for sequences of real-valued random variables $(X_n)_{n\in\mathbb{N}}$ and obtain the so-called "mod-Gaussian convergence" under certain assumptions on cumulants.

We will first describe and then provide a simple proof of their bound by a different approach using the tree-graph bound. Let $\mathrm{ST}(H)$ denote the number of spanning trees of a graph G. Consider now a graph H with vertex set $[r]$ and a set partition π of $[r]$. For each $P \in \pi$, we use $\mathrm{ST}(H[P])$ to denote the number of spanning trees of the subgraph of H induced by the vertex set P. We will also consider the contraction H/π of H with respect to π. It is a multigraph defined as follows. The vertex set of H/π is the index set $[|\pi|]$ of the elements $P \in \pi$, and for distinct $P_i, P_j \in \pi$, there are as many edges between i and j as edges between a vertex of P_i and a vertex of P_j in H. Denote $\mathrm{ST}(H/\pi)$ the number of spanning trees of this contracted graph with the consideration of multiple edges.

The cumulant bound by Féray, Méliot and Nikeghbali is the following.

Theorem 2.13 ([26, Eq. (47)])**.** *Let $\{X_i\}_{i\in[n]}$ be G-dependent variables that are uniformly bounded by M. Then*

$$|\kappa(V(G))| \leqslant (2M)^{r-1}\,\mathbf{E}\left[|X_1|\right]\mathrm{ST}(G) \leqslant n2^{r-1}M^r\mathrm{ST}(G). \tag{2.4.1}$$

To give an alternative proof of this, we need a simple combinatorial lemma.

Lemma 2.14 ([26, Eq. (43)]). *For any graph H, we have*

$$2^{v_H-1}\mathrm{ST}(H) = \sum_{\pi\in\Pi(V(H))} \mathrm{ST}(H/\pi)\prod_{P\in\pi}\mathrm{ST}(H[P]). \tag{2.4.2}$$

This is by noting that the union of a spanning tree $\overline{T}$ of H/π and of spanning trees T_i of $H[P]$ for all $P\in\pi$ gives a spanning tree T of H. Conversely, take a spanning tree T on H and a bicoloration of its edges. Edges of color 1 can be seen as a subgraph of H with the same vertex set $[r]$. This graph is of course acyclic. Its connected components define a partition π of $[r]$ and edges of color 1 correspond to a collection of spanning trees T_i of $H[P]$ for $P\in\pi$. Besides, edges of color 2 define a spanning tree $\overline{T}$ on H/π.

Proof of Theorem 2.13. To avoid triviality, we assume that G has a single connected component, otherwise, the cumulant is simply zero. By the definition of the cumulant, we have that

$$\kappa(V(G)) = \sum_{\pi\in\Pi(V(G))}(-1)^{|\pi|-1}(|\pi|-1)!\prod_{P\in\pi}\mu(P) = \sum_{\pi\in\Pi(V(G))}\sum_{G\in\mathrm{CSpan}(K_{|\pi|})}(-1)^{e_G}\prod_{P\in\pi}\mu(P),$$

where the last equality is by using the identity (2.2.8).

By considering the polymer partition $\Pi_c(\cdot)$ introduced in the proof of Lemma 2.4, we represent the partition π as a set of pairwise non-adjacent polymers in G. Then we have that

$$\begin{aligned}\kappa(V(G)) &= \sum_{\{C_1,\dots,C_n\}\in\Pi_c(V)}\sum_{\pi\in\Pi'}\sum_{G\in\mathrm{CSpan}(K_{|\pi|})}(-1)^{e_G}\prod_{i\in[n]}\mu(C_i)\\ &= \sum_{\{C_1,\dots,C_n\}\in\Pi_c(V)}\sum_{\pi\in\Pi'}\sum_{G\in\mathrm{CSpan}(K_{|\pi|})}(-1)^{e_G}\prod_{P\in\pi}\mu(P)\,\mathbf{1}_{\{P\in\mathcal{I}(\mathbb{G}(C_1,\dots,C_n))\}},\end{aligned} \tag{2.4.3}$$

where we note that

$$\prod_{P\in\pi}(-1)^{|P|}\mu(P) = \prod_{i\in[n]}(-1)^{|C_i|}\mu(C_i)$$

if and only if every element of the partition $\pi\in\Pi'$ is an independent set of $\mathbb{G}(C_1,\dots,C_n)$, recalling Π' is defined by (2.3.15). By Lemma 2.5, we have

$$\sum_{\pi\in\Pi'}\sum_{G\in\mathrm{CSpan}(K_{|\pi|})}(-1)^{e_G}\prod_{P\in\pi}\mathbf{1}_{\{P\in\mathcal{I}(\mathbb{G}(C_1,\dots,C_n))\}} = \sum_{H\in\mathrm{CSpan}(\mathbb{G}(C_1,\dots,C_n))}(-1)^{e_H}.$$

Combining above and the tree-graph bound in Corollary 2.11, we obtain, from (2.4.3), that

$$\begin{aligned}\kappa(V(G)) &= \sum_{\{C_1,\dots,C_n\}\in\Pi_c(V)}\sum_{H\in\mathrm{CSpan}(\mathbb{G}(C_1,\dots,C_n))}(-1)^{e_H}\prod_{i\in[n]}\mu(C_i)\\ &\leqslant \sum_{\{C_1,\dots,C_n\}\in\Pi(V)}|\mathcal{ST}(\mathbb{G}(C_1,\dots,C_n))|\prod_{i\in[n]}\mu(C_i)\,\mathbf{1}_{\{C_i\in\mathcal{C}(G)\}}.\end{aligned}$$

Note that for all $\pi = \{C_1, \dots, C_n\} \in \Pi(V)$, we have that

$$\begin{aligned} \left|\mathcal{ST}(\mathbb{G}(C_1, \dots, C_n))\right| &\leqslant \mathrm{ST}(G/\pi), \\ \mathbf{1}_{\{C_i \in \mathcal{C}(G)\}} &\leqslant \mathrm{ST}(G[C_i]). \end{aligned}$$

The reason is that $\mathrm{ST}(H/\pi)$ is the number of spanning trees of the contracted graph where multiple edges are considered, and $\mathbb{G}(C_1, \dots, C_n)$ is simply a quotient graph, that can be obtained from $\mathrm{ST}(H/\pi)$ by replacing parallel edges with a single edge.

Since variables are assumed to be all bounded by M, we also have

$$\prod_{i \in [n]} \mu(C_i) \leqslant M^{n-1} \mathbf{E}\left[|X_1|\right],$$

Therefore we conclude

$$\begin{aligned} \kappa(V(G)) &\leqslant M^{n-1}\, \mathbf{E}\left[|X_1|\right] \sum_{\{C_1,\dots,C_n\} \in \Pi(C)} \left|\mathcal{ST}(\mathbb{G}(C_1, \dots, C_n))\right| \prod_{i \in [n]} \mathbf{1}_{\{C_i \in \mathcal{C}(G)\}} \\ &= M^{n-1}\, \mathbf{E}\left[|X_1|\right] \sum_{\pi \in \Pi(V)} \mathrm{ST}(G/\pi) \prod_{P \in \pi} \mathrm{ST}(G[P]) = M^{n-1}\, \mathbf{E}\left[|X_1|\right] 2^{n-1} \mathrm{ST}(G). \end{aligned}$$

This completes the proof. □

Next we bound the cumulants of the summation of graph-dependent variables, which leads to a normality criterion.

Theorem 2.15. *Let $\{X_v\}_{v \in V(G)}$ be G-dependent random variables that are uniformly bounded by M. Then*

$$\left|\kappa_r\left(\sum_{v \in V(G)} X_v\right)\right| \leqslant (2M)^{r-1} \sum_{v \in V(G)} \mathbf{E}\left[|X_v|\right] \sum_{(v_1,\dots,v_{r-1}) \in V(G)^{r-1}} \mathrm{ST}(\mathbb{G}(v, v_1, \dots, v_{r-1})).$$

Moreover, let $X = \sum_{v \in V(G)} X_v$ and $\sigma^2 = \mathrm{var}(X) \neq 0$. If for $r \geqslant 3$,

$$\frac{1}{M}\left(\frac{M}{\sigma}\right)^r \sum_{v \in V(G)} \mathbf{E}\left[|X_v|\right] \sum_{(v_1,\dots,v_{r-1}) \in V(G)^{r-1}} \mathrm{ST}(\mathbb{G}(v, v_1, \dots, v_{r-1})) \to 0, \tag{2.4.4}$$

then

$$\frac{X - \mathbf{E}[X]}{\sigma} \to \mathcal{N}(0, 1).$$

The proof is by directly applying Theorem 2.9 by verifying its assumption using (2.4.4). To obtain a more explicit criterion, we introduce the following lemma that bounds the number of spanning trees using the maximum degree and Cayley's formula.

Lemma 2.16. *[26, Corollary 9.17] Let G be a graph on n vertices and maximal degree Δ and $r \geqslant 1$. Fix a vertex v_1 of G. The number of pairs $\big((v_1, \dots, v_r), T\big)$ where each v_i is a vertex of V and T a* *spanning tree of the induced subgraph $G[\{v_1, \ldots, v_r\}]$ is bounded above by*

$$r^{r-2}\,(\Delta+1)^{r-1}.$$

Using above lemma to simplify (2.4.4) yields a more explicit criterion,

$$\frac{1}{M}\left(\frac{M}{\sigma}\right)^r r^{r-2}\,(\Delta+1)^{r-1}\sum_{v\in V(G)}\mathbf{E}\left[|X_v|\right]\to 0.$$

By noting that $r^{r-2} = O\left(1\right)$ for $r = \Theta(1)$, and $\sum_{v\in V(G)}\mathbf{E}\left[|X_v|\right]\leqslant nM$, our criterion implies the following criterion by Janson.

Theorem 2.17 ([54, Theorem 2])**.** *Let $\{X_i\}_{i\in[n]}$ be G-dependent variables that are uniformly bounded by M. Let $X = \sum_{v\in V(G)} X_v$ and $\sigma^2 = \mathrm{var}(X)\neq 0$. If for $r\geqslant 3$,*

$$\left(\frac{(\Delta+1)M}{\sigma}\right)^r\frac{n}{\Delta+1}\to 0,$$

then

$$\frac{X-\mathbf{E}[X]}{\sigma}\to\mathcal{N}(0,1).$$

2.5 A convergent cluster expansion series via Kotecký-Preiss criterion

By showing the absolute convergence of the cluster expansion under certain conditions, we use the truncated series to approximate the logarithm of the asymptotic probability of non-occurrences $\mathbf{P}\left(X=0\right)$. For convenience, given G-dependent random indicators $\{X_v\}_{v\in V(G)}$, for every positive integer k, we define

$$L_{G,k} := \sum_{\gamma\in\Gamma(G):\|\gamma\|=k}\frac{\phi(\gamma)}{|\gamma|!}(-1)^{\|\gamma\|}\prod_{C\in\gamma}\mu\left(C\right)\quad\text{and}\quad T_{G,k} := \sum_{i\in[k-1]} L_{G,i}. \tag{2.5.1}$$

The truncated series gives the approximation of $\log\mathbf{P}\left(X=0\right)$ up to arbitrary accuracy.

Theorem 2.18. *Let $\{X_v\}_{v\in V(G)}$ be G-dependent random indicators, Δ be the maximum degree of G, and $p_{\max} := \max_{i\in V(G)}\mathbf{E}\left[X_i\right]$. Let $\theta\geqslant 0$ be an integer. If there exists $\beta=\beta(\theta)>0$ such that*

(1) $e^3\Delta\beta<1$, and

(2) for all $C\in\mathcal{C}(G)$ with $|C|\geqslant\theta+1$, we have

$$\mu\left(C\right)\leqslant p_{\max}\beta^{|C|-1}\quad\text{and}\quad(e^3\Delta)^{\theta+1}(\Delta+1)p_{\max}\left(\frac{1}{e^3\Delta-1}+\frac{\beta^\theta}{1-e^3\Delta\beta}\right)\leqslant 1. \tag{2.5.2}$$

Then given any $\delta>0$, for all $k\geqslant\log\left(v_G/\delta\right)$, we have

$$\left|\log\mathbf{P}\left(X=0\right)-T_{G,k}\right|\leqslant\delta. \tag{2.5.3}$$

Note that the number of terms in the expansion may not be a constant. The absolute convergence of the cluster expansion is by using the Koteckŷ-Preiss criterion.

Lemma 2.19 (Koteckŷ-Preiss criterion [61]). *Let functions $f, g : \mathcal{C}(G) \to [0, \infty)$ be such that*

$$\sum_{C \in \mathcal{C}(G): C \sim C_0} e^{f(C)+g(C)} \mu(C) \leqslant f(C_0) \tag{2.5.4}$$

for all $C_0 \in \mathcal{C}(G)$. Then the cluster expansion (2.2.1) converges absolutely. Moreover, let $g(\gamma) := \sum_{C \in \gamma} g(C)$. Then for all $C_0 \in \mathcal{C}(G)$,

$$\sum_{\gamma \in \Gamma(G): \gamma \sim C_0} \left| \frac{\phi(\gamma)}{|\gamma|!} \prod_{C \in \gamma} \mu(C) \right| e^{g(\gamma)} \leqslant f(C_0), \tag{2.5.5}$$

where we write $\gamma \sim C_0$ if there exists $C \in \gamma$ such that $C \sim C_0$.

To bound the number of polymers, we introduce a simple combinatorial lemma that is well-known.

Lemma 2.20 ([34, Lemma 2.1]). *In a graph with maximum degree Δ, the number of connected induced subgraphs of order t containing a fixed vertex v is at most $(e\Delta)^t$.*

Proof of Theorem 2.18. We first verify the Koteckŷ-Preiss criterion by choosing $f(C) = g(C) = |C|$ for all $C \in \mathcal{C}(G)$. Then for every $v \in V(G)$,

$$\begin{aligned}
\sum_{C \in \mathcal{C}(G): C \sim v} e^{2|C|} \mu(C) &\leqslant \sum_{t=1}^{\infty} \sum_{C \in \mathcal{C}_t(G): C \sim v} e^{2|C|} \mu(C) \\
&\leqslant \sum_{t=1}^{\theta} \sum_{C \in \mathcal{C}_t(G): C \sim v} e^{2|C|} \mu(C) + \sum_{t=\theta+1}^{\infty} \sum_{C \in \mathcal{C}_t(G): C \sim v} e^{2|C|} \mu(C) \\
&\leqslant \sum_{t=1}^{\theta} e^{2t} (e\Delta)^t p_{\max} + \sum_{t=\theta+1}^{\infty} e^{2t} (e\Delta)^t p_{\max} \beta^{t-1},
\end{aligned}$$

where we bound the number of polymers using Lemma 2.20 and the bound on the joint moment $\mu(C)$ with $|C| \geqslant \theta + 1$ is by the first assumption in (2.5.2).

Therefore we have

$$\begin{aligned}
\sum_{C \in \mathcal{C}(G): C \sim v} e^{2|C|} \mu(C) &\leqslant p_{\max} \sum_{t=1}^{\theta} (e^3 \Delta)^t + \frac{p_{\max}}{\beta} \sum_{t=\theta+1}^{\infty} (e^3 \Delta \beta)^t \\
&\leqslant p_{\max} \frac{(e^3\Delta)^{\theta+1}}{e^3\Delta - 1} + \frac{p_{\max}}{\beta} \frac{(e^3 \Delta\beta)^{\theta+1}}{1 - e^3\Delta\beta} \\
&= p_{\max} (e^3 \Delta)^{\theta+1} \left(\frac{1}{e^3\Delta - 1} + \frac{\beta^\theta}{1 - e^3\Delta\beta} \right) \leqslant \frac{1}{\Delta+1},
\end{aligned}$$

where the last inequality is due to the second assumption in (2.5.2). Hence for every $C_0 \in \mathcal{C}(G)$,

$$\sum_{C \in \mathcal{C}(G): C \sim C_0} e^{2|C|} \mu(C) \leqslant \sum_{v \in N^+(C_0)} \sum_{C \in \mathcal{C}(G): C \sim v} e^{2|C|} \mu(C) \leqslant \sum_{v \in N^+(C_0)} \frac{1}{\Delta+1} \leqslant |C_0|,$$

by noting that $|N^+(C_0)| \leqslant (\Delta+1)|C_0|$. This verifies the Kotecký-Preiss criterion (2.5.4), thus the expansion convergences absolutely.

Next, we bound the truncation error. Summing (2.5.5) over $C_0 = \{v\}$ for all $v \in V(G)$ gives

$$\sum_{\gamma \in \Gamma(G)} \left| \frac{\phi(\gamma)}{|\gamma|!} \prod_{C \in \gamma} \mu(C) \right| e^{\|\gamma\|} \leqslant v_G.$$

Therefore we have

$$e^k \sum_{\gamma \in \Gamma(G): \|\gamma\| \geqslant k} \frac{\phi(\gamma)}{|\gamma|!} \prod_{C \in \gamma} (-1)^{|C|} \mu(C) \leqslant \sum_{\gamma \in \Gamma(G): \|\gamma\| \geqslant k} \left| \frac{\phi(\gamma)}{|\gamma|!} \prod_{C \in \gamma} \mu(C) \right| e^{\|\gamma\|} \leqslant v_G.$$

Hence if we have $e^k \geqslant v_G/\delta$, then

$$\left| \log \mathbf{P}(X=0) - \sum_{\gamma \in \Gamma(G): \|\gamma\| \in [k-1]} \frac{\phi(\gamma)}{|\gamma|!} (-1)^{\|\gamma\|} \prod_{C \in \gamma} \mu(C) \right| \leqslant \delta.$$

This completes the proof. □

Assumptions in (2.5.2) get simplified under various conditions.

(A1) **Truncation under negative association**

Suppose the G-dependent random indicators are negatively associated [60], in particular, for all disjoint $U, V \subseteq V(G)$ and all non-decreasing functions f, g, we have

$$\mathbf{E}\left[f(X_i, i \in U) g(X_i, i \in V)\right] \leqslant \mathbf{E}\left[f(X_i, i \in U)\right] \mathbf{E}\left[g(X_i, i \in V)\right]. \tag{2.5.6}$$

Since negative association implies a negative correlation, we have $\mu(C) \leqslant p_{\max}^{|C|}$ for all $C \in \mathcal{C}(G)$. Hence we can choose $\beta = p_{\max}$ and Theorem 2.18 implies the following.

Corollary 2.21. *Let $\{X_v\}_{v \in V(G)}$ be negatively associated G-dependent random indicators. If we have*

$$p_{\max} \leqslant \frac{1}{4e^3 \Delta(\Delta+1)}, \tag{2.5.7}$$

then given any $\delta > 0$, for all $m \geqslant \log(v_G/\delta)$, the bound (2.5.3) holds.

It is straightforward to check that choosing $\theta = 0$ and $\beta = p_{\max}$, we have

$$(e^3\Delta)^{\theta+1}(\Delta+1) p_{\max} \frac{1}{e^3\Delta - 1} \leqslant \frac{1}{4(e^3\Delta-1)} \leqslant \frac{1}{2e^3\Delta},$$

and

$$(e^3\Delta)^{\theta+1}(\Delta+1) p_{\max} \frac{\beta^\theta}{1-e^3\Delta\beta} \leqslant \frac{1}{4(1-e^3\Delta\beta)} \leqslant \frac{1}{4-(\Delta+1)^{-1}}.$$

This suffices to verify bounds in (2.5.2).

(A2) **Bounding joint moments via independence**

A simple combinatorial upper bound on the joint moment is via the maximal independent set of the induced subgraph $G[C]$ by the definition of dependency graph, specifically, for all $C \in \mathcal{C}(G)$,

$$\mu(C) \leqslant \min_{U \in \mathcal{I}(G[C])} p_{\max}^{|U|}.$$

This provides a feasible choice of β in (2.5.2) for a given θ:

$$\beta(\theta) = p_{\max}^{\xi} \quad \text{with} \quad \xi = \inf_{C \in \mathcal{C}(G): |C| \geqslant \theta+1} \frac{\max_{U \in \mathcal{I}(G[C])} |U| - 1}{|C| - 1}. \tag{2.5.8}$$

2.6 Limiting distribution of extremes under m-dependence

In this section, we consider the distribution of extremes under m-dependence, by using cluster expansion series. A sequence of random variables $\{X_i\}_{i\in[n]}$ is said to be $f(n)$-dependent if all sets of variables separated by at least the distance $f(n)$ are independent. This notion was introduced by Hoeffding and Robbins [41], and has been studied extensively. Let integer $m \geqslant 0$. A special case of $f(n)$-dependence when $f(n) = m$ is the following m-dependent model.

Definition 2.22 (m-dependence [41])**.** A sequence of random variables $\{X_i\}_{i\in[n]}$ is m-dependent for some integer $m \geqslant 0$ if $(X_j)_{j=1}^{i}$ are independent of $(X_j)_{j=i+m+1}^{n}$ for all $i > 0$.

The m-dependent sequences usually appear as block factors. Let $k > 0$ be an integer. The sequence $(X_i)_i$ is an k-block factor if there is an independent identically distributed sequence $(Y_j)_{-\infty}^{\infty}$ and a function $g : \mathbb{R}^k \to \mathbb{R}$ such that $X_i = g(Y_i, \ldots, Y_{i+k-1})$. Note that every such sequence $(X_i)_i$ is $(k-1)$-dependent, and there are m-dependent sequences that are not block factors, see, for example, [14].

The following limiting distribution of maxima of under m-dependence is by Newell [79].

Lemma 2.23 ([79])**.** *Let (Z_i) be a stationary sequence of m-dependent random variables and (y_n) be a sequence. If $\mathbf{P}(Z_i > y_n) = O(1/n)$, then*

$$\left| \mathbf{P}\left(\max_{i\in[n]} Z_i \leqslant y_n \right) - \exp\left(-n\mathbf{P}\left(B_1(y_n)\right)\right) \right| = O\left(\frac{1}{n}\right), \tag{2.6.1}$$

where

$$B_1(y) := \{Z_1 > y\} \cap \bigcap_{k=1}^{m} \{Z_{1+k} \leqslant y\}. \tag{2.6.2}$$

Newell's result extends a result by Watson [101], whose y_n is chosen such that $n\mathbf{P}(Z_i > y_n)$ has a finite limit as $n \to \infty$. Newell mentioned he did not specify that $n\mathbf{P}(B_i(y_n))$ must have a finite limit, nevertheless, we have

$$n\mathbf{P}(B_i(y_n)) \leqslant n\mathbf{P}(Z_i > y_n) = O(1).$$

We generalize Newell's result and consider a combinatorial dependent setting via dependency graphs. The following graph $D_{n,m}$ is a dependency graph for m-dependent random variables $\{X_i\}_{i\in[n]}$,

$$D_{n,m} := \left([n], \left\{ \{i,j\} \in \binom{[n]}{2} : |i-j| \in [m] \right\} \right). \tag{2.6.3}$$

Given a sequence of random variables $(Z_i)_i$, we have

$$\mathbf{P}\left(\max_{i\in[n]} Z_i \leqslant y\right) = \mathbf{P}\left(\sum_{i\in[n]} X_i = 0\right),$$

where $X_i := \mathbf{1}_{\{Z_i>y\}}$ denotes the random indicator of exceedance for all $i \in [n]$. Let $X := \sum_{i\in[n]} X_i$ count the exceedances.

From (2.6.2), we have

$$\mathbf{P}(B_i(y)) = \mathbf{E}\left[X_i \prod_{k=1}^{m}(1 - X_{i+k})\right] = \sum_{S\subseteq[i+1,i+m]} (-1)^{|S|}\mu(S \cup \{i\}),$$

where $[i + 1, i + m]$ denotes the integer set $\{i + 1, i + 2, \dots, i + m\}$. Then Newell's limiting distribution of maxima under m-dependence (2.6.1) can be reformulated as an approximation of $\mathbf{P}(X = 0)$ for $D_{n,m}$-dependent indicators using cliques of $D_{n,m}$ with size at most $m + 1$, specifically,

$$\mathbf{P}(X = 0) = \exp\left(\sum_{C\in\bigcup_{t\in[m+1]}\mathcal{K}_t(D_{n,m})} (-1)^{|C|}\mu(C) + O\left(n^{-1}\right)\right), \tag{2.6.4}$$

where the set of cliques on t vertices in graph G is denoted by

$$\mathcal{K}_t(G) := \left\{C \in \binom{V(G)}{t} : G[C] = K_{|C|}\right\}. \tag{2.6.5}$$

The extremal index is the most popular approach to study the clustering of extremes in random processes. This notion, originated by Newell [79], Loynes [70] and O'Brien [80], was given a firm definition by Leadbetter [64]. Formally, a stationary sequence $(X_i)_i$ has extremal index $\theta \in [0, 1]$ if for every $\tau > 0$,

(EI1) there exists $y_n(\tau)$ such that $\lim_{n\to\infty} n\mathbf{P}(X_i > y_n(\tau)) = \tau$,

(EI2) $\lim_{n\to\infty} \mathbf{P}\left(\max_{i\in[n]} X_i \leqslant y_n(\tau)\right) = e^{-\theta\tau}$.

Moreover, if a mixing stationary sequence $(X_i)_i$ satisfies (EI1) and $\lim_{n\to\infty} \mathbf{P}\left(\max_{i\in[n]} X_i \leqslant y_n(\tau)\right)$ exists, then (EI2) holds with some $\theta \in [0, 1]$. For more details, see [62, subsection 3.7]. Newell's result in Lemma 2.23 also gives the extremal index for stationary m-dependent sequences

$$\theta = \lim_{n\to\infty} \mathbf{P}\left(\bigcap_{k=1}^{m}\{X_{i+k} \leqslant y_n\} \mid X_i > y_n\right),$$

see also [91, Eq. (1.2)].

The discussion of extremal index is usually restricted to stationary sequences. Although non-stationary sequences are of practical interest, an exact analog of the extremal index may not exist in this case, see, for example, [43].

The extremal index is essentially some constant adjustment factor of the first order term. Our general framework handles the clustering via taking into account the explicit large clusters, and our formula gives the asymptotic probability up to arbitrary accuracy. Moreover, stationarity is not necessary for our method.

2.6.1 Probability of non-occurrences under m-dependence

The cluster expansion gives an asymptotic expansion of the probability of non-occurrences under m-dependence with high accuracy. Note that for m-dependent random variables, recalling the corresponding dependency graph (2.6.3), the maximum degree is $\Delta = 2m$. By setting $\theta = m + 1$, Theorem 2.18 gives the following.

Theorem 2.24. *Let $\{X_i\}_{i\in[n]}$ be m-dependent random indicators and $p_{\max} := \max_{i\in[n]} \mathbf{E}[X_i]$. If there exists $\beta > 0$ such that*

(1) $2e^3 m\beta < 1$, and

(2) for all $C \in \mathcal{C}(D_{n,m})$ with $|C| \geqslant m+2$, we have

$$\mu(C) \leqslant p_{\max}\beta^{|C|-1} \quad \text{and} \quad \frac{p_{\max}}{2me^3-1} + \frac{\beta^{m+1}p_{\max}}{1-2me^3\beta} \leqslant \frac{1}{(2m+1)(2me^3)^{m+2}}, \tag{2.6.6}$$

then given any $\delta > 0$, for all $k \geqslant \log(n/\delta)$, we have

$$\left|\log \mathbf{P}(X=0) - T_{D_{n,m},k}\right| \leqslant \delta. \tag{2.6.7}$$

Moreover, by Corollary 2.21, if $\{X_i\}_{i\in[n]}$ are negatively associated, then (1) and (2) can simply be replaced by

$$p_{\max} \leqslant \frac{1}{24e^3m^2}. \tag{2.6.8}$$

Remark 2.25. *If we set*

$$\beta := p_{\max}^{1/(m+1)(m+2)}$$

as in (2.5.8), and assume that

$$\frac{p_{\max}}{2me^3-1} + \frac{p_{\max}^{1/(m+2)+1}}{1-2me^3 p_{\max}^{1/(m+1)(m+2)}} \leqslant \frac{1}{(2m+1)(2me^3)^{m+2}}, \tag{2.6.9}$$

then both assumptions in (2.6.6) hold since for all C with $|C| \geqslant m+2$, we have

$$\left\lceil \frac{|C|}{m+1} \right\rceil - 1 \geqslant \frac{|C|-m-1}{m+1} = \frac{|C|-1}{m+1}\left(1 - \frac{m}{|C|-1}\right) \geqslant \frac{|C|-1}{(m+1)(m+2)},$$

and therefore,

$$\mu(C) \leqslant p_{\max}^{\lceil |C|/(m+1)\rceil} \leqslant p_{\max}^{(|C|-1)/(m+1)(m+2)+1} = p_{\max}\beta^{|C|-1}.$$

An even more explicit sufficient condition is

$$p_{\max} \leqslant \left(\frac{1}{me^4}\right)^{(m+2)^3}. \tag{2.6.10}$$

This condition implies (2.6.6) *by noting*

$$(2m+1)(2me^3)^{m+2}\frac{p_{\max}}{2me^3-1} \leqslant \frac{3m(2me^3)^{m+2}}{me^3}\left(\frac{1}{me^4}\right)^{(m+2)^3} \leqslant \frac{(me^4)^{m+2}}{e(me^4)^{(m+2)^3}} \leqslant \frac{1}{e},$$

and

$$(2m+1)(2me^3)^{m+2}\frac{p_{\max}^{1/(m+2)+1}}{1-2me^3p_{\max}^{1/(m+1)(m+2)}} \leqslant 3m(me^4)^{m+2}\frac{\left(\frac{1}{me^4}\right)^{(m+2)^2}}{1-me^4\left(\frac{1}{me^4}\right)^{(m+2)^2/(m+1)}}$$
$$\leqslant \left(\frac{1}{me^4}\right)^{(m+2)(m+1)}\frac{3m}{1-\left(\frac{1}{me^4}\right)^{(m+2)^2/(m+1)-1}} \leqslant \frac{6m}{(me^4)^{(m+2)(m+1)}} \leqslant \frac{6}{e^4}.$$

2.6.2 The asymptotic distribution of maxima under m-dependence

Here we use the cluster expansion to give the asymptotic distribution of maxima. Theorem 2.24 and Remark 2.25 give the following.

Theorem 2.26. *Let $\{X_i\}_{i\in[n]}$ be m-dependent random variables such that*

$$\max_{i\in[n]} \mathbf{P}\left(X_i > y\right) \leqslant \left(\frac{1}{me^4}\right)^{(m+2)^3}. \tag{2.6.11}$$

Then given any $k>0$, there exists $K=K(k)$ such that for all $M \geqslant K\log n$, we have

$$\mathbf{P}\left(\max_{i\in[n]} X_i \leqslant y\right) = \exp\left(T_{D_{n,m},M} + O\left(n^{-k}\right)\right), \tag{2.6.12}$$

where the series is for the random indicators $\{\mathbf{1}_{\{X_i>y\}}\}_{i\in[n]}$. Moreover, if $\{\mathbf{1}_{\{X_i>y\}}\}_{i\in[n]}$ are negatively associated, then by (2.6.8), we can replace (2.6.11) *by*

$$\max_{i\in[n]} \mathbf{P}\left(X_i > y\right) \leqslant \frac{1}{24e^3m^2}.$$

Remark 2.27. *Note that the assumption* (2.6.11) *permits $p_{\max}$ to be a fixed constant that does not depend on n, whereas Newell [79] requires $p_{\max} = O\left(1/n\right) = o(1)$ in Lemma 2.23, thus we significantly relax the restriction on the probability. The terms in Newell's series* (2.6.4) *involves some polymers $C \in \bigcup_{t\in[m+1]}\mathcal{K}_t(D_{n,m})$, by noting $\mathcal{K}_t(D_{n,m}) \subseteq \mathcal{C}(D_{n,m})$ in view of its definition* (2.6.5)*, and our series considers more clusters.*

Notice that the clusters in (2.6.12) *are for indicator variables $\{\mathbf{1}_{\{X_i>y\}}\}_{i\in[n]}$, rather than $\{X_i\}_{i\in[n]}$. Also, note that stationarity is not assumed in Corollary 2.26, moreover, if the sequence is stationary, then we have the extremal index*

$$\theta = -\lim_{n\to\infty}\frac{T_{D_{n,m},M}}{n\mathbf{P}(X_i>y)}. \tag{2.6.13}$$

One common negative association is the Negative Orthant Dependence (NOD) by Joag-Dev and Proschan [60], which is a weaker notion. Random variables $\{X_i\}_{i\in[n]}$ are negative orthant dependent if *for all real* $(x_i)_{i\in[n]}$*, we have*

$$\mathbf{P}\left(\bigcap_{i\in[n]}\{X_i > x_i\}\right) \leqslant \prod_{i\in[n]} \mathbf{P}(X_i > x_i). \tag{2.6.14}$$

A similar dependent setting appears in [16, Example 2.2], where they obtain a compound Poisson approximation via Stein's method. Let $\{X_i\}_{i\in[n]}$ *be* m*-dependent and negative orthant dependent random variables. Then the indicators of exceedance* $\{\mathbf{1}_{\{X_i>y\}}\}_{i\in[n]}$ *are negative associated, this is due to (2.6.14), the negative orthant dependence of* $\{X_i\}_{i\in[n]}$*.*

Recently, Newell's results are extended to stationary random fields on $\mathbb{Z}^d$ in [53, 91]. Formally, a d-dimensional stationary random field $(X_{\mathbf{i}} : \mathbf{i} \in \mathbb{Z}^d)$ is m-dependent if $(X_{\mathbf{i}})_{\mathbf{i}\in A}$ and $(X_{\mathbf{j}})_{\mathbf{j}\in B}$ are independent for every pair of finite sets $A, B \subset \mathbb{Z}^d$ such that

$$\min_{\mathbf{i}\in A,\mathbf{j}\in B} \|\mathbf{i}-\mathbf{j}\| > m,$$

where $\|\mathbf{i}-\mathbf{j}\| := \max_{k\in[d]} |i_k - j_k|$.

Let $\mathbf{N}(n) := (\mathbf{N}(n) : n \in \mathbb{N}) \subset \mathbb{N}^d$ be such that $\prod_{i\in[d]} N_i(n) = O\left(n^d\right)$. Jakubowski and Soja-Kukieła [53], Soja-Kukieła [91] extend Lemma 2.23 and obtain the asymptotic distribution of maxima of d-dimensional m-dependent stationary random field,

$$\left|\mathbf{P}\left(M_{\mathbf{N}(n)} \leqslant y_n\right) - \exp\left(-n^d \mathbf{P}\left(X_{\mathbf{0}} \geqslant y_n, M_{A(m)} \leqslant y_n\right)\right)\right| = o(1),$$

where $M_A := \sup\{X_{\mathbf{i}} : \mathbf{i} \in A\}$ for all finite $A \subset \mathbb{Z}^d$, and

$$A(m) := \{(i_1, \dots, i_d) \in \mathbb{Z}^d : \forall j \in [d] : -m \leqslant i_j \leqslant m\}.$$

All our results are valid for the maxima of random fields with the substitution of $D_{n,m}$ with the following graph $\mathbf{D}_{n,m}$ for d-dimensional vectors

$$\mathbf{D}_{n,m} := \left(\mathbf{N}(n), \left\{\{\mathbf{i},\mathbf{j}\} \in \binom{\mathbf{N}(n)}{2} : \|\mathbf{i}-\mathbf{j}\| \in [m]\right\}\right).$$

2.6.3 Example: maxima of moving minima

We consider the limiting distribution of the maxima of moving minima process as an application of Theorem 2.26.

Corollary 2.28. *Let* $\{Z_i\}_i$ *be i.i.d. with* $\mathbf{P}(Z_i > y) = p$*. Let* $X_i = \min(Z_i, \dots, Z_{i+m})$*. If*

$$n\mathbf{P}(X_i > y)^2 = np^{2(m+1)} = o(1),$$

then given any $k > 0$*, there exists* K *depending on* k *such that for all* $M \geqslant K \log n$*, we have*

$$\mathbf{P}\left(\max_{i\in[n]} X_i \leqslant y\right) = \exp\left(T_{D_{n,m,M}} + O\left(n^{-k}\right)\right), \tag{2.6.15}$$

where the terms in the exponent of (2.6.15) are for indicator variables $\{\mathbf{1}_{\{X_i>y\}}\}_i$*.*

For illustration, we compute explicitly the terms in (2.6.15) of Corollary 2.28 with $m=1$. We denote $p := \mathbf{P}(Z_i > y)$ and assume that $np^4 = o(1)$. Then we have the joint moment for every $C \in \mathcal{C}(P_n)$,

$$\mu(C) = \mathbf{P}\left(\bigcap_{i \in C} \{\min(Z_i, Z_{i+1}) > y\}\right) = p^{|C|+1}.$$

First, we list contributing clusters up to size 4 in P_n in Table 2.1.

Cluster types	Count (ordered)	$\phi(\gamma)/\lvert\gamma\rvert! \cdot (-1)^{\lVert\gamma\rVert}$
$\{\{i\}\}$	n	$1 \cdot (-1)$
$\{\{i, i+1\}\}$	n	$1 \cdot (-1)^2$
$\{\{i\}, \{i+1\}\}$	$2n$	$-1/2 \cdot (-1)^2$
$\{\{i\}, \{i\}\}$	n	$-1/2 \cdot (-1)^2$
$\{\{i, i+1, i+2\}\}$	n	$1 \cdot (-1)^3$
$\{\{i\}, \{i+1, i+2\}\}$	$2n$	$-1/2 \cdot (-1)^3$
$\{\{i+1\}, \{i+1, i+2\}\}$	$2n$	$-1/2 \cdot (-1)^3$
$\{\{i+2\}, \{i+1, i+2\}\}$	$2n$	$-1/2 \cdot (-1)^3$
$\{\{i+3\}, \{i+1, i+2\}\}$	$2n$	$-1/2 \cdot (-1)^3$
$\{\{i\}, \{i+1\}, \{i+2\}\}$	$6n$	$2/3! \cdot (-1)^3$
$\{\{i+1\}, \{i+1\}, \{i+2\}\}$	$3n$	$2/3! \cdot (-1)^3$
$\{\{i+1\}, \{i+2\}, \{i+2\}\}$	$3n$	$2/3! \cdot (-1)^3$
$\{\{i\}, \{i\}, \{i\}\}$	n	$2/3! \cdot (-1)^3$
$\{\{i, i+1, i+2, i+3\}\}$	n	$1 \cdot (-1)^4$
$\{\{i, i+1\}, \{i, i+1\}\}$	n	$-1/2 \cdot (-1)^4$
$\{\{i, i+1\}, \{i+1, i+2\}\}$	$2n$	$-1/2 \cdot (-1)^4$
$\{\{i, i+1\}, \{i+2, i+3\}\}$	$2n$	$-1/2 \cdot (-1)^4$

Table 2.1: A list of small clusters of P_n

Then we compute the terms in the cluster expansion up to the error $o(1)$ as follows:

$$\begin{aligned}
L^{\emptyset}_{P_n,1} &= -n\mu(\{i\}) = -np^2, \\
L^{\emptyset}_{P_n,2} &= n\mu(\{i,i+1\}) - n\mu(\{i\})\,\mu(\{i+1\}) = np^3 - np^4, \\
L^{\emptyset}_{P_n,3} &= -n\mu(\{i,i+1,i+2\}) + n\mu(\{i\})\,\mu(\{i+1,i+2\}) + n\mu(\{i+3\})\,\mu(\{i+1,i+2\}) \\
&\quad - 2n\mu(\{i\})\,\mu(\{i+1\})\,\mu(\{i+2\}) + o(1) = -np^4 + o(1).
\end{aligned}$$

The Koteckỳ-Preiss criterion can also be used to show that the remaining terms in the expansion (2.6.12) are negligible if $np^4 = o(1)$, more precisely, we will show that

$$\sum_{i \geqslant 4} L_{P_n,i} = \sum_{\gamma \in \Gamma(P_n): \|\gamma\| \geqslant 4} \frac{\phi(\gamma)}{|\gamma|!} (-1)^{\|\gamma\|} \prod_{C \in \gamma} \mu(C) = o(1). \tag{2.6.16}$$

This is by choosing $f(C) = 1$ and $g(C) = -|C| \log p/k$ for all $C \in \mathcal{C}(P_n)$ in Lemma 2.19, where $k > 2$ is some fixed constant. Then for every $v \in [n]$,

$$\begin{aligned}
\sum_{C \in \mathcal{C}(P_n): C \sim v} e^{1-|C|\log p/k} \mu(C) &\leqslant e \sum_{t=1}^{\infty} \sum_{C \in \mathcal{C}_t(P_n): C \sim v} p^{-t/k} \mu(C) \\
&\leqslant e \sum_{t=1}^{\infty} p^{-t/k} (2e)^t p^{t+1} \leqslant ep \sum_{t=1}^{\infty} (2ep^{1-1/k})^t = \frac{2e^2 p^{2-1/k}}{1 - 2ep^{1-1/k}} = o(1).
\end{aligned}$$

This verifies the Koteckỳ-Preiss criterion since P_n has the maximum degree 2. Similar to the proof of Theorem 2.18, summing (2.5.5) over $v \in [n]$ gives

$$\sum_{\gamma \in \Gamma(P_n)} \left| \frac{\phi(\gamma)}{|\gamma|!} \prod_{C \in \gamma} \mu(C) \right| e^{-\sum_{C \in \gamma} |C| \log p / k} = \sum_{\gamma \in \Gamma(P_n)} \left| \frac{\phi(\gamma)}{|\gamma|!} \prod_{C \in \gamma} \mu(C) \right| p^{-\|\gamma\|/k} \leqslant n.$$

Therefore we have

$$\sum_{\gamma \in \Gamma(P_n) : \|\gamma\| \geqslant 4k} \frac{\phi(\gamma)}{|\gamma|!} (-1)^{\|\gamma\|} \prod_{C \in \gamma} \mu(C) \leqslant np^4 = o(1). \tag{2.6.17}$$

It is not hard to show that the differences of the sum in (2.6.16) and (2.6.17) is $o(1)$, since $k > 2$ is some fixed constant.

Now we have the asymptotic cumulative distribution function

$$\mathbf{P}\left(\max_{i \in [n]} X_i \leqslant y \right) = \exp\left(\sum_{i \in [3]} L^{\emptyset}_{P_n, i} + o(1) \right) = \exp\left(-np^2 + np^3 + o(1) \right). \tag{2.6.18}$$

Formula (2.6.13) gives the extremal index

$$-\lim_{n \to \infty} \frac{-np^2 + np^3 + o(1)}{np^2} = 1.$$

Our expansion is valid for $np^4 = o(1)$, in which case the coefficient of the term np^3 cannot be determined by the extremal index approach or Newell's expansion (2.6.4). Furthermore, stationarity is not necessary and is used just for simplification of the calculation of joint moments; variables $(Z_i)_i$ can be not independent or not identically distributed or neither.

Chapter 3

Extremal independence under graphical mixing

3.1 Introduction

Fisher–Tippett–Gnedenko theorem is central in the extreme value theory; it was discovered first by Fisher and Tippett [28] and later proved in full generality by Gnedenko [37]. This theorem states that if the maximum of the first n terms of a sequence of independent and identically distributed (i.i.d.) random variables has a non-degenerate limit distribution after a proper normalisation, then it belongs to either the Gumbel, the Fréchet, or the Weibull families of distributions.

The FTG theorem generalises to stationary random sequences of dependent random variables under the additional assumptions that its distant terms are independent [101] or weakly dependent [70]. Leadbetter in [63] significantly relaxed the assumptions of [101, 70]. The analogues of Leadbetter's conditions were also found for non-stationary sequences [42, 43] and for random fields [65, 83]. In fact, the behaviour of maxima for non-stationary sequences is more complicated than that for the stationary case. Even in the simplest case when the variables are independent, the limit distribution might not belong to any of the Gumbel, the Frechet or the Weibull families, see [24, Section 8.3]. That is, it is impossible to classify all possible limit distributions for general random systems with dependencies. Nevertheless, extremal characteristics of such systems always attracted significant attention of researchers in computer science, statistical physics, financial mathematics and network studies.

In this chapter, we focus on the following extremal independence property that helps to reduce general random systems to the independent case, where standard statistical techniques apply. Let $\mathbf{X}(n) = (X_1(n), \ldots, X_d(n))^T \in \mathbb{R}^d$ be a sequence of random vectors, where $d = d(n)$ be a sequence of positive integers. We give sufficient conditions for the property that

$$\left| \mathbf{P}\left(\max_{i\in[d]} X_i \leqslant x \right) - \prod_{i\in[d]} \mathbf{P}\left(X_i \leqslant x \right) \right| \to 0 \qquad \text{for any fixed } x \in \mathbb{R}. \tag{3.1.1}$$

Allowing arbitrary sequences of vectors $\mathbf{X}(n)$ in (3.1.1) encapsulates several similar questions arising in the studies of sequences of random variables, triangular arrays, random fields, and so on. For example, for a sequence $\xi_1, \xi_2, \ldots$ of identically distributed (i.d.) random variables, one can set

$$X_i(n) := \frac{\xi_i - a_n}{b_n},$$

where a_n and b_n are the normalising constants from the FTG theorem. This immediately extends the FTG theorem to the sequences of dependent i.d. random variables that satisfy our sufficient conditions.

Clearly, the extremal independence property (3.1.1) is equivalent to

$$\left| \mathbf{P}\left(\bigcap_{i\in[d]} \overline{A_i} \right) - \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right) \right| \to 0, \tag{3.1.2}$$

where the system of events $\mathbf{A}$ is defined by

$$\mathbf{A} = \mathbf{A}(n,x) := (A_i)_{i\in[d]}, \quad A_i := \{X_i > x\}, \tag{3.1.3}$$

and $\overline{A_i}$ is the complement event of A_i. Estimates for the probability of non-occurrence of events appear in many applications in probabilistic combinatorics and number theory. In particular, to justify the existence of a certain object, it is sufficient to show that the related probability (over all places where this object might appear) is positive; see, for example, [3, Section 5].

In this chapter, we establish new bounds for (3.1.2) by developing the idea proposed by Galambos [32, 31] and Arratia, Goldstein, Gordon [4, 5]: the weak and strong dependencies between events $(A_i)_{i\in[d]}$ are considered separately, and the bounds do not incorporate the computation of moments of the number of occurrences $Z = \sum_{i\in[d]} \mathbf{1}_{\{A_i\}}$ higher than the second one. This allows to overcome the disadvantages of classical bounds. In particular, bounds based on dependency graphs (Lovász local lemma [21], Janson's inequality [57], Suen's inequality [99, 56]) allow complicated dependence structures, but often fail to characterise the relations quantitatively. Applying the method of moments gets complicated when high factorial moments diverge or are hard to compute. Both the Chen-Stein method [4, 5, 6] and the method of moments often give suboptimal bounds in (3.1.2) as they deal with the whole distribution of Z instead of focusing on the probability at 0. Our bounds for (3.1.2) do not require computation of high moments and the proofs are based on elementary techniques inspired by local lemma. To demonstrate the simplicity in application and effectiveness of our bounds, we derive new results on distributions of extremal characteristics of Gaussian systems and of maximal pattern extensions counts in random network models.

The chapter is organised as follows. Our new bounds for the extremal independence property (3.1.1) are stated in Section 3.2 as Theorem 3.1. In Section 3.2.1, we give a detailed comparison of Theorem 3.1 to the related results including aforementioned papers [4, 5, 32, 31]. In Section 3.2.2, we give two useful lemmas that facilitate verifying the assumptions. We prove Theorem 3.1 in Section 3.3: the upper and lower bounds are treated separately in Section 3.3.1 and Section 3.3.3, respectively. In Section 3.4, we apply Theorem 3.1 for finding the asymptotic distribution of maximum number of pattern extensions in binomial random graphs.

3.2 Sufficient conditions for extremal independence

Let $\mathbf{A} := (A_i)_{i\in[d]}$ be a system of events. Everywhere below we assume that $\mathbf{P}(A_i) \neq 0$. Clearly, this assumption does not lead the loss of the generality since the events of zero probability can be excluded from $\mathbf{A}$ without affecting the expression in (3.1.2). We represent the dependencies among the events of

$\mathbf{A}$ by a graph $\mathbf{D}$ on the vertex set $[d]$ with edges indicating the pairs of 'strongly dependent' events, while non-adjacent vertices correspond to 'weakly dependent' events. One can think of $\mathbf{D}$ as a set system $(D_i)_{i\in[d]}$, where $D_i \subseteq [d]$ is the closed neighbourhood of vertex i in graph $\mathbf{D}$. Moreover, we allow $\mathbf{D}$ to be a directed graph, that is, there might exist $i, j \in [d]$, such that $i \in D_j$ and $j \notin D_i$.

To measure the quality of the representation of the dependencies for $\mathbf{A}$ by a graph $\mathbf{D}$, we introduce the following mixing coefficient:

$$\varphi(\mathbf{A},\mathbf{D}) := \max_{i\in[d]} \left| \mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \;\middle|\; A_i \right) - \mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \right) \right|. \tag{3.2.1}$$

This is a special case of ϕ-mixing coefficient widely used in the probability theory; see, for example, survey [13].

The influence of 'strongly dependent' events is measured by declustering coefficients Δ_1 and Δ_2 defined by

$$\Delta_1(\mathbf{A},\mathbf{D}) := \sum_{i\in[d]} \mathbf{P}\left(A_i \cap \bigcup_{j\in[i-1]\cap D_i} A_j \right) \prod_{k\in[d]\setminus[i]} \mathbf{P}\left(\overline{A_k}\right), \tag{3.2.2}$$

$$\Delta_2(\mathbf{A},\mathbf{D}) := \sum_{i\in[d]} \mathbf{P}\left(A_i\right) \mathbf{P}\left(\bigcup_{j\in[i-1]\cap D_i} A_j \right) \prod_{k\in[d]\setminus[i]} \mathbf{P}\left(\overline{A_k}\right). \tag{3.2.3}$$

In our model, the choice of graph $\mathbf{D}$ is arbitrary, and therefore the flexibility may leads to the trade-off between the mixing coefficient $\varphi(\mathbf{A},\mathbf{D})$ and declustering coefficients $\Delta_1(\mathbf{A},\mathbf{D})$ and $\Delta_2(\mathbf{A},\mathbf{D})$ for different applications, since $\Delta_1(\mathbf{A},\mathbf{D})$ and $\Delta_2(\mathbf{A},\mathbf{D})$ increase as $\mathbf{D}$ gets denser, and $\varphi(\mathbf{A},\mathbf{D})$ may decreases.

By the inclusion-exclusion principle, we write Δ_1 and Δ_2 via clusters as

$$\Delta_1(\mathbf{A},\mathbf{D}) = \sum_{i\in[d]} \sum_{\emptyset\neq S\subseteq[i-1]\cap D_i} (-1)^{|S|-1} \mathbf{P}\left(A_i \cap \bigcap_{j\in S} A_j \right) \prod_{k\in[d]\setminus[i]} \mathbf{P}\left(\overline{A_k}\right),$$

$$\Delta_2(\mathbf{A},\mathbf{D}) = \sum_{i\in[d]} \sum_{\emptyset\neq S\subseteq[i-1]\cap D_i} (-1)^{|S|-1} \mathbf{P}\left(A_i\right) \mathbf{P}\left(\bigcap_{j\in S} A_j \right) \prod_{k\in[d]\setminus[i]} \mathbf{P}\left(\overline{A_k}\right).$$

We are ready to state our sufficient condition for satisfying (3.1.2).

Theorem 3.1. *For any system of events $\mathbf{A} = (A_i)_{i\in[d]}$ and graph $\mathbf{D}$ with vertex set $[d]$, the following bound holds*

$$\left| \mathbf{P}\left(\bigcap_{i\in[d]} \overline{A_i} \right) - \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right) \right| \leqslant \left(1 - \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right) \right) \varphi + \max\{\Delta_1, \Delta_2\}, \tag{3.2.4}$$

where $\varphi = \varphi(\mathbf{A},\mathbf{D})$, $\Delta_1 = \Delta_1(\mathbf{A},\mathbf{D})$, and $\Delta_2 = \Delta_2(\mathbf{A},\mathbf{D})$.

Although the proof of Theorem 3.1 is elementary (see Section 3.3), it gives a very useful and convenient tool to prove extremal independence property (3.1.1) stated below.

Corollary 3.2.1. Let $d = d(n) \in \mathbb{N}$, $\mathbf{X}(n) = (X_1, \ldots, X_d)^T \in \mathbb{R}^d$, and $\mathbf{A}$ is defined in (3.1.3). If for every fixed $x \in \mathbb{R}$, there is a graph $\mathbf{D} = \mathbf{D}(n, x)$ such that

$$\varphi(\mathbf{A}, \mathbf{D}) = o(1), \quad \Delta_1(\mathbf{A}, \mathbf{D}) = o(1), \quad \Delta_2(\mathbf{A}, \mathbf{D}) = o(1), \tag{3.2.5}$$

then, (3.1.1) holds.

Corollary 3.2.1 can be applied to extremal problems. We extend Bollobás result [10] on the limit distribution of the maximum degree of binomial random graph $\mathcal{G}(n, p)$ to the hypergraph setting; see Section 3.4.1. Our result on the distribution of maximum extension counts implies the law of large numbers by Spencer [95] and optimizes the denominator for clique extensions; see Sections 3.4.2–3.4.4. Corollary 3.2.1 simplifies the arguments of [84] for the maximum number of h-neighbours and extends it to unbounded h; see Section 3.4.3.

Recent results [98, 77, 104] derive more accurate estimates for $\mathbf{P}\left(\bigcap_{i\in[d]} \overline{A_i}\right)$ using truncated cumulant series and investigating clusters of dependent random variables. It will be interesting to obtain similar extensions of Theorem 3.1 relying on bounds for clusters of strongly dependent random variables.

3.2.1 Related results

By the union bound, it is easy to see that

$$\begin{aligned}
\Delta_1(\mathbf{A}, \mathbf{D}) \leqslant \Delta_1'(\mathbf{A}, \mathbf{D}) &:= \sum_{i\in[d]} \sum_{j\in[i-1]\cap D_i} \mathbf{P}\left(A_i \cap A_j\right), \\
\Delta_2(\mathbf{A}, \mathbf{D}) \leqslant \Delta_2'(\mathbf{A}, \mathbf{D}) &:= \sum_{i\in[d]} \sum_{j\in[i-1]\cap D_i} \mathbf{P}\left(A_i\right) \mathbf{P}\left(A_j\right).
\end{aligned}$$

The declustering assumption $\Delta_1'(\mathbf{A}, \mathbf{D}) = o(1)$ is typical in the study of extremal characteristics of random systems. It guarantees that the clusters of exceedances A_i are negligible. The assumption $\Delta_2'(\mathbf{A}, \mathbf{D}) = o(1)$ is easy to verify. For example, if all probabilities $\mathbf{P}\left(A_i\right)$ are of the same order n^{-1}, then this assumption is equivalent to the graph $\mathbf{D}$ to be sparse, which usually happens in applications. In addition, $\Delta_2'(\mathbf{A}, \mathbf{D})$ can be bounded above by $\Delta_1'(\mathbf{A}, \mathbf{D}) = o(1)$ if the events are monotone. The most innovative part of Corollary 3.2.1 is the remaining assumption $\varphi(\mathbf{A}, \mathbf{D}) = o(1)$, which is often easier to check and less restrictive than other mixing assumptions known in the literature. The detailed comparisons are given below.

First, we consider a stationary sequence of random variables. If its distant terms are 'weakly dependent', then we can construct the graph $\mathbf{D}$ by connecting vertices that are close to each other. Then, omitting some details, the following corresponds to Leadbetter's mixing condition D:

$$\left| \mathbf{P}\left(\bigcap_{i\in I\cup J} \overline{A_i}\right) - \mathbf{P}\left(\bigcap_{i\in I} \overline{A_i}\right) \mathbf{P}\left(\bigcap_{i\in J} \overline{A_i}\right) \right| = o(1) \tag{3.2.6}$$

for all disjoint $I, J \subset [d]$ with no edges from $\mathbf{D}$ between them, see [63, Eq. (1.2)]. Although, (4.3.5) looks similar to our assumption $\varphi(\mathbf{A}, \mathbf{D}) = o(1)$, none of them implies the other. One advantage of our assumption in comparison with (4.3.5) is that one only needs to check the mixing condition for considerably fewer pairs of sets I and J, namely for $I = [i-1] \setminus D_i$ and $J = \{i\}$ for all $i \in [d]$. The same conclusion remains valid for the extensions of Leadbetter's mixing condition D for non-stationary sequences and random fields on $\mathbb{Z}_+^2$, see, Hüsler [43, Theorem 1.1] and Pereira, Ferreira [83, Proposition

3.2], respectively. In fact, our framework is much more flexible since one can arbitrarily choose the graph $\mathbf{D}$, without relying on the distances between indices.

Second, we consider the case when $\varphi(\mathbf{A},\mathbf{D}) = 0$. For this case, under some additional requirement, Dubickas [20, Theorem 1] proved the following bound:

$$\mathbf{P}\left(\bigcap_{i\in[d]} \overline{A_i}\right) \geqslant \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right) - \Delta_2(\mathbf{A},\mathbf{D}). \tag{3.2.7}$$

Thus, in this case, (3.2.7) gives the lower bound for $\mathbf{P}\left(\bigcap_{i\in[d]} \overline{A_i}\right)$ similar to Theorem 3.1. In the binomial subset setting and under condition $\Delta_1'(\mathbf{A},\mathbf{D}) = o(1)$, the matching upper bound for $\mathbf{P}\left(\cap_{i\in[d]}\overline{A_i}\right)$ can be derived from Janson's inequality [57]. Our graph-dependent model is also related to the notions of lopsided (negative) dependency graph [22] and ϵ-near-positive dependency graph [71]. Those are models with one-sided mixing conditions sufficient for the lower and upper bounds respectively.

Next, we compare Corollary 3.2.1 with the results by Galambos [32, 31]. To our knowledge, he was the first to represent the weak and strong dependencies among $(A_i)_{i\in[d]}$ by a graph. Galambos established the extremal independence property (3.1.1) using the so-called graph-sieve method; see, for example, [33] for detailed overview. In particular, Galambos' mixing assumptions require that, for a fixed graph $\mathbf{D}$,

$$\sum_S \left|\mathbf{P}\left(\bigcap_{i\in S} A_i\right) - \prod_{i\in S} \mathbf{P}\left(A_i\right)\right| = o(1), \tag{3.2.8}$$

where the sum in (3.2.8) is over all $S \subseteq [d]$ with no edges of $\mathbf{D}$. Assumption (3.2.8) is very restrictive for many applications since such set S can be large. For example, in some of the applications that we consider in Section 3.4, the graph $\mathbf{D}$ is empty so the results in [32, 31] is of little use, since assumption (3.2.8) is equivalent to the extremal independence property (3.1.1) that we wish to establish.

To illustrate the advantage of our approach with respect to the methods of moments, we briefly consider the following example. Let A_i, where $i \in [d]$ and $d = \binom{n}{h}$, to be the event that the number of common neighbors of one of the corresponding h-subsets of vertices in $\mathcal{G}(n,p)$ is greater than $a_n + b_n x$ (for some appropriately chosen a_n, b_n). In Section 3.4.3, we show that this system of events obey the asymptotic independence property (3.1.2) despite the fact that the second moment of $Z = \sum_{i\in[d]} \mathbf{1}_{\{A_i\}}$ approaches infinity when p is a sufficiently large constant (depending on h). In fact, one can get around this difficulty and modify the random variables so the second moment converges to the desired limit by conditioning on a certain event $\mathcal{E}_n$ that holds with probability $1 - o(1)$. However, it does not help a lot even for the third moment, and it is not evident that the convergence of the higher moments can be established directly by a careful choice of random variables.

The aforementioned difficulty in applying the method of moments was also pointed out by Arratia, Goldstein and Gordon in [4, 5]. Based on the Chen-Stein method they discovered that the computation of two moments is sufficient for Poisson approximation under a certain mixing condition for weakly dependent random variables. For the rest of this section, we compare [5, Theorem 3] with our Theorem 3.1 as these results have very similar setups.

Arratia et al. [4, 5] introduced another mixing coefficient different from our φ:

$$\widetilde{\varphi} := \sum_{i\in[d]} \mathbf{P}\left(A_i\right) \sum_{k=0}^{d} \left|\mathbf{P}\left(Z^i = k \mid A_i\right) - \mathbf{P}\left(Z^i = k\right)\right|,$$

where $Z^i = \sum_{j\notin D_i} \mathbf{1}_{\{A_i\}}$. Their result [5, Theorem 3] states that

$$\left|\mathbf{P}\left(\bigcap_{i\in[d]} \overline{A_i}\right) - \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right)\right| \leqslant 2\widetilde{\varphi} + 4\Delta_1'' + 4\Delta_2'' + 4\sum_{i\in[d]} \mathbf{P}\left(A_i\right)^2, \tag{3.2.9}$$

where

$$\Delta_1'' = \sum_{i\in[d]}\sum_{j\in D_i} \mathbf{P}\left(A_i \cap A_j\right) \geqslant \Delta_1', \quad \Delta_2'' = \sum_{i\in[d]}\sum_{j\in D_i} \mathbf{P}\left(A_i\right)\mathbf{P}\left(A_j\right) \geqslant \Delta_2'.$$

To compare $\widetilde{\varphi}$ with our mixing coefficient φ, we observe that

$$\widetilde{\varphi} \geqslant \sum_{i\in[d]} \mathbf{P}\left(A_i\right) \left|\mathbf{P}\left(\bigcup_{j\notin D_i} A_j \,\middle|\, A_i\right) - \mathbf{P}\left(\bigcup_{j\notin D_i} A_j\right)\right|, \tag{3.2.10}$$

In the typical case when $\sum_{i\in[d]} \mathbf{P}\left(A_i\right) = \Theta(1)$, $\sum_{i\in[d]} (\mathbf{P}\left(A_i\right))^2 = o(1)$ (and up to ordering of vertices in $\mathbf{D}$) the RHS of (3.2.10) has the same order of magnitude (or even bigger) as $\left(1 - \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right)\right)\varphi$. Thus, our bound is at least as efficient as (3.2.9) for such applications. Moreover, the lower bound (3.2.10) on $\widetilde{\varphi}$ could be far from being sharp, that is, the actual value of the mixing coefficient $\widetilde{\varphi}$ could be much bigger. Furthermore, Theorem 3.1 surpasses [5, Theorem 3] in several important instances listed below.

(1) **Slowly decreasing $\sum_{i\in[n]} \left(\mathbf{P}\left(A_i\right)\right)^2$.** Clearly, Theorem 3.1 does not have this error term. Thus, our results partially answer the question formulated by Arratia et al. [4, 5] about the extremal independence property (3.1.1) in case when Poisson approximation is not good enough.

(2) **Slowly growing $\sum_{i\in[n]} \mathbf{P}\left(A_i\right)$.** The term $\left(1 - \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right)\right)\varphi$ has additional advantage for upper tail estimates where $\prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right) \to 1$.

(3) **Inhomogenous random graphs.** For example, consider the random graph model with vertex set $[n]$ and independent adjacencies, where all adjacencies happen with probability p excluding adjacencies incident to one special vertex. The edges incident to this vertex appear with a slightly higher probability $p' = \frac{a_n + b_n x}{n} = p + (1 - o(1))\sqrt{\frac{2p(1-p)\ln n}{n}}$, where a_n, b_n and constant $x \in \mathbb{R}$ are appropriately chosen. Defining A_i as the event that vertex i in the considered random graph has degree more than $a_n + b_n x$, our inequality gives the upper bound $O(n^{-1/2})$ in (3.2.4) while [5, Theorem 3] gives a useless bound $O(1)$.

3.2.2 Bridging sequences

Here, we state two helpful lemmas in applying Theorem 3.1 to study the extremal characteristics of random combinatorial structures. It will be convenient to work with non-scaled random variables $\{X_i\}$. Everywhere in this section, we assume the following:

- $\mathbf{X}(n) = (X_1, \ldots, X_d)^T \in \mathbb{R}^d$ is a sequence of random vectors, where $d = d(n) \in \mathbb{N}$;

- F is a continuous cdf on $\mathbb{R}$ and $\mathcal{X}$ is the set of all $x \in \mathbb{R}$ such that $0 < F(x) < 1$;
- there exist a_n and b_n such that $\prod_{i=1}^d \mathbf{P}\left(X_i \leqslant a_n + b_n x\right) \to F(x)$ for any $x \in \mathcal{X}$;
- for all $i \in [d]$, denote $A_i := A_i(x) = \{X_i > a_n + b_n x\}$.

The first lemma shows that $\varphi(\mathbf{A}, \mathbf{D}) \to 0$ as $n \to \infty$ provided that, for all $i \in [d]$ and $j \in [i-1] \setminus D_i$, the random variables X_j are approximated by some random variables $X_j^{(i)}$, which are independent of X_i. We will use this lemma to derive the distribution of the maximum codegrees in random hypergraphs.

Lemma 3.2.2. *Let $x \in \mathcal{X}$. Let sets $D_i \subseteq [d] \setminus \{i\}$ and random variables $X_j^{(i)}$ be such that, for all $j \in [i-1] \setminus D_i$, $X_j^{(i)}$ is independent of X_i and, for any fixed $\varepsilon > 0$,*

$$\mathbf{P}\left(\max_{j \in [i-1] \setminus D_i} \left| X_j - X_j^{(i)} \right| > \varepsilon b_n \right) = o(1) \mathbf{P}\big(A_i\big), \tag{3.2.11}$$

uniformly over $i \in [d]$. Then $\varphi(\mathbf{A}, \mathbf{D}) \to 0$.

Proof of Lemma 3.2.2. Find $\delta > 0$ such that $0 < F(x-\delta) \leqslant F(x+\delta) < 1$. Let $\varepsilon \in (0, \delta/2)$. We may assume that n is so large that $\mathbf{P}\left(A_i\right) \leqslant \mathbf{P}(A_i(x-2\varepsilon)) < 1$ for all $i \in [d]$ (otherwise, $\prod_{i=1}^d \mathbf{P}(\overline{A_i(x-2\varepsilon)})$ can not approach $F(x-2\varepsilon)$). For $i \in [d]$ and $j \in [i-1] \setminus D_i$, consider the events $A_i^{\varepsilon} := A_i(x+2\varepsilon)$ and $U_{ji}^{\varepsilon} := \{X_j^{(i)} > a_n + (x+\varepsilon) b_n\}$. Then, from (3.2.11), we get that uniformly over all $i \in [d]$

$$\mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} \left(U_{ji}^{\varepsilon} \setminus A_j\right)\right) = o(1) \mathbf{P}\left(A_i\right) \quad \text{and} \quad \mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} \left(A_j^{\varepsilon} \setminus U_{ji}^{\varepsilon}\right)\right) = o(1) \mathbf{P}\left(A_i\right).$$

The events U_{ji}^{ε} and A_i are independent since $X_j^{(i)}$ is independent of X_i. Therefore,

$$\begin{aligned} \mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} A_j \,\middle|\, A_i\right) &\geqslant \mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} U_{ji}^{\varepsilon} \,\middle|\, A_i\right) - o(1) \\ &= \mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} U_{ji}^{\varepsilon}\right) - o(1) \geqslant \mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} A_j^{\varepsilon}\right) - o(1). \end{aligned}$$

By the union bound, we get that

$$\begin{aligned} &\mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} A_j\right) - \mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} A_j \,\middle|\, A_i\right) \\ &\leqslant \mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} A_j\right) - \mathbf{P}\left(\bigcup_{j \in [i-1] \setminus D_i} A_j^{\varepsilon}\right) + o(1) \\ &\leqslant \sum_{j \in [i-1] \setminus D_i} \mathbf{P}\left(A_j \setminus \bigcup_{s \in [i-1] \setminus D_i} A_s^{\varepsilon}\right) + o(1) \leqslant \sum_{j \in [i-1] \setminus D_i} \mathbf{P}(A_j \setminus A_j^{\varepsilon}) + o(1). \end{aligned}$$

Using the inequality $\sum_{i \in [d]} t_i \leqslant -1 + \prod_{i \in [d]}(1+t_i)$, where $t_i := \frac{\mathbf{P}(A_i) - \mathbf{P}(A_i^{\varepsilon})}{1 - \mathbf{P}(A_i)} \geqslant 0$, and recalling that

$F(x) > 0$, we estimate

$$\sum_{i\in[d]} \mathbf{P}(A_i \setminus A_i^\varepsilon) \leqslant \sum_{i\in[d]} \frac{\mathbf{P}(A_i)-\mathbf{P}(A_i^\varepsilon)}{1-\mathbf{P}(A_i)} \leqslant -1 + \prod_{i\in[d]} \frac{1-\mathbf{P}(A_i^\varepsilon)}{1-\mathbf{P}(A_i)} \to \frac{F(x+2\varepsilon)}{F(x)} - 1.$$

Recalling that F is continuous at x and that the above holds for any $\varepsilon \in (0, \delta/2)$, we conclude that

$$\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j\right) - \mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \;\middle|\; A_i\right) \leqslant o(1).$$

The lower bound

$$\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j\right) - \mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \;\middle|\; A_i\right) \geqslant o(1)$$

is obtained similarly by using the events $A_j^{-\varepsilon} := A_i(x - 2\varepsilon)$, $U_{ji}^{-\varepsilon} := \{X_j^{(i)} > a_n + (x - \varepsilon)b_n\}$ and the relations

$$\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} \left(A_j \setminus U_{ji}^{-\varepsilon}\right)\right) = o(1)\mathbf{P}\left(A_i\right), \quad \mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} \left(U_{ji}^{-\varepsilon} \setminus A_j^{-\varepsilon}\right)\right) = o(1)\mathbf{P}\left(A_i\right),$$

that hold uniformly over all $i \in [d]$. This completes the proof of Lemma 3.2.2. $\square$

The second lemma allows us to transfer the asymptotic distribution of the maximum component of $\mathbf{X}(n)$ to any random vector $\mathbf{Y}(n) \in \mathbb{R}^d$ that 'approximates' $\mathbf{X}(n)$. Using this lemma, we will derive the distribution of the maximum clique-extension count in random graphs from the results on the maximum degree.

Lemma 3.2.3. *Let $\mathbf{Y}(n) \in \mathbb{R}^d$ be a sequence of random vectors. Assume that, for any $x \in \mathcal{X}$,*

(i) $\mathbf{P}\left(\max_{i\in[d]} X_i \leqslant a_n + b_n x\right) \to F(x)$;

(ii) for any fixed $\varepsilon > 0$,

$$\mathbf{P}(|X_i - Y_i| > \varepsilon b_n) = o(1)\mathbf{P}(X_i > a_n + b_n x),$$

uniformly over all $i \in [d]$.

Then $\mathbf{P}\left(\max_{i\in[d]} Y_i \leqslant a_n + b_n x\right) \to F(x)$ for all $x \in \mathcal{X}$.

Proof of Lemma 3.2.3. Find $\delta > 0$ such that $0 < F(x - \delta) \leqslant F(x + \delta) < 1$. Let $\varepsilon \in (0, \delta)$. Let $A_i^\varepsilon := A_i(x + \varepsilon)$, $B_i := \{Y_i > a_n + b_n x\}$. From assumption (i), we get

$$1 - \mathbf{P}\left(\bigcup_{i\in[d]} A_i^\varepsilon\right) \to F(x + \varepsilon).$$

Also by the third assumption from the list of preliminary assumptions at the beginning of Subsection 3.2.2, we obtain

$$\prod_{i\in[d]} (1 - \mathbf{P}(A_i^\varepsilon)) \to F(x + \varepsilon).$$

Therefore, we conclude

$$1-\mathbf{P}\left(\bigcup_{i\in[d]} A_i^\varepsilon\right) \sim \prod_{i\in[d]} (1-\mathbf{P}(A_i^\varepsilon)) \to F(x+\varepsilon). \tag{3.2.12}$$

Since $F(x+\varepsilon) \geqslant F(x-\varepsilon) > 0$, we get

$$\sum_{i\in[d]} \mathbf{P}(A_i^\varepsilon) \leqslant -\sum_{i\in[d]} \log\left(1-\mathbf{P}(A_i^\varepsilon)\right) = O(1). \tag{3.2.13}$$

From (ii), we find that $\mathbf{P}(A_i^\varepsilon \setminus B_i) = o(1)\mathbf{P}(A_i^\varepsilon)$. Then the relations

$$\begin{aligned}\mathbf{P}\left(\bigcup_{i\in[d]} A_i^\varepsilon\right) - \mathbf{P}\left(\bigcup_{i\in[d]} B_i\right) &\leqslant \mathbf{P}\left(\bigcup_{i\in[d]} A_i^\varepsilon \setminus \bigcup_{i\in[d]} B_i\right)\\ &\leqslant \sum_{i\in[d]} \mathbf{P}\left(A_i^\varepsilon \setminus \bigcup_{j\in[d]} B_j\right) \leqslant \sum_{i\in[d]} \mathbf{P}\left(A_i^\varepsilon \setminus B_i\right)\end{aligned}$$

imply

$$\mathbf{P}\left(\bigcup_{i\in[d]} B_i\right) \geqslant \mathbf{P}\left(\bigcup_{i\in[d]} A_i^\varepsilon\right) - o(1)\sum_{i\in[d]} \mathbf{P}(A_i^\varepsilon) = 1 - F(x+\varepsilon) - o(1).$$

The last equality follows from (3.2.12) and (3.2.13). Recalling that F is continuous and that the above holds for any $\varepsilon \in (0,\delta)$, we conclude that $1-\mathbf{P}\left(\bigcup_{i\in[d]} B_i\right) \leqslant F(x)+o(1)$. The lower bound $1-\mathbf{P}\left(\bigcup_{i\in[d]} B_i\right) \geqslant F(x)-o(1)$ is obtained similarly, using the events $A_i^{-\varepsilon} = A_i(x-\varepsilon)$ and the relations $\mathbf{P}(B_i \setminus A_i^{-\varepsilon}) = o(1)\mathbf{P}(A_i^{-\varepsilon})$ that follow directly from (ii). □

3.3 The probability of non-occurrence under graphical φ-mixing

In this section, we give new lower and upper bounds that allow to make a classification of dependencies between events flexible and that do not require the implication from pairwise to mutual independence. Our bounds are follow-up to the inequalities of Arratia, Goldstein, Gordon [4, 5] and give a certain improvement for applications in various settings (see Section 3.2.1). However, the proofs are elementary and inspired by the proof of the Lovász Local Lemma. Note that our lower bound given in Section 3.3.2 is a strict generalisation of Dubickas' inequality [20].

3.3.1 Upper bound

Here and in the next section, we use the notations $\Delta_1(\mathbf{A},\mathbf{D})$ and $\Delta_2(\mathbf{A},\mathbf{D})$ that are defined in (3.2.2) and (3.2.3) respectively.

Lemma 3.3.1. *Let $\varphi \geqslant 0$. If events $(A_i)_{i\in[d]}$ with non-zero probabilities and sets $(D_i \subset [d]\setminus\{i\})_{i\in[d]}$ satisfy*

$$\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \,\middle|\, A_i\right) - \mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j\right) \leqslant \varphi, \tag{3.3.1}$$

for all $i \in [d]$, then

$$\mathbf{P}\left(\bigcap_{i\in[d]} \overline{A_i}\right) \leqslant \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right) + \varphi\left(1 - \prod_{i\in[d]} \mathbf{P}\left(\overline{A_i}\right)\right) + \Delta_1(\mathbf{A}, \mathbf{D}). \tag{3.3.2}$$

Proof. Let us prove that, for every $s \in [d]$,

$$\mathbf{P}\left(\bigcap_{i\in[s]} \overline{A_i}\right) \leqslant (1-\varphi)\prod_{i\in[s]} \mathbf{P}\left(\overline{A_i}\right) + \varphi + \sum_{i\in[s]} \mathbf{P}\left(A_i \cap \bigcup_{j\in[i-1]\cap D_i} A_j\right) \prod_{k\in[s]\setminus[i]} \mathbf{P}\left(\overline{A_k}\right) \tag{3.3.3}$$

by induction on s. The required bound (3.3.2) is exactly (3.3.3) when $s = d$.

For $s = 1$, (3.3.3) follows from $\varphi \geqslant 0$. Assume that (3.3.3) holds for some $s \in [d-1]$. Let

$$B := \bigcup_{j\in[s]\setminus D_{s+1}} A_j, \qquad C := \bigcup_{j\in[s]\cap D_{s+1}} A_j. \tag{3.3.4}$$

Note that

$$1 - \mathbf{P}\left(\overline{A_{s+1}} \mid \bigcap_{i\in[s]} \overline{A_i}\right) = \mathbf{P}\left(A_{s+1} \mid \overline{B} \cap \overline{C}\right) \geqslant \mathbf{P}\left(A_{s+1} \mid \overline{B}\right)\left(1 - \mathbf{P}\left(C \mid A_{s+1} \cap \overline{B}\right)\right). \tag{3.3.5}$$

By (3.3.1), we have $\mathbf{P}\left(\overline{B} \mid A_{s+1}\right) \geqslant \mathbf{P}\left(\overline{B}\right) - \varphi$. Therefore,

$$\mathbf{P}\left(A_{s+1} \mid \overline{B}\right) = \frac{\mathbf{P}\left(\overline{B}|A_{s+1}\right)}{\mathbf{P}(\overline{B})}\mathbf{P}\left(A_{s+1}\right) \geqslant \left(1 - \frac{\varphi}{\mathbf{P}\left(\overline{B}\right)}\right)\mathbf{P}\left(A_{s+1}\right).$$

We also find that

$$\mathbf{P}\left(C \mid A_{s+1} \cap \overline{B}\right) = \frac{\mathbf{P}\left(C \cap \overline{B} \mid A_{s+1}\right)}{\mathbf{P}\left(\overline{B} \mid A_{s+1}\right)} \leqslant \frac{\mathbf{P}(C|A_{s+1})}{\mathbf{P}\left(\overline{B}\right) - \varphi}.$$

Using the above two bounds in (3.3.5), we derive that

$$\mathbf{P}\left(\overline{A_{s+1}} \mid \bigcap_{i\in[s]} \overline{A_i}\right) \leqslant 1 - \left(1 - \frac{\varphi}{\mathbf{P}\left(\overline{B}\right)}\right)\mathbf{P}\left(A_{s+1}\right) + \frac{\mathbf{P}(A_{s+1}\cap C)}{\mathbf{P}\left(\overline{B}\right)}.$$

Then, since $\mathbf{P}\left(\overline{B}\right) \geqslant \mathbf{P}\left(\bigcap_{i\in[s]} \overline{A_i}\right)$, we get

$$\mathbf{P}\left(\bigcap_{i\in[s+1]} \overline{A_i}\right) \leqslant \mathbf{P}\left(\bigcap_{i\in[s]} \overline{A_i}\right)\mathbf{P}\left(\overline{A_{s+1}}\right) + \varphi\mathbf{P}\left(A_{s+1}\right) + \mathbf{P}(A_{s+1} \cap C).$$

By (3.3.3), we have

$$\mathbf{P}\left(\bigcap_{i\in[s+1]}\overline{A_i}\right) \leqslant (1-\varphi)\prod_{i\in[s+1]}\mathbf{P}\left(\overline{A_i}\right)+\varphi$$
$$+\sum_{i\in[s+1]}\mathbf{P}\left(A_i\cap\bigcup_{j\in[i-1]\cap D_i}A_j\right)\prod_{k\in[s+1]\setminus[i]}\mathbf{P}\left(\overline{A_k}\right).$$

This completes the proof. □

Remark 3.3.2. Lu and Székely introduce the notion of ϵ-near-positive dependency graphs in [71]. Their assumptions are

$$\Delta_1'(\mathbf{A},\mathbf{D}) := \sum_{i\in[d]}\sum_{j\in[i-1]\cap D_i}\mathbf{P}\left(A_i\cap A_j\right)=0,$$

and the following mixing condition to hold for all $i\in[d]$ for some epsilon $0<\epsilon<1$:

$$\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\,\middle|\,A_i\right)-\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\right)\leqslant\epsilon\,\mathbf{P}\left(\bigcap_{j\in[i-1]\setminus D_i}\overline{A_j}\right).$$

Under these assumptions, they obtained

$$\mathbf{P}\left(\bigcap_{i\in[d]}\overline{A_i}\right)\leqslant\prod_{i\in[d]}\mathbf{P}\left(\overline{A_i}\right)\left(1+\epsilon\frac{\mathbf{P}(A_i)}{\mathbf{P}\left(\overline{A_i}\right)}\right).$$

3.3.2 Lower bound

Lemma 3.3.3 (Generalised Dubickas' inequality)**.** *Let $\varphi\geqslant 0$. If events $(A_i)_{i\in[d]}$ with non-zero probabilities and sets $D_i\subset[d]\setminus\{i\}$ satisfy*

$$\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\right)-\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\,\middle|\,A_i\right)\leqslant\varphi, \tag{3.3.6}$$

for all $i\in[d]$, then

$$\mathbf{P}\left(\bigcap_{i\in[d]}\overline{A_i}\right)\geqslant\prod_{i\in[d]}\mathbf{P}\left(\overline{A_i}\right)-\varphi\left(1-\prod_{i\in[d]}\mathbf{P}\left(\overline{A_i}\right)\right)-\Delta_2(\mathbf{A},\mathbf{D}). \tag{3.3.7}$$

Proof. Let us prove that, for every $s\in[d]$,

$$\mathbf{P}\left(\bigcap_{i\in[s]}\overline{A_i}\right)\geqslant(1+\varphi)\prod_{i\in[s]}\mathbf{P}\left(\overline{A_i}\right)-\varphi-\sum_{i\in[s]}\mathbf{P}\left(A_i\right)\mathbf{P}\left(\bigcup_{j\in[i-1]\cap D_i}A_j\right)\prod_{k\in[s]\setminus[i]}\mathbf{P}\left(\overline{A_k}\right) \tag{3.3.8}$$

by induction on s. The required bound (3.3.7) is exactly (3.3.8) when $s=d$.

For $s=1$, (3.3.8) is straightforward since $\varphi \geqslant 0$. Assume that (3.3.8) holds for $s \in [d-1]$. Consider the events B and C defined in (3.3.4). Then

$$\mathbf{P}\left(\overline{A_{s+1}} \mid \bigcap_{i\in[s]} \overline{A_i}\right) = 1 - \frac{\mathbf{P}\left(A_{s+1}\cap\overline{B}\cap\overline{C}\right)}{\mathbf{P}\left(\overline{B}\cap\overline{C}\right)} \geqslant 1 - \frac{\mathbf{P}\left(\overline{B}|A_{s+1}\right)}{\mathbf{P}\left(\overline{B}\cap\overline{C}\right)}\mathbf{P}\left(A_{s+1}\right).$$

From (3.3.6), we have $\mathbf{P}(\overline{B}|A_{s+1}) \leqslant \mathbf{P}(\overline{B}) + \varphi$. Therefore,

$$\mathbf{P}\left(\overline{A_{s+1}} \mid \bigcap_{i\in[s]} \overline{A_i}\right) \geqslant 1 - \frac{\mathbf{P}\left(\overline{B}\right)+\varphi}{\mathbf{P}\left(\overline{B}\cap\overline{C}\right)}\mathbf{P}\left(A_{s+1}\right). \tag{3.3.9}$$

Moreover,

$$\mathbf{P}\left(\overline{B}\right) = \mathbf{P}\left(\overline{B}\cap\overline{C}\right) + \mathbf{P}\left(\overline{B}\cap C\right) \leqslant \mathbf{P}\left(\bigcap_{i\in[s]} \overline{A_i}\right) + \mathbf{P}\left(C\right). \tag{3.3.10}$$

Combining (3.3.8), (3.3.9) and (3.3.10), we get

$$\begin{aligned}\mathbf{P}\left(\bigcap_{i\in[s+1]} \overline{A_i}\right) &\geqslant \mathbf{P}\left(\bigcap_{i\in[s]} \overline{A_i}\right)\mathbf{P}\left(\overline{A_{s+1}}\right) - \varphi\mathbf{P}\left(A_{s+1}\right) - \mathbf{P}\left(\bigcup_{j\in[s]\cap D_{s+1}} A_j\right)\mathbf{P}\left(A_{s+1}\right)\\ &\geqslant (1+\varphi)\prod_{i\in[s+1]}\mathbf{P}\left(\overline{A_i}\right) - \varphi - \sum_{i\in[s+1]}\mathbf{P}\left(A_i\right)\mathbf{P}\left(\bigcup_{j\in[i-1]\cap D_i} A_j\right)\prod_{k\in[s+1]\setminus[i]}\mathbf{P}\left(\overline{A_k}\right).\end{aligned}$$

This completes the proof. □

Remark 3.3.4. As mentioned in Section 3.2.1, the special case of (3.3.7) with $\varphi = 0$ proves Dubickas' inequality (3.2.7). Note also that our condition

$$\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j\right) \leqslant \mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i} A_j \,\middle|\, A_i\right),$$

is weaker than the Dubickas' requirement on the connection between pairwise and mutual independencies.

3.4 Applications in binomial random graphs

Let us recall that $\mathcal{G}(n,p)$ is a random graph on the vertex set $[n] = \{1,\ldots,n\}$ distributed as

$$\mathbf{P}(\mathcal{G}(n,p) = G) = p^{e(G)}(1-p)^{\binom{n}{2}-e(G)},$$

where $e(G)$ is the number of edges of a graph G with the vertex set $[n]$ (that is, every pair of distinct vertices of $[n]$ is adjacent with probability p independently of all the others).

In [10], Bollobás proved that, for $p = \Theta(1)$, the maximum degree Δ of $\mathcal{G}(n,p)$ after appropriate rescaling converges to Gumbel distribution. More formally, there exist sequences a_n and b_n (the exact values are known) such that $\frac{\Delta - a_n}{b_n}$ converges in distribution to a standard Gumbel random variable.

Ivchenko proved [51] that the same holds for p such that $p(1-p) \gg \frac{\log n}{n}$. In other words, for the rescaled degree sequence of $\mathcal{G}(n,p)$, the extremal independence property (3.1.1) holds. This is not unexpected since the dependence of degrees of two vertices of the random graph is 'focused' in the only edge between these vertices. In Section 3.4.1, we show that Theorem 3.1 implies the same result for the maximum degree of binomial random *hypergraph* that can not be obtained by the approach of Bollobás and Ivchenko directly.

The results of Bollobás and Ivchenko can be viewed as a particular case of the following problem suggested by Spencer in [95]. Let G be a graph, and H be its subgraph on h vertices. Define

$$d(H,G) = \frac{|E(G)|-|E(H)|}{|V(G)|-|V(H)|}$$

(here, as usual, $V(G)$ and $E(G)$ are the set of vertices and the set of edges of G respectively). Let the pair (H,G) be *strictly balanced* and *grounded*, that is,

- for every S such that $H \subsetneq S \subsetneq G$, we have $d(H,S) < d(H,G)$, and
- there is an edge between $V(H)$ and $V(G) \setminus V(H)$ in G.

For brevity, we denote by $[n]_h$ and $\binom{[n]}{h}$ the set of all h-tuples of distinct vertices from $[n]$ and the set of all h-subsets of $[n]$, respectively. For an h-tuple $\mathbf{x} = (x_1, \ldots, x_h) \in [n]_h$, denote by $X_{\mathbf{x}}$, the number of (H,G)*-extensions* of $\mathbf{x}$ in $\mathcal{G}(n,p)$ (that is, the number of copies of $(V(G), E(G) \setminus E(H))$ in $\mathcal{G}(n,p)$ in which each vertex v_j, $j \in [h]$, of H maps onto x_j). For example, the degree of a vertex u equals X_u when $h = 1$ and $G = K_2$ (as usual, we denote by K_r a complete graph on r vertices and call it an r*-clique*). Spencer raised the question about the deviation of $X_{\mathbf{x}}$ from its expectation and proved that

$$\frac{\max_{\mathbf{x} \in [n]_h} |X_{\mathbf{x}} - \mu|}{\mu} \xrightarrow{\mathbf{P}} 0 \tag{3.4.1}$$

whenever $\mu := \mathbf{E}\left[X_{(1,\ldots,h)}\right] = \Theta\left(n^{|V(G)|-|V(H)|} p^{|E(G)|-|E(H)|}\right) \gg \log n$. In Section 3.4.2, we show that Theorem 3.1 results in a tight lower bound of a possible denominator in the law of large numbers (3.4.1) for a slightly more narrow range of p and some specific strictly balanced and grounded (H,G). More precisely, for $h = 1$ and G being a clique (its size may depend on n), we prove that $\max_{u \in [n]} X_u$ after appropriate rescaling converges to Gumbel distribution. Moreover, as we discuss in Sections 3.4.3 and 3.4.4, these techniques can be applied for $h > 1$ as well.

3.4.1 Maximum degree and codegree

Let $H_k(n,p)$ be the k-uniform binomial random hypergraph with the vertex set $[n]$. Recall that every k-set from $\binom{[n]}{k}$ appears as an edge in $H_k(n,p)$ with probability p independently. For a set $S \subseteq [n]$ with $|S| < k$ let X_S be the codegree of S in $H_k(n,p)$ (that is, the number of edges of $H_k(n,p)$ containing S). In particular, X_i is the degree of a vertex i. Note that

$$X_S \sim \mathrm{Bin}\left(\binom{n-|S|}{k-|S|}, p\right). \tag{3.4.2}$$

In this section, using Theorem 3.1, we show that, under some assumptions on the parameters k and p (in terms of n), the asymptotic distribution of $\max_S X_S$ is the same as if the variables X_S were independent. For independent random variables, the asymptotic distribution is given by the following lemma.

Lemma 3.4.1. *Let $d = d(n) \in \mathbb{N}$, $N = N(n) \in \mathbb{N}$, and $p = p(n) \in (0, 1)$ satisfy*

$$Np(1-p) \gg \log^3 d \gg 1.$$

If $\xi_1, \ldots, \xi_d$ are $\mathrm{Bin}(N, p)$ independent random variables, then $\left[\max_{i\in[d]} \xi_i - a_n\right]/b_n$ converges in distribution to a standard Gumbel random variable with a_n and b_n defined by

$$\begin{aligned} a_n = a_n(d, N, p) &:= pN + \sqrt{2Np(1-p)\log d}\left(1 - \frac{\log\log d}{4\log d} - \frac{\log(2\sqrt{\pi})}{2\log d}\right), \\ b_n = b_n(d, N, p) &:= \sqrt{\frac{Np(1-p)}{2\log d}}. \end{aligned} \tag{3.4.3}$$

Proof. For p bounded away from 0 and 1 we refer to [78, Theorem 3]. For $p \to 0$, $p \gg \frac{\log^3 d}{N}$, we find by [51, Lemmas 4 and 5] that

$$d\,\mathbf{P}(\xi_1 > a_n + b_n x) \to e^{-x}. \tag{3.4.4}$$

Since $d \to \infty$,

$$\mathbf{P}\left(\max_{i\in[d]} \xi_i \leqslant a_n + b_n x\right) = \left(\mathbf{P}(\xi_1 \leqslant a_n + b_n x)\right)^d = \left(1 - \frac{e^{-x}+o(1)}{d}\right)^d \to e^{-e^{-x}}.$$

Finally, if $\frac{\log^3 d}{N} \ll 1 - p = o(1)$, then (3.4.4) can be obtained similarly by applying de Moivre–Laplace theorem (see, e.g., [12, Theorem 1.6]). $\square$

Remark 3.4.2. In fact, Lemma 3.4.1 can be extended to the range $Np(1-p) \gg \log d$. This would involve a more complicated expression for a_n, while b_n remains the same, see [51, Lemma 5]. However, such a generalisation is not needed for the applications we consider.

In the next theorem, we show that, under certain assumptions, the maximum degree in the random hypergraph $H_k(n, p)$ converges to the Gumbel distribution.

Theorem 3.2. *Assume $p = p(n) \in (0, 1)$ and $k = k(n) \in \{2, \ldots, n\}$ are such that*

$$\binom{n-1}{k-1} p(1-p) \gg \log^3 n, \qquad k \ll n/\log^2 n. \tag{3.4.5}$$

Then $\left[\max_{i\in[n]} X_i - a_n\right]/b_n$ converges in distribution to a standard Gumbel random variable, where $a_n = a_n\left(n, \binom{n-1}{k-1}, p\right)$ and $b_n = b_n\left(n, \binom{n-1}{k-1}, p\right)$ are defined in (3.4.3).

Proof. Take any $x \in \mathbb{R}$. For all $i \in [n]$, let $A_i := \{X_i > a_n + x b_n\}$. Let $d := n$ and $N := \binom{n-1}{k-1}$. By Lemma 3.4.1, we find that

$$\prod_{i\in[n]} \mathbf{P}(\overline{A}_i) \to e^{-e^{-x}}. \tag{3.4.6}$$

For $i \in [d]$, let $D_i = \emptyset$. Then $\Delta_1(\mathbf{A}, \mathbf{D}) = \Delta_2(\mathbf{A}, \mathbf{D}) = 0$. By Corollary 3.2.1, we only need to show that $\varphi(\mathbf{A}, \mathbf{D}) = o(1)$. We employ Lemma 3.2.2 to verify it. Note that $F(x) := e^{-e^{-x}}$ is the cdf of the standard Gumbel distribution. In particular, F is continuous and $0 < F(x) < 1$ for all $x \in \mathbb{R}$. To apply Lemma 3.2.2, it remains to construct random variables $X_j^{(i)}$. For $j \in [d] \setminus i$, define $X_j^{(i)} := \mathbf{E}\left[X_j \mid H_i\right]$, where H_i is the set of edges of $H_k(n, p)$ that does not contain the vertex i. Clearly, $X_j^{(i)}$ is independent of X_i because the random set H_i is independent of X_i. Recalling that $X_{i,j} := X_{\{i,j\}} \sim \mathrm{Bin}\left(\binom{n-2}{k-2}, p\right)$ is the number of edges of $H_k(n,p)$ containing both i and j, we get

$$X_j - X_j^{(i)} = X_{i,j} - \mathbf{E}\left[X_{i,j} \mid H_i\right] = X_{i,j} - \mathbf{E}\left[X_{i,j}\right]. \tag{3.4.7}$$

Next, we estimate the probability that $|X_{i,j} - \mathbf{E}\left[X_{i,j}\right]| > \varepsilon b_n$. Here, without loss of the generality, we may assume that $p \leqslant \frac{1}{2}$. Otherwise, we can consider the random variable $\binom{n-2}{k-2} - X_{i,j} \sim \mathrm{Bin}\left(\binom{n-2}{k-2}, 1-p\right)$ and repeat the arguments. By the assumptions, we get that $b_n = \sqrt{\frac{\binom{n-1}{k-1}p(1-p)}{2\log n}}$ satisfies

$$b_n \gg \log n \qquad \text{and} \qquad \frac{b_n^2}{\mathbf{E}[X_{i,j}]} = \frac{\binom{n-1}{k-1}(1-p)}{2\binom{n-2}{k-2}\log n} \gg \log n.$$

Applying the Chernoff bound (see, for example, [58, Theorem 2.1]), we find that, for any fixed $\varepsilon > 0$,

$$\mathbf{P}\left(|X_{i,j} - \mathbf{E}\left[X_{i,j}\right]| > \varepsilon b_n\right) \leqslant 2\exp\left(-\frac{(\varepsilon b_n)^2}{2\mathbf{E}[X_{i,j}]+\varepsilon b_n}\right) = e^{-\omega(\log n)}. \tag{3.4.8}$$

Combining (3.4.7), (3.4.8) and applying the union bound for all $j \in [i-1]$, we get that

$$\mathbf{P}\left(\max_{j\in[i-1]}\left|X_j - X_j^{(i)}\right| > \varepsilon b_n\right) \leqslant ne^{-\omega(\log n)} = e^{-\omega(\log n)}.$$

From (3.4.6), we find that $\mathbf{P}(X_i > a_n + b_n x) = \Omega(n^{-1}) \gg e^{-\omega(\log n)}$ uniformly over all $i \in [n]$. Thus, we get the desired $X_j^{(i)}$ satisfying all conditions of Lemma 3.2.2. This completes the proof. □

Remark 3.4.3. The binomial random graph $\mathcal{G}(n,p)$ is a special case of $H_k(n,p)$ for $k = 2$. In the particular case, Theorem 3.3 gives the asymptotic distribution of the maximum degree of $\mathcal{G}(n,p)$. This result was obtained for the first time by Bollobás [10] and Ivchenko [51] using the method of moments. For every $i \in [n]$, they consider the Bernoulli random variable η_i that equals 1 if and only if its degree is bigger than $a_n + b_n x$. Letting $\eta = \eta_1 + \ldots + \eta_n$, they easily get that $\mathbf{E}\left[\eta\right] \to e^{-x}$ as $n \to \infty$. Thus, it is sufficient to prove that η converges in distribution to a Poisson random variable as $n \to \infty$. For $k = 2$, one can derive that $\mathbf{E}\left[\binom{\eta}{r}\right] \to e^{-xr}/r!$ for any fixed $r \in \mathbb{N}$. However, when $k > 2$ the dependencies are stronger so the computation of factorial moments becomes much more technically involved. In contrast, our method does not require any computations aside from the single application of the Chernoff bound in (3.4.8) for all k.

Remark 3.4.4. Another advantage of our approach is that it gives an estimate of the rate of convergence to the Gumbel distribution. A careful investigation of the proofs of Theorem 3.1, Lemma 3.2.2, and Theorem 3.2 shows that

$$\left|\mathbf{P}\left(\max_{i\in[n]} X_i \leqslant a_n + xb_n\right) - \prod_{i\in[n]}\mathbf{P}\left(X_i \leqslant a_n + xb_n\right)\right| = O\left(\sqrt{\frac{\log^3 n}{\binom{n-1}{k-1}p(1-p)}} + \sqrt{\frac{k\log^2 n}{n}}\right).$$

That is, the rate of convergence is governed by the rate of decrease of ε, for which $\mathbf{P}\left(\max_{j\in[i-1]}\left|X_j - X_j^{(i)}\right| > \varepsilon b_n\right)$ remains very small. In addition, for a_n and b_n defined by (3.4.3), the convergence rate of $\prod_{i\in[n]}\mathbf{P}\left(X_i \leqslant a_n + xb_n\right)$ to the Gumbel distribution is $O\left(\frac{\log\log n}{\log n}\right)$. However, this convergence rate can be improved by using a more precise expression for the scaling parameter a_n.

Our approach applied to codegrees X_S in the random hypergraph $H_k(n,p)$ leads to the following result.

Theorem 3.3. *Assume $p = p(n) \in (0,1)$, $s = s(n) \in [n-1]$, and $k = k(n) \in [n] \setminus [s]$ are such that*

$$\binom{n-s}{k-s} p(1-p) \gg s^3 \log^3 n, \qquad (k-s)s^2 \ll (n-s)/\log^2 n.$$

Then $\left[\max_{S \in \binom{[n]}{s}} X_S - a_n\right]/b_n$ converges in distribution to a standard Gumbel random variable, where $a_n = a_n\left(\binom{n}{s}, \binom{n-s}{k-s}, p\right)$ and $b_n = b_n\left(\binom{n}{s}, \binom{n-s}{k-s}, p\right)$ are defined in (3.4.3).

Proof. Theorem 3.3 is proved in exactly the same way as Theorem 3.2. Take any $x \in \mathbb{R}$. For all $S \in \binom{[n]}{s}$, let $A_S := \{X_S > a_n + x b_n\}$. Let $d := \binom{n}{s}$ and $N := \binom{n-s}{k-s}$. Since $\binom{n}{s} \leqslant n^s$, the assumptions imply $Np(1-p) \gg \log^3 d$. Recalling that $X_S \sim \mathrm{Bin}(N,p)$ and using Lemma 3.4.1, we find that

$$\prod_{S \in \binom{[n]}{s}} \mathbf{P}(\overline{A}_S) \to e^{-e^{-x}}.$$

Again, we can take $D_S = \emptyset$ for all $S \in \binom{[n]}{s}$. Thus, we only need to show that $\varphi(\mathbf{A}, \mathbf{D}) = o(1)$. The key fact needed to apply Lemma 3.2.2 is that, for any fixed $\varepsilon > 0$,

$$\mathbf{P}\left(\left| X_{U \cup S} - \mathbf{E}\left[X_{U \cup S}\right] \right| \leqslant \varepsilon b_n \text{ for all distinct } U, S \in \binom{[n]}{s} \right) \geqslant 1 - e^{-\omega(\log d)}.$$

Similarly to (3.4.8), this is a straightforward application of the Chernoff bound. $\square$

3.4.2 Maximum clique-extension counts

Let $k \geqslant 3$ be an integer. In this section, we find the asymptotic distribution of the maximum number of k-clique extensions in the random graph $\mathcal{G}(n,p)$. For $i \in [n]$, let X_i be the number of k-cliques containing vertex i. Below, we show that Theorem 3.1 implies the asymptotic distribution of the maximum value of X_i over $i \in [n]$.

Let

$$\begin{aligned} a_{k,n} &:= \frac{(pn)^{k-2} p^{\binom{k-1}{2}}}{(k-1)!} \left[pn + (k-1)\sqrt{2np(1-p)\log n} \left(1 - \frac{\log\log n}{4 \log n} - \frac{\log(2\sqrt{\pi})}{2\log n} \right) \right], \\ b_{k,n} &:= \frac{1}{(k-2)!} (pn)^{k-2} p^{\binom{k-1}{2}} \sqrt{\frac{np(1-p)}{2\log n}}. \end{aligned}$$

Theorem 3.4. *Let $p = p(n) \in (0,1)$, $k = k(n) \in \{3, \ldots, n\}$ be such that*

$$\log^3 n = o(np(1-p)), \quad \log^2 n = o\left(\frac{np^{\binom{k-1}{2}+1}(1-p)}{k^3} \right). \tag{3.4.9}$$

Then $\left[\max_{i \in [n]} X_i - a_n^k\right]/b_n^k$ converges in distribution to a standard Gumbel random variable.

Proof. Let d_i be the degree of the vertex i, and $Y_i = \mathbf{E}[X_i \mid d_i] = \binom{d_i}{k-1}p^{\binom{k-1}{2}}$. Note that

$$\max_{i\in[n]} Y_i = \binom{\max_{i\in[n]} d_i}{k-1}p^{\binom{k-1}{2}}.$$

Let $x \in \mathbb{R}$. By Theorem 3.2, we have

$$\mathbf{P}\left(\max_{i\in[n]} Y_i \leqslant \binom{a_n + b_n x}{k-1}p^{\binom{k-1}{2}}\right) = \mathbf{P}\left(\max_{i\in[n]} d_i \leqslant a_n + b_n x\right) \to e^{-e^{-x}},$$

where $a_n = a_n(n, n-1, p)$ and $b_n = b_n(n, n-1, p)$ are defined in (3.4.3) by

$$a_n = (n-1)p + \sqrt{2(n-1)p(1-p)\log n}\left(1 - \frac{\log\log n}{4\log n} - \frac{\log(2\sqrt{\pi})}{2\log n}\right), \qquad b_n = \sqrt{\frac{(n-1)p(1-p)}{2\log n}}.$$

Computing directly, we get

$$\begin{aligned}\binom{a_n + b_n x}{k-1}p^{\binom{k-1}{2}} &= (1+o(1))\frac{(a_n+b_n x)^{k-1}}{(k-1)!}p^{\binom{k-1}{2}} \\ &= (1+o(1))\frac{a_n^{k-1}}{(k-1)!}p^{\binom{k-1}{2}} + (1+o(1))\frac{a_n^{k-2}}{(k-2)!}p^{\binom{k-1}{2}}b_n x = a_{k,n} + b_{k,n}x(1+o(1)).\end{aligned}$$

Therefore we have

$$\mathbf{P}\left(\max_{i\in[n]} Y_i \leqslant a_{k,n} + b_{k,n}x\right) \to e^{-e^{-x}}. \tag{3.4.10}$$

Set $\widetilde{X}_i = \frac{X_i - a_{k,n}}{b_{k,n}}$, $\widetilde{Y}_i = \frac{Y_i - a_{k,n}}{b_{k,n}}$. It remains to show that

$$\mathbf{P}\left(\left|\widetilde{X}_i - \widetilde{Y}_i\right| > \varepsilon\right) = o\left(\mathbf{P}\left(Y_i > a_n^k + x b_n^k\right)\right) = o\left(\frac{1}{n}\right), \tag{3.4.11}$$

and apply Lemma 3.2.3.

The de Moivre–Laplace theorem and the relation $\int_x^\infty e^{-t^2/2}dt = \frac{1}{x}e^{-x^2/2}(1+o(1))$ (see, e.g., [11, Relation (1')]) imply

$$\mathbf{P}\left(|d_i - np| > \sqrt{2np(1-p)\log n}\right) = \frac{1+o(1)}{n\sqrt{\pi\log n}}.$$

Therefore,

$$\begin{aligned}\mathbf{P}\left(\left|\widetilde{X}_i - \widetilde{Y}_i\right| > \varepsilon\right) &= \mathbf{P}\left(|X_i - Y_i| > \varepsilon b_n^k\right) \\ &= \sum_{|j-np|\leqslant\sqrt{2np(1-p)\log n}} \mathbf{P}\left(\left|X_i - \binom{j}{k-1}p^{\binom{k-1}{2}}\right| > \varepsilon b_n^k,\, d_i = j\right) + o\left(\frac{1}{n}\right).\end{aligned}$$

It remains to bound from above $\mathbf{P}\left(X_i - \binom{j}{k-1}p^{\binom{k-1}{2}} > \varepsilon b_n^k \;\middle|\; d_i = j\right)$ and $\mathbf{P}\left(X_i - \binom{j}{k-1}p^{\binom{k-1}{2}} < -\varepsilon b_n^k \;\middle|\; d_i = j\right)$. For the lower tail, we apply Janson's inequality [58, Theorem 2.14] that does not work, in general, for upper tails. However, a weaker bound [58, Proposition 2.44] can be applied for that. To apply the bounds, we need to compute the number of $(k-1)$-cliques that are not edge-disjoint with a given $(k-1)$-clique in K_j (which is denoted by Δ below) and the expected number of pairs of non-edge-disjoint $(k-1)$-cliques in $\mathcal{G}_{j,p}$ (which is denoted by $\overline{\Delta}$ below).

For $j\in\mathbb{N}$ such that $|j-np|\leqslant\sqrt{2np(1-p)\log n}$, denote the number of $(k-1)$-subsets of $[j]$ having at least 2 common element with $[k-1]$ by Δ. Clearly,

$$\Delta=\binom{j}{k-1}-\binom{j-k+1}{k-1}-(k-1)\binom{j-k+1}{k-2}=\binom{k-1}{2}\frac{j^{k-3}}{(k-3)!}(1+o(1)).$$

Moreover, let $\overline{\Delta}$ be the expected number of pairs of (not necessarily distinct) k-cliques with non-empty edge intersections:

$$\begin{aligned}\overline{\Delta}&=\binom{j}{k-1}\sum_{\ell=2}^{k-1}\binom{k-1}{\ell}\binom{j-k+1}{k-1-\ell}p^{(k-1)(k-2)-\binom{\ell}{2}}\\&=\frac{j^{2k-4}}{(k-1)!(k-3)!}\binom{k-1}{2}p^{(k-1)(k-2)-1}(1+o(1)).\end{aligned}$$

By (3.4.9) and [58, Proposition 2.44], uniformly over all $j\in\mathbb{N}$ such that $|j-np|\leqslant\sqrt{2np(1-p)\log n}$, we have

$$\begin{aligned}\mathbf{P}\left(X_i-\binom{j}{k-1}p^{\binom{k-1}{2}}>\varepsilon b_n^k\,\middle|\,d_i=j\right)&\leqslant(\Delta+1)\exp\left(-\frac{\varepsilon^2\left[b_n^k\right]^2}{4(\Delta+1)(\mathbf{E}[X_i|d_i=j]+\varepsilon b_n^k/3)}\right)\\&=\exp\left(-\frac{\varepsilon^2p^{\binom{k-1}{2}}np(1-p)}{4(k-2)^2\log n}(1+o(1))\right)=o\left(\frac{1}{n}\right).\end{aligned}\tag{3.4.12}$$

Moreover, by (3.4.9) and Janson's inequality [58, Theorem 2.14], uniformly over all $j\in\mathbb{N}$ such that $|j-np|\leqslant\sqrt{2np(1-p)\log n}$,

$$\begin{aligned}\mathbf{P}\left(X_i-\binom{j}{k-1}p^{\binom{k-1}{2}}<-\varepsilon b_n^k\,\middle|\,d_i=j\right)&\leqslant\exp\left(-\frac{\varepsilon^2\left[b_n^k\right]^2}{2\overline{\Delta}}\right)\\&=\exp\left(-\frac{\varepsilon^2np^2(1-p)}{2(k-2)^2\log n}(1-o(1))\right)=o\left(\frac{1}{n}\right).\end{aligned}\tag{3.4.13}$$

Finally, combining (3.4.12) and (3.4.13), we get

$$\sum_{|j-np|\leqslant\sqrt{2np(1-p)\log n}}\mathbf{P}\left(\left|X_i-\binom{j}{k-1}p^{\binom{k-1}{2}}\right|>\varepsilon b_n^k,\,d_i=j\right)=o\left(\frac{1}{n}\right).$$

□

3.4.3 Maximum number of h-neighbours

The particular case of the following result for constant h was proved in [84]. Let us show that it is a more or less direct corollary of Theorem 3.1.

For $h\in\mathbb{N}$ and $\mathbf{x}\in\binom{[n]}{h}$, denote the number of common neighbours of vertices in $\mathbf{x}$ in $\mathcal{G}(n,p)$ by $X_{\mathbf{x}}$. Set $a_{h,n}:=a_n(\binom{n}{h},n,p^h)$, $b_{h,n}:=b_n(\binom{n}{h},n,p^h)$, where a_n and b_n are defined in (3.4.3).

Theorem 3.5. *Let $h=h(n)=o(\log n/\log\log n)$ and $p=p(n)\in(0,1)$ be such that*

$$\frac{p^h}{h^3}\gg\frac{\log^3 n}{n},\quad 1-p\gg\sqrt{\frac{\log\log n}{\log n}}.\tag{3.4.14}$$

Then $\left[\max_{\mathbf{x}\in\binom{[n]}{h}} X_{\mathbf{x}} - a_{h,n}\right]/b_{h,n}$ *converges in distribution to a standard Gumbel random variable.*

Proof. For any $\mathbf{x} \in \binom{[n]}{h}$, $X_{\mathbf{x}}$ follows $\mathrm{Bin}\left(n-h, p^h\right)$. Then, by Lemma 3.4.1,

$$\prod_{\mathbf{x}\in\binom{[n]}{h}} \mathbf{P}\left(X_{\mathbf{x}} \leqslant a_{h,n} + b_{h,n}x\right) \to e^{-e^{-x}}. \tag{3.4.15}$$

Let us label the h-subsets of $[n]$ by positive integers $1, 2, \ldots, \binom{n}{h}$, that is $\{\mathbf{x} : \mathbf{x} \in \binom{[n]}{h}\} = \{\mathbf{x}_i : i \in [\binom{n}{h}]\}$. For simplicity, we use X_i to denote $X_{\mathbf{x}_i}$. Set $d = \binom{n}{h}$ and fix $x \in \mathbb{R}$. For $i \in [d]$, set $A_i = \{X_i > a_{h,n} + b_{h,n}x\}$ and $D_i = [d] \setminus D_i^*$, where D_i^* is the set of labels of h-subsets $\binom{[n]\setminus \mathbf{x}_i}{h}$.

Let us first verify that $\varphi(\mathbf{A}, \mathbf{D}) = o(1)$. Let $i \in [d]$. We denote by H_i the set of edges of $\mathcal{G}(n,p)$ that do not contain any vertices of $\mathbf{x}_i$. For any $j \in [i-1] \setminus D_i$, let $X_{j,i}$ be the number of vertices in $\mathbf{x}_i$ adjacent to all vertices in $\mathbf{x}_j$ in $\mathcal{G}(n,p)$ (notice that $\mathbf{x}_i$ and $\mathbf{x}_j$ are disjoint). Since $X_{j,i}$ is independent of H_i, we get $X_j - \mathbf{E}\left[X_j \mid H_i\right] = X_{j,i} - \mathbf{E}\left[X_{j,i}\right]$. Set $\widetilde{X}_i = \frac{X_i - a_{h,n}}{b_{h,n}}$ and $\widetilde{X}_j^{(i)} = \mathbf{E}\left[\widetilde{X}_j \;\middle|\; H_i\right]$. Since $X_{j,i} \sim \mathrm{Bin}\left(h, p^h\right)$, we get by (3.4.14) and the Chernoff bound (see, e.g., [58, Theorem 2.1]) that, for every $\varepsilon > 0$,

$$\begin{aligned}\mathbf{P}\left(\left|\widetilde{X}_j - \widetilde{X}_j^{(i)}\right| > \varepsilon\right) &= \mathbf{P}\left(\left|X_{j,i} - hp^h\right| > \varepsilon b_{h,n}\right) \leqslant 2\exp\left(-\frac{(\varepsilon b_{h,n})^2}{2(hp^h + \varepsilon b_{h,n}/3)}\right) \\ &= \exp\left(-\frac{3\varepsilon b_{h,n}}{2}(1+o(1))\right) \leqslant \exp\left(-\frac{3\varepsilon}{2}\sqrt{\frac{np^h(1-p)}{2h\log n}}(1+o(1))\right) = o\left(\frac{1}{n^{2h}}\right).\end{aligned} \tag{3.4.16}$$

By (3.4.15), we get $\mathbf{P}\left(\widetilde{X}_i > x\right) = \binom{n}{h}^{-1} e^{-x}(1+o(1))$. Therefore, by the union bound,

$$\begin{aligned}\mathbf{P}\left(\max_{j\in[i-1]\setminus D_i}\left|\widetilde{X}_j - \widetilde{X}_j^{(i)}\right| > \varepsilon\right) &\leqslant \binom{n}{h}\mathbf{P}\left(\left|\widetilde{X}_j - \widetilde{X}_j^{(i)}\right| > \varepsilon\right) \\ &= o\left(\frac{1}{\binom{n}{h}}\right) = o(1)\mathbf{P}\left(\widetilde{X}_i > x\right).\end{aligned}$$

Lemma 3.2.2 implies $\varphi(\mathbf{A}, \mathbf{D}) = o(1)$.

By Corollary 3.2.1, it remains to verify the conditions $\Delta_1(\mathbf{A}, \mathbf{D}) = o(1)$ and $\Delta_2(\mathbf{A}, \mathbf{D}) = o(1)$. Unfortunately, these conditions do not hold. Nevertheless, the events $(A_i)_{i\in[d]}$ can be modified slightly to make the desired relations hold. Define

$$E = \bigcap_{\ell=1}^{h-1} \bigcap_{\mathbf{u}\in\binom{[n]}{\ell}} \left\{X_{\mathbf{u}} \leqslant np^\ell + \sqrt{2\ell np^\ell(1-p^\ell)\log n}\right\}.$$

For $i \in [d]$, let $\widetilde{A}_i = A_i \cap E$ and $\widetilde{\mathbf{A}} = (\widetilde{A}_i)_{i\in[d]}$.

The following lemma is proven in [84] for constant h; for $h = o(\log n/\log\log n)$, the same proof works.

Lemma 3.4.5 ([84]). *The following relations hold*

1. $\mathbf{P}(E) = 1 - o(1)$,
2. *for every* $x \in \mathbb{R}$, $\mathbf{P}\left(\widetilde{A}_i\right) = (1 - o(1))\mathbf{P}(A_i)$ *uniformly over all* $i \in [d]$,

3. $\sum_{i\in[d]}\sum_{j\in[i-1]\cap D_i}\mathbf{P}\left(\widetilde{A}_i\cap\widetilde{A}_j\right)=o\left(1\right)$.

We have shown that $\varphi(\mathbf{A},\mathbf{D})=o(1)$ uniformly over all $i\in[d]$, now we consider $\varphi(\widetilde{\mathbf{A}},\mathbf{D})$. For any $i\in[d]$, we have that

$$\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\,\middle|\,\widetilde{A}_i\right)=\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\,\middle|\,A_i\right)\frac{\mathbf{P}(A_i)}{\mathbf{P}(\widetilde{A}_i)}-\mathbf{P}\left(\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\right)\cap A_i\cap\overline{E}\right)\frac{1}{\mathbf{P}(\widetilde{A}_i)},$$

and therefore,

$$\begin{aligned}&\left|\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\,\middle|\,\widetilde{A}_i\right)-\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\right)\right|\\&\leqslant\left|\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\,\middle|\,A_i\right)\frac{\mathbf{P}(A_i)}{\mathbf{P}(\widetilde{A}_i)}-\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\right)\right|+\frac{\mathbf{P}(A_i\cap\overline{E})}{\mathbf{P}(\widetilde{A}_i)}.\end{aligned}$$

Note that

$$\begin{aligned}\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\,\middle|\,A_i\right)&=\mathbf{P}\left(E\cap\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\right)\,\middle|\,A_i\right)\\&=\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\,\middle|\,A_i\right)-\mathbf{P}\left(\overline{E}\cap\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\right)\,\middle|\,A_i\right).\end{aligned}$$

By Lemma 3.4.5, we obtain

$$\begin{aligned}&\left|\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\,\middle|\,\widetilde{A}_i\right)-\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}\widetilde{A}_j\right)\right|\\&\leqslant\left|\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\,\middle|\,A_i\right)(1+o(1))-\mathbf{P}\left(\bigcup_{j\in[i-1]\setminus D_i}A_j\right)\right|+\frac{2\mathbf{P}(A_i\cap\overline{E})}{\mathbf{P}(\widetilde{A}_i)},\end{aligned}$$

and $\varphi(\widetilde{\mathbf{A}},\mathbf{D})=o(1)$ follows from

$$\frac{2\mathbf{P}(A_i\cap\overline{E})}{\mathbf{P}(\widetilde{A}_i)}=\frac{2\mathbf{P}(A_i)}{\mathbf{P}(\widetilde{A}_i)}-2=o(1).$$

Notice that the third statement of Lemma 3.4.5 is exactly $\Delta_1'(\widetilde{\mathbf{A}},\mathbf{D})=o(1)$ (see the definitions of Δ_1' and Δ_2' in Section 3.2.1). It remains to prove that $\Delta_2'(\widetilde{\mathbf{A}},\mathbf{D})=o(1)$. But this is straightforward:

$$\begin{aligned}\Delta_2'(\widetilde{\mathbf{A}},\mathbf{D})&=\sum_{i\in[d]}\sum_{j\in[i-1]\cap D_i}\mathbf{P}(\widetilde{A}_i)\mathbf{P}(\widetilde{A}_j)\leqslant\sum_{i\in[d]}\sum_{j\in[i-1]\cap D_i}\mathbf{P}\left(A_i\right)\mathbf{P}(A_j)\\&\leqslant\binom{n}{h}^{-1}e^{-2x}(1+o(1))\max_{i\in[d]}|D_i|=\frac{\binom{n}{h}-\binom{n-h}{h}}{\binom{n}{h}}e^{-2x}(1+o(1))=o(1).\end{aligned}$$

By Theorem 3.1, we get that (3.1.2) holds for $\widetilde{\mathbf{A}}$. The first two statements of Lemma 3.4.5 imply that

(3.1.2) also holds for $\mathbf{A}$. Indeed, $\mathbf{P}(E)=1-o(1)$ implies that

$$\mathbf{P}\left(\cap_{i\in[d]}\overline{A_i}\right)=\mathbf{P}\left(\cap_{i\in[d]}\overline{\widetilde{A}_i}\right)-\mathbf{P}\left(\overline{E}\setminus\left(\cap_{i\in[d]}\overline{A_i}\right)\right)=\mathbf{P}\left(\cap_{i\in[d]}\overline{\widetilde{A}_i}\right)+o(1);$$

and $\mathbf{P}(\widetilde{A}_i)=(1-o(1))\mathbf{P}\left(A_i\right)$ implies

$$\prod_{i\in[d]}\mathbf{P}\left(\overline{\widetilde{A}_i}\right)=\prod_{i\in[d]}\left[1-\mathbf{P}\left(A_i\right)(1+o(1))\right]=\left(1-\frac{e^{-x}+o(1)}{d}\right)^d\to e^{-e^{-x}}.$$

This completes the proof. □

3.4.4 Further results in maximum extensions counts

As we discussed in the beginning of Section 3.4, the above results are in the framework of extensions counting. Given a strictly balanced grounded pair (H,G) with $|V(H)|=h$, we are interested in the asymptotic behaviour of $\max_{\mathbf{x}\in[n]_h}X_{\mathbf{x}}$. Recall that, in [95], Spencer proved the law of large numbers (3.4.1). In recent paper [90], Šileikis and Warnke studied the validity of this law when $\mu=\Theta(\log n)$.

In Section 3.4.2, we found an optimal denominator in the law of large numbers for $h=1$, $G=K_k$ and p satisfying (3.4.9) (that is, far from the threshold value):

$$\frac{\max_{i\in[n]}X_i-\mu}{\mu(k-1)\sqrt{2(1-p)\log n/(pn)}}\xrightarrow{\mathbf{P}}1.$$

Notice that the result holds for the numerator $\max_{i\in[n]}|X_i-\mu|=\max\{\max_{i\in[n]}X_i-\mu,\mu-\min_{i\in[n]}X_i\}$ as well. Indeed, let d_i be the degree of the vertex i. Theorem 3.2 implies the asymptotic distribution of the minimum degree of $\mathcal{G}(n,p)$ since it equals in distribution to $n-\max_{i\in[n]}d_i[G_{n,1-p}]$. Thus,

$$\mathbf{P}\left(\min_{i\in[n]}\mathbf{E}\left[X_i|d_i\right]\geqslant\widetilde{a}_n^k-b_n^kx\right)\to e^{-e^{-x}}$$

where $\widetilde{a}_n^k=\frac{1}{(k-1)!}(pn)^{k-2}p^{\binom{k-1}{2}}n-a_n^k$. To get the distribution of the minimum degree, it remains to reformulate Lemma 3.2.3 for the events $A_i:=\{X_i<\widetilde{a}_n^k-b_n^kx\}$ and probabilities $\mathbf{P}(\min X_i\geqslant\widetilde{a}_n^k-b_n^kx)$, $\mathbf{P}(\min Y_i\geqslant\widetilde{a}_n^k-b_n^kx)$ (clearly, the same proof works) and follow absolutely the same steps as in the proof of Theorem 3.4.

Our method works not only in the case $h=1$. In Section 3.4.3, we have found the asymptotic distribution of $X_{\mathbf{x}}$ when $h\geqslant 2$ and G contains a unique vertex outside H which is adjacent to all vertices in H. Our arguments should work even in the case when H,G are both cliques of arbitrary size. Indeed, the result for cliques G such that $|V(G)|-|V(H)|\geqslant 2$ can be obtained from Theorem 3.5 using Lemma 3.2.3 in the same way as we obtain Theorem 3.4 from Theorem 3.2 in Section 3.4.2.

Chapter 4

Cumulant expansion for Eulerian orientation count

4.1 Introduction

An *Eulerian orientation* of an undirected graph is an assignment of directions to all the edges such that, for any vertex, the number of edges directed towards equals the number of edges directed out. Let $\mathrm{EO}(G)$ denote the number of Eulerian orientations of graph G. It is well known that $\mathrm{EO}(G) > 0$ if and only if all degrees of G are even. Computing $\mathrm{EO}(G)$ corresponds to evaluating the Tutte polynomial at point $(0, -2)$, see [52].

Schrijver [87] pointed out that the computation of $\mathrm{EO}(G)$ can be reduced to the evaluation of the permanent of a certain matrix M associated with graph G, that is,

$$\mathrm{EO}(G) = \frac{\operatorname{perm}(M)}{\prod_{i \in V(G)} (d_i/2)!}, \tag{4.1.1}$$

where d_i denotes the degree of vertex i in G. As Bethe permanent can be used both for a lower and an upper bound of permanent, it further gives bounds on $\mathrm{EO}(G)$.

Unfortunately, there is no efficient algorithm known to find $\mathrm{EO}(G)$ exactly. Mihail and Winkler [75] noticed that the Markov Chain Monte Carlo method approximating permanent, combining with (4.1.1), gives an approximation algorithm of computing the number of Eulerian orientations, and it is #P-complete in general.

Wide interest in counting Eulerian orientations is due to the equivalence to the partition function of so-called ice-type models in statistical physics; see, for example, [102] and references therein. Lieb [66] and Baxter [7] have derived asymptotic expressions for the number of Eulerian orientations of the square and triangular lattice by the transfer matrix method [66, 7] using eigenvalues of matrices. No extension of this to higher dimensions is known, to the best of our knowledge.

A *regular tournament* on n vertices is an Eulerian orientation of the complete graph K_n. Clearly, n must be odd for the existence of a regular tournament. McKay [73] established that, for odd $n \to \infty$,

$$\mathrm{EO}(K_n) \sim \left(\frac{n}{e}\right)^{1/2} \left(\frac{2^{n+1}}{\pi n}\right)^{(n-1)/2}. \tag{4.1.2}$$

Adopting McKay's approach to dense graphs with strong mixing properties, Isaev and Isaeva [44] obtained

$$\mathrm{EO}(G) \sim \widehat{\mathrm{EO}}(G) := \frac{2^{|E(G)|}}{\sqrt{\mathrm{ST}(G)}} \left(\frac{2}{\pi}\right)^{(n-1)/2} \exp\left(-\frac{1}{4}\sum_{jk\in G}\left(\frac{1}{d_j}+\frac{1}{d_k}\right)^2\right), \tag{4.1.3}$$

where $\mathrm{ST}(G)$ denotes the number of spanning trees of G. More precisely, graphs considered in [44] satisfy $h(G) \geqslant \gamma n$ for some fixed γ, where $h(G)$ is the *Cheeger constant* (also known as isoperimetric number) of a graph G. Recall that

$$h(G) := \min\left\{\frac{|\partial_G U|}{|U|} : U \subset V(G), 1 \leqslant |U| \leqslant \frac{1}{2}|V(G)|\right\}, \tag{4.1.4}$$

where $\partial_G U$ is the set of edges of G with one end in U and the other end in $V(G) \setminus U$. The asymptotic result (4.1.3) was further extended in [45, Corollary 3]: if $h(G) \geqslant \gamma d$ and $d \geqslant n^{1/3+\varepsilon}$, then

$$\mathrm{EO}(G) = \widehat{\mathrm{EO}}(G) \exp\left(O\left(\frac{n}{d^2}\log\frac{2n}{d}\right)\right), \tag{4.1.5}$$

where d is the maximum degree of G.

Note that the error term $\frac{n}{d^2}\log\frac{2n}{d}$ in (4.1.5) grows as $n \to \infty$ if $d < n^{1/2}$. In this chapter, we give an asymptotic expansion series that gives the value of $\mathrm{EO}(G)$ up to precision $O(n^{-c})$ for any arbitrary constant $c > 0$ under the assumptions that $h(G) \geqslant \gamma d$ and $d \gg \log^8 n$. As a corollary, we prove that the asymptotic bound of (4.1.5) holds for such graphs.

4.1.1 Main results

Our estimates for $\mathrm{EO}(G)$ rely on cumulant expansion with respect to a certain Gaussian random vector associated with graph G. First, we recall the definition of cumulants. Let $X_1, \ldots, X_r$ be random variables with finite moments, the *joint cumulant* (or *mixed cumulant*) is defined by

$$\kappa(X_1, \ldots, X_r) := [t_1 \ldots t_r] \log\left(\mathbf{E}\left[\exp\left(\sum_{i=1}^r t_i X_i\right)\right]\right), \tag{4.1.6}$$

where $[t_1 \ldots t_r]$ stands for the coefficient of $t_1 \ldots t_r$ in the series expansion. If all random variables $X_1, \ldots, X_r$ are the same variable X, then we write

$$\kappa_r(X) := \kappa(X, \ldots, X),$$

which becomes the *cumulant* of order r for random variable X.

Given a graph G, the *Laplacian Matrix* $L = L(G)$ is defined by

$$\boldsymbol{x}^{\mathrm{T}} L \boldsymbol{x} = \sum_{jk\in G} (x_j - x_k)^2, \tag{4.1.7}$$

for $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$ and the summation is over all (unordered) edges $jk \in G$. Clearly, L is symmetric positive-semidefinite. The vector $\mathbf{1} = (1, \ldots, 1)^T$ is an eigenvector of L with eigenvalue 0. If G is connected then all other eigenvalues of L are positive. In this case, let $\mathbf{X}_G$ denote n-dimensional singular Gaussian random vector on the subspace

$$\mathcal{V} := \{\boldsymbol{x} \in \mathbb{R}^n : x_1 + \cdots + x_n = 0\}$$

with density proportional to $\exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}L\boldsymbol{x}\right)$.

Let $(c_{2\ell})_{\ell\geqslant 1}$ denote the coefficients of the Taylor expansion of $\log\cos x$ at $x=0$,

$$\log\cos x = \sum_{\ell\geqslant 1} c_{2\ell}x^{2\ell} = -\frac{1}{2}x^2 - \frac{1}{12}x^4 - \frac{1}{45}x^6 - \frac{17}{2520}x^8 - \cdots,$$

where

$$c_{2\ell} = -\frac{4^\ell(4^\ell-1)|B_{2\ell}|}{2\ell(2\ell)!}, \tag{4.1.8}$$

with $B_{2\ell}$ denoting the Bernoulli number, see, for example, [38, 1.518]. Note that $|B_{2\ell}| < 4(2\ell)!/(2\pi)^{2\ell}$, and therefore,

$$|c_{2\ell}| < \frac{2^{2\ell+1}(4^\ell-1)}{\ell(2\pi)^{2\ell}} < \frac{2}{\ell}\left(\frac{2}{\pi}\right)^{2\ell}. \tag{4.1.9}$$

For an integer $K \geqslant 2$, define multivariate polynomial f_K by

$$f_K(\mathbf{x}) = \sum_{\ell=2}^{K} c_{2\ell}\sum_{jk\in G}(x_j-x_k)^{2\ell}. \tag{4.1.10}$$

Our main result is the following theorem.

Theorem 4.1.1. *Let $G = G(n)$ be a graph with even degrees. Assume that*

(A1) $d \gg \log^8 n$, where d is the maximum degree of G;

(A2) the Cheeger constant $h(G) \geqslant \gamma d$ for some constant $\gamma > 0$.

Let $c > 0$ be a constant and $M = M(c)$ and $K = K(c)$ be defined by

$$K := \left\lceil \frac{(c+1)\log n}{\log d - 4\log\log n}\right\rceil, \qquad M := \left\lceil \frac{2(c+1)\log n}{\log d - 8\log\log n}\right\rceil.$$

Then as $n \to \infty$,

$$\mathrm{EO}(G) = \frac{2^{|E(G)|}}{\sqrt{\mathrm{ST}(G)}}\left(\frac{2}{\pi}\right)^{(n-1)/2}\exp\left(\sum_{s=1}^{M}\frac{1}{s!}\kappa_s\left(f_K(\mathbf{X}_G)\right) + O\left(n^{-c}\right)\right), \tag{4.1.11}$$

where $\mathrm{ST}(G)$ denotes the number of spanning trees of graph G.

As a corollary of Theorem 4.1.1, we get the following result.

Corollary 4.1.2. *The asymptotic bound of (4.1.5) holds under the assumptions of Theorem 4.1.1.*

For $G = K_n$ with odd n, computing a few cumulant terms in (4.1.11) establishes the following extension of McKay's formula (4.1.2) for the number of regular tournaments.

Corollary 4.1.3. *For odd $n \to \infty$,*

$$\begin{aligned}\mathrm{EO}(K_n) = n^{1/2}\left(\frac{2^{n+1}}{\pi n}\right)^{(n-1)/2}&\exp\Big(-\frac{1}{2} + \frac{1}{4n} + \frac{1}{4n^2} + \frac{7}{24n^3} + \frac{37}{120n^4} + \frac{31}{60n^5} + \frac{81}{28n^6}\\ &+ \frac{5981}{336n^7} + \frac{22937}{240n^8} + \frac{90031}{180n^9} + \frac{1825009}{660n^{10}} + \frac{4344847}{264n^{11}} + O\left(n^{-12}\right)\Big).\end{aligned} \tag{4.1.12}$$

4.1.2 Chapter structure

The chapter is structured as follows. First, in Section 4.2, we include some interesting observations that formula (4.1.3) works reasonably well even in the range of graphs far beyond that considered in Theorem 4.1.1.

In Section 4.3, we review basic facts about cumulants in general and cumulants of Gaussian random variables in particular. Most notably, we give a new asymptotic estimate on the tail of the cumulant series of a function of a growing number of independent random variables; see Theorem 4.3.2. This estimate can be applied to functions of Gaussian vectors after a proper rotation; see Theorem 4.3.6.

The proof of Theorem 4.1.1 is given in Section 4.4. We represent $\mathrm{EO}(G)$ in terms of a high-dimensional integral, which we estimate using a variation of the Laplace method and then applying Theorem 4.3.6.

In Section 4.5, we give the details of cumulant computation for the number of regular tournaments, that is, the case of the complete graph $G = K_n$. We provide comparisons with exact values. Also, we include two more asymptotic formulas for the number of Eulerian digraphs, and Eulerian oriented graphs, whose proofs are very similar.

Section 4.6 contains the proof of Theorem 4.3.2. The proof is based on the estimates for the variation of conditional cumulants with respect to revealing one variable at a time.

4.2 Pauling's estimate, and ice-entropy comparisons for regular graphs

The study of "ice models" motivates the following definition. For a graph G on n vertices, let

$$eo(G) := \frac{1}{n} \log \mathrm{EO}(G). \tag{4.2.1}$$

We call $eo(G)$ the *ice-entropy* of G. Determining the asymptotics of $eo(G)$ is a key question in the area, see for example [7, Chapter 8] and [67]. In particular, it is known for the square lattice L_n and the triangular lattice T_n; see [66, 7]. However, the question remains largely open for other graphs. Even for the cubic lattice C_n, there are only the upper and lower bounds by estimating permanent in (4.1.1) and a heuristic estimate by Pauling around 90 years ago in his seminal paper [81].

4.2.1 Pauling's estimate

Pauling's idea is to orient the edges of graph, and to ignore the dependencies among the events that vertices are unbalanced (this is obviously not right, except for the trivial case that G is an empty graph). Formally, if every edge of a graph G is oriented independently with probability $1/2$, then

$$\mathbf{P}\left(\sum_{v\in V(G)} X_v = 0\right) = 2^{-|E(G)|}\,\mathrm{EO}(G),$$

where X_j is the indicator of the event that vertex j is unbalanced: the numbers of incoming and outgoing arcs are different.

This leads to the following estimate,

$$\widehat{\mathrm{EO}}_{\mathrm{Pauling}}(G) := 2^{|E(G)|} \prod_{v \in V(G)} 2^{-d_v} \binom{d_v}{d_v/2} = 2^{-|E(G)|} \prod_{v \in V(G)} \binom{d_v}{d_v/2}.$$

For a d-regular graph G on n vertices, we have

$$\widehat{\mathrm{EO}}_{\mathrm{Pauling}}(G) = 2^{-nd/2} \binom{d}{d/2}^n = \left(\frac{d!}{(d/2)!^2 2^{d/2}} \right)^n.$$

For large d, in view of (4.2.1), using Stirling's formula gives

$$\widehat{eo}_{\mathrm{Pauling}}(G) = \log \left(\frac{d!}{(d/2)!^2 2^{d/2}} \right) = \frac{\log 2}{2} d - \frac{1}{2} \log d + \frac{1}{2} \log \left(\frac{2}{\pi} \right).$$

Let

$$\mathrm{Pauling}(G) := \log \binom{d}{d/2} - \frac{d}{2} \log 2. \tag{4.2.2}$$

Schrijver [87] showed that

$$\mathrm{Pauling}(G) \leqslant eo(G) \leqslant \frac{1}{2} \log \binom{d}{d/2}. \tag{4.2.3}$$

In this section, we compare the ice entropy of a d-regular graph G and Pauling's estimate (4.2.2) with

$$\widehat{eo}(G) := \frac{1}{n} \log \widehat{\mathrm{EO}}(G) = \frac{d}{2} \log 2 - \frac{1}{2n} \log \mathrm{ST}(G) - \frac{n-1}{2n} \log \frac{\pi}{2} - \frac{1}{2d}. \tag{4.2.4}$$

The quantity $\frac{1}{n} \log \mathrm{ST}(G)$ is called the *spanning tree entropy* in the literature. Even though the graphs with constant degrees are beyond the reach of our results, this simple formula gives a surprisingly good estimate of the ice-entropy, slightly above the exact value (where it is known). Furthermore, we believe that computing more cumulants as in Theorem 4.1.1 can improve the precision even further.

4.2.2 Square lattice L_n and Lieb's constant

For a square lattice L_n on n vertices, Lieb's square ice constant [66] is

$$\lim_{n\to\infty} \mathrm{EO}(L_n)^{1/n} = \frac{8\sqrt{3}}{9} \approx 1.540.$$

Therefore, $eo(L_n) = \log \left(\frac{8\sqrt{3}}{9} \right) \approx 0.431$. Pauling's estimate gives

$$\mathrm{Pauling}(L_n) = \log 1.5 \approx 0.405.$$

From [36], we know that

$$\frac{1}{n} \log \mathrm{ST}(L_n) = \frac{4}{\pi} \sum_{i \geqslant 1} \frac{\sin(i\pi/2)}{i^2} \approx 1.166.$$

Then, our estimate (4.2.4) gives

$$\widehat{eo}(G) = 2 \log 2 - \frac{2}{\pi} \sum_{i \geqslant 1} \frac{\sin(i\pi/2)}{i^2} - \frac{1}{2} \log \frac{\pi}{2} - \frac{1}{8} \approx 0.453.$$

4.2.3 Triangular lattice T_n and Baxter's constant

For a triangular lattice T_n on n vertices, Baxter's constant [7] is

$$\lim_{n\to\infty} \mathrm{EO}(T_n)^{1/n} = \frac{3\sqrt{3}}{2} \approx 2.598.$$

Therefore, $eo(T_n) = \log\left(\frac{3\sqrt{3}}{2}\right) \approx 0.955$. Pauling's estimate gives

$$\mathrm{Pauling}(T_n) = \log 2.5 \approx 0.916.$$

For triangular lattice, from [36], we know that

$$\frac{1}{n}\log \mathrm{ST}(T_n) = \frac{4}{\pi}\sum_{i\geqslant 1}\frac{\sin(i\pi/3)}{i^2} \approx 1.615.$$

Then, our estimate (4.2.4) gives

$$\widehat{eo}(T_n) = 3\log 2 - \frac{2}{\pi}\sum_{i\geqslant 1}\frac{\sin(i\pi/3)}{i^2} - \frac{1}{2}\log\frac{\pi}{2} - \frac{1}{12} \approx 0.963.$$

4.2.4 Cubic lattice C_n

The asymptotic value of $eo(C_n)$ is unknown. The upper and lower bounds by (4.2.3) are

$$0.916 \approx \mathrm{Pauling}(C_n) \leqslant eo(C_n) \leqslant \frac{1}{2}\log 20 \approx 1.498.$$

For cubic lattice C_n, from [85] we know that

$$\frac{1}{n}\log \mathrm{ST}(C_n) \approx 1.673.$$

Therefore, we get by our estimate (4.2.4),

$$\widehat{eo}(C_n) \approx 3\log 2 - \frac{1}{2}1.673 - \frac{1}{2}\log\frac{\pi}{2} - \frac{1}{12} \approx 0.934.$$

4.2.5 Hypercube Q_d

For a d-dimensional hypercube Q_d on $n = 2^d$ vertices, the asymptotics value of $eo(Q_d)$ is also unknown. The exact values for d up to 6 are given in OEIS (the On-Line Encyclopedia of Integer Sequences) as the sequence A358177. Using formula (4.1.1), we also estimated $\mathrm{EO}(Q_8)$. Taking the logarithm and dividing it by the number of vertices gives

$$eo(Q_4) \approx 0.500, \qquad eo(Q_6) \approx 0.955, \qquad eo(Q_8) \approx 1.489.$$

The upper and lower bounds by (4.2.3) are

$$0.405 \approx \text{Pauling}(Q_4) \leqslant eo(Q_4) \leqslant \tfrac{1}{2}\log 6 \approx 0.896.$$
$$0.916 \approx \text{Pauling}(Q_6) \leqslant eo(Q_6) \leqslant \tfrac{1}{2}\log 20 \approx 1.498.$$
$$1.476 \approx \text{Pauling}(Q_8) \leqslant eo(Q_8) \leqslant \tfrac{1}{2}\log 70 \approx 2.124.$$

For hypercube Q_d, from [9], we know that

$$\text{ST}(Q_d) = \frac{1}{n}\prod_{i=1}^{d}(2i)^{\binom{d}{i}}.$$

Then, our estimate (4.2.4) gives

$$\widehat{eo}(Q_4) \approx 0.501, \qquad \widehat{eo}(Q_6) \approx 0.960, \qquad \widehat{eo}(Q_8) \approx 1.495.$$

Such precision is very surprising given that both degree and the number of vertices are relatively small.

4.2.6 Large degrees

Alon [2] proved that the number of spanning trees in any connected d-regular graph with n vertices lies in $[((1-\varepsilon(d))d)^n, d^n]$ for some function $\varepsilon(d)$ that goes to 0 as $d \to \infty$. Then, using Stirling's formula, it is straightforward to check that our estimate $\widehat{eo}(G)$ is asymptotically equivalent to Pauling's estimate. We believe that it actually gives the correct asymptotics of $eo(G)$.

Conjecture 4.2.1. *If $\{G_i\}$ is a sequence of connected regular graphs with degree going to infinity, then*

$$eo(G_i) \sim \text{Pauling}(G_i).$$

as $i \to \infty$.

As a consequence of Theorem 4.1.1, Corollary 4.1.2, and [2, Theorem 1.1], we get the following result.

Corollary 4.2.2. *Conjecture 4.2.1 is true if $d \gg \log^8 n$ and $h(G) \geqslant \gamma d$ for some fixed $\gamma > 0$.*

This is by using Theorem 4.1.1 and Corollary 4.1.2 to obtain, from (4.1.3), that

$$\begin{aligned} eo(G_i) &= \frac{1}{n}\log\left(\frac{2^{nd/2}}{\sqrt{\text{ST}(G)}}\left(\frac{2}{\pi}\right)^{(n-1)/2}\right) - \frac{1}{2d} + O\left(\frac{1}{d^2}\log\frac{2n}{d}\right) \\ &\sim \frac{d}{2}\log 2 + \frac{1}{2}\log\left(\frac{2}{\pi}\right) - \frac{1}{2}\log d - \frac{1}{2d} + O\left(\frac{1}{d^2}\log\frac{2n}{d}\right), \end{aligned}$$

and noting that

$$\text{Pauling}(G_i) = \log\binom{d}{d/2} - \frac{d}{2}\log 2 \sim \log\left(\frac{\sqrt{2\pi d}\left(\frac{d}{e}\right)^d}{\pi d\left(\frac{d}{2e}\right)^d}\right) - \frac{d}{2}\log 2 = \frac{d}{2}\log 2 + \frac{1}{2}\log\left(\frac{2}{\pi}\right) - \frac{1}{2}\log d.$$

Our Theorem 4.1.1 also gives that

$$\log \mathbf{P}\left(X=0\right) = -\tfrac{1}{2}\mathrm{ST}(G) + \tfrac{n-1}{2}\log\left(\tfrac{2}{\pi}\right) + \sum_{r=1}^{K}\frac{1}{r!}\,\kappa_r\left(f_M(\mathbf{X})\right) + O\left(n^{-c}\right). \qquad (4.2.5)$$

Its connection to the cluster expansion series (2.2.1) is unclear. Recall that the dependency graph for the indicators $\{X_v\}_{v\in V(G)}$ of unbalanced vertices can be chosen to be G itself. Therefore, our asymptotic expansion (4.2.5) is for moderately dense dependency graph with some expansion properties. This result is quite rare as the common applications of dependency graphs is usually for sparse graphs!

4.3 Cumulants and expansion of Laplace-type integrals

In this section, we develop tools for estimating integrals that typically appear in applications of the Laplace method. Isaev and McKay [46] proved that, for random vector $\mathbf{X}$ with the normal density $\pi^{-n/2}|A|^{1/2}e^{-\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}}$,

$$\int_{\Omega} e^{-\boldsymbol{x}^T A\boldsymbol{x}+f(\boldsymbol{x})}d\boldsymbol{x} \approx \pi^{n/2}|A|^{-1/2}\exp\Big(\mathbf{E}\left[f(\mathbf{X})\right] + \frac{1}{2}\operatorname{Var} f(\mathbf{X})\Big),$$

under some smoothness conditions on f that limits variations with respect to changing one or two coordinates, and the existence of a proper rotation diagonalising matrix A with bounded infinity norm. To achieve better accuracy, we extend this result to allow more terms using a cumulant expansion; see Theorem 4.3.6.

4.3.1 Cumulants and cumulant series

Let $[n]$ denote the integer set $\{1,2,\ldots,n\}$ for every integer $n\geqslant 1$. The combinatorial definition of joint cumulant of random variables $\{X_i\}_{i\in[n]}$, equivalent to (4.1.6), is

$$\kappa\left(X_1,\ldots,X_n\right) = \sum_{\tau\in P_n}(-1)^{|\tau|-1}(|\tau|-1)!\prod_{B\in\tau}\mathbf{E}\left[\prod_{i\in B}X_i\right], \qquad (4.3.1)$$

where P_s denotes the set of unordered partitions τ of $[s]$ (with non-empty blocks) and $|\tau|$ denotes the number of blocks in the partition τ. Recall the useful multi-linearity property of cumulants, that is, for $r\geqslant 1$,

$$\kappa\left(\sum_{i_1\in[n_1]}X_{i_1}, \sum_{i_2\in[n_2]}Y_{i_2},\ldots,\sum_{i_r\in[n_r]}Z_{i_r}\right) = \sum_{i_1\in[n_1]}\sum_{i_2\in[n_2]}\cdots\sum_{i_r\in[n_r]}\kappa\left(X_{i_1},Y_{i_2},\ldots,Z_{i_r}\right),$$

and in particular,

$$\kappa_r\left(\sum_{i\in[n]}X_i\right) = \sum_{(i_1,\ldots i_r)\in[n]^r}\kappa(X_{i_1},\ldots,X_{i_r}).$$

The joint cumulant can be regarded as a measure of the mutual dependences of the variables. An important property of the joint cumulant $\kappa(X_1,\ldots,X_n)$ is that if $[n]$ can be partitioned into two subsets S_1 and S_2 such that the variables $\{X_i\}_{i\in S_1}$ are independent of the variables $\{X_j\}_{j\in S_2}$, then

$\kappa(X_1,\ldots,X_n)=0$ (see, for example, [93]).

Next, we introduce a lemma that is useful to bound cumulants.

Lemma 4.3.1.

$$\sum_{\tau\in P_s}(|\tau|-1)!\leqslant\left(\tfrac{3}{2}\right)^s(s-1)!. \tag{4.3.2}$$

Proof. Recall that P_s is the set of unordered partitions of $[s]$, therefore,

$$\sum_{\tau\in P_s}(|\tau|-1)!=\sum_{\tau\in P_s}\frac{|\tau|!}{|\tau|}=\sum_{\tau'}|\tau'|^{-1},$$

where the last sum is over ordered partitions of $[s]$. We can bound the sum over τ' as follows. Take any permutation of $[s]$ and cut it at some number of distinct places (e.g. $3|2,4|1,5$ for $s=5$). The number of ways to place k cuts is $\binom{s-1}{k}$ and each ordered partition with $k+1$ parts is made in

$$b_1!\cdots b_{k+1}!\geqslant 2^{b_1-1}\cdots 2^{b_{k+1}-1}=2^{s-k-1}$$

ways by this procedure. Allowing $\frac{1}{k+1}$ for the weight $|\tau'|^{-1}$, we get

$$\sum_{\tau\in P_s}(|\tau|-1)!\leqslant s!\sum_{k=0}^{s-1}\frac{2^{k-s+1}}{k+1}\binom{s-1}{k}=s!\,\frac{3^s-1}{s\,2^s}\leqslant\left(\tfrac{3}{2}\right)^s(s-1)!.$$

This completes the proof. □

4.3.2 Cumulant expansion for Laplace transform

Let $\mathbf{X}=(X_1,\ldots,X_n)$ be a random vector with independent components taking values in $\mathbf{S}:=S_1\times\cdots\times S_n$. For $\boldsymbol{y}\in\mathbf{S}$, let $R^j_{\boldsymbol{y}}$ denote the operator that replaces the j-th coordinate with $y_j\in S_j$:

$$R^j_{\boldsymbol{y}}[f](\boldsymbol{x})=f(x_1,\ldots,x_{j-1},y_j,x_{j+1},\ldots x_n).$$

For $V=\{v_1,\ldots,v_k\}\subset[n]$, define

$$R^V_{\boldsymbol{y}}=R^{v_1}_{\boldsymbol{y}}\cdots R^{v_k}_{\boldsymbol{y}}\qquad\text{and}\qquad \partial^V_{\boldsymbol{y}}:=\partial^{v_1}_{\boldsymbol{y}}\cdots\partial^{v_k}_{\boldsymbol{y}}, \tag{4.3.3}$$

where $\partial^j_{\boldsymbol{y}}:=I-R^j_{\boldsymbol{y}}$ and I is the identity operator. Let

$$\Delta_V(f)=\Delta_V(f,\mathbf{S}):=\sup_{\boldsymbol{x},\mathbf{y}\in\mathbf{S}}\left|\partial^V_{\boldsymbol{y}}[f](\boldsymbol{x})\right|. \tag{4.3.4}$$

We also set $\Delta_\emptyset(f):=\sup_{\boldsymbol{x}\in\mathbf{S}}|f(\boldsymbol{x})|$.

Theorem 4.3.2. *Let $m>0$ be an integer and $\alpha\geqslant 0$. Let $\mathbf{X}=(X_1,\ldots,X_n)$ be a random vector with independent components taking values in $\mathbf{S}$. Suppose f is a bounded function on $\mathbf{S}$ such that*

$$\max_{\substack{v\in[m]\\ j\in[n]}}\sum_{V\in\binom{[n]}{v}:j\in V}\Delta_V(f)\leqslant\alpha. \tag{4.3.5}$$

Then, we have

$$\mathbf{E}\left[e^{f(\mathbf{X})}\right] = (1+\delta)^n \exp\left(\sum_{s=1}^{m} \frac{\kappa_s(f(\mathbf{X}))}{s!}\right),$$

where $\delta > -1$ *satisfies* $|\delta| \leqslant e^{(200\alpha)^{m+1}} - 1$. *Furthermore, for any* $s \in [m]$,

$$|\kappa_s(f(\mathbf{X}))| \leqslant 0.011\, n \frac{(s-1)!}{s} (100\alpha)^s.$$

We will prove Theorem 4.3.2 in Section 4.6. The case $m = 2$ is similar to the second-order approximation of the Laplace transform considered by Catoni in [15]. A significant improvement in comparison with [15, Theorem 1.1.] (apart from having more cumulants) is that we only require averages of $\Delta_V(f)$ to be small while some of them can be quite big.

4.3.3 Cumulants of truncated Gaussians

We introduce a lemma giving errors in approximating cumulants of truncated Gaussian by their values for the unrestricted Gaussian.

Lemma 4.3.3. *Let* A *be an* $n \times n$ *symmetric positive-definite real matrix. Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a measurable function satisfying*

$$|f(\boldsymbol{x})| \leqslant \exp\left(\frac{b}{n} \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}\right) \tag{4.3.6}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$ *and some* $b \geqslant 0$. *Let* $\mathbf{X} : \mathbb{R} \to \mathbb{R}$ *be a random vector with density*

$$\pi^{-n/2} |A|^{1/2} e^{-\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}}.$$

Suppose Ω *is a measurable subset of* $\mathbb{R}^n$ *and define* $p = \mathbf{P}(\mathbf{X} \notin \Omega)$. *Then, if* $p \leqslant \frac{3}{4}$ *and for fixed* s, *we have* $n \geqslant sb + s^2 b^2$, *then*

$$\left|\kappa_s\big(f(\mathbf{X}) \mid \mathbf{X} \in \Omega\big) - \kappa_s f(\mathbf{X})\right| \leqslant 4s! 6^s e^{s^2 b/2 + s/4} p^{1 - sb/n}.$$

The following simple estimate will be useful in the proof of Lemma 4.3.3.

Lemma 4.3.4. *If we have that* $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in [-1, 1]$, *then*

$$\left|\prod_{i \in [n]} a_i - \prod_{i \in [n]} b_i\right| \leqslant \sum_{i \in [n]} |a_i - b_i|. \tag{4.3.7}$$

Proof. We prove it by induction on n. The base case when $n = 1$ holds trivially. Suppose (4.3.7) holds for some $n \geqslant 2$. Let A_n and B_n denote $\prod_{i \in [n]} a_i$ and $\prod_{i \in [n]} b_i$ respectively. Then

$$\begin{aligned} |A_{n+1} - B_{n+1}| = |a_{n+1} A_n - b_{n+1} B_n| &= |(a_{n+1} - b_{n+1}) A_n - b_{n+1}(A_n - B_n)| \\ &\leqslant |a_{n+1} - b_{n+1}| + |A_n - B_n| \leqslant \sum_{i \in [n+1]} |a_i - b_i|. \end{aligned}$$

This completes the proof. $\square$

Proof of Lemma 4.3.3. We use the following estimate from [46, Lemma 4.1].

$$\left|\mathbf{E}\left[f(\mathbf{X}) : \mathbf{X} \in \Omega\right] - \mathbf{E}\left[f(\mathbf{X})\right]\right| \leqslant 15e^{b/2}p^{1-b/n}, \tag{4.3.8}$$

which implies our estimate for the case $s = 1$. Using (4.3.1), we find that

$$\begin{aligned}&\kappa_s(f(\mathbf{X}) : \mathbf{X} \in \Omega) - \kappa_s f(\mathbf{X})\\ &= \sum_{\tau \in P_s} (-1)^{|\tau|-1}(|\pi|-1)! \left(\prod_{B\in\tau} \mathbf{E}\left[f(\mathbf{X})^{|B|}\right] - \prod_{B\in\tau} \mathbf{E}\left[f(\mathbf{X})^{|B|} : \mathbf{X} \in \Omega\right]\right).\end{aligned} \tag{4.3.9}$$

For any s, we can bound the terms in (4.3.9) as follows:

$$|f(\boldsymbol{x})^{|B|}| \leqslant |f(\boldsymbol{x})|^s \leqslant e^{\frac{sb}{n}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}.$$

Therefore, using (4.3.8) (with $f(x)$ replaced by $f(\mathbf{X})^{|B|}$),

$$\left|\mathbf{E}\left[f(\boldsymbol{x})^{|B|}\right] - \mathbf{E}\left[f(\boldsymbol{x})^{|B|} \mid \mathbf{X} \in \Omega\right]\right| \leqslant 15\, e^{sb/2}p^{1-sb/n}.$$

Furthermore, we can bound

$$\left|\mathbf{E}\left[f(\mathbf{X})^{|B|}\right]\right| \leqslant \mathbf{E}\left[\left|f(\mathbf{X})^{|B|}\right|\right] \leqslant \frac{\int_{\mathbb{R}^n} e^{-(1-\frac{sb}{n})\boldsymbol{x}^T A\boldsymbol{x}} d\boldsymbol{x}}{\int_{\mathbb{R}^n} e^{-\boldsymbol{x}^T A\boldsymbol{x}} d\boldsymbol{x}} = (1-sb/n)^{-n/2}$$

and, similarly,

$$\left|\mathbf{E}\left[f(\mathbf{X})^{|B|} : \mathbf{X} \in \Omega\right]\right| \leqslant (1-p)^{-1}(1-sb/n)^{-n/2} \leqslant 4(1-sb/n)^{-n/2}.$$

For $n \geqslant sb + s^2b^2$, the function $(1-sb/n)^{-n/2}e^{-sb/2} < e^{1/4}$ is increasing in b and nonincreasing in n, so

$$(1-sb/n)^{-n/2} < e^{sb/2+1/4}. \tag{4.3.10}$$

Then, since $|\tau| \leqslant s$ for any $\tau \in P_s$, dividing by $4^{|\tau|}(1-sb/n)^{-|\tau|n/2}$ and applying Lemma 4.3.4, we get that

$$\begin{aligned}\left|\prod_{B\in\tau} \mathbf{E}\left[f(\mathbf{X})^{|B|}\right] - \prod_{B\in\tau} \mathbf{E}\left[f(\mathbf{X})^{|B|} : \mathbf{X} \in \Omega\right]\right| &\leqslant 4^{|\tau|}(1-sb/n)^{-|\tau|n/2} \sum_{B\in\tau} \frac{15\, e^{sb/2}p^{1-sb/n}}{4(1-sb/n)^{-n/2}}\\ &\leqslant 15\cdot 4^{s-1}(1-sb/n)^{-(s-1)n/2}e^{sb/2}sp^{1-sb/n}\\ &\leqslant 4^{s+1}e^{s^2b/2+s/4}sp^{1-sb/n}.\end{aligned}$$

Substituting the above bound into (4.3.9) and using Lemma 4.3.1, we derive that

$$\begin{aligned}|\kappa_s(f(\mathbf{X}) : \mathbf{X} \in \Omega) - \kappa_s f(\mathbf{X})| &\leqslant 4^{s+1}e^{s^2b/2+s/4}sp^{1-sb/n} \sum_{\tau\in P_s} (|\pi|-1)!\\ &\leqslant \left(\tfrac{3}{2}\right)^s (s-1)!4^{s+1}e^{s^2b/2+s/4}sp^{1-sb/n}\\ &\leqslant 4\, s!6^s e^{s^2b/2+s/4}p^{1-sb/n},\end{aligned}$$

as required. $\square$

4.3.4 Cumulant expansion for Laplace-type integrals

Recall that our aim is to estimate integrals of the type

$$\int_{\Omega} e^{-\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}+f(\boldsymbol{x})}\,d\boldsymbol{x},$$

which appear in applications of Laplace's method. To apply Theorem 4.3.2, we consider $g(\boldsymbol{y}) := f(T^{-1}\boldsymbol{x})$, where T is a linear transformation such that the components of $\mathbf{Y} = T^{-1}\mathbf{X}$ are independent. We bound the quantities $\Delta_V(g)$ required in Theorem 4.3.2 in terms of the mixed derivatives of f using the following lemma.

For $\Omega \subseteq \mathbb{R}^n$ and some continuous $f : \Omega \to \mathbb{R}$, for $\ell \in [n]$, define

$$\overline{\Delta}_\ell(f,\Omega) := \max_{u_1\in[n]} \sum_{u_2,\ldots,u_\ell\in[n]} \sup_{\boldsymbol{x}\in\Omega} \left|\frac{\partial^\ell f(\boldsymbol{x})}{\prod_{r\in[\ell]}\partial x_{u_r}}\right|. \tag{4.3.11}$$

For $\rho \geqslant 0$, let

$$U_n(\rho) := \{\boldsymbol{x}\in\mathbb{R}^n : \|\boldsymbol{x}\|_\infty \leqslant \rho\}.$$

Lemma 4.3.5. *Suppose $g : U_n(\rho) \to \mathbb{R}$ is an infinitely smooth function. Let $f : \Omega \to \mathbb{R}$ be defined by $f(\boldsymbol{x}) := g(T^{-1}\boldsymbol{x})$, where $\Omega = T(U_n(\rho))$ and T is a real $n\times n$ invertible matrix. Then, for all $\ell \in [n]$,*

$$\max_{j\in[n]} \sum_{V\in\binom{[n]}{\ell}:j\in V} \Delta_V(g) \leqslant \frac{\|T\|_\infty^{\ell-1}\,\|T\|_1}{(\ell-1)!}\rho^\ell\,\overline{\Delta}_\ell(f,\Omega),$$

where $\Delta_V(g) = \Delta_V(g, U_n(\rho))$ is defined according to (4.3.4).

Proof. Applying the mean value theorem, we get that

$$\Delta_V(g) \leqslant \sup_{\boldsymbol{y}\in U_n(\rho)} \left|\frac{\partial^\ell g(\boldsymbol{y})}{\prod_{r\in V}\partial y_r}\right|\rho^\ell.$$

Therefore, it suffices to show that

$$\widehat{\Delta}_\ell(g) \leqslant \frac{\|T\|_\infty^{\ell-1}\,\|T\|_1}{(\ell-1)!}\,\overline{\Delta}_\ell(f,\Omega),$$

where

$$\widehat{\Delta}_\ell(g) := \max_{j\in[n]} \sum_{V\in\binom{[n]}{\ell}:j\in V} \sup_{\boldsymbol{y}\in U_n(\rho)} \left|\frac{\partial^\ell g(\boldsymbol{y})}{\prod_{r\in V}\partial y_r}\right|.$$

Let $T = (t_{jk})$. Since $\boldsymbol{x} = T\boldsymbol{y}$, by the Chain rule, for any distinct $j_1,\ldots,j_\ell \in [n]$, we have that

$$\frac{\partial^\ell g(\boldsymbol{y})}{\partial y_{j_1}\cdots\partial y_{j_\ell}} = \frac{\partial^\ell f(T\boldsymbol{y})}{\partial y_{j_1}\cdots\partial y_{j_\ell}} = \sum_{u_1,\ldots,u_\ell\in[n]} t_{u_1j_1}\cdots t_{u_\ell j_\ell}\,\frac{\partial^\ell f(\boldsymbol{x})}{\partial x_{u_1}\cdots\partial x_{u_\ell}}. \tag{4.3.12}$$

Consequently,

$$\widehat{\Delta}_\ell(g) \leqslant \max_{j\in[n]} \sum_{\substack{\{j_2,\dots,j_\ell\}\subseteq[n] \\ j\notin\{j_2,\dots,j_\ell\}}} \sum_{u_1,\dots,u_\ell\in[n]} |t_{u_1 j} t_{u_2 j_2} \cdots t_{u_\ell j_\ell}| \sup_{\boldsymbol{x}\in\Omega} \left| \frac{\partial^\ell f(\boldsymbol{x})}{\partial x_{u_1} \cdots \partial x_{u_\ell}} \right|.$$

Summing over $\{j_2, \dots, j_\ell\}$ yields

$$\begin{aligned}\widehat{\Delta}_\ell(g) &\leqslant \frac{\|T\|_\infty^{\ell-1}}{(\ell-1)!} \max_{j\in[n]} \sum_{u_1,\dots,u_\ell\in[n]} |t_{u_1 j}| \sup_{\boldsymbol{x}\in\Omega} \left| \frac{\partial^\ell f(\boldsymbol{x})}{\partial x_{u_1} \cdots \partial x_{u_\ell}} \right| \\ &\leqslant \frac{\|T\|_\infty^{\ell-1}}{(\ell-1)!}\, \overline{\Delta}_\ell(f,\Omega) \max_{j\in[n]} \sum_{u_1\in[n]} |t_{u_1 j}| \leqslant \frac{\|T\|_\infty^{\ell-1}\,\|T\|_1}{(\ell-1)!}\, \overline{\Delta}_\ell(f,\Omega).\end{aligned}$$

This completes the proof. ☐

Finally, we are ready to prove the main result of this section.

Theorem 4.3.6. *Let $c_1, c_2, c_3, \varepsilon$ be non-negative real constants with $c_1, \varepsilon > 0$. Let A be an $n \times n$ positive-definite symmetric real matrix and let T be a real matrix such that $T^{\mathrm{T}}AT = I$. Suppose the following assumptions hold for some $m \in [n]$, measurable set $\Omega \subseteq \mathbb{R}^n$, measurable function $f : \mathbb{R}^n \to \mathbb{R}$, and numbers $\rho_1, \rho_2, \alpha \in \mathbb{R}$.*

(i) $U_n(\rho_1) \subseteq T^{-1}(\Omega) \subseteq U_n(\rho_2)$, where $\rho_2 \geqslant \rho_1 \geqslant 2mc_2^{1/2} + c_1(\log n)^{1/2+\varepsilon}$.

(ii) For any $\ell \in [m]$, we have $\rho_2^\ell \frac{\|T\|_\infty^{\ell-1}\,\|T\|_1}{(\ell-1)!}\, \overline{\Delta}_\ell(f, T(U_n(\rho_2))) \leqslant \alpha < \frac{1}{400}$, where $\overline{\Delta}_\ell(\cdot)$ is defined by (4.3.11).

(iii) $n \geqslant m^2c_2^2 + mc_2$ and, for any $\boldsymbol{x} \in \mathbb{R}^n$, we have $|f(\boldsymbol{x})| \leqslant n^{c_3} e^{c_2 \frac{\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}}{n}}$.

Then there is $n_0 = n_0(c_1, c_2, c_3, \varepsilon)$ such that, for any $n \geqslant n_0$, we have

$$\int_\Omega e^{-\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}+f(\boldsymbol{x})}\, d\boldsymbol{x} = \pi^{n/2}|A|^{-1/2} \exp\left(\sum_{s=1}^m \frac{\kappa_s[f(\mathbf{X})]}{s!} + \delta \right),$$

where $\mathbf{X}$ is a random vector with the normal density $\pi^{-n/2}|A|^{1/2}e^{-\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}}$, and

$$\delta = \delta(\Omega, A, f) \leqslant 2n(200\alpha)^{m+1} + e^{-\rho_1^2/2}.$$

Proof of Theorem 4.3.6. Let $\boldsymbol{y} = T^{-1}\boldsymbol{x}$. Since $T^TAT = I$, we have $|T| = |A|^{-1/2}$ and

$$\int_\Omega e^{-\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}+f(\boldsymbol{x})}\, d\boldsymbol{x} = |A|^{-1/2} \int_{T^{-1}(\Omega)} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}+f(T\boldsymbol{y})}\, d\boldsymbol{y}. \tag{4.3.13}$$

Let $\rho \in \{\rho_1, \rho_2\}$. Let $\mathbf{Y}$ have the normal density $\pi^{-n/2}e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}$ and define $p := \mathbf{P}(\mathbf{Y} \notin U_n(\rho))$. In view of the assumption (ii), combining Lemma 4.3.5 and Theorem 4.3.2, we get

$$\log\left(\frac{1}{(1-p)\pi^{n/2}} \int_{U_n(\rho)} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}+f(T\boldsymbol{y})}\, d\boldsymbol{y} \right) = \sum_{s=1}^m \frac{1}{s!}\kappa_s\Big[f(T\mathbf{Y}) \mid \mathbf{Y} \in U_n(\rho) \Big] + \delta',$$

where, since $\alpha < \frac{1}{400}$,

$$|\delta'| = n|\log(1+K)| \leqslant 2n\log(1+|\delta|) \leqslant 2n(200\alpha)^{m+1}.$$

By standard bounds on the tail of the normal distribution, we have $p \leqslant ne^{-\rho^2}/(1+\rho)$. Under our assumptions, there is $n_0 = n_0(c_1, c_2, c_3, \varepsilon)$ such that for $n \geqslant n_0$, we have

$$p \leqslant \tfrac{3}{4}, \qquad mc_2/n \leqslant n^{-1/2}, \qquad 1-p \geqslant \tfrac{1}{2}e^{-\rho^2/2},$$

and

$$\begin{aligned}\sum_{s=1}^{m} \frac{1}{s!}\left|\kappa_s\Big[f(T\mathbf{Y}) \mid \mathbf{Y} \in U_n(\rho)\Big] - \kappa_s[F(\mathbf{Y})]\right| &\leqslant 4\sum_{s=1}^{m} 6^s e^{s^2c_2/2+s/4} p^{1-sc_2/n} \\ &\leqslant 4m\,6^m e^{m^2c_2/2+m/4}\left(\frac{ne^{-\rho^2}}{1+\rho}\right)^{1-mc_2/n} \leqslant e^{-\rho^2/2}.\end{aligned}$$

Then, applying Lemma 4.3.3 to the function $n^{-c_3}f(T\boldsymbol{y})$, we obtain that

$$\int_{U_n(\rho)} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}+f(T\boldsymbol{y})}\,d\boldsymbol{y} = \pi^{n/2}\exp\left(\sum_{s=1}^{m}\frac{\kappa_s(F(\mathbf{Y}))}{s!} + \delta_\rho\right), \tag{4.3.14}$$

where $|\delta_\rho| \leqslant 2n(200\alpha)^{m+1} + e^{-\rho^2/2}$. By assumption (i), we get that

$$\int_{U_n(\rho_1)} e^{f(T\boldsymbol{y})-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\,d\boldsymbol{y} \leqslant \int_{T^{-1}(\Omega)} e^{f(T\boldsymbol{y})-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\,d\boldsymbol{y} \leqslant \int_{U_n(\rho_2)} e^{f(T\boldsymbol{y})-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\,d\boldsymbol{y}.$$

Using (4.3.13) and applying (4.3.14) twice with $\rho = \rho_1$ and $\rho = \rho_2$, we complete the proof. □

4.3.5 Cumulants of Gaussian random variables

If S is a set of even size, a *pairing* of S is a partition of S into $|S|/2$ disjoint pairs. We will write the pairs as (i_1, i_2), here each pair is unordered. Recall the following result of Isserlis [50].

Lemma 4.3.7. *Let A be a positive-definite real symmetric matrix of order n and let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random variable with the Gaussian density $\pi^{-n/2}|A|^{1/2}e^{-\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}}$. Let $\varSigma = (\sigma_{jk}) = (2A)^{-1}$ be the corresponding covariance matrix. Consider a product $Z = X_{j_1}X_{j_2}\cdots X_{j_k}$, where the subscripts do not need to be distinct. If k is odd, then $\mathbf{E}[Z] = 0$. If k is even, then*

$$\mathbf{E}[Z] = \sum_{\{(i_1,i_2),(i_3,i_4),\ldots,(i_{k-1},i_k)\}} \sigma_{j_{i_1}j_{i_2}}\cdots\sigma_{j_{i_{k-1}}j_{i_k}},$$

where the sum is over all pairings of $\{1, \ldots, k\}$. The number of terms in the sum is $(k-1)!! = (k-1)(k-3)\cdots 3\cdot 1$.

In particular, we have

$$\mathbf{E}\left[X_1^2\right] = \sigma_{11}, \quad \mathbf{E}\left[X_1^4\right] = 3\sigma_{11}^2, \quad \mathbf{E}\left[X_1^2X_2^2\right] = \sigma_{11}\sigma_{22} + 2\sigma_{12}^2, \quad \mathbf{E}\left[X_1^6\right] = 15\sigma_{11}^3, \quad \mathbf{E}\left[X_1^8\right] = 105\sigma_{11}^4.$$

In quantum field theory, pairings are known as Feynman graphs, and Lemma 4.3.7 is known as

Wick's formula after a later discoverer.

Theorem 4.3.8. *Assume the conditions of Lemma 4.3.7 for an even k and let $\{P_1, \ldots, P_r\}$ be a partition of set $[k]$. If π is a pairing of $[k]$, define the graph H_π as follows: $V(H_\pi) = [r]$, and for $\ell \neq m$, $\ell m \in E(H_\pi)$ iff π has a pair ij such that $i \in P_\ell$ and $j \in P_m$. Let $\Pi = \Pi(P_1, \ldots, P_m)$ be the set of all pairings π of $[k]$ such that H_π is connected. Then*

$$\kappa\left(\prod_{i\in P_1} Z_{v_i}, \ldots, \prod_{i\in P_r} Z_{v_i}\right) = \sum_{\pi\in\Pi}\prod_{ij\in\pi}\sigma_{v_i v_j},$$

where $v_1, \ldots, v_k \in [N]$ may not necessarily be distinct.

Proof. For $B \subseteq \{1, \ldots, r\}$, define $P(B) = \bigcup_{j\in B} P_j$. So from the definition of joint cumulant (4.3.1), we have that

$$\kappa\left(\prod_{i\in P_1} Z_{v_i}, \ldots, \prod_{i\in P_r} Z_{v_i}\right) = \sum_{B_1\cup\cdots\cup B_t=\{1,\ldots,r\}} (-1)^{t-1}(t-1)! \prod_{u=1}^{t} \mathbf{E}\left[\prod_{i\in P(B_u)} Z_{v_i}\right], \tag{4.3.15}$$

where the sum is over all partitions of $\{1, \ldots, r\}$. By Theorem 4.3.7, we have

$$\prod_{u=1}^{t}\mathbf{E}\left[\prod_{i\in P(B_u)} Z_{v_i}\right] = \sum_{\pi\in\Pi_B}\prod_{ij\in\pi}\sigma_{v_i,v_j},$$

where Π_B is the set of pairings such that no pair spans two of the sets $P(B_u)$.

Now consider any particular product $\prod_{ij\in\pi}\sigma_{v_i v_j}$, corresponding to pairing π. The total weight with which this occurs in (4.3.15) is

$$\sum_{k=1}^{n}(-1)^{k-1}(k-1)!\left\{{m \atop k}\right\}, \tag{4.3.16}$$

where m is the number of components of H_π and $\left\{{m \atop k}\right\}$ denotes the Stirling number of the second kind (number of partitions of an m-set into k parts). A standard identity for the Stirling numbers is that (4.3.16) equals 1 for $m = 1$ and 0 for $m \geqslant 2$, which completes the proof. $\square$

Corollary 4.3.9. *Assume the conditions of Lemma 4.3.7 and that the covariance matrix $\Sigma = (\sigma_{uv})_{u,v\in[N]}$ satisfies $|\sigma_{u,v}| \leqslant 1$ for all $u, v \in [N]$. Then, for any $k_1, \ldots, k_r \in \mathbb{N}$ with even sum,*

$$\sum_{v_1,\ldots,v_r\in[N]} |\kappa(Z_{v_1}^{k_1}, \ldots, Z_{v_r}^{k_r})| \leqslant |\Pi| \cdot \|\Sigma\|_\infty^{r-1} N,$$

where $\Pi = \Pi(P_1, \ldots, P_r)$ is defined as in Theorem 4.3.8 for some fixed partition of $[k]$ with $|P_i| = k_i$ for all $i \in [r]$ and $k = k_1 + \cdots + k_r$. In particular, $|\Pi| \leqslant (k-1)!! = (k-1)(k-3)\cdots 3 \cdot 1$.

Proof. Define $\tau : [k] \to [r]$ as follows: for any $i \in [k]$ if $i \in P_j$ then $\tau(i) := j$. Applying Theorem 4.3.8, we find that

$$\sum_{v_1,\ldots,v_r\in[N]} |\kappa(Z_{v_1}^{k_1}, \ldots, Z_{v_r}^{k_r})| \leqslant \sum_{\pi\in\Pi}\sum_{v_1,\ldots,v_r\in[N]}\prod_{ij\in\pi}|\sigma_{v_{\tau(i)},v_{\tau(j)}}|.$$

It is sufficient to show that, for any $\pi \in \Pi$,

$$\sum_{v_1,\dots,v_r\in[N]} \prod_{ij\in\pi} |\sigma_{v_{\tau(i)},v_{\tau(j)}}| \leqslant \|\Sigma\|_\infty^{r-1} N,$$

We prove it by induction on r. If $r = 1$ then, by assumptions,

$$\sum_{v_1\in[N]} \prod_{ij\in\pi} \sigma_{v_1,v_1} \leqslant N.$$

For the induction step, assume $r \geqslant 2$ and observe that since H_π is connected there is a vertex in $V(H_\pi)$ such that the graph remains connected if we remove it from H_π. Without loss of generality, we can assume that this vertex is r adjacent to vertex x. Let π' be the pairing obtained from π by removing all pairs containing a point from the part P_r corresponding to vertex r of H_π. Estimating other covariances by 1, we find that

$$\sum_{v_r\in[N]} \prod_{ij\in\pi\setminus\pi'} |\sigma_{v_{\tau(i)},v_{\tau(j)}}| \leqslant \sum_{v_r\in[N]} |\sigma_{v_x,v_r}| \leqslant \|\Sigma\|_\infty.$$

Using the induction hypothesis, we get that

$$\sum_{v_1,\dots,v_r\in[N]} \prod_{ij\in\pi} |\sigma_{v_{\tau(i)},v_{\tau(j)}}| \leqslant \sum_{v_1,\dots,v_{r-1}\in[N]} \|\Sigma\|_\infty \prod_{ij\in\pi'} |\sigma_{v_{\tau(i)},v_{\tau(j)}}| \leqslant \|\Sigma\|_\infty^{r-1} N,$$

This proves the induction step and completes the proof the the corollary. □

4.3.6 Bounds on cumulants in Theorem 4.1.1

In the next theorem, we prove bounds on the cumulants of $f_K(\mathbf{X}_G)$ in terms of the infinity norm of a non-singular Gaussian random variable $\mathbf{X}$ that projects into $\mathbf{X}_G$. Let J_n denote the matrix with every entry one.

Lemma 4.3.10. *Let $L = L(G)$ be the Laplacian matrix of a connected graph G. Let Σ_w be defined by $\Sigma_w^{-1} = L + wJ_n$ for some $w > 0$. If $\mathbf{X}$ is a Gaussian random vector with density $(2\pi)^{-n/2}|\Sigma_w|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\Sigma_w^{-1}\boldsymbol{x}\right)$, then $g(\mathbf{X}_G) \overset{d}{=} g(\mathbf{X})$ for any function g such that $g(\boldsymbol{x}+\theta\mathbf{1}) = g(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}$.*

Proof. Since G is connected, we get that $L + wJ_n$ is positive definite matrix for all $w > 0$. Note that $I - J_n/n$ is the projector operator to the space

$$\mathcal{V} = \{\boldsymbol{x} \in \mathbb{R}^n : x_1 + \dots + x_n = 0\}.$$

Applying this linear transformation to the Gaussian vector $\mathbf{X}$, we find that $(I - J_n/n)\mathbf{X}$ has the same distribution as $\mathbf{X}_G$. Therefore, $g(\mathbf{X}_G) \overset{d}{=} g((I - J_n/n)\mathbf{X}) = g(\mathbf{X})$ as claimed. □

Theorem 4.3.11. *Let G be a connected graph with maximal degree d, minimal degree δ. Assume that, for some $w \in (0, 1]$,*

$$\Sigma_w^{-1} = L + wJ_n, \qquad \|\Sigma_w\|_\infty \leqslant 1/2,$$

where $L = L(G)$ is the Laplacian matrix of G. Then the following bounds hold.

(a) For $K \leqslant \delta/2$, we have

$$\mathbf{E}\left[f_K(\mathbf{X}_G)\right] = -\left(\frac{1}{4} + O(\delta^{-1})\right) \sum_{jk \in G} \left(\frac{1}{d_j} + \frac{1}{d_j}\right)^2 + O\left(\frac{n}{\delta}\|\Sigma_w\|_\infty + \frac{nd}{\delta^2}\|\Sigma_w\|_\infty^2\right).$$

where $d_1, \ldots, d_n$ are the degrees of G.

(b) For r, K such that $r(K+2) \leqslant \delta/4$, we have

$$|\kappa_r(f_K(\mathbf{X}_G))| \leqslant \frac{n}{2\delta}\left(\frac{5d}{\delta}\right)^r \|\Sigma_w\|_\infty^{r-1}(4r-1)!!\,.$$

(c) For r, K, K' such that $K' > K$ and $(r+1)(K'+2) \leqslant \delta/4$, we have

$$|\kappa_r(f_K(\mathbf{X}_G)) - \kappa_r(f_{K'}(\mathbf{X}_G))| \leqslant \frac{rn}{2\delta}\left(\frac{5d}{\delta}\right)^r \left(\frac{2}{\delta}\right)^{K-1} \|\Sigma_w\|_\infty^{r-1}(2K+4r-3)!!\,.$$

Proof. Recalling the definition of f_K and applying Lemma 4.3.10, we get that

$$f_K(\mathbf{X}_G) \equiv f_k(\mathbf{X}),$$

where $\mathbf{X}$ is a Gaussian random vector with density $(2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\Sigma_w^{-1}\boldsymbol{x}\right)$. Thus, it is sufficient to prove the bounds for $\mathbf{E}\left[f_K(\mathbf{X})\right]$ and $\kappa_r[f_k(\mathbf{X})]$.

First, we estimate the entries of the matrix $\Sigma_w = (\sigma_{jk})$. Define the diagonal matrix D by $D_{jj} = d_j$, where d_j is the degree of vertex j in G. Observing also that all entries of $\Sigma_w^{-1} - D = L + wJ - D$ are in $[-1, 1]$, we get that

$$\begin{aligned}\|\Sigma_w - D^{-1}\|_{\max} &= \|\Sigma_w(D - \Sigma_w^{-1})D^{-1}\|_{\max} \\ &\leqslant \|\Sigma_w\|_\infty \|(D - \Sigma_w^{-1})D^{-1}\|_{\max} \leqslant \|\Sigma_w\|_\infty \max_{j\in[n]} \frac{1}{d_j} \leqslant \frac{1}{\delta}\|\Sigma_w\|_\infty,\end{aligned}$$

where $\|\cdot\|_{\max}$ denotes the maximum absolute value of entries. This gives

$$\left|\sigma_{jj} - \frac{1}{d_j}\right| \leqslant \frac{1}{\delta}\|\Sigma_w\|_\infty, \qquad |\sigma_{jk}| \leqslant \frac{1}{\delta}\|\Sigma_w\|_\infty, \tag{4.3.17}$$

Let $X_{jk} := X_j - X_k$ for distinct $j, k \in [n]$. Let

$$\sigma_{jk,st} = \operatorname{Cov}(X_{jk}, X_{st}) = \sigma_{js} - \sigma_{jt} - \sigma_{ks} + \sigma_{kt}. \tag{4.3.18}$$

Using 4.3.17, for all distinct $j, k \in [n]$,

$$\sigma_{jk,jk} = \frac{1}{d_j} + \frac{1}{d_k} \pm \frac{4}{\delta}\|\Sigma_w\|_\infty.$$

Then, by the assumption that $\|\Sigma_w\|_\infty \leqslant 1/2$, we get

$$|\sigma_{jk,st}| \leqslant \begin{cases} \frac{4}{\delta}, & \text{if } |\{j,k\} \cap \{s,t\}| \geqslant 1, \\ \frac{4}{\delta}\|\Sigma_w\|_\infty, & \text{if } |\{j,k\} \cap \{s,t\}| = 0. \end{cases} \tag{4.3.19}$$

From Lemma 4.3.7, we know that

$$\mathbf{E}\left[X_{jk}^{2\ell}\right] = (2\ell-1)!!\,\sigma_{jk}^{\ell}.$$

Recalling the definition of $f_K(\boldsymbol{x})$ from (4.1.10) and using (4.1.9), we obtain that

$$\begin{aligned}\mathbf{E}\left[f_K(\mathbf{X})\right] &= \sum_{\ell=2}^{K} c_{2\ell} \sum_{jk\in G} \mathbf{E}\left[X_{jk}^{2\ell}\right] = -\frac{1}{12}\sum_{jk\in G}\mathbf{E}\left[X_{jk}^{4}\right] + \sum_{\ell=3}^{K} c_{2\ell}\sum_{jk\in G}\mathbf{E}\left[X_{jk}^{2\ell}\right]\\ &= -\frac{1}{4}\sum_{jk\in G}\left(\frac{1}{d_j}+\frac{1}{d_j}\pm\frac{4}{\delta}\|\Sigma_w\|_\infty\right)^2\left(1+O(1)\sum_{l=3}^{K}\left(\frac{2}{\pi}\right)^{2\ell}(2\ell-1)!!\left(\frac{4}{\delta}\right)^{\ell-2}\right).\end{aligned}$$

Since $x(2\ell-x)\leqslant \ell^2$, we can bound $(2\ell-1)!!\leqslant \ell^\ell$. Recalling the assumption that $K\leqslant\delta/2$, we estimate

$$\sum_{\ell=3}^{K}\left(\frac{2}{\pi}\right)^{2\ell}(2\ell-1)!!\left(\frac{4}{\delta}\right)^{\ell-2} = O(\delta^{-1}).$$

We also have

$$\sum_{jk\in G}\left(\frac{1}{d_j}+\frac{1}{d_j}\pm\frac{4}{\delta}\|\Sigma_w\|_\infty\right)^2 = \sum_{jk\in G}\left(\frac{1}{d_j}+\frac{1}{d_j}\right)^2 \pm \frac{16n}{\delta}\|\Sigma_w\|_\infty \pm \frac{16nd}{\delta^2}\|\Sigma_w\|_\infty^2.$$

This proves (a).

We proceed to part (b). Using (4.3.18), for any distinct $j,k\in[n]$, we have that

$$\sum_{st\in G}\left|\sigma_{jk,st}\right| \leqslant 2\sum_{st\in G}\left|\sigma_{js}\right| + 2\sum_{st\in G}\left|\sigma_{ks}\right| \leqslant 4d\|\Sigma_w\|_\infty. \tag{4.3.20}$$

By linearity,

$$\begin{aligned}\kappa_r\left(f_K(\mathbf{X})\right) &= \kappa_r\left(\sum_{\ell=2}^{K} c_{2\ell}\sum_{jk\in G}(X_j-X_k)^{2\ell}\right)\\ &= \sum_{\ell_1=2}^{K}\cdots\sum_{\ell_r=2}^{K}\sum_{e_1,\ldots,e_r\in G}\left(\prod_{s\in[r]} c_{2\ell_s}\right)\kappa\left(X_{e_1}^{2\ell_1},\ldots,X_{e_r}^{2\ell_r}\right).\end{aligned} \tag{4.3.21}$$

Applying Corollary 4.3.9 for $(Z_v)_{v\in[N]} := \dfrac{\delta^{1/2}}{2}(X_{jk})_{jk\in G}$ and $N=nd/2$. From (4.3.19), we get that the covariances of $(Z_v)_{v\in[N]}$ are bounded by 1 as required. Using (4.3.20) to bound the infinity norm of the covariance matrix, we get that

$$\sum_{e_1,\ldots,e_r\in G}\left|\kappa\left(X_{e_1}^{2\ell_1},\ldots,X_{e_r}^{2\ell_r}\right)\right| \leqslant \frac{nd}{2}\left(\frac{4}{\delta}\right)^{\ell}(\delta d\|\Sigma_w\|_\infty)^{r-1}(2\ell-1)!!, \tag{4.3.22}$$

where $\ell=\ell_1+\cdots+\ell_r$. Since $\ell_i\in\{2,\ldots,K\}$, we can bound

$$(2\ell-1)!! \leqslant (2r+\ell)^{\ell-2r}(4r-1)!! \leqslant ((K+2)r)^{\ell-2r}(4r-1)!!$$

Substituting the above bounds and also (4.1.9) into (4.3.21) and recalling that $r(K+2) \leqslant \delta/4$, we get

$$\begin{aligned}
|\kappa_r(f_K(\mathbf{X}))| &\leqslant \tfrac{nd}{2}(\delta d\|\Sigma_w\|_\infty)^{r-1}(4r-1)!! \sum_{\ell_1=2}^{K}\cdots\sum_{\ell_r=2}^{K}\left(\tfrac{2}{\pi}\right)^{2\ell}\left(\tfrac{4}{\delta}\right)^{\ell}((K+2)r)^{\ell-2r} \\
&\leqslant \tfrac{nd}{2}(\delta d\|\Sigma_w\|_\infty)^{r-1}\left(\tfrac{16}{\pi^2\delta}\right)^{2r}(4r-1)!!\left(\sum_{i=0}^{K-2}\left(\tfrac{16r(K+2)}{\pi^2\delta}\right)^{i}\right)^{r} \\
&\leqslant \tfrac{n}{2\delta}\left(\tfrac{d}{\delta}\right)^{r}\|\Sigma_w\|_\infty^{r-1}(4r-1)!!\left(\tfrac{256}{\pi^4}\sum_{i=0}^{\infty}\left(\tfrac{4}{\pi^2}\right)^{i}\right)^{r}.
\end{aligned}$$

Computing $\frac{256}{\pi^4}\sum_{i=0}^{\infty}\left(\frac{4}{\pi^2}\right)^i \approx 4.42 < 5$, part (b) follows.

For part (c), using (4.1.9), (4.3.21) and (4.3.22), we estimate

$$\begin{aligned}
&\Big|\kappa_r\left(f_K(\mathbf{X})\right)-\kappa_r\left(f_{K'}(\mathbf{X})\right)\Big| \\
&\qquad\leqslant r\sum_{\ell_1=K+1}^{K'}\sum_{\ell_2=2}^{K'}\cdots\sum_{\ell_r=2}^{K'}\sum_{e_1,\dots,e_r\in G}\left(\prod_{s\in[r]}c_{2\ell_s}\right)\Big|\kappa\left(X_{e_1}^{2\ell_1},\dots,X_{e_r}^{2\ell_r}\right)\Big| \\
&\qquad\leqslant r\sum_{\ell_1=K+1}^{K'}\sum_{\ell_2=2}^{K'}\cdots\sum_{\ell_r=2}^{K'}\tfrac{nd}{2}\left(\tfrac{16}{\delta\pi^2}\right)^{\ell}(\delta d\|\Sigma_w\|_\infty)^{r-1}(2\ell-1)!!,
\end{aligned} \tag{4.3.23}$$

where $\ell=\ell_1+\cdots+\ell_r$ as before. Similarly, we get

$$(2\ell-1)!! \leqslant (\ell+K+2r-1)^{\ell-K-2r+1}(2K+4r-3)!! \leqslant ((r+1)(K'+2))^{\ell-K-2r+1}(2K+4r-3)!!.$$

Then, using $(r+1)(K'+2)\leqslant \delta/4$, we can bound

$$\begin{aligned}
&\sum_{\ell_1=K+1}^{K'}\sum_{\ell_2=2}^{K'}\cdots\sum_{\ell_r=2}^{K'}\left(\tfrac{16}{\delta\pi^2}\right)^{\ell}(2\ell-1)!! \\
&\qquad\leqslant \left(\tfrac{16}{\delta\pi^2}\right)^{K+2r-1}(2K+4r-3)!!\left(\sum_{i=0}^{K'}\left(\tfrac{16(r+1)(K'+2)}{\delta\pi^2}\right)^{i}\right)^{r} \\
&\qquad\leqslant 5^r\left(\tfrac{16}{\pi^2\delta}\right)^{K-1}\delta^{-2r}(2K+4r-3)!!
\end{aligned}$$

Substituting this bound into (4.3.23) completes the proof. □

4.4 Estimating the integral for Eulerian orientations

Throughout this section, we work under the assumptions of Theorem 4.1.1. For $\boldsymbol{\theta}\in\mathbb{R}^n$, define

$$F(\boldsymbol{\theta}) := \prod_{jk\in G}\cos(\theta_j-\theta_k). \tag{4.4.1}$$

Using generating functions, see, for example, [45, Section 3], one can show that

$$\mathrm{EO}(G) = 2^{|E(G)|}\pi^{-n}J', \quad \text{where} \quad J' := \int_{(\mathbb{R}/\pi)^n} F(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \tag{4.4.2}$$

The integral J' in (4.4.2) is dominated by the region Ω_0 defined below, where all θ_j are approximately the same, which makes it possible to estimate the integral asymptotically.

Given $x \in \mathbb{R}$, an *interval* of $\mathbb{R}/\pi$ of *length* $\rho \geqslant 0$ is a set of the form

$$I(x, \rho) := \{\theta \in \mathbb{R}/\pi : |x - \theta|_\pi \leqslant \tfrac{1}{2}\rho\}, \quad \text{where} \quad |x|_\pi := \min\{|x - k\pi| : k \in \mathbb{Z}\}.$$

By the assumption $d \gg \log^8 n$, we choose some $\zeta > 0$ such that

$$\Big(\frac{\log n}{d^{1/8}} + \frac{1}{\log\log n}\Big)^{1/2} \ll 1/\zeta \ll 1, \tag{4.4.3}$$

and choose

$$\rho_0 := \zeta^2 d^{-1/2} \log^{3/2} n \tag{4.4.4}$$

to satisfy a finite number of inequalities in the proof.

Define

$$J_0 := \int_{\Omega_0} F(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{4.4.5}$$

where Ω_0 is the region consists of those $\boldsymbol{\theta} \in (\mathbb{R}/\pi)^n$ such that all components θ_j can be covered by an interval of $\mathbb{R}/\pi$ of length at most ρ_0:

$$\Omega_0 := \big\{\boldsymbol{\theta} \in (\mathbb{R}/\pi)^n : \text{there exists } x \in \mathbb{R}/\pi \text{ such that } \boldsymbol{\theta} \in I(x, \rho_0)^n\big\}. \tag{4.4.6}$$

The proof of Theorem 4.1.1 consists of two parts. First, we estimate J_0 using Theorem 4.3.6 and some preliminary lemmas given in Section 4.4.1. Then, we show that the integral over the region $(\mathbb{R}/\pi)^n \setminus \Omega_0$ is negligible in comparison with J_0.

4.4.1 Preliminaries

Here we state two lemmas from [46, 45] that will be useful in the proof of Theorem 4.3.6. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator, let $\ker T = \{\mathbf{x} \in \mathbb{R}^n : T\mathbf{x} = 0\}$.

The first lemma helps to deal with integrals over a subspace of $\mathbb{R}^n$.

Lemma 4.4.1 ([46, Lemma 4.6])**.** *Let $Q, W : \mathbb{R}^n \to \mathbb{R}^n$ be linear operators such that $\ker Q \cap \ker W = \{\mathbf{0}\}$ and $\operatorname{span}(\ker Q, \ker W) = \mathbb{R}^n$. Let $n_\perp$ denote the dimension of $\ker Q$. Suppose $\Omega \subseteq \mathbb{R}^n$ and $F : \Omega \cap Q(\mathbb{R}^n) \to \mathbb{C}$. For any $\eta > 0$, define*

$$\Omega_\eta = \big\{\boldsymbol{x} \in \mathbb{R}^n : Q\boldsymbol{x} \in \Omega \text{ and } W\boldsymbol{x} \in U_n(\eta)\big\}.$$

Then, if the integrals exist,

$$\int_{\Omega \cap Q(\mathbb{R}^n)} F(\boldsymbol{y})\, d\boldsymbol{y} = (1-\delta)^{-1}\, \pi^{-n_\perp/2} \left|Q^{\mathrm{T}}Q + W^{\mathrm{T}}W\right|^{1/2} \int_{\Omega_\eta} F(Q\boldsymbol{x})\, e^{-\boldsymbol{x}^{\mathrm{T}}W^{\mathrm{T}}W\boldsymbol{x}}\, d\boldsymbol{x}, \tag{4.4.7}$$

where

$$0 \leqslant \delta < \min(1, ne^{-\eta^2/\kappa^2}), \quad \kappa = \sup_{W\mathbf{x} \neq 0} \frac{\|W\boldsymbol{x}\|_\infty}{\|W\boldsymbol{x}\|_2} \leqslant 1.$$

Moreover, if $U_n(\eta_1) \subseteq \Omega \subseteq U_n(\eta_2)$ for some $\eta_2 \geqslant \eta_1 > 0$ then

$$U_n\left(\min\left(\frac{\eta_1}{\|Q\|_\infty}, \frac{\eta}{\|W\|_\infty}\right)\right) \subseteq \Omega_\eta \subseteq U_n\left(\|P\|_\infty\, \eta_2 + \|R\|_\infty\, \eta\right)$$

for any linear operators $P, R : \mathbb{R}^n \to \mathbb{R}^n$ such that $PQ + RW$ is equal to the identity operator on $\mathbb{R}^n$.

Let

$$A := \frac{d}{n} J_n + \frac{1}{2} L, \tag{4.4.8}$$

where $L = L(G)$ is the Laplacian matrix of G defined by (4.1.7), and recall that J_n denotes the matrix with every entry one. The next lemma will be useful for the norm bounds required in the application of Theorem 4.3.6.

Lemma 4.4.2 ([45, Lemma 12]). *Under the assumptions of Theorem 4.1.1, the following estimates hold.*

(a) $\|A^{-1}\|_\infty = O\big(d^{-1} \log \frac{2n}{\Delta}\big)$.

(b) If $A^{-1} = (a_{jk})$, then $a_{jj} = O(d^{-1})$ and $a_{jk} = O\big(d^{-2} \log \frac{2n}{\Delta}\big)$ uniformly for $1 \leqslant j \neq k \leqslant n$.

(c) There exists a symmetric positive-definite matrix $T = A^{-1/2}$ such that $T^{\mathrm{T}} A T = I$. Moreover, $\|T\|_\infty = O(d^{-1/2} \log^{1/2} n)$ and $\|T^{-1}\|_\infty = O(d^{1/2})$.

The above lemma is derived under the assumption in [45], its proof works under our slightly stronger assumption in Theorem 4.1.1.

4.4.2 The integral inside Ω_0

In this section, we obtain the estimates of J_0 in Lemma 4.4.3 and Lemma 4.4.7. By the definition of Ω_0 in (4.4.6), we have that

$$\{\boldsymbol{\theta} \in (\mathbb{R}/\pi)^n : \max_{j \in [n]} |\theta_j - \theta_n|_\pi \leqslant \rho_0\} \subseteq \Omega_0 \subseteq \{\boldsymbol{\theta} \in (\mathbb{R}/\pi)^n : \max_{j \in [n]} |\theta_j - \theta_n|_\pi \leqslant 2\rho_0\}.$$

Observe from the definition of $F(\boldsymbol{\theta})$ (4.4.1) that, for any $\boldsymbol{\theta} \in \mathbb{R}^n$,

$$F(\boldsymbol{\theta}) = F(\boldsymbol{\theta} - \theta_n \mathbf{1}), \tag{4.4.9}$$

where $\mathbf{1} := (1, \ldots, 1)^T \in \mathbb{R}^n$. Therefore,

$$J_0 = \pi \int_{\Omega_0 \cap \mathcal{L}} F(\boldsymbol{\theta}')\, d\boldsymbol{\theta}', \tag{4.4.10}$$

where $\mathcal{L} := \{\boldsymbol{\theta}' \in \mathbb{R}^n : \theta'_n = 0\}$ and

$$U_{n-1}(\rho_0) \subseteq (\Omega_0 \cap \mathcal{L}) \subseteq U_{n-1}(2\rho_0).$$

Next, we lift the integral back to the full dimension using Lemma 4.4.1. Let M be the matrix with one in the last column and zero elsewhere. Define

$$P = I - \frac{1}{n} J, \quad Q = I - M, \quad R = d^{-1/2} I, \quad \text{and} \quad W = d^{1/2} n^{-1} J,$$

where J denotes the matrix with every entry one. One can easily check that $PQ + RW = I$, and also that $\ker Q \cap \ker W = \{\mathbf{0}\}$, $\ker Q$ has dimension 1 and $\langle \ker Q, \ker W \rangle = \mathbb{R}^n$. From [45, Section 3.1], we know that

$$|Q^{\mathrm{T}}Q + W^{\mathrm{T}}W| = nd, \qquad \|P\|_\infty \leqslant 2 \qquad \|Q\|_\infty = 2, \qquad \|R\|_\infty = d^{-1/2}, \qquad \|W\|_\infty = d^{1/2}.$$

Then applying Lemma 4.4.1 with $\kappa = 1$, $\eta_1 = \rho_0$, $\eta_2 = 2\rho_0$, we get, from (4.4.7), that

$$J_0 = \left(1 + e^{-\omega(\log n)}\right) \pi^{1/2} (dn)^{1/2} \int_{\Omega} \widehat{F}(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{4.4.11}$$

where Ω is some region such that

$$U_n\left(\rho_0/2\right) \subseteq \Omega \subseteq U_n(5\rho_0) \tag{4.4.12}$$

and

$$\widehat{F}(\boldsymbol{\theta}) := \exp\left(-\tfrac{d}{n} \Big(\sum_{i \in [n]} \theta_i \Big)^2 + \sum_{jk \in G} \log\cos(\theta_j - \theta_k) \right).$$

Recall from the assumptions in Theorem 4.1.1 that

$$K = \left\lceil \frac{(c+1)\log n}{\log d - 4\log\log n} \right\rceil, \qquad M = \left\lceil \frac{2(c+1)\log n}{\log d - 8\log\log n} \right\rceil.$$

In view of the definition of f_K in (4.1.10), and the definition of matrix A in (4.4.8), using Taylor's series, we get that

$$\begin{aligned} \log \widehat{F}(\boldsymbol{\theta}) &= -\tfrac{d}{n} \Big(\sum_{i \in [n]} \theta_i \Big)^2 - \tfrac{1}{2} \sum_{jk \in G} (\theta_j - \theta_k)^2 + \sum_{\ell \geqslant 2} c_{2\ell} \sum_{jk \in G} (\theta_j - \theta_k)^{2\ell} \\ &= -\boldsymbol{\theta}^{\mathrm{T}} A \boldsymbol{\theta} + f_K(\boldsymbol{\theta}) + O(n^{-c}). \end{aligned} \tag{4.4.13}$$

where for all $\boldsymbol{\theta} \in \Omega$, noting the bound on $c_{2\ell}$ in (4.1.9), the choice of K, and the condition on ϵ in (4.4.3), we have

$$\sum_{\ell > K} c_{2\ell} \sum_{jk \in G} (\theta_j - \theta_k)^{2\ell} = \sum_{jk \in G} O\left(\rho_0^{2(K+1)}\right) = O\left(nd \left(\zeta^2 d^{-1/2} \log^{3/2} n \right)^{2K+2} \right) = O(n^{-c}).$$

Now we are ready to derive the main estimate of this section.

Lemma 4.4.3. *Under the assumptions of Theorem 4.1.1, we have*

$$J_0 = \pi^{(n+1)/2} d^{1/2} n^{1/2} |A|^{-1/2} \exp\left(\sum_{s=1}^{M} \frac{1}{s!} \kappa_s\left(f_K(\mathbf{X})\right) + O\left(n^{-c}\right) \right), \tag{4.4.14}$$

where $\mathbf{X}$ is a random vector with the normal density $\pi^{-n/2}|A|^{1/2}e^{-\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}}$.

Proof. The proof is by combining (4.4.13) and Theorem 4.3.6. First, we verify its assumptions (i)–(iii). By Lemma 4.4.2(c), there is a symmetric positive definite matrix T such that $T^T A T = I$ and

$$\|T\|_\infty = O(d^{-1/2} \log^{1/2} n), \qquad \|T^{-1}\|_\infty = O(d^{1/2}). \tag{4.4.15}$$

Recall from (4.4.4) that $\rho_0 = \Theta(\zeta^2 d^{-1/2}\log^{3/2} n)$. Then, combining (4.4.12) and (4.4.15), we get

$$U_n(\widehat{\rho}_1) \subseteq T^{-1}(\varOmega) \subseteq U_n(\widehat{\rho}_2),$$

where

$$\widehat{\rho}_1 = \Theta(\zeta^2 \log n) \quad \text{ and } \quad \widehat{\rho}_2 = \Theta(\zeta^2 \log^{3/2} n).$$

This is because for $\boldsymbol{\theta} \in U_n(O(\rho_0))$, we have that

$$\|T^{-1}\boldsymbol{\theta}\|_\infty \leqslant \|T^{-1}\|_\infty \|\boldsymbol{\theta}\|_\infty = O\left(\zeta^2 \log^{3/2} n\right),$$

and

$$\|T\|_\infty^{-1}\|\boldsymbol{\theta}\|_\infty = O\left(\frac{\zeta^2 d^{-1/2}\log^{3/2} n}{d^{-1/2}\log^{1/2} n}\right) = O\left(\zeta^2 \log n\right).$$

Since $M = o(\log n)$, assumption (i) of Theorem 4.3.6 holds.

For assumption (ii), we need to estimate $\overline{\Delta}_\ell(f_M, T(U_n(\widehat{\rho}_2)))$ for $\ell \in [M]$. Using (4.4.15), we get that, for any $\boldsymbol{\theta} \in T(U_n(\widehat{\rho}_2))$,

$$\|\boldsymbol{\theta}\|_\infty = O\left(\widehat{\rho}_2 d^{-1/2}\log^{1/2} n\right) = O\left(\zeta^2 d^{-1/2}\log^2 n\right).$$

Then, for any integer $i \in [M]$,

$$\frac{\partial^i f_K(\boldsymbol{\theta})}{\partial^i \theta_j} = \sum_{\ell=2}^{K} (2\ell)_i c_{2\ell} \sum_{k: jk\in G} (\theta_j - \theta_k)^{(2\ell-i)_+} = O\left(d\|\boldsymbol{\theta}\|_\infty^{(4-i)_+}\right),$$

where $(\cdot)_i$ denotes the falling factorial and $(x)_+$ denotes $\max\{x, 0\}$. Similarly, for $jk \in G$ and $p, q \in [K]$,

$$\frac{\partial^{p+q} f_K(\boldsymbol{\theta})}{\partial^p \theta_j \partial^q \theta_k} = \sum_{\ell=2}^{K} (2\ell)_{p+q} c_{2\ell} (\theta_j - \theta_k)^{(2\ell-p-q)_+} = O\left(\|\boldsymbol{\theta}\|_\infty^{(4-p-q)_+}\right).$$

All higher-order mixed derivatives with three or more distinct indices are zeros. Since T is symmetric, we have $\|T\|_1 = \|T\|_\infty$. Then, using (4.4.15), we obtain

$$\begin{aligned}
&\widehat{\rho}_2^{\ell} \frac{\|T\|_\infty^{\ell-1}\|T\|_1}{(\ell-1)!} \max_{u_1\in[n]} \sum_{u_2,\dots,u_\ell\in[n]} \sup_{\boldsymbol{\theta}\in T(U_n(\widehat{\rho}_2))} \left| \frac{\partial^\ell f_K(\boldsymbol{\theta})}{\prod_{r\in[\ell]}\partial\theta_{u_r}} \right| \\
&= O\left((\widehat{\rho}_2 \|T\|_\infty)^\ell \, d\|\boldsymbol{\theta}\|_\infty^{(4-\ell)_+} \right) = O\left(d \left(\zeta^2 d^{-1/2}\log^2 n\right)^{\ell+(4-\ell)_+} \right) = O(d^{-1}\zeta^8 \log^8 n).
\end{aligned}$$

Thus, assumption (ii) of Theorem 4.3.6 holds with some $\alpha = O(d^{-1}\zeta^8\log^8 n)$.

Finally, using Lemma 4.4.2(a), we find that

$$\boldsymbol{\theta}^T A \boldsymbol{\theta} \geqslant \frac{\|\boldsymbol{\theta}\|_2^2}{\|A^{-1}\|_2} \geqslant \frac{\|\boldsymbol{\theta}\|_2^2}{\|A^{-1}\|_\infty} = \Omega\left(d\|\boldsymbol{\theta}\|_2^2 \log\frac{2n}{\varDelta}\right).$$

Then, assumption (iii) trivially holds with $c_3 = K+1$ and $c_2 = 1$.

Combining (4.4.11), (4.4.13) and applying Theorem 4.3.6, we find that

$$J_0 = \pi^{(n+1)/2} d^{1/2} n^{1/2} |A|^{-1/2} \exp\left(\sum_{r=1}^{M} \frac{1}{r!}\, \kappa_r \left(f_K(\mathbf{X}) \right) + \delta \right),$$

where

$$|\delta| \leqslant \exp\left(e^{-\widehat{\rho}_1^2/2} + n(200\alpha)^{M+1} \right) - 1 = \exp\left(n^{-\omega(\log n)} + O\left(n^{-c}\right) \right) - 1 = O\left(n^{-c}\right).$$

This completes the proof. □

Next, we obtain an explicit bound on J_0 starting with a few useful lemmas whose variants appeared in [45].

Lemma 4.4.4. *For $jk \in G$, $|\cos(x)|$ is a decreasing function of $|x|_\pi$ with $\cos(0) = 1$ and*

$$|\cos(x)|^2 = 1 - \sin^2 x \leqslant \exp\left(-\tfrac{1}{4} |x|_\pi^2 \right). \tag{4.4.16}$$

In addition, for any $|y|_\pi \leqslant |x|_\pi$, we have

$$|\cos(x)| \leqslant |\cos(y)| \, \exp\left(-\tfrac{1}{4\pi} (|x|_\pi^2 - |y|_\pi^2)(\pi - |x|_\pi - |y|_\pi) \right). \tag{4.4.17}$$

Proof. The first part of (4.4.16) follows from the definition of $\cos(x)$ and implies that $|\cos(x)| = \cos(|x|_\pi)$ for all x. Therefore we can assume that $0 \leqslant y \leqslant x \leqslant \frac{1}{2}\pi$, which implies that $|x|_\pi = x$ and $|y|_\pi = y$.

By the concavity of $\cos x$ on $[0, \frac{\pi}{2}]$, we have $\cos x \geqslant 1 - \frac{2x}{\pi}$ on this range, which in turn implies (by symmetry about the line $x = \frac{\pi}{2}$) that

$$\sin x \geqslant \frac{1}{\pi} x(\pi - x), \quad x \in [0, \pi]. \tag{4.4.18}$$

Therefore, for $0 \leqslant x \leqslant \frac{1}{2}\pi$, we have that

$$|\cos(x)|^2 = 1 - \sin^2 x \leqslant 1 - \tfrac{1}{\pi^2} x^2 (\pi - x)^2 \leqslant 1 - \tfrac{1}{4} x^2 \leqslant \exp\left(-\tfrac{x^2}{4} \right).$$

Inequality (4.4.17) is trivial if $x = y$, so assume that $0 \leqslant y < x \leqslant \frac{1}{2}\pi$. In that case, $\cos(y) \neq 0$ and

$$\tfrac{\cos^2(x)}{\cos^2(y)} \leqslant \tfrac{1 - \sin^2 x}{1 - \sin^2 y} \leqslant \exp\left(-\tfrac{\sin^2 x - \sin^2 y}{1 - \sin^2 y} \right) \leqslant \exp\bigl(-(\sin^2 x - \sin^2 y) \bigr).$$

Finally, by (4.4.18), we have

$$\sin^2 x - \sin^2 y = \sin(x+y) \sin(x-y) \geqslant \tfrac{1}{\pi^2}(x^2 - y^2)(\pi - x + y)(\pi - x - y) \geqslant \tfrac{1}{2\pi}(x^2 - y^2)(\pi - x - y)$$

for $0 \leqslant y \leqslant x \leqslant \frac{1}{2}\pi$, which completes the proof of (4.4.17). □

Now, we introduce a generalisation of Lemma 4.4.2 (c) following essentially the same proof in [45].

Lemma 4.4.5. *Given some constant $c > 0$, let $B := \frac{c}{n} J + L$. Then we have $\|B^{-1/2}\|_\infty = O(d^{-1/2} \log^{1/2} n)$ and $\|B^{1/2}\|_\infty = O(d^{1/2})$.*

To prove Lemma 4.4.5, we need the following lemma.

Lemma 4.4.6. *[45, Corollary 28] Let L be a symmetric matrix with nonpositive off-diagonal elements and zero row sums. Suppose the eigenvalues of L are $0 = \mu_1 < \mu_2 \leqslant \cdots \leqslant \mu_n$. Let $X := I - (2\|L\|_\infty)^{-1}L$ with eigenvalues $1 = \nu_1 > \nu_2 \geqslant \cdots \geqslant \nu_n$, where $\nu_j = 1 - (2\|L\|_\infty)^{-1}\mu_j$ for each j. For $c > 0$, let $B := \frac{c}{n}J + L$. Then, for any real $\alpha \geqslant -1$, the positive-definite power B^α satisfies*

$$\|B^\alpha\|_\infty \leqslant c^\alpha + (2\|L\|_\infty)^\alpha \left(2 \sum_{k=0}^{N-1} \left| \binom{\alpha}{k} \right| + \frac{1}{(1-\nu_2)n^{1/2}} \right),$$

where $N = \lceil |\alpha| + \log_{\nu_2} n^{-1} \rceil$.

Proof of Lemma 4.4.5. We follow the proof of Lemma 12 in [45]. By noting that $\left|\binom{-1/2}{k}\right| < k^{-1/2}$ for $k \geqslant 1$, we have, by Lemma 4.4.6, that

$$\begin{aligned}
\|B^{-1/2}\|_\infty &= \|(\tfrac{c}{n}J + L)^{-1/2}\|_\infty \leqslant c^{-1/2} + (2\|L\|_\infty)^{-1/2} \left(2 \sum_{k=0}^{N-1} \left| \binom{-1/2}{k} \right| + \frac{1}{(1-\nu_2)n^{1/2}} \right) \\
&\leqslant c^{-1/2} + (2\|L\|_\infty)^{-1/2} \left(2 + 2 \sum_{k=1}^{N-1} k^{-1/2} + \frac{2\|L\|_\infty}{\mu_2 n^{1/2}} \right) \\
&\leqslant c^{-1/2} + (2\|L\|_\infty)^{-1/2} \left(2 + 2 \sum_{k=1}^{N-1} k^{-1/2} \right) + \frac{(2\|L\|_\infty)^{1/2}}{\mu_2 n^{1/2}} \\
&\approx c^{-1/2} + d^{-1/2}\sqrt{N} + \frac{2d^{1/2}}{\mu_2 n^{1/2}} = c^{-1/2} + d^{-1/2}\sqrt{-\frac{1}{2} + \frac{\log n}{\log(1/\nu_2)}} + \frac{2d^{1/2}}{\mu_2 n^{1/2}} = O(d^{-1/2}\log^{1/2} n).
\end{aligned}$$

Similarly, we have, by $\left|\binom{1/2}{k}\right| < k^{-3/2}$ for $k \geqslant 1$, that

$$\begin{aligned}
\|B^{1/2}\|_\infty = \|(\tfrac{c}{n}J + L)^{1/2}\|_\infty &\leqslant c^{1/2} + (2\|L\|_\infty)^{1/2} \left(2 \sum_{k=0}^{N-1} \left| \binom{1/2}{k} \right| + \frac{1}{(1-\nu_2)n^{1/2}} \right) \\
&\leqslant c^{1/2} + (2\|L\|_\infty)^{1/2} \left(2 + 2 \sum_{k=1}^{N-1} k^{-3/2} \right) + \frac{(2\|L\|_\infty)^{3/2}}{\mu_2 n^{1/2}} \\
&\approx c^{1/2} + 2d^{1/2}(1 - \frac{1}{2\sqrt{N}}) + \frac{8d^{3/2}}{\mu_2 n^{1/2}} = O(d^{1/2}).
\end{aligned}$$

This completes the proof in view of the constraint on d. □

We are ready to bound J_0.

Lemma 4.4.7. *We have*

$$\int_{\Omega_0} |F(\boldsymbol{\theta})|\, d\boldsymbol{\theta} = J_0 = e^{O(n \log n)}.$$

Proof of Lemma 4.4.7. Since $\boldsymbol{\theta} \in I(x, \rho_0)^n$ for $\boldsymbol{\theta} \in \Omega_0$ with some $x \in \mathbb{R}/\pi$, then

$$\int_{\Omega_0} |F(\boldsymbol{\theta})|\, d\boldsymbol{\theta} = \int_{\Omega_0} F(\boldsymbol{\theta})\, d\boldsymbol{\theta} = J_0.$$

By the cosine inequality, we have, for all x, that

$$\cos(x) \geqslant 1 - \frac{1}{2}x^2.$$

Therefore for $\boldsymbol{\theta} \in \Omega_0$, we have, for any $jk \in G$, that $\theta_j - \theta_k = o(1)$, and in view of the definition (4.4.1), we obtain that

$$F(\boldsymbol{\theta}) = \prod_{jk \in G} \cos(\theta_j - \theta_k) \geqslant \prod_{jk \in G} \left(1 - \tfrac{1}{2}(\theta_j - \theta_k)^2\right) \geqslant \exp\left(- \sum_{jk \in G} (\theta_j - \theta_k)^2\right),$$

where we use $1 - x/2 \geqslant e^{-x}$ for $x \in [0, 1.59]$. We also get that

$$F(\boldsymbol{\theta}) \leqslant \exp\left(-\frac{1}{8} \sum_{jk \in G} (\theta_j - \theta_k)^2\right),$$

since $\cos(x) \leqslant \exp\left(-x^2/8\right)$ by Lemma 4.4.4. Let

$$F_0(\boldsymbol{\theta}) := \exp\left(- \sum_{jk \in G} (\theta_j - \theta_k)^2\right).$$

Then for $\boldsymbol{\theta} \in \Omega_0$,

$$F_0(\boldsymbol{\theta}) \leqslant F(\boldsymbol{\theta}) \leqslant F_0^{1/8}(\boldsymbol{\theta}).$$

Now we have, by a similar reducing and lifting dimension argument, that

$$\int_{\Omega_0} F_0(\boldsymbol{\theta}) d\boldsymbol{\theta} = \left(1 + o(1)\right) \pi^{1/2} (dn)^{1/2} \int_{\Omega} \widehat{F}(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

where Ω is a region with some constants $c_2 > c_1 > 0$ such that $U_n\left(c_1 \rho_0\right) \subseteq \Omega \subseteq U_n(c_2 \rho_0)$, and in view of (4.1.7) and (4.4.8),

$$\widehat{F}(\boldsymbol{\theta}) = \exp\left(-\frac{d}{n}\left(\sum_{i \in [n]} \theta_i\right)^2 - \sum_{jk \in G} (\theta_j - \theta_k)^2\right) = \exp\left(-\boldsymbol{\theta}^{\mathrm{T}}\left(\frac{d}{n} J + L\right)\boldsymbol{\theta}\right).$$

Let $T = \left(\frac{d}{n} J + L\right)^{-1/2}$ and $\boldsymbol{x} = T\boldsymbol{y}$. Then

$$\int_{\Omega} \widehat{F}(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\Omega} \exp\left(-\boldsymbol{x}^{\mathrm{T}}\left(\frac{d}{n} J + L\right)\boldsymbol{x}\right)\, d\boldsymbol{x} = \left|\frac{d}{n} J + L\right|^{-1/2} \int_{T^{-1}(\Omega)} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\, d\boldsymbol{y}.$$

Suppose $\rho \geqslant c(\log n)^{1/2+\varepsilon}$ for some constant $c > 0$. Let $\mathbf{Y}$ be a random vector with normal density $\pi^{-n/2} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}$. By standard bounds on the tail of the normal distribution, we have $\mathbf{P}(\mathbf{Y} \notin U_n(\rho)) \leqslant n e^{-\rho^2}/(1+\rho)$.

Then we have for n sufficiently large, by Lemma 4.4.5, there are constants $c_1', c_2' > 0$ such that

$U_n(\widehat{\rho}_1') \subseteq T^{-1}\Omega \subseteq U_n(\widehat{\rho}_2')$, where $\widehat{\rho}_1' := c_1'\zeta^2 \log n$ and $\widehat{\rho}_2' := c_2'\zeta^2 \log^{3/2} n$. Then

$$\begin{aligned}\int_{T^{-1}(\Omega)} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\, d\boldsymbol{y} &= \int_{U_n(\rho_1')} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\, d\boldsymbol{y} + \int_{T^{-1}(\Omega)\setminus U_n(\rho_1)} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\, d\boldsymbol{y} \\ &= \int_{U_n(\rho_1')} e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\, d\boldsymbol{y} + C\Big(\int_{U_n(\rho_2')} - \int_{U_n(\rho_1')}\Big) e^{-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\, d\boldsymbol{y} = (1+K)\pi^{n/2},\end{aligned}$$

where we have constants $|C| \leqslant 1$ and $|K| = o(1)$.

Therefore, we have that

$$\int_{\Omega_0} F_0(\boldsymbol{\theta}) d\boldsymbol{\theta} = \big(1+o(1)\big)\, \pi^{1/2}(dn)^{1/2}\pi^{n/2} \left|\tfrac{d}{n}J + L\right|^{-1/2}, \tag{4.4.19}$$

and similarly,

$$\int_{\Omega_0} F_0^{1/8}(\boldsymbol{\theta}) d\boldsymbol{\theta} = \big(1+o(1)\big)\, \pi^{1/2}(dn)^{1/2}\pi^{n/2} \left|\tfrac{d}{n}J + \tfrac{1}{8}L\right|^{-1/2}. \tag{4.4.20}$$

For any constant c, in view of the definition of J, we have

$$|A + cL| = \left|\tfrac{d}{n}J + (\tfrac{1}{2}+c)L\right| = (c+\tfrac{1}{2})^n \left|\tfrac{d}{n(c+1/2)}J + L\right|.$$

Note that the eigenvalues of $\frac{d}{n(c+1/2)}J + L$ are $\frac{d}{c+1/2}$ together with the non-zero eigenvalues of L. Therefore $|A+cL| = \Theta((c+1/2)^n|A|)$. Note that all of the eigenvalues of A^{-1} are bounded below by $\|A\|_\infty^{-1}$ and bounded above by $\|A^{-1}\|_\infty$. Therefore we have that $|\frac{d}{n}J + L|^{-1/2} = e^{O(n\log n)}$ and $\left|\frac{d}{n}J + \frac{1}{8}L\right|^{-1/2} = e^{O(n\log n)}$. This completes the proof by sandwiching J_0 between (4.4.19) and (4.4.20). □

4.4.3 The integral outside Ω_0

It remains to show that the integral of $F(\boldsymbol{\theta})$ is negligible outside Ω_0 following the approach in [45, Section 3.3] with slight modifications and improvements. Recall the definition of J' and J_0 in (4.4.2) and (4.4.5) respectively. Define

$$\rho_{\text{small}} := \sqrt{\rho_0 d^{-1/2}\log^{-1/2} n} = \zeta d^{-1/2}\log^{1/2} n.$$

Then

$$\frac{\rho_0}{\rho_{\text{small}}} = \sqrt{\frac{\zeta d^{-1/2}\log^{3/2} n}{d^{-1/2}\log^{-1/2} n}} = \zeta \log n. \tag{4.4.21}$$

First, we bound the integral of $|F(\boldsymbol{\theta})|$ in a "scattered" region

$$\Omega_1 := \big\{\boldsymbol{\theta} \in (\mathbb{R}/\pi)^n : \text{for every } \xi \in \mathbb{R}/\pi \text{ we have } |\{j : \theta_j \in I(\xi, \rho_{\text{small}})\}| < \tfrac{4}{5}n\big\}.$$

Lemma 4.4.8. *We have*

$$\int_{\Omega_1} |F(\boldsymbol{\theta})|\, d\boldsymbol{\theta} = e^{-\omega(n\log n)} J_0.$$

We need the following lemmas to prove Lemma 4.4.8.

Lemma 4.4.9 ([45])**.** *Suppose* $0 < t < \frac{1}{3}\pi$ *and* $q \leqslant \frac{1}{5}n$. *Let* $X = \{x_1, \dots, x_n\}$ *be a multisubset of* $\mathbb{R}/\pi$ *such that no interval of length* $3t$ *contains* $n - q$ *or more elements of* X. *Then there is some interval* $I(x, \rho)$, $\rho < \frac{1}{3}\pi$, *such that both* $I(x, \rho)$ *and* $\mathbb{R}/\pi - I(x, \rho + t)$ *contain at least* q *elements of* X.

Lemma 4.4.10 ([45])**.** *Let* G *be a graph of maximum degree* d. *Assume also that* $h(G) \geqslant \gamma d$ *for some* $\gamma > 0$. *Let* U_1, U_2 *be two disjoint sets of vertices. Then, there exist at least*

$$\gamma d \frac{\min\{|U_1|,|U_2|\}}{2\ell(U_1,U_2)}$$

pairwise edge-disjoint paths in G *with one end in* U_1 *and the other end in* U_2, *and of lengths bounded above by*

$$\ell(U_1, U_2) := 2 + 2\log_{1+\gamma/2}\left(\frac{|V(G)|}{\min\{|U_1|,|U_2|\}+\gamma d/2}\right).$$

Proof of Lemma 4.4.8. If $\boldsymbol{\theta} \in \Omega_1$, the definition of Ω_1 implies that every interval of $\mathbb{R}/\pi$ of length ρ_{small} has fewer than $\frac{4}{5}n$ components of $\boldsymbol{\theta}$. Applying Lemma 4.4.9 with $t = \frac{1}{3}\rho_{\text{small}}$, $q = \frac{1}{5}n$, and $X = \boldsymbol{\theta}$ tells us that there exist $p \in \mathbb{R}/\pi$ and $s < \frac{\pi}{3}$ such that both $I(p, s)$ and $\mathbb{R}/\pi - I(p, s + t)$ contain at least $\frac{1}{5}n$ components of $\boldsymbol{\theta}$.

For such $\boldsymbol{\theta}$, let U, U' denote the indices of the elements of $\boldsymbol{\theta}$ belonging to $I(p, s)$ and $\mathbb{R}/\pi - I(p, s+t)$ respectively. Then $|\theta_j - \theta_k|_\pi \geqslant t$ whenever $j \in U, k \in U'$. Note that $t = o(1)$.

Consider any of the paths $v_0, v_1, \dots, v_\ell$ provided by Lemma 4.4.10. Note that we have $\ell = \ell(U_1, U_2) = O(1)$. By assumption, $|\theta_{v_0} - \theta_{v_\ell}|_\pi \geqslant t$. Since $|\cdot|_\pi$ is a seminorm, we have

$$\sum_{j=1}^{\ell} |\theta_{v_j} - \theta_{v_{j-1}}|_\pi^2 \geqslant \frac{1}{\ell}\left(\sum_{j=1}^{\ell} |\theta_{v_j} - \theta_{v_{j-1}}|_\pi\right)^2 \geqslant \frac{t^2}{\ell}.$$

Multiplying the bound (4.4.16) over all the edges of all the paths given by Lemma 4.4.10 gives that

$$|F(\boldsymbol{\theta})| \leqslant \exp\left(-\Omega\left(\gamma d \frac{\min\{|U_1|,|U_2|\}}{2\ell}\frac{t^2}{\ell}\right)\right) = \exp\left(-\Omega\left(\frac{t^2 dn}{\ell^2}\right)\right).$$

Recalling $t = \frac{1}{3}\rho_{\text{small}}$ and $\rho_{\text{small}} = \zeta d^{-1/2} \log^{1/2} n$, using π^n to bound the volume of Ω_1 yields

$$\int_{\Omega_1} |F(\boldsymbol{\theta})|\, d\boldsymbol{\theta} \leqslant \pi^n \exp\left(-\Omega\left(t^2 dn\right)\right) = \pi^n \exp\left(-\Omega\left(\zeta n \log n\right)\right) \leqslant \pi^n \exp\left(-\omega\left(n \log n\right)\right) = e^{-\omega(n \log n)}.$$

Then the result follows from Lemma 4.4.7. $\square$

Next, we bound the integral of $|F(\boldsymbol{\theta})|$ in a "concentrated" region

$$\Omega_2 := \left\{\boldsymbol{\theta} \in (\mathbb{R}/\pi)^n : \text{for some } x \in \mathbb{R}/\pi \text{ we have } |\{j : \theta_j \in I(x, e^{-\zeta \log n})\}| \geqslant \tfrac{4}{5}n\right\}.$$

Lemma 4.4.11. *We have*

$$\int_{\Omega_2} |F(\boldsymbol{\theta})|\, d\boldsymbol{\theta} = e^{-\omega(n \log n)}\, J_0.$$

Proof. The volume of Ω_2 is only $e^{-\omega(n \log n)}$, so the bound $|F(\boldsymbol{\theta})| \leqslant 1$ is adequate in conjunction with Lemma 4.4.7. $\square$

For disjoint $U, W \subseteq V(G)$ define by $\Omega_{U,W}$ the set of $\boldsymbol{\theta} \in (\mathbb{R}/\pi)^n$ for which there exists some $x \in \mathbb{R}/\pi$ and ρ with $\rho_{\text{small}} \leqslant \rho \leqslant \rho_0$ such that the following hold:

(i) $\theta_j \in I(x, \rho_{\text{small}})$ for at least $4n/5$ components θ_j.

(ii) $\theta_j \in I(x, \rho + \rho_{\text{small}})$ if and only if $j \notin U$.

(iii) $\theta_j \in I(x, \rho + \rho_{\text{small}}) - I(x, \rho)$ if and only if $j \in W$.

Lemma 4.4.12. *We have*

$$(\mathbb{R}/\pi)^n - \Omega_0 - \Omega_1 \subseteq \bigcup_{\substack{1 \leqslant |U| \leqslant n/5 \\ |W| \leqslant |U|/\zeta}} \Omega_{U,W}.$$

Proof. Any $\boldsymbol{\theta} \in (\mathbb{R}/\pi)^n - \Omega_1$ is such that at least $4n/5$ of its components θ_j lie in some interval $I(x, \rho_{\text{small}})$. Suppose it is not covered by any $\Omega_{U,W}$. For $1 \leqslant k \leqslant \rho_0/\rho_{\text{small}} = \zeta \log n$, take $\rho = k\rho_{\text{small}} \leqslant \rho_0$ and let U correspond to the components not in $I(x, \rho + \rho_{\text{small}})$. Since (iii) cannot hold, we get

$$\frac{|\{j : \theta_j \notin I(x, k\rho_{\text{small}})\}|}{|\{j : \theta_j \notin I(x, (k+1)\rho_{\text{small}})\}|} = 1 + \frac{|\{j : \theta_j \in I(x, \rho+\rho_{\text{small}}) - I(x,\rho)\}|}{|\{j : \theta_j \notin I(x, \rho+\rho_{\text{small}})\}|} > 1 + \zeta^{-1}.$$

Recalling that $|\{j : \theta_j \notin I(x, \rho_{\text{small}})\}| \leqslant n/5$, we can apply this ratio repeatedly starting with $k = 1$ to find that

$$|\{j : \theta_j \notin I(x, \rho_0)\}| \leqslant \tfrac{1}{5} n \Big(1 + \zeta^{-1}\Big)^{-\zeta \log n + 1} < 1.$$

This implies that $\boldsymbol{\theta} \in \Omega_0$, which completes the proof. $\square$

Lemma 4.4.13 ([45]). *For any disjoint $U, W \subset V(G)$ with $|U| \leqslant n/5$ and $|W| \leqslant \epsilon^{1/2}|U|$, we have*

$$\int_{\Omega_{U,W} - \Omega_2} |F(\boldsymbol{\theta})|\, d\boldsymbol{\theta} = e^{-\omega(|U| \log n)} J_0.$$

Proof. Let $X := V(G) - (U \cup W)$ and define the map $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n) : \Omega_{U,W} \to \Omega_0$ as follows. By the definition of $\Omega_{U,W}$, for any $\boldsymbol{\theta} \in \Omega_{U,W}$ there is some interval of length at most ρ_0 that contains $\{\theta_j\}_{j \in X}$. Let $I(z, \xi)$ be the unique shortest such interval. We can ignore parts of $\Omega_{U,W}$ that lie in Ω_2, which means that we can assume $\xi \geqslant e^{-\zeta \log n}$.

Identifying $\mathbb{R}/\pi$ with $(z - \frac{1}{2}\xi, z - \frac{1}{2}\xi + \pi]$, define

$$\phi_j = \phi_j(\boldsymbol{\theta}) := \begin{cases} z + \frac{1}{2}\xi - \frac{\xi}{\pi - \xi}\big(\theta_j - z - \frac{1}{2}\xi\big), & \text{if } j \in U \cup W; \\ \theta_j, & \text{if } j \in X. \end{cases}$$

For $j \in U \cup W$, $\theta_j \notin I(z, \xi)$ and ϕ_j maps the complementary interval $I(z + \frac{1}{2}\pi, \pi - \xi)$ linearly onto $I(z, \xi)$ (reversing and contracting with $z \pm \frac{1}{2}\xi$ fixed). For $j \in X$, $\theta_j \in I(z, \xi)$ and $\phi_j = \theta_j$.

Thus $|\phi_j - \phi_k|_\pi \leqslant |\theta_j - \theta_k|_\pi$ for all j, k. From Lemma 4.4.4, we find that

$$|\cos(\theta_j - \theta_k)| \leqslant |\cos(\phi_j - \phi_k)|.$$

Moreover, for $j \in U$ and $k \in X$, we get that $|\phi_j - \phi_k|_\pi \leqslant |\theta_j - \theta_k|_\pi - \frac{1}{2}\rho_{\text{small}}$. Observing also that $|\phi_j - \phi_k|_\pi \leqslant \xi = o(1)$ and using (4.4.17), we find that

$$\frac{|\cos(\theta_j - \theta_k)|}{|\cos(\phi_j - \phi_k)|} \leqslant e^{-\Omega(\rho_{\text{small}}^2)} = e^{-\Omega(\zeta d^{-1} \log n)}.$$

By assumption (2) of Theorem 4.1.1 and the definition of the Cheeger constant, this bound applies to at least

$$h(G)|U| - d|W| \geqslant (\gamma + o(1))d|U|$$

pairs $jk \in \partial_G U$, thus

$$|F(\boldsymbol{\theta})| = \left| \prod_{jk \in G} \cos(\theta_j - \theta_k) \right| = e^{-\Omega(\zeta^2 |U| \log n)} |F(\boldsymbol{\phi}(\boldsymbol{\theta}))|.$$

Note that the map $\boldsymbol{\phi}$ is injective, since $I(z, \xi)$ can be determined from $\{\phi_j\}_{j \in X} = \{\theta_j\}_{j \in X}$. Also, $\boldsymbol{\phi}$ is analytic except at places where the map from $\{\theta_j\}_{j \in X}$ to (z, ξ) is non-analytic, which happens only when two distinct components $\theta_j, \theta_{j'}$ for $j, j' \in X$ lie at the same endpoint of $I(z, \xi)$. Thus, the points of non-analyticity of $\boldsymbol{\phi}$ lie on a finite number of hyperplanes, which contribute nothing to the integral. To complete the calculation, we need to bound the Jacobian of the transformation $\boldsymbol{\phi}$ in the interior of a domain of analyticity.

We have

$$\frac{\partial \phi_j}{\partial \theta_k} = \begin{cases} 1, & \text{if } j = k \in X; \\ \pm \frac{\xi}{\pi - \xi}, & \text{if } j = k \notin X; \\ 0, & \text{if } j \neq k \text{ and either } j \in X \text{ or } k \notin X. \end{cases}$$

Although we have not specified all the entries of the matrix, these entries show that the matrix is triangular, and hence the determinant has the absolute value

$$\left(\frac{\xi}{\pi - \xi} \right)^{|U| + |W|} = e^{-O(\zeta |U| \log n)},$$

by noting $\xi \geqslant e^{-\zeta \log n} = o(1)$. □

Lemma 4.4.14. $J_0 = (1 + e^{-\omega(\log n)}) J'$.

Proof of Lemma 4.4.14. The integral in $\Omega_1 \cup \Omega_2$ is by Lemmas 4.4.8 and Lemma 4.4.11.

The remaining parts of J' are bounded by the sum of Lemma 4.4.13 over disjoint $U, W \subset V(G)$ with $1 \leqslant |U| \leqslant n/5$ and $|W| \leqslant |U|/\zeta$. The number of choices of W for given U is less than $2^{|U|}$, so the total contribution here is

$$\begin{aligned} \sum_{\substack{1 \leqslant |U| \leqslant n/5 \\ |W| \leqslant |U|/\zeta}} \int_{\Omega_{U,W} - \Omega_2} |F(\boldsymbol{\theta})| \, d\boldsymbol{\theta} &= \sum_{\substack{1 \leqslant |U| \leqslant n/5 \\ |W| \leqslant |U|/\zeta}} e^{-\omega(|U| \log n)} J_0 = J_0 \sum_{1 \leqslant |U| \leqslant n/5} 2^{|U|} e^{-\omega(|U| \log n)} \\ &= J_0 \sum_{t=1}^{n/5} \binom{n}{t} e^{-\omega(t \log n)} \leqslant \left(\left(1 + e^{-\omega(\log n)}\right)^n - 1 \right) J_0 = O\left(n e^{-\omega(\log n)}\right) J_0. \end{aligned}$$

This completes the proof of Lemma 4.4.14. □

4.4.4 Proof of Theorem 4.1.1 and Corollary 4.1.2

Now, we are ready to prove Theorem 4.1.1. Let A be the matrix defined in (4.4.8):

$$A = \tfrac{1}{2} L + \tfrac{d}{n} J_n.$$

The set of eigenvalues of matrix A contains element d (from the term $\frac{d}{n}J_n$), together with all the non-zero eigenvalues of $L/2$. Therefore, using the Matrix-Tree Theorem (see, for example, [76, Theorem 5.2]), we get

$$d^{1/2}n^{1/2}|A|^{-1/2} = \mathrm{ST}(G)^{-1/2}.$$

Note that A^{-1} is of the form specified in Theorem 4.3.11. Combining (4.4.2), (4.4.5), Lemma 4.4.3, Lemma 4.4.14, and Theorem 4.3.11, prove Theorem 4.1.1.

Corollary 4.1.2 follows from Theorem 4.1.1 on applying the cumulant bounds in Theorem 4.3.11 together with Lemma 4.4.2(a).

4.5 Regular tournaments, Eulerian digraphs, and Eulerian oriented graphs

Let $\mathrm{RT}(n)$ be the number of labeled regular tournaments with n vertices. It is easy to see that $\mathrm{RT}(n) = 0$ if n is even since the sum of in-degree and out-degree is odd.

An Eulerian digraph is a directed graph such that the in-degree is equal to the out-degree for each vertex. Note that an oriented graph is a directed graph having no symmetric pair of directed edges, that is, at most one of the edges (u, v) and (v, u) is permitted for any distinct vertices u and v. Let $\mathrm{ED}(n)$ be the number of labeled loop-free simple Eulerian digraphs with n vertices. Allowing simple loops would multiply $\mathrm{ED}(n)$ by exactly 2^n, since loops do not affect the Eulerian property.

Let $\mathrm{EOG}(n)$ be the number of labeled loop-free simple Eulerian oriented graphs, that is, Eulerian digraphs with no symmetric pair of directed edges.

We are interested in the asymptotic value of $\mathrm{RT}(n)$, $\mathrm{ED}(n)$, $\mathrm{EOG}(n)$, and $\mathrm{EO}(G)$.

Spencer [94] evaluates $\mathrm{RT}(n)$ to within a factor of $(1+o(1))^n$. Later, McKay [73] obtained asymptotic values of $\mathrm{RT}(n)$, $\mathrm{ED}(n)$, and $\mathrm{EOG}(n)$.

Theorem 4.5.1 ([73]). *For any $\varepsilon > 0$ and $n \to \infty$,*

$$\mathrm{ED}(n) = \left(1 + O\left(n^{-1/2+\varepsilon}\right)\right) \frac{n^{1/2}}{e^{1/4}} \left(\frac{4^n}{\pi n}\right)^{(n-1)/2}. \tag{4.5.1}$$

$$\mathrm{EOG}(n) = \left(1 + O\left(n^{-1/2+\varepsilon}\right)\right) \frac{n^{1/2}}{e^{3/8}} \left(\frac{3^{n+1}}{4\pi n}\right)^{(n-1)/2}, \tag{4.5.2}$$

and for odd $n \to \infty$,

$$\mathrm{RT}(n) = \left(1 + O\left(n^{-1/2+\varepsilon}\right)\right) \left(\frac{n}{e}\right)^{1/2} \left(\frac{2^{n+1}}{\pi n}\right)^{(n-1)/2}. \tag{4.5.3}$$

The goal of this section is to give more accurate asymptotics of $\mathrm{RT}(n)$, $\mathrm{ED}(n)$, $\mathrm{EOG}(n)$. We first introduce some definitions that are useful for the statement of results.

Define the constants $c_{\mathrm{ED},2\ell}, c_{\mathrm{EOG},2\ell}$ by the Taylor series

$$\log\left(\tfrac{1}{2}\left(1 + \cos x\right)\right) = \sum_{\ell \geqslant 1} c_{\mathrm{ED},2\ell}\, x^{2\ell} = -\frac{x^2}{4} - \frac{x^4}{96} - \frac{x^6}{1440} - \frac{17x^8}{322560} - \cdots,$$

$$\log\left(\tfrac{1}{3}\left(1 + 2\cos x\right)\right) = \sum_{\ell \geqslant 1} c_{\mathrm{EOG},2\ell}\, x^{2\ell} = -\frac{x^2}{3} - \frac{x^4}{36} - \frac{13x^6}{3240} - \frac{41x^8}{60480} - \cdots.$$

Let $\mathbf{X}_{\mathrm{RT}} = (X_1, \ldots, X_n)$ be a random vector with the Gaussian density $(2\pi)^{-n/2} n^{n/2} e^{-n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}/2}$. For some fixed integer $K \geqslant 2$, define $f_{\mathrm{RT},K}(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^n$ by

$$f_{\mathrm{RT},K}(\mathbf{x}) = \sum_{\ell=2}^{K} c_{2\ell} \sum_{1 \leqslant j < k \leqslant n} (x_j - x_k)^{2\ell}, \tag{4.5.4}$$

where $c_{2\ell}$ is defined by (4.1.8); and similarly, let $\mathbf{X}_{\mathrm{ED}}$ and $\mathbf{X}_{\mathrm{EOG}}$ be random vectors with the Gaussian density $(2\pi)^{-n/2}(n/2)^{n/2} e^{-n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}/4}$ and $(2\pi)^{-n/2}(2n/3)^{n/2} e^{-n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}/3}$ respectively, and define $f_{\mathrm{ED},M}(\mathbf{x})$ and $f_{\mathrm{EOG},M}(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^n$ by

$$f_{\mathrm{ED},K}(\mathbf{x}) = \sum_{\ell=2}^{K} c_{\mathrm{ED},2\ell} \sum_{1 \leqslant j < k \leqslant n} (x_j - x_k)^{2\ell}, \tag{4.5.5}$$

$$f_{\mathrm{EOG},K}(\mathbf{x}) = \sum_{\ell=2}^{K} c_{\mathrm{EOG},2\ell} \sum_{1 \leqslant j < k \leqslant n} (x_j - x_k)^{2\ell}. \tag{4.5.6}$$

Theorem 4.5.2. *Let $c > 0$ be a constant. There exist $M = M(c)$ and $K = K(c)$ such that as $n \to \infty$,*

$$\mathrm{ED}(n) = n^{1/2} \left(\frac{4^n}{\pi n}\right)^{(n-1)/2} \exp\left(\sum_{r=1}^{M} \frac{1}{r!} \kappa_r \left(f_{\mathrm{ED},K}(\mathbf{X}_{\mathrm{ED}})\right) + O\left(n^{-c}\right)\right), \tag{4.5.7}$$

$$\mathrm{EOG}(n) = n^{1/2} \left(\frac{3^{n+1}}{4\pi n}\right)^{(n-1)/2} \exp\left(\sum_{r=1}^{M} \frac{1}{r!} \kappa_r \left(f_{\mathrm{EOG},K}(\mathbf{X}_{\mathrm{EOG}})\right) + O\left(n^{-c}\right)\right), \tag{4.5.8}$$

and as odd $n \to \infty$,

$$\mathrm{RT}(n) = n^{1/2} \left(\frac{2^{n+1}}{\pi n}\right)^{(n-1)/2} \exp\left(\sum_{r=1}^{M} \frac{1}{r!} \kappa_r \left(f_{\mathrm{RT},K}(\mathbf{X}_{\mathrm{RT}})\right) + O\left(n^{-c}\right)\right). \tag{4.5.9}$$

Moreover, the summation of cumulants for each case is a polynomial in n^{-1} of degree $c - 1$.

Theorem 4.5.2 gives the more precise $\mathrm{RT}(n)$, $\mathrm{ED}(n)$ and $\mathrm{EOG}(n)$. Next, we compute the first few terms of the exponent for each case explicitly for illustration purposes. This extends the asymptotic formulae (4.5.1), (4.5.2), and (4.5.3).

Corollary 4.5.3. *For $n \to \infty$,*

$$\begin{aligned} \mathrm{ED}(n) = n^{1/2} \left(\frac{4^n}{\pi n}\right)^{(n-1)/2} \exp\Bigg(&-\frac{1}{4} + \frac{3}{16n} + \frac{1}{8n^2} + \frac{47}{384n^3} + \frac{371}{1920n^4} + \frac{1807}{3840n^5} + \frac{655}{448n^6} \\ &+ \frac{435581}{86016n^7} + \frac{1145941}{61440n^8} + \frac{13318871}{184320n^9} + \frac{99074137}{337920n^{10}} + \frac{1339710847}{1081344n^{11}} + O\left(n^{-12}\right)\Bigg), \end{aligned} \tag{4.5.10}$$

$$\begin{aligned} \mathrm{EOG}(n) = n^{1/2} \left(\frac{3^{n+1}}{4\pi n}\right)^{(n-1)/2} \exp\Bigg(&-\frac{3}{8} + \frac{11}{64n} + \frac{7}{64n^2} + \frac{233}{2048n^3} + \frac{497}{2560n^4} + \frac{27583}{61440n^5} + \frac{55463}{43008n^6} \\ &+ \frac{33678923}{7340032n^7} + \frac{101414573}{5242880n^8} + \frac{1882520759}{20971520n^9} + \frac{101145677531}{230686720n^{10}} + \frac{2469157786549}{1107296256n^{11}} + O\left(n^{-12}\right)\Bigg). \end{aligned} \tag{4.5.11}$$

Since every regular tournament on n vertices is an Eulerian orientation of the complete graph K_n. So $\mathrm{RT}(n) = \mathrm{EO}(K_n)$. The cases for Eulerian digraphs and Eulerian oriented graphs are also essentially identical, requiring minor changes, and therefore omitted. For completeness, we record an intermediate step.

Claim 4.1. *We have*

$$\mathrm{RT}(n) = \frac{2^{n(n-1)/2}}{(2\pi)^n} I_{\mathrm{RT}} \quad \textit{with} \quad I_{\mathrm{RT}} = \int_{U_n(\pi)} \prod_{1 \leqslant j < k \leqslant n} \cos(\theta_j - \theta_k)\, d\boldsymbol{\theta}$$

$$\mathrm{ED}(n) = \frac{2^{n(n-1)}}{(2\pi)^n} I_{\mathrm{ED}} \quad \textit{with} \quad I_{\mathrm{ED}} = \int_{U_n(\pi)} \prod_{1 \leqslant j < k \leqslant n} \left(\tfrac{1}{2} + \tfrac{1}{2} \cos(\theta_j - \theta_k) \right) d\boldsymbol{\theta}$$

$$\mathrm{EOG}(n) = \frac{3^{n(n-1)/2}}{(2\pi)^n} I_{\mathrm{EOG}} \quad \textit{with} \quad I_{\mathrm{EOG}} = \int_{U_n(\pi)} \prod_{1 \leqslant j < k \leqslant n} \left(\tfrac{1}{3} + \tfrac{2}{3} \cos(\theta_j - \theta_k) \right) d\boldsymbol{\theta}.$$

Let $c > 0$ be a constant. There exist $M = M(c)$ and $K = K(c)$ such that as $n \to \infty$,

$$I_{\mathrm{ED}} = n\sqrt{\pi} \left(\frac{4\pi}{n} \right)^{n/2} \exp\left(\sum_{r=1}^{M} \frac{1}{r!} \kappa_r \left(f_{\mathrm{ED},K}(\mathbf{X}_{\mathrm{ED}}) \right) + O\left(n^{-c}\right) \right),$$

$$I_{\mathrm{EOG}} = \frac{2}{\pi} \sqrt{\frac{1}{3n}} \left(\frac{3\pi}{n} \right)^{n/2} \exp\left(\sum_{r=1}^{M} \frac{1}{r!} \kappa_r \left(f_{\mathrm{EOG},K}(\mathbf{X}_{\mathrm{EOG}}) \right) + O\left(n^{-c}\right) \right),$$

and as odd $n \to \infty$,

$$I_{\mathrm{RT}} = 2^n n \sqrt{\frac{\pi}{2}} \left(\frac{2\pi}{n} \right)^{n/2} \exp\left(\sum_{r=1}^{M} \frac{1}{r!} \kappa_r \left(f_{\mathrm{RT},K}(\mathbf{X}_{\mathrm{RT}}) \right) + O\left(n^{-c}\right) \right).$$

What remains is to show that the cumulant terms are of the claimed form.

Lemma 4.5.4. *For any fixed integer $K \geqslant 2$ and $r \geqslant 1$, we have $\kappa_r\left(f_{\mathrm{RT},K}(\mathbf{X})\right)$ $\kappa_r\left(f_{\mathrm{ED},K}(\mathbf{X})\right)$, and $\kappa_r\left(f_{\mathrm{EOG},K}(\mathbf{X})\right)$ are polynomials in n^{-1} and are of order $O\left(n^{1-r}\right)$.*

Proof. We prove for the regular tournament only, the cases of Eulerian digraphs and Eulerian oriented graphs are essentially the same, and therefore omitted.

A simple expansion leads to

$$f_{\mathrm{RT},K}(\mathbf{x}) = \sum_{\ell=2}^{K} c_{2\ell} \sum_{1 \leqslant j < k \leqslant n} (x_j - x_k)^{2\ell} = \sum_{\ell=2}^{K} c_{2\ell} \sum_{t=0}^{2\ell} (-1)^t \binom{2\ell}{t} \sum_{1 \leqslant j < k \leqslant n} x_j^{2\ell - t} x_k^t.$$

The multi-linearity of cumulants gives

$$\begin{aligned}
&\kappa_r\left(f_{\mathrm{RT},K}(\mathbf{X})\right) \\
&= \kappa_r \left(\sum_{\ell=2}^{K} c_{2\ell} \sum_{t=0}^{2\ell} (-1)^t \binom{2\ell}{t} \sum_{1 \leqslant j < k \leqslant n} X_j^{2\ell - t} X_k^t \right) \\
&= \sum_{\boldsymbol{\ell} \in [2,K]^r} \left(\prod_{s \in [r]} c_{2\ell_s} \right) \sum_{\boldsymbol{t} \in [0,2\boldsymbol{\ell}]^r} (-1)^{\|t\|_1} \left(\prod_{s \in [r]} \binom{2\ell_s}{t_s} \right) \sum_{\boldsymbol{e} \in E(G)^r} \kappa\left(X_{j_1}^{2\ell_1 - t_1} X_{k_1}^{t_1}, \ldots, X_{j_r}^{2\ell_r - t_r} X_{k_r}^{t_r} \right),
\end{aligned} \tag{4.5.12}$$

where both vectors $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_r)$ and $\boldsymbol{t} = (t_1, \ldots, t_r)$ may contain repeated entries, and $\boldsymbol{e} = (j_1 k_1, \ldots, j_r k_r)$ denotes a vector of edges of G that may repeat.

Then it suffices to deal with $\kappa\left(X_{j_1}^{2\ell_1-t_1}X_{k_1}^{t_1},\ldots,X_{j_r}^{2\ell_r-t_r}X_{k_r}^{t_r}\right)$, and show it admits the claimed properties. To do this, we apply Theorem 4.3.8 to each cumulant in (4.5.12) by choosing, for each $i\in[r]$, the multiset $P_i := \{\{j_i,\ldots,j_i,k_i,\ldots,k_i\}\}$ containing $2\ell_i - t_i$ copies of j_i and t_i copies of k_i.

Let π be a pairing of $[2\|\boldsymbol{\ell}\|_1]$, define the graph G_π as follows: $V(G_\pi)=[r]$, and for $\ell\neq m$, $\{\ell,m\}\in E(G_\pi)$ iff π has a pair (i_1,i_2) such that $i_1\in P_\ell$ and $i_2\in P_m$. Let Π be the set of all pairings π such that G_π is connected and $j_p=j_q$ for every pair $(p,q)\in\Pi$.

Recall that $\mathbf{X}_{\mathrm{RT}}=(X_1,\ldots,X_n)$ is a random vector with the Gaussian density $(2\pi)^{-n/2}n^{n/2}e^{-n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}/2}$, and therefore the covariance matrix of $\mathbf{X}_{\mathrm{RT}}$ is $(\sigma_{jk})=(nI)^{-1}=\frac{1}{n}I$. Then

$$\begin{aligned}
\kappa\left(X_{j_1}^{2\ell_1-t_1}X_{k_1}^{t_1},\ldots,X_{j_r}^{2\ell_r-t_r}X_{k_r}^{t_r}\right) &= \sum_{\{(i_1,i_2),(i_3,i_4),\ldots,(i_{2\|\boldsymbol{\ell}\|_1-1},i_{2\|\boldsymbol{\ell}\|_1})\}\in\Pi} \sigma_{j_{i_1}j_{i_2}}\cdots\sigma_{j_{i_{2\|\boldsymbol{\ell}\|_1-1}}j_{i_{2\|\boldsymbol{\ell}\|_1}}}\\
&= \sum_{\{(i_1,i_2),(i_3,i_4),\ldots,(i_{2\|\boldsymbol{\ell}\|_1-1},i_{2\|\boldsymbol{\ell}\|_1})\}\in\Pi} n^{-\|\boldsymbol{\ell}\|_1}.
\end{aligned}$$

Therefore, in view of (4.5.12), we conclude that $\kappa_r\left(f_{\mathrm{RT},K}(\mathbf{X})\right)$ is a polynomial of n^{-1}.

Since we only consider pairings π such that G_π is connected, there are at most $r+1$ distinct X_i in (4.5.12), because $\sigma_{ij}\neq 0$ unless $X_i=X_j$. Therefore, the total number of pairings is at most $(2\|\boldsymbol{\ell}\|_1-1)!!$, and we then have

$$\begin{aligned}
&\kappa_r\left(f_{\mathrm{RT},K}(\mathbf{X})\right)\\
&= \sum_{\boldsymbol{\ell}\in[2,K]^r}\left(\prod_{s\in[r]}c_{2\ell_s}\right)\sum_{\boldsymbol{t}\in[0,2\boldsymbol{\ell}]^r}(-1)^{\|\boldsymbol{t}\|_1}\left(\prod_{s\in[r]}\binom{2\ell_s}{t_s}\right)\sum_{\boldsymbol{e}\in E(G)^r}\ \sum_{\{(i_1,i_2),(i_3,i_4),\ldots,(i_{2\|\boldsymbol{\ell}\|_1-1},i_{2\|\boldsymbol{\ell}\|_1})\}\in\Pi} n^{-\|\boldsymbol{\ell}\|_1}\\
&= \sum_{\boldsymbol{\ell}\in[2,K]^r}\left(\prod_{s\in[r]}c_{2\ell_s}\right)\sum_{\boldsymbol{t}\in[0,2\boldsymbol{\ell}]^r}(-1)^{\|\boldsymbol{t}\|_1}\left(\prod_{s\in[r]}\binom{2\ell_s}{t_s}\right)O\left((2\|\boldsymbol{\ell}\|_1-1)!!\frac{[n]_{r+1}r!}{n^{\|\boldsymbol{\ell}\|_1}}\right)\\
&\leqslant \sum_{\boldsymbol{\ell}\in[2,K]^r}\sum_{\boldsymbol{t}\in[0,2\boldsymbol{\ell}]^r}\left(\prod_{s\in[r]}\binom{2\ell_s}{t_s}\right)O\left((2\|\boldsymbol{\ell}\|_1-1)!!\frac{[n]_{r+1}r!}{n^{2r}}\right)=O\left(n^{1-r}\right).
\end{aligned}$$

This completes the proof. □

Proof of Theorem 4.5.2. Theorem 4.5.2 follows from Claim 4.1 and Lemma 4.5.4. □

4.5.1 Moments of symmetric functions of i.i.d. Gaussians

Let $\mathbf{X}=(X_1,\ldots,X_n)$ be a Gaussian random vector with independent and identically distributed components with each having mean 0 and variance n^{-1}. For $k\geqslant 0$, define the power sum symmetric functions

$$\mu_k=\mu_k(\mathbf{X})=\sum_{j=1}^{n}X_j^k.$$

Note that $\mu_0=n$. In this section, we provide some estimates of the expectation of monomials in $\{\mu_k\}$.

Lemma 4.5.5. *(a) For any integer $m\geqslant 0$,*

$$\mathbf{E}\left[X_1^m\right]=\begin{cases}0, & \textit{if } m \textit{ is odd;}\\ \frac{(m-1)!!}{n^{m/2}}, & \textit{if } m \textit{ is even,}\end{cases}$$

where $(m-1)!!$ *denotes the double factorial, that is,* $(m-1)!! = (m-1)(m-3)\cdots 1$.
(b) Suppose $m = O(1)$ *and* $j_i = O(1)$ *for all* $i \in [m]$. *Then*

$$\mathbf{E}\left[\prod_{i\in[m]} \mu_{j_i}\right] = \mathbf{E}\left[\mu_{j_1}\cdots\mu_{j_m}\right] = \begin{cases} 0, & \text{if } j_1+\cdots+j_m \text{ is odd;} \\ O\left(n^J\right), & \text{if } j_1+\cdots+j_m \text{ is even,} \end{cases}$$

where

$$J = J(j_1,\ldots,j_m) = |\{k : j_k \text{ is even}\}| + \frac{1}{2}|\{k : j_k \text{ is odd}\}| - \frac{1}{2}\sum_{i\in[m]} j_i.$$

Proof. Part (a) is the standard property of moments of Gaussian by noting $\sigma^2(X_1) = 1/n$. For (b), first note that

$$\mathbf{E}\left[\mu_{j_1}\cdots\mu_{j_m}\right] = \sum_{1\leqslant t_1,\ldots,t_m\leqslant n} \mathbf{E}\left[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m}\right],$$

in which $t_1,\ldots,t_m$ may not be distinct. Since $X_1,\ldots,X_n$ are identical and independent, then the expectation inside the sum depends only on the equalities amongst the values $t_1,\ldots,t_m$.

Any particular sequence $t_1,\ldots,t_m$ defines a partition of $q = q(t_1,\ldots,t_m)$ non-empty disjoint cells (sets), whose union is $\{1,\ldots,m\}$, that is,

$$\boldsymbol{\Pi}(t_1,\ldots,t_m) = \{\Pi_1,\ldots,\Pi_q\}$$

such that for any $1\leqslant a < b \leqslant m$, two indices t_a and t_b are equal if and only if $a\in\Pi$ and $b\in\Pi$.

Then, by the independence of the $\{X_j\}$, we have

$$\mathbf{E}\left[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m}\right] = \prod_{\Pi\in\boldsymbol{\Pi}(t_1,\ldots,t_m)} \mathbf{E}\left[X_1^{\sum_{s\in\Pi} j_s}\right], \tag{4.5.13}$$

where the expectations on the right are provided by Lemma 4.5.5 (a).

If $j_1+\cdots+j_m$ is odd, then by contradiction, there exists $\Pi\in\boldsymbol{\Pi}(t_1,\ldots,t_m)$ such that $\sum_{s\in\Pi} j_s$ is odd. Hence from (a), we have that $\mathbf{E}\left[\mu_{j_1}\cdots\mu_{j_m}\right] = 0$.

For now, we assume that $j_1+\cdots+j_m$ is even. For any $t_1,\ldots,t_m$ such that $\sum_{s\in\Pi} j_s$ is even for all $\Pi\in\boldsymbol{\Pi}(t_1,\ldots,t_m)$, we have that

$$\begin{aligned}\mathbf{E}\left[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m}\right] &= \prod_{\Pi\in\boldsymbol{\Pi}(t_1,\ldots,t_m)} \mathbf{E}\left[X_1^{\sum_{s\in\Pi} j_s}\right] = \prod_{\Pi\in\boldsymbol{\Pi}(t_1,\ldots,t_m)} \left(\sum_{s\in\Pi} j_s - 1\right)!! n^{-\frac{1}{2}\sum_{s\in\Pi} j_s} \\ &= n^{-\frac{1}{2}\sum_{i\in[m]} j_i} \prod_{\Pi\in\boldsymbol{\Pi}(t_1,\ldots,t_m)} \left(\sum_{s\in\Pi} j_s - 1\right)!! = O\left(n^{-\frac{1}{2}\sum_{i\in[m]} j_i}\right). \end{aligned} \tag{4.5.14}$$

It then suffices to count $t_1,\ldots,t_m$ such that $\sum_{s\in\Pi} j_s$ is even for all $\Pi\in\boldsymbol{\Pi}(t_1,\ldots,t_m)$. Let (M_e, M_o) be a partition of $[m]$ such that j_i is even for all $i\in M_e$ and j_i is odd for all $i\in M_o$. For each $i\in M_e$, we have n choices of t_i. For each $i\in M_o$, there must exist $j\in M_o$ such that $i\neq j$ and $t_i = t_j$, since $\sum_{s\in\Pi} j_s$ is even for all $\Pi\in\boldsymbol{\Pi}(t_1,\ldots,t_m)$. So that we have $|M_o|/2$ pairs with each having n choices. Therefore in total, there are $|M_e|+|M_o|/2$ ways of choosing with each having n options, then combining with (4.5.14) completes the proof. ☐

Equation (4.5.13), together with the fact that $\boldsymbol{\Pi}(t'_1, \ldots, t'_m) = \boldsymbol{\Pi}(t_1, \ldots, t_m)$ for exactly $(n)_q$ index sets $t'_1, \ldots, t'_m$, allows us to compute $\mathbf{E}\left[\mu_{j_1} \cdots \mu_{j_m}\right]$ by summing over all partitions of $\{1, \ldots, m\}$. However, this is very inefficient since the number of partitions (the Bell numbers) grow very quickly.

Note that many partitions give the same contribution. We define $\mathbb{N} = \{0, 1, \ldots\}$ and $\mathbb{N}_+ = \{1, 2, \ldots\}$. We will represent a multiset as a pair $(S, \nu) = (S, \nu(\cdot))$, where S is a set that contains distinct elements and $\nu = \nu(\cdot) : S \to \mathbb{N}$ is a function that specifies the number of times an element of S belongs to the multiset. We will also use the notation $\{\{z_1, \ldots, z_k\}\}$ for a multiset by listing all its elements. For example, $\{\{2, 2, 2, 3, 3, 4\}\}$ is the multiset $(\mathbb{N}_+, \nu)$ where $\nu(2) = 3$, $\nu(3) = 2$, $\nu(4) = 1$, and $\nu(k) = 0$ otherwise.

Now we introduce a definition that is useful for calculation.

Definition 4.2. A *cell type* is a multiset $(\mathbb{N}_+, \nu)$. A *partition type* (C, η) is a multiset of cell types, where C is the set of all cell types.

Then the expansion of $\mu_{j_1} \cdots \mu_{j_m}$ contains n^m terms such that each term $X_{t_1}^{j_1} \cdots X_{t_m}^{j_m}$ has a unique partition type where each cell type corresponds to the exponents of a set of equal indices. For example, in the expansion of $\mu_1\mu_1\mu_2\mu_2\mu_4$, the term $X_5^1 X_5^1 X_6^2 X_5^2 X_6^4$ has partition type $\{\{\{\{1,1,2\}\}, \{\{2,4\}\}\}\}$, since X_5^1, X_5^1, X_5^2 have the same index and X_6^2, X_6^4 have the same index.

Note that the multiset union of the cell types in the partition type equals the multiset of the indices of the monomial $\mu_1\mu_1\mu_2\mu_2\mu_4$, namely, $\{\{1,1,2,2,4\}\}$. Also note that for instance, the different term $X_3^1 X_3^1 X_3^2 X_2^2 X_2^4$ in the expansion of $\mu_1\mu_1\mu_2\mu_2\mu_4$ has the same partition type even though the positions of the equal indices changed from $\{1,2,4\}, \{3,5\}$ to $\{1,2,3\}, \{4,5\}$.

For $k \in \mathbb{N}_+$, let $\widehat{\nu}(k)$ be the number of times μ_k appears in $\mu_{j_1} \cdots \mu_{j_m}$. Clearly there is no term in the expansion of $\mu_{j_1} \cdots \mu_{j_m}$ with partition type T unless

$$\sum_{(\mathbb{N}_+, \nu) \in T} \nu(k) = \widehat{\nu}(k), \tag{4.5.15}$$

for all $k \in \mathbb{N}_+$.

Lemma 4.5.6. *We have that*

$$\mathbf{E}\left[\mu_{j_1} \cdots \mu_{j_m}\right] = \sum_{T} A_T B_T \prod_{\tau \in T} \mathbf{E}\left[X_1^{\sum_{j \in \tau} j}\right], \tag{4.5.16}$$

where the sum is over all partition types satisfying (4.5.15),

$$A_T = A_{(C,\eta)} = \left(\left(n - \sum_{\tau \in C} \eta(\tau)\right)! \prod_{\tau \in C} \eta(\tau)!\right)^{-1} n!,$$

and

$$B_T = \prod_{k \in \mathbb{N}_+} \left(\prod_{(\mathbb{N}_+, \nu) \in T} \nu(k)!\right)^{-1} \left(\sum_{(\mathbb{N}_+, \nu) \in T} \nu(k)\right)!.$$

Proof. Suppose $T = (C, \eta)$ is a partition type satisfying (4.5.15). Since some of the cells in T are the same, then the number of ways to assign distinct indices from $\{1, \ldots, n\}$ is A_T. The number of ways to assign the m positions of $\mu_{j_1} \cdots \mu_{j_m}$ to the cells of T is B_T. This then completes the proof. $\square$

Here we include an example for illustration. If $T = \{\{\{\{1,1,2\}\}, \{\{1,1,2\}\}, \{\{1,2,2\}\}\}\}$, then $\eta(\{\{1,1,2\}\}) = 2$ and $\eta(\{\{1,2,2\}\}) = 1$. So

$$A_T = \frac{n!}{(n-3)!2!1!} = \frac{1}{2}n(n-1)(n-2),$$

since we need to choose three distinct indices from $[n]$ for each cell type such that two of them are of the same type. We also have

$$B_T = \frac{5!}{2!2!}\frac{4!}{2!},$$

since we need to choose the exponents for each cell type, and that is $\binom{5}{2}\binom{3}{2}$ for one and $\binom{4}{2}\binom{2}{1}$ for two.

This is the key to our method because the contribution of a term to $\mathbf{E}\left[\mu_1\mu_1\mu_2\mu_2\mu_4\right]$ only depends on its partition type T, namely

$$\prod_{\tau \in T} \mathbf{E}\left[X_1^{\sum_{j\in\tau} j}\right].$$

Thus, to evaluate $\mathbf{E}\left[\mu_{j_1}\cdots\mu_{j_m}\right]$, we can sum over partition types, rather than over partitions. This is a very large improvement, since the number of partition types is much smaller. As an example, for $\mu_2^{10}\mu_3^2\mu_4^{10}$ the number of partitions is 4,506,715,738,447,323, but the number of partition types is only 360,847.

Proof of Corollary 4.1.3 and Corollary 4.5.3. We will only show the computation of Corollary 4.1.3 here. Similar computations are used in Corollary 4.5.3 and are therefore omitted.

We choose $M = 12$ and $K = 13$. From (4.1.6), we know that

$$\sum_{r=1}^{M} \frac{t^r}{r!}\kappa_r f_K(\mathbf{X}) = \sum_{r=1}^{K}[t^r]\log\left(\sum_{r=0}^{K}\frac{t^r}{r!}\mathbf{E}\left[f_M(\mathbf{X})^r\right]\right)t^r, \tag{4.5.17}$$

where $[t^r]$ signifies extraction of the coefficient of t^r in the Taylor expansion of what follows. We applied this using a `C` program to find $\mathbf{E}\left[\mu_{j_1}\cdots\mu_{j_m}\right]$ for all required cases (about 268,000 cases with up to 24 factors), then using (4.5.17) in Maple to find cumulants.

For the record, we give the first five cumulants to lesser precision.

$$\begin{aligned}
\kappa_1\left(f_{\mathrm{RT},K}(\mathbf{X}_{\mathrm{RT}})\right) &= -\frac{1}{2} - \frac{5}{6n} - \frac{13}{3n^2} - \frac{137}{5n^3} - \frac{9568}{45n^4} + O\left(n^{-5}\right),\\
\kappa_2\left(f_{\mathrm{RT},K}(\mathbf{X}_{\mathrm{RT}})\right) &= \frac{13}{6n} + \frac{35}{2n^2} + \frac{6871}{45n^3} + \frac{66428}{45n^4} + O\left(n^{-5}\right),\\
\kappa_3\left(f_{\mathrm{RT},K}(\mathbf{X}_{\mathrm{RT}})\right) &= -\frac{25}{n^2} - \frac{1261}{3n^3} - \frac{88042}{15n^4} + O\left(n^{-5}\right),\\
\kappa_4\left(f_{\mathrm{RT},K}(\mathbf{X}_{\mathrm{RT}})\right) &= \frac{1541}{3n^3} + \frac{42215}{3n^4} + O\left(n^{-5}\right),\\
\kappa_5\left(f_{\mathrm{RT},K}(\mathbf{X}_{\mathrm{RT}})\right) &= -\frac{15988}{n^4} + O\left(n^{-5}\right).
\end{aligned}$$

As evidence for the correctness of our calculations, all the coefficients up to the n^{-9} term were predicted in advance using the sequence acceleration method of Wynn [100] applied to the exact numbers in the Appendix. The last two coefficients also match the numerical evidence. $\square$

4.6 Tail bounds for cumulant expansions

We prove Theorem 4.3.2 in this section. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector with independent components taking values in $\mathbf{S} := S_1 \times \cdots \times S_n$. Let $\mathcal{F}_\infty(\mathbf{X})$ be the space of bounded real functions on $\mathbf{S}$, measurable with respect to $\mathbf{X}$, equipped with the infinity norm

$$\|f\|_\infty = \sup_{\boldsymbol{x} \in \mathbf{S}} |f(\boldsymbol{x})|, \qquad f \in \mathcal{F}_\infty(\mathbf{X}).$$

For a linear operator F on $\mathcal{F}_\infty(\mathbf{X})$, we consider the standard induced operator norm

$$\|F\|_\infty := \sup_{\substack{f \in \mathcal{F}_\infty(\mathbf{X}) \\ \|f\|_\infty > 0}} \frac{\|F[f]\|_\infty}{\|f\|_\infty}.$$

4.6.1 Expectation and difference operators

For any $V \subseteq [n]$, define the operator $\mathbf{E}^V$ on $\mathcal{F}_\infty(\mathbf{X})$ by

$$\mathbf{E}^V[f](\boldsymbol{x}) := \mathbf{E}\left[f(\mathbf{X}) \mid X_j = x_j \text{ for } j \notin V\right],$$

Informally, $\mathbf{E}^V$ corresponds to "averaging" with respect to all X_j with $j \in V$. Since expectation can not exceed the supremum, we get that $\|\mathbf{E}^V\|_\infty \leqslant 1$. Observe also that the operators $\mathbf{E}^V$ and $\mathbf{E}^{V'}$ commute for any $V, V' \subseteq [n]$:

$$\mathbf{E}^V \mathbf{E}^{V'} = \mathbf{E}^{V'} \mathbf{E}^V = \mathbf{E}^{V \cup V'}.$$

If $V = \{j\}$, we write $\mathbf{E}^j := \mathbf{E}^{\{j\}}$.

Recall from (4.3.3) that

$$R_{\boldsymbol{y}}^V := R_{\boldsymbol{y}}^{v_1} \cdots R_{\boldsymbol{y}}^{v_k}, \qquad \partial_{\boldsymbol{y}}^V := \partial_{\boldsymbol{y}}^{v_1} \cdots \partial_{\boldsymbol{y}}^{v_k},$$

where $\partial_{\boldsymbol{y}}^j := I - R_{\boldsymbol{y}}^j$ and I is the identity operator. This definition does not depend on the order of elements in V since the operators $R_{\boldsymbol{y}}^j$ and $R_{\boldsymbol{y}}^{j'}$ commute for any $j, j' \in [n]$. Clearly $\|R_{\boldsymbol{y}}^V\|_\infty \leqslant 1$. We also have that

$$\partial_{\boldsymbol{y}}^V R_{\boldsymbol{y}}^{V'} = R_{\boldsymbol{y}}^{V'} \partial_{\boldsymbol{y}}^V. \tag{4.6.1}$$

From (4.3.4), we have that

$$\Delta_V(f) = \sup_{\mathbf{y} \in \mathbf{S}} \|\partial_{\boldsymbol{y}}^V[f]\|_\infty$$

By definition, Δ_V satisfies the triangle inequality

$$\Delta_V(f + g) \leqslant \Delta_V(f) + \Delta_V(g). \tag{4.6.2}$$

Lemma 4.6.1. *For any $f \in \mathcal{F}_\infty(\mathbf{X})$, $V \subseteq [n]$, and $j \in [n]$, we have*

$$\Delta_V(\mathbf{E}^j[f]) \leqslant \Delta_V(f), \qquad \Delta_V(f - \mathbf{E}^j[f]) \leqslant \Delta_{V \cup \{j\}}(f).$$

Furthermore, if $j \in V$ then $\Delta_V(\mathbf{E}^j[f]) = 0$ and $\Delta_V(f - \mathbf{E}^j[f]) = \Delta_V(f)$.

Proof. We start with the second part, which is the case when $j \in V$. For any $\boldsymbol{y} \in \mathbf{S}$, we have that

$R^j_{\mathbf{y}}\mathbf{E}^j[f] = \mathbf{E}^j[f]$ because $\mathbf{E}^j[f](\boldsymbol{x})$ does not depend on jth component of $\mathbf{x}$. This implies that

$$\partial^V_{\mathbf{y}}\mathbf{E}^j[f] = 0 \quad \text{and} \quad \partial^V_{\mathbf{y}}[f - \mathbf{E}^j[f]] = \partial^V_{\mathbf{y}}[f].$$

The second part follows.

Now we assume that $j \notin V$. Let $\boldsymbol{y} \in \mathbf{S}$ and $\mathbf{Y} = (y_1, \ldots, y_{j-1}, X_j, y_{j+1}, y_n)$. For any $W \subset [n] \setminus \{j\}$, we have

$$R^W_{\boldsymbol{y}}[f] = \mathbf{E}\left[R^W_{\mathbf{Y}}[f]\right] \quad \text{and} \quad (R^W_{\boldsymbol{y}}\mathbf{E}^j)[f] = \mathbf{E}\left[R^{W\cup\{j\}}_{\mathbf{Y}}[f]\right].$$

Therefore, we find that

$$\partial^V_{\boldsymbol{y}}[\mathbf{E}^j[f]] = \sum_{W\subseteq V}(-1)^{|W|}(R^W_{\boldsymbol{y}}\mathbf{E}^j)[f] = \mathbf{E}\left[\sum_{W\subseteq V}(-1)^{|W|}R^{W\cup\{j\}}_{\mathbf{Y}}[f]\right] = \mathbf{E}\left[\partial^W_{\mathbf{Y}}R^j_{\mathbf{Y}}[f]\right]$$

and

$$\begin{aligned}\partial^V_{\boldsymbol{y}}[f - \mathbf{E}^j[f]] &= \sum_{W\subseteq V}(-1)^{|W|}R^W_{\boldsymbol{y}}[f - \mathbf{E}^j[f]] \\ &= \mathbf{E}\left[\sum_{W\subseteq V}(-1)^{|W|}\left(R^W_{\mathbf{Y}}[f] - R^{W\cup\{j\}}_{\mathbf{Y}}[f]]\right)\right] = \mathbf{E}\left[\partial^{V\cup\{j\}}_{\mathbf{Y}}[f]\right].\end{aligned}$$

Since the expectation can not exceed the supremum, we get

$$\begin{aligned}\partial^V_{\boldsymbol{y}}[\mathbf{E}^j[f]] &\leqslant \sup_{\mathbf{y}\in\mathbf{S}}\|\partial^V_{\boldsymbol{y}}[f]\|_\infty = \Delta_V(f), \\ \|\partial^V_{\boldsymbol{y}}[f - \mathbf{E}^j[f]]\| &\leqslant \sup_{\mathbf{y}\in\mathbf{S}}\|\partial^{V\cup\{j\}}_{\boldsymbol{y}}[f]\|_\infty = \Delta_{V\cup\{j\}}(f).\end{aligned}$$

Taking the supremum over $\boldsymbol{y}$ completes the proof. □

Let $D_s(V)$ denote the set of all dissections of $V \subseteq [n]$ into ordered collection of s subsets $(V_1, \ldots, V_s)$, that is, the sets V_j are disjoint (possibly empty) and $V = V_1 \cup \cdots \cup V_s$.

Lemma 4.6.2. *Let $f_1, \ldots, f_s \in \mathcal{F}_\infty(\mathbf{X})$ and $V \subseteq [n]$. Then*

$$\Delta_V(f_1 \cdots f_s) \leqslant \sum_{(V_1,\ldots,V_s)\in D_s(V)}\prod_{j=1}^{s}\Delta_{V_j}(f_j).$$

Proof. The statement is trivial for $s = 1$. We proceed to the case when $s = 2$. For any $\boldsymbol{y} \in \mathbf{S}$, $j \in [n]$, and $f, g \in \mathcal{F}_\infty(X)$, observe that

$$\partial^j_{\boldsymbol{y}}[f_1f_2] = f_1 \cdot \partial^j_{\boldsymbol{y}}[f_2] + \partial^j_{\boldsymbol{y}}[f_1] \cdot R^j_{\boldsymbol{y}}[f_2].$$

Applying this analog of the product rule of differentiation several times and using (4.6.1), we get that

$$\partial^V_{\boldsymbol{y}}[f_1f_2] = \sum_{W\subset V}\partial^W_{\boldsymbol{y}}[f_1] \cdot R^W_{\boldsymbol{y}}[\partial^{V\setminus W}_{\boldsymbol{y}}f_2].$$

Recalling that $\|R_{\boldsymbol{y}}^W\| \leqslant 1$ and using the triangle inequality and definition (4.3.4), we find that

$$\|\partial_{\boldsymbol{y}}^V[f_1 f_2]\|_\infty \leqslant \sum_{W \in V} \|\partial_{\boldsymbol{y}}^W[f_1]\|_\infty \cdot \|\partial_{\boldsymbol{y}}^{V\setminus W}[f_2]\|_\infty \leqslant \sum_{W \subset V} \Delta_W(f_1)\Delta_{V\setminus W}(f_2).$$

Taking supremum over $\boldsymbol{y}$ completes the proof for the case when $s = 2$.

The statement for $s > 2$ follows from the bound above for $s = 2$ by bounding

$$\Delta_V(f_1 \cdots f_s) \leqslant \sum_{W \in V} \Delta_W(f_1)\Delta_{V\setminus W}(f_2 \cdots f_s)$$

and using a simple inductive argument. □

4.6.2 Cumulant identities and bounds

For each $j \in [n]$ let

$$\mathbf{E}^{\geqslant j} := \mathbf{E}^{(\{j,\dots,n\})}.$$

For $f_1, \dots, f_s \in \mathcal{F}_\infty(\mathbf{X})$, define the joint cumulant

$$\kappa^{\geqslant j}[f_1, \dots, f_s] = \sum_{\tau \in P_s} (|\tau| - 1)!\,(-1)^{|\tau|-1} \prod_{B \in \tau} \mathbf{E}^{\geqslant j}\Big[\prod_{k \in B} f_k\Big], \tag{4.6.3}$$

where P_s denotes the set of unordered partitions τ of $[s]$ (with non-empty blocks) and $|\tau|$ denotes the number of blocks in the partition τ. We also set

$$\mathbf{E}^{(\geqslant n+1)}[f] = \kappa^{(\geqslant n+1)}[f] := f \quad \text{and} \quad \kappa^{(\geqslant n+1)}[f_1, \dots, f_s] := 0 \text{ for } s \geqslant 2.$$

Lemma 4.6.3. *Let $f_1, \dots, f_s \in \mathcal{F}_\infty(\mathbf{X})$. The following hold.*

(a) $\kappa^{\geqslant j}$ is a symmetric function and also a multilinear function, that is,

$$\kappa^{\geqslant j}[c_1 f_1 + c_2 f_1', f_2, \dots, f_s] = c_1 \kappa^{\geqslant j}[f_1, \dots, f_s] + c_2 \kappa^{\geqslant j}[f_1', f_2, \dots, f_s]$$

for any $c_1, c_2 \in \mathbb{R}$ and $f_1' \in \mathcal{F}_\infty(\mathbf{X})$. Furthermore, if $s \geqslant 2$ and $f_s = \mathbf{E}^{\geqslant j} g$ for some $g \in \mathcal{F}_\infty(\mathbf{X})$, then $\kappa^{\geqslant j}[f_1, \dots, f_s] \equiv 0$.

(b) Let $\log(1+t) = \sum_{k=1}^\infty \frac{(-1)^k}{k} t^k$. We have

$$\kappa^{\geqslant j}[f_1, \dots, f_s] = [t_1 \cdots t_s] \log\Bigg(1 + \sum_{k=1}^\infty \frac{\kappa^{\geqslant j}[t_1 f_1 + \cdots + t_s f_s]^k}{k!}\Bigg),$$

where $t_1, \dots, t_s$ are real indeterminants, and $[t_1 \cdots t_s]$ indicates coefficient extraction in the formal series expansion.

(c) Let $k \in \{j, \dots, n+1\}$. Then

$$\kappa^{\geqslant j}[f_1, \dots, f_s] = \sum_{p=1}^{s} \sum_{\{B_1,\dots,B_p\} \in P_s} \kappa^{\geqslant j}\left[\kappa^{\geqslant k}[f_j : j \in B_1], \dots, \kappa^{\geqslant k}[f_j : j \in B_p]\right],$$

where

$$\kappa^{\geqslant k}[f_j : j \in \{i_1, \dots, i_\ell\}] := \kappa^{\geqslant k}[f_{i_1}, \dots, f_{i_\ell}].$$

(d) *For any set* $V \subseteq [j-1]$*, we have*

$$\Delta_V\big(\kappa^{\geqslant j}[f_1, \dots, f_s]\big) \leqslant \big(\tfrac{3}{2}\big)^s (s-1)! \sum_{(V_1,\dots,V_s)\in D_s(V)} \prod_{k=1}^{s} \Delta_{V_k}(f_k).$$

Proof. The fact that K_F is symmetric and multilinear follows immediately from the definition. We proceed to the second part of (a). Consider the terms of the defining summation in (4.6.3) that correspond to the partition τ' of $[s-1]$ that results from disregarding s. For any partition $\tau' = \{B_1, \dots, B_k\}$, there are exactly $k+1$ corresponding terms. One has s by itself, $\tau_0 = \{\{s\}, B_1, \dots, B_k\}$ with coefficient $(-1)^k k!$, and k have the form $\tau_j = \{B_1, \dots, B_{j-1}, B_j \cup \{s\}, B_{j+1}, \dots, B_k\}$ with coefficient $(-1)^{k-1}(k-1)!$. Moreover for $0 \leqslant j \leqslant k$ we have $\prod_{B\in\tau_j} \mathbf{E}^{\geqslant j}\Big[\prod_{k\in B} f_k\Big] = f_s \prod_{i=1}^{k} \mathbf{E}^{\geqslant j}\Big[\prod_{k\in B_i} f_k\Big]$. Since the coefficients have a zero sum, we get $\kappa^{\geqslant j}[f_1, \dots, f_s] \equiv 0$.

Parts (b) and (c) are proved by Speed [93] for random variables when F is the expectation operator and G is a conditional expectation operator. It is easy to check that, in addition to the combinatorial properties of the partition lattice, only the linearity of expectation and the law of total expectation are used, so the same proofs work here also.

For (d), combining triangle inequality (4.6.2) and definition (4.6.3), we get

$$\Delta_V(\kappa^{\geqslant j}[f_1, \dots, f_s]) \leqslant \sum_{\tau\in P_s} (|\tau|-1)!\, \Delta_V\left(\prod_{B\in\tau} \mathbf{E}^{\geqslant j}\Big(\prod_{k\in B} f_k\Big)\right).$$

Applying Lemma 4.6.1 several times, we get that, for any $W \subseteq V$,

$$\Delta_W\left(\mathbf{E}^{\geqslant j}\left[\prod_{j\in B} f_j\right]\right) \leqslant \Delta_W(f_j).$$

Then, using Lemma 4.6.2 twice, we get that

$$\begin{aligned}\Delta_V\left(\prod_{B\in\tau} \mathbf{E}^{\geqslant j}\left[\prod_{j\in B} f_j\right]\right) &\leqslant \sum_{(U_B)_{B\in\tau}\in D_{|\tau|}(V)} \prod_{B\in\tau} \Delta_{U_B}\left(\mathbf{E}^{\geqslant j}\left[\prod_{k\in B} f_k\right]\right)\\ &\leqslant \sum_{(U_B)_{B\in\tau}\in D_{|\tau|}(V)} \prod_{B\in\tau} \Delta_{U_B}\left(\prod_{k\in B} f_k\right) \leqslant \sum_{(V_1,\dots,V_s)\in D_s(V)} \prod_{k=1}^{s} \Delta_{V_k}(f_k).\end{aligned}$$

Applying Lemma 4.3.1 then completes the proof. □

The conditional cumulant of order s is defined by

$$\kappa_s^{\geqslant j}[f] = \kappa^{(\geqslant j)}[\underbrace{f, \dots, f}_{s \text{ times}}].$$

Applying Lemma 4.6.3, we derive the following properties of $\kappa_s^{\geqslant j}$.

Lemma 4.6.4. *If* $f \in \mathcal{F}_\infty(\mathbf{X})$*,* $j \in [n]$*, and* $s \in \mathbb{N}$*, then the following hold.*

(a) For any $k \in [n+1]$, $k \geqslant j$, we have

$$\kappa_s^{\geqslant j}[f] = \sum_{p=1}^{s} \sum_{\{B_1,\ldots,B_p\} \in P_s} \kappa^{\geqslant j}\left[\kappa_{|B_1|}^{\geqslant k}[f], \ldots, \kappa_{|B_p|}^{\geqslant k}[f]\right].$$

(b) For any $V \subseteq [j-1]$ and $s \geqslant 2$, we have

$$\Delta_V\left(\kappa_s^{\geqslant j}[f]\right) \leqslant \sum_{k=j}^{n} \sum_{p=2}^{s} \sum_{\{B_1,\ldots,B_p\} \in P_s} \left(\tfrac{3}{2}\right)^p (p-1)! \sum_{(V_1,\ldots,V_p) \in D_p(V)} \prod_{r=1}^{p} \Delta_{V_r \cup \{k\}}\left(\kappa_{|B_r|}^{\geqslant k+1}[f]\right),$$

Proof. Part (a) is just a special case of Lemma 4.6.3(c). For (b), applying part (a) with $k = j+1$ and using triangle inequality (4.6.2), we get

$$\Delta_V\left(\kappa_s^{\geqslant j}[f]\right) \leqslant \Delta_V\left(\mathbf{E}^{\geqslant j} \kappa_s^{\geqslant j+1}[f]\right) + \sum_{p=2}^{s} \sum_{\{B_1,\ldots,B_p\} \in P_s} \Delta_V\left(\kappa^{\geqslant j}\left[\kappa_{|B_1|}^{\geqslant j+1}[f], \ldots, \kappa_{|B_p|}^{\geqslant j+1}[f]\right]\right).$$

From Lemma 4.6.1, we know that

$$\Delta_V\left(\mathbf{E}^{\geqslant j} \kappa_s^{\geqslant j+1}[f]\right) \leqslant \Delta_V\left(\kappa_s^{\geqslant j+1}[f]\right).$$

Using Lemma 4.6.3(a), we find that

$$\begin{aligned} \kappa^{\geqslant j}\left[\kappa_{|B_1|}^{\geqslant j+1}[f], \ldots, \kappa_{|B_p|}^{\geqslant j+1}[f]\right] &= \kappa^{\geqslant j}\left[(I - \mathbf{E}^j)\kappa_{|B_1|}^{\geqslant j+1}[f], \kappa_{|B_2|}^{\geqslant j+1}[f], \ldots, \kappa_{|B_p|}^{\geqslant j+1}[f]\right] \\ \cdots &= \kappa^{\geqslant j}\left[(I - \mathbf{E}^j)\kappa_{|B_1|}^{\geqslant j+1}[f], \ldots, (I - \mathbf{E}^j)\kappa_{|B_p|}^{\geqslant j+1}[f]\right]. \end{aligned} \tag{4.6.4}$$

Then, applying Lemma 4.6.3(d) and using Lemma 4.6.1 to estimate

$$\Delta_{V_r}\left((I - \mathbf{E}^j)\kappa_{|B_r|}^{\geqslant j+1}[f]\right) \leqslant \Delta_{V_r \cup \{j\}}\left(\kappa_{|B_r|}^{\geqslant j+1}[f]\right),$$

we obtain that

$$\begin{aligned} \Delta_V\left(\kappa_s^{\geqslant j}[f]\right) &\leqslant \Delta_V\left(\kappa_s^{\geqslant j+1}[f]\right) \\ &\quad + \sum_{p=2}^{s} \sum_{\{B_1,\ldots,B_p\} \in P_s} \left(\tfrac{3}{2}\right)^p (p-1)! \sum_{(V_1,\ldots,V_p) \in D_p(V)} \prod_{r=1}^{p} \Delta_{V_r \cup \{j\}}\left(\kappa_{|B_r|}^{\geqslant j+1}[f]\right). \end{aligned}$$

Estimating similarly $\Delta_V\left(\kappa_s^{\geqslant j}[f]\right)$ for $k = j+1, \ldots, n$ and recalling that $\kappa_s^{\geqslant n+1}[f] \equiv 0$ for $s \geqslant 2$, we prove part (b). □

4.6.3 Estimates when the sums of Δ_V are bounded

For $v \in [n]$, let

$$S_v(f) := \max_{j \in [n]} \sum_{V \in \binom{[n]}{v} : j \in V} \Delta_V(f).$$

Throughout this section, we assume that $S_v(f)$ is not very big. Namely, for $\alpha \geqslant 0$ and positive integer m, let

$$\mathcal{F}_m^{\alpha}(\mathbf{X}) := \{f \in \mathcal{F}_\infty(\mathbf{X}) : S_v(f) \leqslant \alpha \text{ for all } v \in [m]\}. \tag{4.6.5}$$

In particular, for any $f \in \mathcal{F}_m^\alpha(\mathbf{X})$, we have

$$\max_{j\in[n]} \Delta_j(f) = S_1(f) \leqslant \alpha.$$

Lemma 4.6.5. *Suppose $f \in \mathcal{F}_m^\alpha(\mathbf{X})$ for some $\alpha \geqslant 0$ and positive integer m. Then, for any $s \in [m]$ and $j \in [n]$, we have*

$$\left\|\kappa_s^{\geqslant j+1}[f] - \mathbf{E}^j \kappa_s^{\geqslant j+1}[f]\right\|_\infty \leqslant 100^{s-1}\frac{(s-1)!}{s}\alpha^s.$$

Proof. First, recalling $\kappa_1^{\geqslant j+1} = \mathbf{E}^{\geqslant j+1}$ and using Lemma 4.6.1, we get for any $V \in [j]$

$$\Delta_V\left(\kappa_1^{\geqslant j+1}[f]\right) = \Delta_V\left(\mathbf{E}^{j+1}\cdots\mathbf{E}^n[f]\right) \leqslant \Delta_V(f). \tag{4.6.6}$$

From Lemma 4.6.1, we also get that if $V \cap \{j+1\ldots n\} \neq \emptyset$ then

$$\Delta_V\left(\kappa_s^{\geqslant j+1}[f]\right) = \Delta_V\left(\mathbf{E}^{\geqslant j+1}\kappa_s^{\geqslant j+1}[f]\right) = 0. \tag{4.6.7}$$

We prove the following statement by induction on $s \in [m]$: for any $j \in [n]$ and $v \in [m-s+1]$, we have

$$S_v(\kappa_s^{\geqslant j+1}[f]) \leqslant \hbar(s,v) := 25^{s-1}\frac{(s-1)!}{s}\binom{v+2s-3}{s-1}\alpha^s. \tag{4.6.8}$$

If $s=1$ then we get from (4.6.6) and (4.6.7) that

$$S_v(\kappa_1^{\geqslant j+1}[f]) \leqslant S_v[f] \leqslant \alpha = \hbar(1,v).$$

Thus, we verified the base of induction.

For the induction step, from Lemma 4.6.4(b), for any $V \in \binom{[n]}{v}$, we find that

$$\Delta_V\left(\kappa_s^{\geqslant j+1}[f]\right) \leqslant \sum_{k=j+1}^{n}\sum_{p=2}^{s}\sum_{\{B_1,\ldots,B_p\}\in P_s}\left(\tfrac{3}{2}\right)^p(p-1)!\sum_{(V_1,\ldots,V_p)\in D_p(V)}\prod_{t=1}^{p}\Delta_{V_p\cup\{k\}}\left(\kappa_{|B_t|}^{\geqslant k+1}[f]\right).$$

Applying the induction hypothesis, we find that, for any $i \in [j]$,

$$\begin{aligned}
&\sum_{V\in\binom{[n]}{v}:i\in V}\ \sum_{\substack{(V_1,\ldots,V_p)\in D_p(V)\\ i\in V_1}}\ \sum_{k=j+1}^{n}\prod_{t=1}^{p}\Delta_{V_p\cup\{k\}}\left(\kappa_{|B_t|}^{\geqslant k+1}[f]\right)\\
&\leqslant \sum_{\substack{v_1,\ldots,v_p\in\mathbb{N}\\ v_1+\cdots+v_p=v}}\ \sum_{k=j+1}^{n}\ \sum_{\substack{V_1\in\binom{[n]}{v_1+1}\\ i,k\in V_1}}\ \sum_{\substack{V_2\in\binom{[n]}{v_2+1}\\ k\in V_2}}\cdots\sum_{\substack{V_p\in\binom{[n]}{v_p+1}\\ k\in V_p}}\prod_{t=1}^{p}\Delta_{V_p\cup\{k\}}\left(\kappa_{|B_t|}^{\geqslant k+1}[f]\right)\\
&\leqslant \sum_{\substack{v_1,\ldots,v_p\in\mathbb{N}\\ v_1+\cdots+v_p=v}}\ \sum_{k=j+1}^{n}\ \sum_{\substack{V_1\in\binom{[n]}{v_1+1}\\ i,k\in V_1}}\Delta_{V_1\cup\{k\}}\left(\kappa_{|B_1|}^{\geqslant k+1}[f]\right)\prod_{t=2}^{p}S_{v_t+1}\left(\kappa_{|B_t|}^{\geqslant k+1}[f]\right)\\
&\leqslant \sum_{\substack{v_1,\ldots,v_p\in\mathbb{N}\\ v_1+\cdots+v_p=v}}\ \sum_{k=j+1}^{n}\ \sum_{\substack{V_1\in\binom{[n]}{v_1+1}\\ i,k\in V_1}}\Delta_{V_1\cup\{k\}}\left(\kappa_{|B_1|}^{\geqslant k+1}[f]\right)\prod_{t=2}^{p}\hbar(v_t+1,|B_t|) \leqslant \sum_{\substack{v_1,\ldots,v_p\in\mathbb{N}\\ v_1+\cdots+v_p=v}}\prod_{t=1}^{p}\hbar(v_t+1,|B_t|).
\end{aligned}$$

Since $p \geqslant 2$, we have that $|B_t| \leqslant s-1$ and

$$1 \leqslant v_t + 1 \leqslant v + 1 + (s - 1 - |B_t|) \leqslant m - b_t + 1.$$

Therefore, the application of induction hypothesis above is correct. Estimating similarly the contribution of the cases when $i \in V_2, \dots, i \in V_p$, we get

$$\begin{aligned}
S_v(\kappa_s^{\geqslant j+1}[f]) &\leqslant \sum_{p=2}^{s} \sum_{\{B_1,\dots,B_p\}\in P_s} (\tfrac{3}{2})^p p! \sum_{\substack{v_1,\dots,v_p\in\mathbb{N}\\ v_1+\dots+v_p=v}} \prod_{t=1}^{p} \hbar(v_t+1, |B_t|) \\
&= \sum_{p=2}^{s} \sum_{\substack{b_1,\dots,b_p\geqslant 1\\ b_1+\dots+b_p=s}} (\tfrac{3}{2})^p \binom{s}{b_1,\dots,b_p} \sum_{\substack{v_1,\dots,v_p\in\mathbb{N}\\ v_1+\dots+v_p=v}} \prod_{t=1}^{p} \hbar(v_t+1, b_t) \\
&= \sum_{p=2}^{s} \sum_{\substack{b_1,\dots,b_p\geqslant 1\\ b_1+\dots+b_p=s}} (\tfrac{3}{2})^p s! \sum_{\substack{v_1,\dots,v_p\in\mathbb{N}\\ v_1+\dots+v_p=v}} \prod_{t=1}^{p} \frac{25^{b_t-1}}{b_t^2} \binom{v_t+2b_t-3}{b_t-1} \alpha^{b_t} \\
&= \hbar(v,s) \frac{s^2}{\binom{v+2s-3}{s-1}} \sum_{p=2}^{s} (\tfrac{3}{2})^p 25^{1-p} \sum_{\substack{b_1,\dots,b_p\geqslant 1\\ b_1+\dots+b_p=s}} \sum_{\substack{v_1,\dots,v_p\in\mathbb{N}\\ v_1+\dots+v_p=v}} \prod_{t=1}^{p} \frac{1}{b_t^2} \binom{v_t+2b_t-3}{b_t-1}.
\end{aligned}$$

We observe that

$$\sum_{\substack{v_1,\dots,v_p\in\mathbb{N}\\ v_1+\dots+v_p=v}} \prod_{t=1}^{p} \binom{v_t+2b_t-3}{b_t-1} = \binom{v+2s-3}{s-1}.$$

Indeed, $\binom{v+2s-3}{s-1}$ is the number of solutions $(x_1,\dots,x_b) \in \mathbb{N}^b$ of the system $x_1+\dots+x_s = v+s-2$. The same count is represented by the LHS if we split according to the sums within corresponding blocks of sizes b_t being equal to v_t for $t \in [p]$.

Next, by induction on $p \geqslant 2$, we show that

$$\sum_{\substack{b_1,\dots,b_p\geqslant 1\\ b_1+\dots+b_p=s}} \prod_{t=1}^{p} \frac{1}{b_t^2} \leqslant \left(\frac{2\pi^2}{3}\right)^{p-1} \frac{1}{s^2}. \tag{4.6.9}$$

Both the base of induction and the induction step for (4.6.9) rely on the following bound:

$$\sum_{t=1}^{s-1} \frac{s^2}{t^2(s-t)^2} = \sum_{t=1}^{s-1} \left(\frac{1}{t} + \frac{1}{s-t}\right)^2 \leqslant 2\sum_{t=1}^{s-1} \left(\frac{1}{t^2} + \frac{1}{(s-t)^2}\right) \leqslant \frac{2\pi^2}{3}.$$

Combining the above estimates, we conclude the

$$\begin{aligned}
S_v(\kappa_s^{\geqslant j+1}[f]) &\leqslant \hbar(v,s) \sum_{p=2}^{s} (\tfrac{3}{2})^p 25^{1-p} \left(\frac{2\pi^2}{3}\right)^{p-1} \\
&= \hbar(v,s) \cdot \left(\frac{3}{2}\right)^2 \cdot \frac{1}{25} \cdot \frac{2\pi^2}{3} \cdot \sum_{p'=0}^{s-2} \left(\frac{\pi^2}{25}\right)^{p'} \\
&\leqslant \hbar(v,s) \cdot \left(\frac{3}{2}\right)^2 \cdot \frac{1}{25} \cdot \frac{2\pi^2}{3} \left(1 - \frac{\pi^2}{25}\right)^{-1} \leqslant \hbar(v,s).
\end{aligned}$$

Thus, we established the induction step and proved (4.6.8).

Finally, using (4.6.8), Lemma 4.6.1, and bounding $\binom{2s-2}{s-1} \leqslant 2^{2s-2}$, we get

$$\|\kappa_s^{\geqslant j+1}[f] - \mathbf{E}^j \kappa_s^{\geqslant j+1}[f]\|_\infty \leqslant \Delta_j(\kappa_s^{\geqslant j+1}[f]) \leqslant \hbar(1,s) \leqslant 100^{s-1}\frac{(s-1)!}{s}\alpha^s$$

as claimed. $\square$

Lemma 4.6.6. *Suppose $f \in \mathcal{F}_m^\alpha(\mathbf{X})$ for some $\alpha \geqslant 0$ and positive integer m. Then, for any $s \in [m]$ and $j \in [n]$, we have*

$$\left\|\kappa_s^{\geqslant j+1}[f] - \kappa_s^{\geqslant j}[f]\right\|_\infty \leqslant 1.1 \cdot 100^{s-1}\frac{(s-1)!}{s}\alpha^s.$$

Proof. From Lemma 4.6.4(a), we have that

$$\kappa_s^{\geqslant j}[f] - \mathbf{E}^j \kappa_s^{\geqslant j+1}[f] = \sum_{p=2}^{s} \sum_{(B_1,\dots,B_p)\in P_s} \kappa^{\geqslant j}\left[\kappa_{|B_1|}^{\geqslant j+1}[f], \dots, \kappa_{|B_p|}^{\geqslant j+1}[f]\right].$$

Recalling from (4.6.4) that

$$\kappa^{\geqslant j}\left[\kappa_{|B_1|}^{\geqslant j+1}[f], \dots, \kappa_{|B_p|}^{\geqslant j+1}[f]\right] = \kappa^{\geqslant j}\left[\kappa_{|B_1|}^{\geqslant j+1}[f] - \mathbf{E}^j\kappa_{|B_1|}^{\geqslant j+1}[f] \dots, \kappa_{|B_p|}^{\geqslant j+1}[f] - \mathbf{E}^j\kappa_{|B_p|}^{\geqslant j+1}[f]\right],$$

applying Lemma 4.6.3(d) with $V = \emptyset$, and using (4.6.9) , we get that

$$\begin{aligned}
\|\kappa_s^{\geqslant j}[f] - \mathbf{E}^j \kappa_s^{\geqslant j+1}[f]\|_\infty &\leqslant \sum_{p=2}^{s} \sum_{\{B_1,\dots,B_p\}\in P_s} \left(\tfrac{3}{2}\right)^p (p-1)! \prod_{t=1}^{p} \|\kappa_{|B_t|}^{\geqslant j+1}[f] - \mathbf{E}^j\kappa_{|B_t|}^{\geqslant j+1}[f]\|_\infty \\
&\leqslant \sum_{p=2}^{s} \sum_{\substack{b_1,\dots,b_p\geqslant 1\\ b_1+\dots+b_p=s}} \binom{s}{b_1,\dots,b_p} \left(\tfrac{3}{2}\right)^p \frac{1}{p} \prod_{t=1}^{p} 100^{b_t-1}\frac{(b_t-1)!}{b_t}\alpha^{b_t} \\
&= 100^{s-1}\frac{(s-1)!}{s}\alpha^s \sum_{p=2}^{s} s^2 \left(\tfrac{3}{2}\right)^p 100^{1-p}\frac{1}{p} \sum_{\substack{b_1,\dots,b_p\geqslant 1\\ b_1+\dots+b_p=s}} \prod_{t=1}^{p} \frac{1}{b_t^2} \\
&\leqslant 100^{s-1}\frac{(s-1)!}{s}\alpha^s \sum_{p=2}^{s} \left(\tfrac{3}{2}\right)^p 100^{1-p}\frac{1}{2}\left(\frac{2\pi^2}{3}\right)^{p-1} \leqslant 0.1 \cdot 100^{s-1}\frac{(s-1)!}{s}\alpha^s.
\end{aligned}$$

Using the bound for $\|\kappa_s^{\geqslant j+1}[f] - \mathbf{E}^j \kappa_s^{\geqslant j+1}[f]\|_\infty$ from Lemma 4.6.5 and the triangle inequality, we complete the proof. $\square$

Lemma 4.6.7. *Suppose $f \in \mathcal{F}_m^\alpha(\mathbf{X})$ for some $\alpha \geqslant 0$ and positive integer m. Then, for any $j \in [n]$, we have*

$$\left\|\mathbf{E}^{\geqslant j} \exp\left(\sum_{s=1}^{m} \frac{\kappa_s^{\geqslant j+1}[f] - \kappa_s^{\geqslant j}[f]}{s!}\right) - 1\right\|_\infty \leqslant e^{(200\alpha)^{m+1}} - 1.$$

Proof. First, if $\alpha \geqslant 1/200$, then using Lemma 4.6.6, we get that

$$\left\|\sum_{s=1}^{m} \frac{\kappa_s^{\geqslant j+1}[f] - \kappa_s^{\geqslant j}[f]}{s!}\right\|_\infty \leqslant 1.1\sum_{s=1}^{m} 100^{s-1}\frac{1}{s^2}\alpha^s \leqslant 1.1 \cdot (\alpha + 100^{m-1}\alpha^m)\sum_{s=1}^{m}\frac{1}{s^2} \leqslant (200\alpha)^{m+1}.$$

Therefore,

$$\left\|\mathbf{E}^{\geqslant j}\exp\left(\sum_{s=1}^{m}\frac{\kappa_s^{\geqslant j+1}[f]-\kappa_s^{\geqslant j}[f]}{s!}\right)-1\right\|_\infty \leqslant \left\|\exp\left(\sum_{s=1}^{m}\frac{\kappa_s^{\geqslant j+1}[f]-\kappa_s^{\geqslant j}[f]}{s!}\right)-1\right\|_\infty$$
$$\leqslant \sum_{j=1}^{\infty}\frac{(200\alpha)^{j(m+1)}}{j!}=e^{(200\alpha)^{m+1}}-1.$$

Thus, in the following, we can assume that $\alpha \leqslant 1/200$.

Let

$$F(z) := \mathbf{E}^{\geqslant j}\exp\left(\sum_{s=1}^{m}\frac{z^s\left(\kappa_s^{\geqslant j+1}[f]-\kappa_s^{\geqslant j}[f]\right)}{k!}\right)-1.$$

and let $f_1, f_2, \ldots \in \mathcal{F}_\infty(\mathbf{X})$ denote the coefficients of its Taylor's expansion:

$$F(z)=\sum_{s=1}^{\infty}z^s f_s. \tag{4.6.10}$$

Due to Lemma 4.6.3(b,d), the series $\sum_{r=1}^{\infty}\frac{z^s\kappa_s^{\geqslant k}[f]}{r!}$ converges for any $k\in[n]$ and $z\in\mathbb{C}$ with $|z|<\frac{2}{3\|f\|_\infty}$ and

$$\exp\left(\sum_{s=1}^{\infty}\frac{z^s\kappa_s^{\geqslant k}(\pi)}{s!}\right)=\mathbf{E}^{\geqslant k}e^{zf}.$$

Taking $k=j,j+1$, we obtain that

$$\mathbf{E}^{\geqslant j}\left[\exp\left(\sum_{s=1}^{\infty}\frac{z^s\left(\kappa_s^{\geqslant j+1}[f]-\kappa_s^{\geqslant j}[f]\right)}{s!}\right)\right]=\exp\left(\sum_{s=1}^{\infty}\frac{-z^s\kappa_s^{\geqslant j}[f]}{s!}\right)\mathbf{E}^{\geqslant j}\left[\mathbf{E}^{\geqslant j+1}[e^{zf}]\right]=1.$$

It implies that first m terms in series of (4.6.10) are trivial: $f_1=\cdots=f_m\equiv 0$. Applying Cauchy's integral theorem, we get that

$$\mathbf{E}^{\geqslant j}\exp\left(\sum_{s=1}^{m}\frac{\left(\kappa_s^{\geqslant j+1}[f]-\kappa_s^{\geqslant j}[f]\right)}{k!}\right)-1=F(1)=\sum_{s=m+1}^{\infty}f_s$$
$$=\frac{1}{2\pi i}\oint\sum_{k>m+1}\frac{1}{z^k}F(z)dz=\frac{1}{2\pi i}\oint\frac{F(z)}{(z-1)z^{m+1}}dz,$$

where the integrals are over any contour encircling the origin. We take the circle $\{z\in\mathbb{C}:|z|=2\}$ as the contour which ensures that $|(z-1)z^{m+1}|\geqslant 1$. Using Lemma 4.6.6 and recalling $\alpha<1/200$, we observe that for any z with $|z|=2$

$$|F(z)|\leqslant 1.1\sum_{s=1}^{m}100^{s-1}\frac{1}{s^2}\alpha^s 2^s\leqslant 0.011\sum_{s=1}^{m}\frac{1}{s^2}<0.02.$$

The required bound follows. $\square$

4.6.4 Proof of Theorem 4.3.2

Recalling that $\kappa_1^{\geqslant n+1}[f] = f$ and $\kappa_s^{\geqslant n+1}[f] = 0$ for all $s \geqslant 2$, we have

$$f(\mathbf{X}) - \sum_{s=1}^{m} \frac{\kappa_s(f(\mathbf{X}))}{s!} = \sum_{j=1}^{n} \sum_{s=1}^{m} \frac{\kappa_s^{\geqslant j+1}[f] - \kappa_s^{\geqslant j}[f]}{s!}.$$

Applying Lemma 4.6.7, we find that

$$\begin{aligned}
&\mathbf{E}\left[\exp\left(\sum_{j=1}^{n}\sum_{s=1}^{m}\frac{\kappa_s^{\geqslant j+1}[f]-\kappa_s^{\geqslant j}[f]}{s!}\right)\right] \\
&= \mathbf{E}\left[\exp\left(\sum_{j=1}^{n-1}\sum_{s=1}^{m}\frac{\kappa_s^{\geqslant j+1}[f]-\kappa_s^{\geqslant j}[f]}{s!}\right)\mathbf{E}^{\geqslant n}\left[\sum_{s=1}^{m}\frac{\kappa_s^{\geqslant n+1}[f]-\kappa_s^{\geqslant n}[f]}{s!}\right]\right] \\
&= (1+K_1)\mathbf{E}\left[\exp\left(\sum_{j=1}^{n-1}\sum_{s=1}^{m}\frac{\kappa_s^{\geqslant j+1}[f]-\kappa_s^{\geqslant j}[f]}{s!}\right)\right] = \cdots = \prod_{j=1}^{n}(1+K_j),
\end{aligned}$$

where $|K_j| \leqslant e^{(200\alpha)^{m+1}} - 1$ for all $j \in [n]$. Since f is real-valued, we also have that $K_j > -1$. Therefore,

$$\mathbf{E}\left[\exp\left(f(\mathbf{X}) - \sum_{s=1}^{m}\frac{\kappa_s(f(\mathbf{X}))}{s!}\right)\right] = (1+K)^n,$$

where $K := \left(\prod_{j=1}^{n}(1+K_j)\right)^{1/n} - 1$. Observing that $\min_{j\in[n]} K_j \leqslant K \leqslant \max_{j\in[n]} K_j$, we establish that $|K| \leqslant e^{(200\alpha)^{m+1}} - 1$ as required.

By Lemma 4.6.6, we have that for any $s \in [m]$ and $j \in [n]$,

$$\left\|\kappa_s^{\geqslant j+1}[f_M(\mathbf{X})] - \kappa_s^{\geqslant j}[f_M(\mathbf{X})]\right\|_\infty \leqslant 1.1 \cdot 100^{s-1}\frac{(s-1)!}{s}\alpha^s.$$

Observing that

$$|\kappa_s\left(f(\mathbf{X})\right)| \leqslant \sum_{j=0}^{n-1}\left\|\kappa_r^{\geqslant j+1}[f(\mathbf{x})] - \kappa_r^{\geqslant j}[f(\mathbf{x})]\right\|_\infty,$$

which completes the proof.

4.7 Appendix

4.7.1 Exact values of $\mathrm{RT}(n)$ for $n \leqslant 37$

Exact values of $\mathrm{RT}(n)$ for $n \leqslant 21$ were found by McKay [72] using a method of summing over roots of unity. We have extended this to $n \leqslant 37$ by using a recurrence for tournaments with given degrees. See Table 4.1.

4.7.2 Alternative calculations of some terms in $\mathrm{RT}(n)$ via Feynman diagrams

In quantum field theory, a Feynman diagram represents a term in the Wick's expansion, which is essentially Isserlis' theorem for the moments of some Gaussian distribution. Here we use Theorem 4.3.8

n	$\mathrm{RT}(n)$
1	1
3	2
5	24
7	2640
9	32 30080
11	4 82515 08480
13	9 30770 06112 92160
15	240 61983 49824 94283 79648
17	85584 72055 41481 49511 79758 79680
19	4271 02683 12628 45202 01657 80015 93666 76480
21	3035 99177 67255 01434 06909 90026 40396 04333 20198 14400
23	31111 25335 58482 03432 16879 55029 99798 94772 74014 27415 01378 56000
25	46 41175 34102 33590 76153 19841 21486 62289 71154 35036 87620 35567 90979 81774 06976
27	1 01613 79494 93595 16286 17799 57707 58654 34480 25582 38882 79881 93794 83077 47797 07683 48315 64800
29	32874 42487 57440 71437 03099 54561 70730 45735 81860 90338 16899 44155 56007 47117 93383 29715 53931 99301 61725 44000
31	1 58080 31329 37879 48113 48365 61225 94846 34453 23284 20116 08717 95271 39910 78379 57102 46216 62189 55225 10623 95890 69751 74000 64000
33	113 55331 66724 13409 50706 27943 25155 74835 60333 94267 70339 22296 31722 43704 16674 54847 32916 31865 21307 04028 02521 10357 42226 17221 58828 67015 68000
35	1 22386 44546 20140 32917 57098 60021 24725 82639 58465 65811 80615 89271 23479 00218 68020 11957 16753 08650 67421 64775 78535 34744 86223 90705 83144 23519 70257 72801 31373 46560
37	19868 18615 30379 61435 68362 02070 63930 19820 07449 69481 15232 49050 85973 12501 88142 39611 14080 38454 82389 53944 91763 75862 81845 56861 43415 69406 38026 42548 47415 17715 34651 86485 95415 04000

Table 4.1: Counts of labelled regular tournaments

as an alternative way of computing some cumulant terms. This method may get complicated for some cumulants of higher orders, and we only compute up to error $O\left(n^{-3}\right)$ for illustration.

Recall that for $x \neq \pm\pi/2$, the Taylor series gives that

$$\log\cos x = -\frac{x^2}{2} - \frac{x^4}{12} - \frac{x^6}{45} - \frac{17x^8}{2520} + O\left(x^{10}\right).$$

So

$$f_{\mathrm{RT},4}(\mathbf{x}) = -\frac{1}{12}\sum_{1\leqslant j<k\leqslant n}(x_j - x_k)^4 - \frac{1}{45}\sum_{1\leqslant j<k\leqslant n}(x_j - x_k)^6 - \frac{17}{2520}\sum_{1\leqslant j<k\leqslant n}(x_j - x_k)^8.$$

Let $\mathbf{Y}$ be a Gaussian vector with density

$$\left(\frac{n}{2\pi}\right)^{n/2}\exp\left(-\frac{n}{2}\mathbf{y}'\mathbf{y}\right),$$

and covariance matrix $\frac{1}{n}I$. Define

$$X_{jk} := Y_j - Y_k.$$

Then $X_{jk} \sim \mathcal{N}(0, 2/n)$. Then we have, by the linearity of the expectation, that

$$
\begin{aligned}
\kappa_1(f_{\mathrm{RT},4}(\mathbf{Y})) &= \mathbf{E}\left[f_{\mathrm{RT},4}(\mathbf{Y})\right] \\
&= -\frac{1}{12}\sum_{1\leqslant j<k\leqslant n}\mathbf{E}\left[X_{jk}^4\right] - \frac{1}{45}\sum_{1\leqslant j<k\leqslant n}\mathbf{E}\left[X_{jk}^6\right] - \frac{17}{2520}\sum_{1\leqslant j<k\leqslant n}\mathbf{E}\left[X_{jk}^8\right] \\
&= -\binom{n}{2}\left(\frac{1}{12}3!!\left(\frac{2}{n}\right)^2 + \frac{1}{45}5!!\left(\frac{2}{n}\right)^3 + \frac{17}{2520}7!!\left(\frac{2}{n}\right)^4\right) = -\frac{1}{2} - \frac{5}{6n} - \frac{13}{3n^2} + O\left(\frac{1}{n^3}\right),
\end{aligned}
$$

where we use

$$
\mathbf{E}\left[X_{jk}^{2\ell}\right] = (2\ell-1)!!\left(\frac{2}{n}\right)^{\ell}
$$

for integer $\ell \geqslant 1$.

For cumulants of higher orders, we use Theorem 4.3.8 to compute by constructing an auxiliary graph, and considering pairings. Note that in order for the cumulants to be non-zero, variables cannot be divided into two subsets that are independent. Therefore, in our case, the indices of the variables need to have some common elements.

Claim 4.3. *We have*

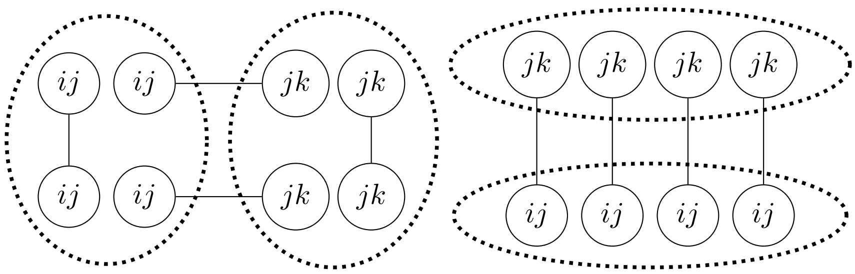
$$
\begin{aligned}
\kappa(X_{ij}^4, X_{jk}^4) &= \left(2^3\binom{4}{2}^2 + 4!\right)\frac{1}{n^4} = \frac{312}{n^4}. \\
\kappa(X_{jk}^4, X_{jk}^4) &= \left(2^4\binom{4}{2}^2 \times 2 + 4! \times 2^4\right)\frac{1}{n^4} = \frac{1536}{n^4}. \\
\kappa(X_{ij}^4, X_{jk}^6) &= \left(\binom{6}{2}\binom{4}{2}^2 2^3 + 2\binom{6}{2}4!\right)\frac{1}{n^5} = \frac{5040}{n^5}.
\end{aligned}
$$

Moreover,

$$
\kappa_2(f_{\mathrm{RT},4}(\mathbf{Y})) = \frac{13}{6n} + \frac{35}{2n^2} + O\left(\frac{1}{n^3}\right).
$$

Proof. In order to evaluate $\kappa(X_{ij}^4, X_{jk}^4)$, we choose P_1 as a multisit containing 4 copies of ij, and P_2 as a multisit containing 4 copies of jk.

Recall the definition of G_π in Theorem 4.3.8, we have its vertex set is $\{1, 2\}$, and a pairing is a matching of elements in $P_1 \cup P_2$ with 4 pairs. There are the following two types.



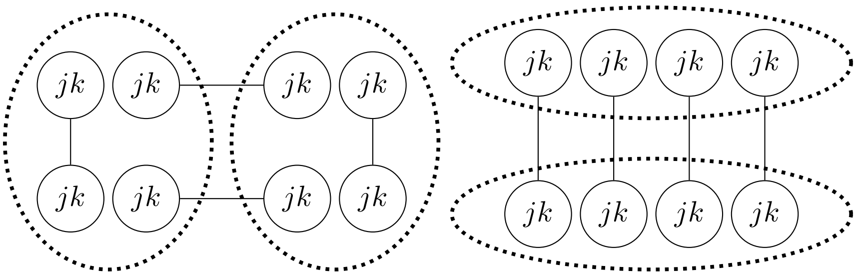
Therefore, we have

$$
\begin{aligned}
\kappa(X_{ij}^4, X_{jk}^4) &= \sum_{\{(i_1,i_2),(i_3,i_4),\ldots,(i_{k-1},i_k)\}\in\Pi} \sigma_{j_{i_1}j_{i_2}}\cdots\sigma_{j_{i_{k-1}}j_{i_k}} \\
&= 2\binom{4}{2}^2 \sigma_{ij,ij}\sigma_{jk,jk}\sigma_{ij,jk}^2 + 4!\sigma_{ij,ij}^4 = 2\binom{4}{2}^2 \frac{1}{n^2}\left(\frac{2}{n}\right)^2 + 4!\frac{1}{n^4},
\end{aligned}
$$

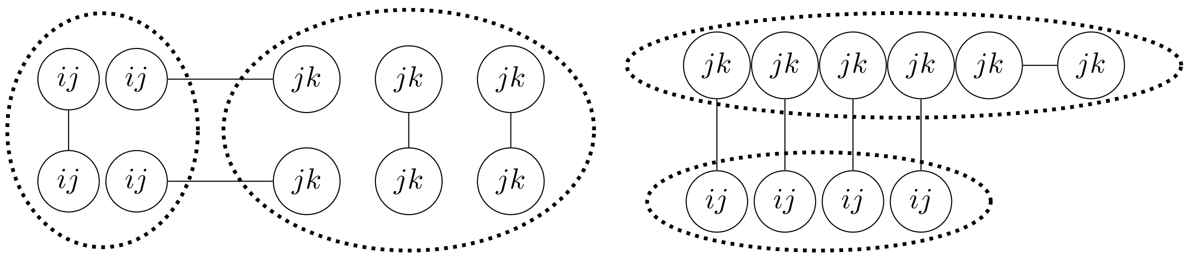
where in the first term: the factor $\binom{4}{2}$ is by choosing pairs within each "blob"; the factor 2 in the front is the way of pairing two ij's with two jk's; and $\sigma_{ij,jk}^2 = 1/n^2$; similarly in the second term we have 4!.

Note that $\kappa(X_{ij}^4, X_{jk}^4)$ and $\kappa(X_{ij}^4, X_{kj}^4)$ are with the isomorphic diagrams, and will not be treated separately in our computation, as $\sigma_{ij,jk}^2 = \sigma_{ij,kj}^2$. However, for the cumulants of higher orders, we may not be able to have this simplification, and the computation of $\kappa_2(f(\mathbf{X}))$ needs to take this into account.

For $\kappa(X_{jk}^4, X_{jk}^4)$, we have the following pairing types.



And similarly for $\kappa(X_{ij}^4, X_{jk}^6)$, we have the following types.



Therefore, we have

$$
\begin{aligned}
&\kappa_2(f(\mathbf{X})) = \kappa_2\left(\frac{1}{12}\sum_{1\leqslant j<k\leqslant n} X_{jk}^4 + \frac{1}{45}\sum_{1\leqslant p<q\leqslant n} X_{pq}^6\right) + O\left(\frac{1}{n^3}\right) \\
&= \frac{1}{12^2}\sum_{1\leqslant j<k\leqslant n}\kappa(X_{jk}^4, X_{jk}^4) + \frac{2}{12^2}\sum_{1\leqslant i<j<k\leqslant n}\kappa(X_{ij}^4, X_{jk}^4) + \frac{2}{12\times 45}\sum_{1\leqslant i<j<k\leqslant n}\kappa(X_{ij}^4, X_{jk}^6) + O\left(\frac{1}{n^3}\right) \\
&= \binom{n}{2}\frac{1}{12^2}\frac{1536}{n^4} + 2\frac{[n]_3}{2}\frac{312}{12^2}\frac{1}{n^4} + 2[n]_3\frac{1}{12\times 45}\frac{5040}{n^5} + O\left(\frac{1}{n^3}\right) = \frac{13}{6n} - \frac{13}{2n^2} + \frac{24}{n^2} + O\left(\frac{1}{n^3}\right).
\end{aligned}
$$

The reason why we ignore the contribution from X_{ij}^8 is that its contribution will be at most with some X_{jk}^4, that is $\kappa(X_{ij}^4, X_{jk}^8)$, which is of order $O\left(n^{-6}\right)$, and summing over i, j, k gives $O\left(n^{-3}\right)$. Similarly, the contribution of $\kappa(X_{ij}^6, X_{jk}^6)$, $\kappa(X_{ij}^6, X_{jk}^8)$, and $\kappa(X_{ij}^8, X_{jk}^8)$, all get absorbed in the error. $\square$

Claim 4.4. *We have*

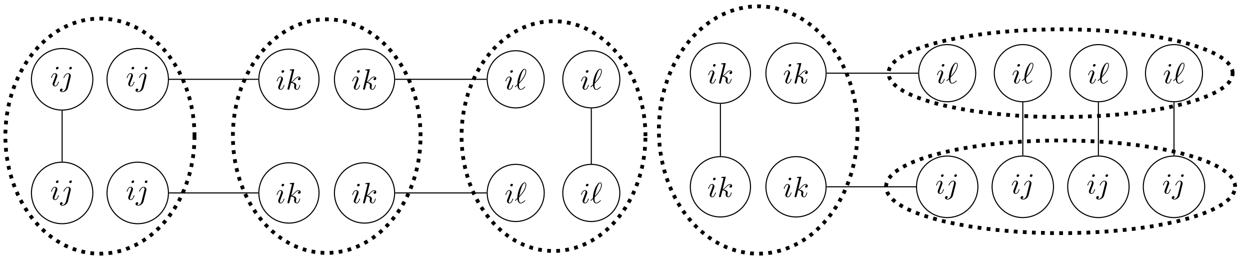
$$\kappa\left(X_{ij}^4, X_{ik}^4, X_{i\ell}^4\right) = \left(\binom{4}{2}^3 2^2 \times 2 \times 2 \times 3 + \binom{4}{2}^3 \times 2^3 \times 2^3 + 4^2 \times 3!\binom{4}{2} 3 \times 2 \times 2 + \binom{4}{2}^3 2^3\right)\frac{1}{n^6}$$

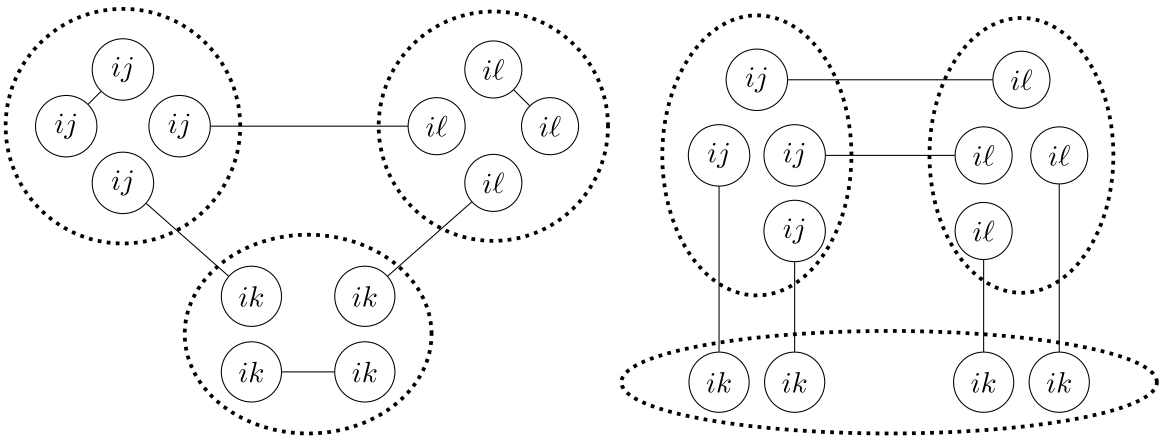
$$\kappa\left(X_{ij}^4, X_{jk}^4, X_{k\ell}^4\right) = \binom{4}{2}^3 2 \times 2 \frac{1}{n^4}\left(\frac{2}{n}\right)^2 = 6^3 \times 2^4 \frac{1}{n^6}.$$

Moreover,

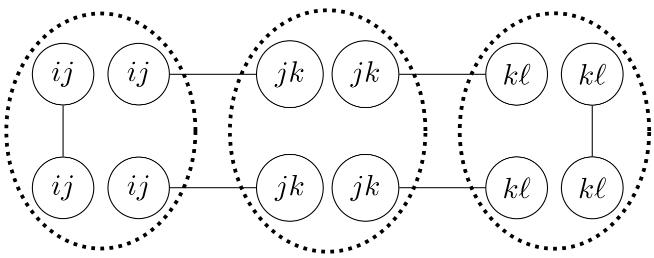
$$\kappa_3(f(\mathbf{X})) = -\frac{25}{n^2} + O\left(\frac{1}{n^3}\right).$$

Proof. For $\kappa\left(X_{ij}^4, X_{ik}^4, X_{i\ell}^4\right)$, we have the following types.





Similarly, for $\kappa\left(X_{ij}^4, X_{jk}^4, X_{k\ell}^4\right)$, we have the following types.



Therefore, we have that

$$
\begin{aligned}
-\kappa_3(f_{\mathrm{RT},4}(\mathbf{Y})) &= \kappa_3\left(\frac{1}{12}\sum_{1\leqslant j<k\leqslant n} X_{jk}^4 + \frac{1}{45}\sum_{1\leqslant p<q\leqslant n} X_{pq}^6\right) + O\left(\frac{1}{n^3}\right) \\
&= \frac{6}{12^3}\sum_{1\leqslant i<j<k<\ell\leqslant n} \kappa\left(X_{ij}^4, X_{ik}^4, X_{i\ell}^4\right) + \frac{6}{12^3}\sum_{1\leqslant i<j<k<\ell\leqslant n} \kappa\left(X_{ij}^4, X_{jk}^4, X_{k\ell}^4\right) + O\left(\frac{1}{n^3}\right) \\
&= \frac{6}{12^3}\left(\frac{[n]_4}{3!}\kappa\left(X_{ij}^4, X_{ik}^4, X_{i\ell}^4\right) + \frac{[n]_4}{2}\kappa\left(X_{ij}^4, X_{jk}^4, X_{k\ell}^4\right)\right) + O\left(\frac{1}{n^3}\right) = \frac{25}{n^2} + O\left(\frac{1}{n^3}\right).
\end{aligned}
$$

□

Combining all above, we have

$$
\begin{aligned}
&\kappa_1(f_{\mathrm{RT},4}(\mathbf{Y})) + \frac{1}{2}\kappa_2(f_{\mathrm{RT},4}(\mathbf{Y})) + \frac{1}{6}\kappa_3(f_{\mathrm{RT},4}(\mathbf{Y})) \\
&= -\frac{1}{2} - \frac{5}{6n} - \frac{13}{3n^2} + \frac{1}{2}\left(\frac{13}{6n} - \frac{13}{2n^2} + \frac{24}{n^2}\right) - \frac{1}{6}\left(\frac{25}{n^2}\right) + O\left(\frac{1}{n^3}\right) = -\frac{1}{2} + \frac{1}{4n} + \frac{1}{4n^2} + O\left(\frac{1}{n^3}\right).
\end{aligned}
$$

This matches the result in Corollary 4.1.3.

Chapter 5

The non-existence of small subhypergraphs via perturbation method

5.1 Introduction

For all integer $n \geqslant 1$, let $[n]$ denote the integer set $\{1, 2, \ldots, n\}$. Define the binomial random r-uniform hypergraph $H_r(n, p)$ to be the r-uniform hypergraph (r-graph for short) on the vertex set $[n]$ such that each r-element subset (r-set for short) is an edge independently with probability p. We use $H_r(n, m)$ to denote the random r-graphs on n vertices obtained by choosing uniformly at random from the $\binom{\binom{n}{r}}{m}$ r-graphs having m hyperedges.

We use ν_H and μ_H to denote the numbers of vertices and hyperedges of a hypergraph H respectively. A hypergraph H_0 is *strictly balanced* if all its subgraphs are strictly less dense than H_0, that is, we have

$$\frac{\mu_{H_0}}{\nu_{H_0}} > \frac{\mu_{H_1}}{\nu_{H_1}}$$

for every proper non-empty subgraph H_1 of H_0. Let $\mathcal{R}$ be some fixed finite set of strictly balanced hypergraphs with each having at least two hyperedges. Our focus is the asymptotic probability that a random hypergraph does not have a subhypergraph that is isomorphic to any hypergraph in $\mathcal{R}$, that is, the probability of $H_r(n, p)$ is $\mathcal{R}$-free. With certain restrictions on p, we show that for $H_r(n, p)$, the probability that there are no copies of any hypergraphs in $\mathcal{R}$ is the exponential of an approperate truncation of a power series in n and p, with error factor $(1 + o(1))$. For $H_r(n, m)$ the non-existence probability is given in the same way, but by a different power series in n and d, where $d = m/\binom{n}{r}$, under corresponding restrictions on d.

We adapt the perturbation method introduced by Wormald [103], and its extension [98], both of which deal with the graph case. We keep track of the distribution of a set of clusters in a random hypergraph, where clusters are edge-overlapping groups of copies of hypergraphs in $\mathcal{R}$; and obtain the ratio of the "adjacent" probabilities of having "perturbed" cluster counts, where the number of a cluster differs by 1 for one cluster type. By deriving recursions for ratios of occurrence probabilities of different types of clusters, we investigate the probability of no occurrences of hypergraphs in $\mathcal{R}$. All our basic work is in $H_r(n, p)$, the hyperedge-independent (binomial) model for random hypergraphs. By considering recursions involving both $\mathcal{R}$ and isolated hyperedges, we obtain results for $H_r(n, m)$.

We follow quite closely the method [103, 98], and extend the framework to the more general hypergraph setting. In particular, the ratios of occurrence probabilities are obtained by a more explicit iterative procedure, and simplifications are made to derive the polynomials and asymptotic formula.

Let $\chi > 0$ be defined by

$$\chi = \chi(\mathcal{R}) = \max_{H \in \mathcal{R}} \chi(H) = \max_{H \in \mathcal{R}} \max_{\substack{H' \subseteq H \\ \mu_{H'} \geqslant 1 \\ \nu_H > \nu_{H'}}} \frac{\nu_H - \nu_{H'}}{\mu_H - \mu_{H'}}. \tag{5.1.1}$$

The *extension value* of a hypergraph H, denoted by x, is defined by

$$x(H, n, p) := \max_{\substack{H' \subseteq H \\ \mu_{H'} \geqslant 1 \\ \nu_H > \nu_{H'}}} n^{\nu_H - \nu_{H'}} p^{\mu_H - \mu_{H'}} = \max_{\substack{H' \subseteq H \\ \mu_{H'} \geqslant 1 \\ \nu_H > \nu_{H'}}} \left(n^{\frac{\nu_H - \nu_{H'}}{\mu_H - \mu_{H'}}} p \right)^{\mu_H - \mu_{H'}}. \tag{5.1.2}$$

We also define $x(\mathcal{R}, n, p) = \max_{H \in \mathcal{R}} x(H, n, p)$ for a set of hypergraphs $\mathcal{R}$. We restrict our consideration, for the rest of this chapter, to $p = p(n)$ such that, for some $\varepsilon > 0$,

$$p = O\left(n^{-\chi - \varepsilon}\right). \tag{5.1.3}$$

Under this assumption, in view of (5.1.2), we have that

$$x = x(\mathcal{R}, n, p) = O\left(n^{-\varepsilon}\right). \tag{5.1.4}$$

The reason for this restriction on p is that our analysis relies on a copy of any $H \in \mathcal{R}$ being unlikely to overlap edge-wise with any other copies of $H' \in \mathcal{R}$, and therefore the set of types of clusters likely to occur and the clusters themselves are of bounded size. For a fixed hypergraph H and any of its subhypergraphs H', let $\Phi(H', H)$ be the expected *extension count* of a fixed "rooted" H' to a copy of H in $H_r(n, p)$, that is the expected number of subgraphs of $G \in H_r(n, p)$ that are isomorphic to H with $E(H') \subseteq E(G)$, conditional on $E(H') \subseteq E(G)$. Then

$$\Phi(H', H) = \Theta\left(\binom{n - \nu_{H'}}{\nu_H - \nu_{H'}} p^{\mu_H - \mu_{H'}}\right) = \Theta\left(n^{\nu_H - \nu_{H'}} p^{\mu_H - \mu_{H'}}\right). \tag{5.1.5}$$

Therefore, if p is larger than $n^{-\chi}$, there will be subhypergraphs consisting of arbitrarily large numbers of copies of H "chained" together by shared hyperedges for some $H \in \mathcal{R}$.

Theorem 5.1. *Let $\mathcal{R}$ be some fixed finite set of hypergraphs. For any $\varepsilon > 0$, if $p = O(n^{-\chi(\mathcal{R}) - \varepsilon})$, then there exist constants $i_\ell > 0$, $j_\ell > 0$ and c_ℓ (all depending only on $\mathcal{R}$), such that*

$$\mathbf{P}\left(H_r(n, p) \text{ is } \mathcal{R}\text{-free}\right) = \exp\left(\sum_{\ell=1}^{M_\varepsilon} c_\ell n^{i_\ell} p^{j_\ell} + o(1)\right), \tag{5.1.6}$$

where the bound implicit in $o(1)$ is uniform over all such p (but depends on ε), and M_ε is a constant depending only on ε and $\mathcal{R}$. A similar result holds for $H_r(n, m)$, for which we set $p = \binom{n}{r}^{-1} m$.

The series in the exponent of (5.1.6) is obtained by considering non-negligible contribution from clusters as corrections of the Poisson approximation, and the terms in the series up to any point can be explicitly computed by the description given in the proof. It is shown by Ruciński [86] that the distribution of subgraph count in random graph is asymptotically Poisson for p up to when $x = o(1)$, which also reveals partially the necessity of the constraint on p. In fact, by keeping track of the numbers of clusters, we obtain stronger results, giving approximation of the conditional probabilities of avoiding certain sets of clusters given the counts for smaller ones.

It may be possible to modify our approach for hypergraphs that are not strictly balanced. In some cases, for instance where H_0 has a unique densest subgraph, the desired result can be deduced immediately from our results. However, other cases are more delicate, with different subhypergraphs of H_0 'competing', for which simply considering the extension value (5.1.2) may not be enough, and more constraints may be needed.

We illustrate our method by obtaining the asymptotic probability of a random hypergraph being linear. Linear hypergraphs have been well studied in many contexts (sometimes under the name 'simple hypergraphs'). A hypergraph is linear if every pair of hyperedges intersects in at most one vertex. We accordingly define a set $\mathcal{H}_0$ of 'forbidden' hypergraphs containing all r-graphs having two distinct hyperedges e_1 and e_2 and vertex set $e_1 \cup e_2$, such that $2 \leqslant |e_1 \cap e_2| < r$. Then the probability that a random hypergraph is linear equals the probability of avoiding all copies of all 'forbidden' hypergraphs in $\mathcal{H}_0$.

The asymptotic probability that there are no copies of any hypergraphs in $\mathcal{R}$ for $H_r(n,p)$ was also obtained by Mousset, Noever, Panagiotou, and Samotij [77], which was utilised to give asymptotic linearity of $H_r(n,p)$ by Zhang [104]. They both need the same restriction (5.1.3) on p. However, their methods cannot give results for $H_r(n,m)$.

We extend the results by McKay and Tian for fixed r to the wider range of p given by constraint (5.1.3). In view of (5.1.3), we have the constraint $p = O\left(n^{2-r-\varepsilon}\right)$ with some $\varepsilon > 0$ for $\mathcal{H}_0$. Then the probability of a random hypergraph being linearity follows as a direct corollary. Recall that $\mathcal{L}_r(n)$ denotes the set of all linear r-uniform hypergraphs on n vertices.

Corollary 5.2. *Let $r \geqslant 3$ be fixed. For any $\varepsilon > 0$, if $p = O(n^{2-r-\varepsilon})$, then there exist constants $i_\ell > 0$, $j_\ell > 0$ and c_ℓ (all depending only on the value of r), such that*

$$\mathbf{P}\left(H_r(n,p) \in \mathcal{L}_r(n)\right) = \exp\left(\sum_{\ell=1}^{M_\varepsilon} c_\ell n^{i_\ell} p^{j_\ell} + o(1)\right), \tag{5.1.7}$$

where the bound implicit in $o(1)$ is uniform over all such p (but depends on ε), and M_ε is a constant depending only on ε and r. A similar result holds for $H_r(n,m)$, for which we set $p = \binom{n}{r}^{-1} m$.

The result by McKay and Tian has thus been extended for all p such that $p = O(n^{2-r-\varepsilon})$. The explicit formula for a particular p requires computations that are described in the proof. Next we consider a specific case, by restricting to $r = 3$, and computing only the first a few terms of the series explicitly for illustration purposes. This extends the range of p for the asymptotic linearity of $H_3(n,p)$ and $H_3(n,m)$ given by McKay and Tian.

Theorem 5.3. *If $p = o\left(n^{-7/5}\right)$, then*

$$\mathbf{P}\left(H_3(n,p) \in \mathcal{L}_3(n)\right) = \exp\left(-\tfrac{1}{4}n^4p^2 + \tfrac{2}{3}n^5p^3 - \tfrac{55}{24}n^6p^4 + \tfrac{3}{2}n^3p^2 + o\left(1\right)\right). \tag{5.1.8}$$

Theorem 5.3 matches the one obtained by Zhang in [104].

Theorem 5.4. *If $m = o\left(n^{8/5}\right)$, then*

$$\mathbf{P}\left(H_3(n,m) \in \mathcal{L}_3(n)\right) = \exp\left(-\frac{1}{4}n^4d^2 - \frac{1}{12}n^5d^3 - \frac{1}{24}n^6d^4 + \frac{3}{2}n^3d^2 + o\left(1\right)\right), \tag{5.1.9}$$

where $d = \binom{n}{3}^{-1}m$.

For $r = r(n) \geqslant 3$, McKay and Tian [74] obtained the probabilities of random hypergraphs $H_r(n,p)$ and $H_r(n,m)$ being linear for $p\binom{n}{r} = o\left(r^{-3}n^{3/2}\right)$, and $m = o(r^{-3}n^{3/2})$, respectively using switching method, whose approach is also to consider ratios of probabilities. For the case when $r = 3$, they obtain the first two terms in the exponent in (5.1.8) with error $O\left(p^{-1/2}\binom{n}{3}^{-1/2}\log^3 n + n^3p^2\right)$ for $p = o\left(n^{-3/2}\right)$; and similarly, for $m = o\left(n^{3/2}\right)$, they obtain the first two terms in the exponent in (5.1.9) with error $O\left(m^2n^{-3}\right)$.

5.2 Clusters and recursions for maximal cluster counts

Our basic setting is an extension of the graph case treated in [98] to hypergraphs. Let Ω be some finite set. A family $\mathcal{K}$ of subsets of Ω is called a *clustering* if $C_1 \in \mathcal{K}$, $C_2 \in \mathcal{K}$ and $C_1 \cap C_2 \neq \emptyset$ imply that $C_1 \cup C_2 \in \mathcal{K}$. The elements of $\mathcal{K}$ are called *clusters*. We will consider the case and assume henceforth in this chapter that $\Omega = \Omega(n)$ is the set of r-subsets of an n-set, or equivalently, Ω is the set of hyperedges of the complete r-uniform hypergraph $K_{n,r}$ on n vertices. Throughout this chapter, we take a fixed finite set of hypergraphs $\mathcal{R}$ with $|E(H)| \geqslant 2$ for all $H \in \mathcal{R}$, and investigate the distribution of the subhypergraph counts of hypergraphs in $\mathcal{R}$ in a random r-uniform hypergraph on n vertices, that is, a random subset of Ω_n.

Every edge set of subhypergraph of $K_{n,r}$ that is isomorphic to any hypergraph $H \in \mathcal{R}$ is called an *elementary* cluster. We deal with the minimal clustering that contains every elementary cluster and call this the $\mathcal{R}$-clustering of Ω. Equivalently, a set of hyperedges $J \subseteq \Omega$ is in the $\mathcal{R}$-clustering if and only if there is a sequence $J_1, \ldots, J_i$ of subsets of Ω such that each J_j is an elementary cluster, $\bigcup_{j=1}^{i} J_j = J$, and $J_k \cap \left(\bigcup_{i=1}^{k-1} J_j\right) \neq \emptyset$ for $2 \leqslant k \leqslant i$. We also consider $\mathcal{R}^*$-clustering, which consists of the clusters of the $\mathcal{R}$-clustering, together with all the 1-element subsets of Ω, recall that we assume that $|E(H)| \geqslant 2$ for all $H \in \mathcal{R}$.

For $H \subseteq \Omega$, a *cluster of H* is any cluster in clustering $\mathcal{K}$ that is contained in H. A *maximal cluster* Q of H is a cluster of H which is contained in no larger cluster of H. Equivalently, if the maximal cluster $Q \in \mathcal{K}$ is a subset of H, then for every $J \in \mathcal{K}$ with $J \subseteq H$, we have either $J \subseteq Q$ or $J \cap Q = \emptyset$.

Being a subset of Ω, a cluster induces a subhypergraph of $K_{n,r}$. The isomorphism class of the subhypergraph is called the *type* of the cluster and also of the subhypergraph. The set of types will be denoted $\mathcal{T}$, and we use τ to denote the function which maps a cluster or the corresponding graph to its type. Given $t \in \mathcal{T}$, we use the notation $\mathfrak{C}_t := \{S \subseteq \Omega : \tau(S) = t\}$.

We will specify a non-empty finite set $\mathbf{S}$ of types of clusters which is closed under taking subsets, that is, if $S, S' \in \mathcal{K}$, $\tau(S) \in \mathbf{S}$ and $S' \subseteq \mathbf{S}$ then $\tau(S) \in \mathbf{S}$. Let $s = |\mathbf{S}|$ be the number of types in $\mathbf{S}$. We will use $t^\star$ to denote the type of the single edge cluster, which appears in the $\mathcal{R}^*$-clustering. The types in $\mathbf{S}$ and clusters of these types are called *small*, and any type or cluster that is not small is called *large*. Let $\mathcal{L}$ be the set of large types. An *unavoidable* cluster is any large cluster which is the union of a small cluster Q and a set of small clusters all pairwise disjoint with each having non-empty intersection with Q. The set of types of unavoidable clusters is denoted by $\mathcal{U}$. Note $\mathcal{U} \subseteq \mathcal{L}$.

To record how many subhypergraphs of every small type are present in a given hypergraph, we consider the set $\mathcal{F}$ of *cluster counts*, by which we mean all non-negative integer functions defined on $\mathbf{S}$. For any $H \subseteq \Omega$, define $s_H(t)$ to be the number of maximal clusters of H of type t for each $t \in \mathbf{S}$. Note that $s_H \in \mathcal{F}$. The function $\delta_t \in \mathcal{F}$ has value 1 at type t, and 0 elsewhere.

For each $f \in \mathcal{F}$, define the set $\mathcal{C}_f$ to consist of every r-graph H on n vertices with no large clusters for which $s_H = f$, that is, with exactly $f(t)$ maximal clusters of type t for each $t \in \mathbf{S}$. We write $\mathbf{P}(f)$ for $\mathbf{P}(H_r(n,p) \in \mathcal{C}_f)$ for simplicity.

For types $u, t \in \mathbf{S}$, a fixed J of type u, and $h \in \mathcal{F}$, we define

$$c(u,t,h) = \sum_{\substack{Q \subseteq J,\ \tau(Q)=t \\ J\setminus Q \subseteq H \subseteq J,\ s_H = h}} p^{|Q\cap H|}(1-p)^{|Q\setminus H|}, \tag{5.2.1}$$

which simplifies, when $u = t$, to

$$c(t,t,h) = \sum_{H \subseteq J,\ s_H = h} p^{|H|}(1-p)^{|J\setminus H|}. \tag{5.2.2}$$

Note that $c(u,t,h) = O(1)$ always; $c(t,t,\mathbf{0}) = 1 + O(p)$; and $c(t,t,h) = O(p)$ for $h \neq \mathbf{0}$.

Recall that ν_G and μ_G denote the numbers of vertices and hyperedges of a hypergraph G. We extend the notation to arbitrary subsets H of Ω, such that ν_H is the number of vertices of the hypergraph induced by H and μ_H is the number of edges. For a cluster H of type t, we also use ν_t and μ_t for the numbers of vertices and hyperedges respectively.

For any $t \in \mathcal{T}$, let Q be a cluster of type t and $|\mathrm{aut}(Q)|$ be the number of automorphisms of the hypergraph induced by Q. Then the expected number of copies of the subhypergraph induced by Q in $H_r(n,p)$ is

$$\lambda_t := |\mathfrak{C}_t| p^{\mu_Q} = \frac{[n]_{\nu_Q}}{|\mathrm{aut}(Q)|} p^{\mu_Q} = \Theta(n^{\nu_Q} p^{\mu_Q}), \tag{5.2.3}$$

for every positive integer $i \leqslant n$, the i-th falling factorial is denoted by $[n]_i := n(n-1)\cdots(n-i+1)$.

For all $f \in \mathcal{F}$ and $t \in \mathbf{S}$, we compare the distribution of cluster counts with the corresponding Poisson distribution to obtain *correction factors* defined by

$$\gamma(f,t) = \frac{f(t)+1}{\lambda_t} \cdot \frac{\mathbf{P}(f+\delta_t)}{\mathbf{P}(f)}. \tag{5.2.4}$$

If the numbers of copies of each small type were independent Poisson variables, then all the γs would be exactly 1. Showing γs are all close to one indicates that the cluster counts are approximately Poisson.

We first introduce a lemma giving approximate recursions for the correction factors. This is an extension to the hypergraph case, and slight reformulation of Proposition 2.1 in [98].

Lemma 5.5. *For all $f \in \mathcal{F}$ and $t \in \mathbf{S}$,*

$$\gamma(f,t) = \frac{1}{c(t,t,\mathbf{0})}\left(1 - \Sigma(f,t) - \frac{\theta(f,\delta_t)}{|\mathfrak{C}_t|\mathbf{P}(f)}\right), \tag{5.2.5}$$

where

(a1) $\Sigma(f,t)$ is defined by

$$\Sigma(f,t) = \sum_{\substack{u\in\mathbf{S}\\ h,f-h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} c(u,t,h)\frac{\lambda_u}{\lambda_t}\gamma(f-h,u)\prod_{i=1}^{k}\frac{f_i(t_i)+1}{\lambda_{t_i}\gamma(f_i,t_i)}, \tag{5.2.6}$$

where t_i and f_i depend on h, and satisfy (a2) and (a3);

(a2) for each h, we have that $t_1,\dots,t_k$ is a sequence in $\mathbf{S}$ such that $h = \sum_{i=1}^{k}\delta_{t_i}$;

(a3) for each $i \in [k]$, $f_i = f - \sum_{j=1}^{i}\delta_{t_j}$;

(a4) we have

$$0 \leqslant \theta(f,\delta_t) \leqslant \sum_{L:\tau(L)\in\mathcal{U}}\ \sum_{\substack{Q\subseteq L,\ \tau(Q)=t\\ L\setminus Q\subseteq H\subseteq L}} \mathbf{P}\left(f-s_H\right)\left(\frac{p}{1-p}\right)^{|H|}. \tag{5.2.7}$$

Proof. For any $f\in\mathcal{F}$ and any cluster $Q\in\mathcal{K}$ of type t, we consider all the possible hypergraphs resulting from adding Q to hypergraphs in $\mathcal{C}_f$. Given a pair (E,Q), where $E=E(G)$ is the edge set of $G\in\mathcal{C}_f$, let J be the maximal cluster of $E\cup Q$ containing Q.

Recall that $\mathfrak{C}_t$ denotes the set of subsets of Ω which can form a cluster of type t. Then classifying $E\cup Q$ according to the type of J, and also according to $h=s_{E\cap J}$, gives that

$$|\mathfrak{C}_t|\mathbf{P}\left(f\right) = \sum_{\substack{u\in\mathbf{S}\\ h\in\mathcal{F}}}(f(u)-h(u)+1)\sum_{Q\subseteq J:\tau(Q)=t}\ \sum_{\substack{J\setminus Q\subseteq H\subseteq J\\ s_H=h}}\left(\frac{1-p}{p}\right)^{|Q\setminus H|}\mathbf{P}\left(f-h+\delta_u\right)+\theta(f,\delta_t), \tag{5.2.8}$$

where the exponent of the factor $(1-p)/p$ arises because we need to remove $|Q\setminus H|$ edges in J to reconstruct the graph $G\in\mathcal{C}_f$ from graph $E\cup Q$. The θ term and the bound (5.2.7) arise from observing that if adding Q to the hypergraph G such that J is a large cluster L, then it must be unavoidable since $G\in\mathcal{C}_f$ has no large clusters, then the resulting graph will be in $f-s_H$, and therefore $|H|$ edges need to be added for reconstruction of G.

Multiplying (5.2.8) by p^{μ_t} gives

$$\lambda_t\mathbf{P}\left(f\right) = \sum_{\substack{u\in\mathbf{S}\\ h\in\mathcal{F}}}(f(u)-h(u)+1)\sum_{\substack{Q\subseteq J:\tau(Q)=t\\ J\setminus Q\subseteq H\subseteq J,\ s_H=h}} p^{|Q\cap H|}(1-p)^{|Q\setminus H|}\mathbf{P}\left(f-h+\delta_u\right)+p^{\mu_t}\theta(f,\delta_t). \tag{5.2.9}$$

Isolating the term with $(u,h)=(t,\mathbf{0})$ in the summation and plugging in $c(u,t,h)$ defined by (5.2.1) yield

$$c(t,t,\mathbf{0})\frac{(f(t)+1)\mathbf{P}(f+\delta_t)}{\lambda_t\mathbf{P}(f)} = 1-\sum_{\substack{u\in\mathbf{S}\\ h,f-h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} c(u,t,h)\frac{\lambda_u}{\lambda_t}\frac{f(u)-h(u)+1}{\lambda_u}\frac{\mathbf{P}(f-h+\delta_u)}{\mathbf{P}(f-h)}\frac{\mathbf{P}(f-h)}{\mathbf{P}(f)}-\frac{\theta(f,\delta_t)}{|\mathfrak{C}_t|\mathbf{P}(f)}.$$

Rewriting using the correction factors defined by (5.2.4) gives

$$c(t,t,\mathbf{0})\gamma(f,t) = 1 - \sum_{\substack{u\in\mathbf{S}\\ h,f-h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} c(u,t,h)\frac{\lambda_u}{\lambda_t}\gamma(f-h,u)\frac{\mathbf{P}(f-h)}{\mathbf{P}(f)} - \frac{\theta(f,\delta_t)}{|\mathfrak{C}_t|\mathbf{P}(f)}. \tag{5.2.10}$$

Note that for each h, there exists H with $s_H = h$ containing pairwise edge-disjoint clusters $J_1,\dots,J_k$ of types $t_1,\dots,t_k$, which may not necessarily be distinct, such that $h=\sum_{i=1}^k \delta_{t_i}$. Here and henceforth, we choose such a sequence $t_1,\dots,t_k$ canonically for each h such that $c(u,t,h)\neq 0$ for $u,t\in\mathbf{S}$. Then the recursive formula (5.2.5) follows by rewriting the ratio as a telescoping product:

$$\frac{\mathbf{P}(f-h)}{\mathbf{P}(f)} = \prod_{i=1}^{k}\frac{\mathbf{P}\big(f-\sum_{j=1}^{i}\delta_{t_j}\big)}{\mathbf{P}\big(f-\sum_{j=1}^{i-1}\delta_{t_j}\big)} = \prod_{i=1}^{k}\frac{(f-\sum_{j=1}^{i}\delta_{t_j})(t_i)+1}{\lambda_{t_i}\gamma(f-\sum_{j=1}^{i}\delta_{t_j},t_i)} = \prod_{i=1}^{k}\frac{f_i(t_i)+1}{\lambda_{t_i}\gamma(f_i,t_i)}, \tag{5.2.11}$$

and therefore completes the proof. □

The assumption on $p=p(n)$ we will make for now is, for some fixed $\varepsilon>0$, that

$$p = n^{-\chi-\varepsilon+o(1)}. \tag{5.2.12}$$

This assumption will be relaxed to obtain asymptotic results that hold uniformly over more general $p=p(n)=O(n^{-\chi-\varepsilon})$ at the end of Section 5.4.

The set of types admits a partial ordering, and it is by defining t to be strictly less than u in the poset $\mathbf{S}$, denoted by $t\prec u$, if and only if any cluster of type u properly contains a cluster of type t. Note that if $t\prec u$, then a cluster of type u can be obtained from a cluster Q of type t by a sequence of non-disjoint unions of elementary clusters $Q_0,\dots,Q_k$, where Q_i is the edge set of a hypergraph that is isomorphic to some $G_i\in\mathcal{R}$, and $Q_i\not\subseteq Q\cup\Big(\bigcup_{j=0}^{i-1}Q_j\Big)$. Then recalling the definition of Φ and bound (5.1.5), for $G\in H_r(n,p)$ and $t\prec u$, by extending from type t, we have the expected number of clusters of type u in $E(G)$ can be bounded above by a finite sum whose terms are all of the form $\lambda_t\prod_{i=0}^k\Phi(H_i,G_i)$ where H_i is a hypergraph with edge set $Q_i\cap\Big(Q\cup\Big(\bigcup_{j=0}^{i-1}Q_j\Big)\Big)$. In view of (5.1.2), $\Phi(H_i,G_i)=O(x)$ for all i, since H_i is a non-empty proper subset of the edge set of an elementary cluster and there is a bounded number of ways to distinguish one of the subgraphs of G_i isomorphic to H_i. Recalling x satisfies (5.1.4), therefore, we have that

$$\text{if } t\prec u,\text{ then } \frac{\lambda_u}{\lambda_t}=O(x). \tag{5.2.13}$$

For a given $\mathcal{R}$ and $\varepsilon>0$ such that p satisfies (5.1.3), we define the set $\mathbf{S}$ of small cluster types to be

$$\mathbf{S}=\{t:\nu_t/\mu_t\geqslant\chi+\varepsilon\}. \tag{5.2.14}$$

Then in view of the assumption on p in (5.2.12), for each $t\in\mathbf{S}$, we have

$$\lambda_t=\Theta\left(n^{\nu_t}p^{\mu_t}\right)=n^{\nu_t-\chi\mu_t-\varepsilon\mu_t+o(1)}, \tag{5.2.15}$$

and therefore $\lambda_t\geqslant n^{o(1)}$ for all $t\in\mathbf{S}$. Note that the set $\mathbf{S}$ depends on ε and $\mathcal{R}$, and its size is finite by (5.1.4) and (5.2.13). Moreover, its size is bounded for ε bounded away from 0.

In view of (5.2.14) and (5.2.15), we have

$$\lambda_{\mathcal{L}} := \sup_{t \notin \mathbf{S}} \lambda_t = O\left(n^{-\varepsilon_{\mathcal{L}}}\right) \tag{5.2.16}$$

for some constant $\varepsilon_{\mathcal{L}} > 0$.

We further define two subtypes of clusters:

(t1) $\mathbf{S}_2 := \{t \in \mathbf{S} : \nu_t/\mu_t = \chi + \varepsilon\}$,

(t2) $\mathbf{S}_1 := \mathbf{S} \setminus \mathbf{S}_2$.

Then we have for all $t \in \mathbf{S}$, that $\lambda_t = \Theta(n^{c+o(1)})$ for some $c = c(t) \geqslant 0$; in particular, $c = 0$ if $t \in \mathbf{S}_2$. For technical reasons, we assume additionally that p satisfies

$$n^r p > n^{\varepsilon''} \tag{5.2.17}$$

for some $\varepsilon'' > 0$. This implies that the expected number of edges in the random hypergraph goes to infinity at a reasonable rate. Without this assumption, each term in the power series goes to zero, since the expected number of copies of any connected hypergraph with at least two hyperedges goes to zero, and therefore the probability is asymptotic to 1. Moreover, this assumption implies that in the case of the $\mathcal{R}^*$-clustering, the single edge cluster is in $\mathbf{S}$.

It would be sufficient to obtain estimates of correction factors for restricted cluster counts in our method. Let $\mathcal{F}_{\mathbf{S}} \subset \mathcal{F}$ be the set containing all $f \in \mathcal{F}$ such that for all $t \in \mathbf{S}$,

$$f(t) \leqslant m_t, \tag{5.2.18}$$

where

$$m_t = \begin{cases} 3\lambda_t & \text{if } t \in \mathbf{S}_1, \\ \lambda_t \log n & \text{if } t \in \mathbf{S}_2. \end{cases} \tag{5.2.19}$$

5.3 Iterative approximations of correction factors

The iterative approximation scheme we use here is essentially the same as that in [98], which is in a less general context. The similar analysis here is in a viewpoint of contractive mappings that is not emphasized there. To recursively calculate the correction factor $\gamma(f, t)$ using its definition, we need to keep track of $\gamma(f, t)$ for each $f \in \mathcal{F}$ and $t \in \mathbf{S}$. Instead we consider a simpler approximation in a compact form, not depending on f explicitly. Let $\mathbf{g} = (g_1, \ldots, g_s)$ denote a vector of variables. (Later g_i is substituted by *scaled cluster count* $f(i)/\lambda_i$.)

Recall that for each h, we have that $t_1, \ldots, t_k$ is a sequence in $\mathbf{S}$ such that $h = \sum_{i=1}^{k} \delta_{t_i}$. In view of the definition of $\gamma(f, t)$ in (5.2.5), we define formal power series $\Gamma_t\,(n, p, \mathbf{g})$ in $n, p, g_1, \ldots, g_s$, for all $t \in \mathbf{S}$, recursively by

$$\Gamma_t = \frac{1}{c(t,t,\mathbf{0})} \left(1 - \sum_{\substack{u \in \mathbf{S} \\ h \in \mathcal{F} \\ (u,h) \neq (t,\mathbf{0})}} c(u,t,h) \frac{\lambda_u}{\lambda_t} \Gamma_u \prod_{i=1}^{k} \frac{g_{t_i}}{\Gamma_{t_i}} \right), \quad \Gamma_t(0,0,\mathbf{0}) = 1. \tag{5.3.1}$$

Here we treat n and p as independent formal indeterminates. To see this properly defines formal power series, recall from the definition that

(1) $c(u,t,h)$ is a polynomial in p;

(2) for all $t \in \mathbf{S}$, λ_t is a polynomial in n and p;

(3) $c(t,t,\mathbf{0}) = 1 + O(p)$;

(4) for $h \neq \mathbf{0}$, $c(t,t,h) = O(p)$;

(5) $c(u,t,h)\lambda_u/\lambda_t$ has no constant term for $u \neq t$. (For $t \prec u$, λ_u/λ_t is a polynomial in n and p with terms of the form $p^{\mu_u - \mu_t} n^i$ since $\mu_u > \mu_t$.)

It follows that there exists a unique formal power series $\Gamma_t(n,p,\mathbf{g})$ for all $t \in \mathbf{S}$ defined by (5.3.1), and it has constant term 1 for each $t \in \mathbf{S}$.

In view of (5.3.1), we accordingly introduce an iterative procedure which is essentially described in [98]. Successive approximations to $\{\Gamma_t\}_{t\in\mathbf{S}}$, in the form of power series in n, p, $\mathbf{g}$, are defined as follows, for each $t \in \mathbf{S}$:

(r1) $\Gamma_t^{(0)} = 1$;

(r2) for every $r \geqslant 0$, given series $\{\Gamma_u^{(r)}\}_{u\in\mathbf{S}}$,

$$\Gamma_t^{(r+1)} = \frac{1}{c(t,t,\mathbf{0})}\left(1 - \overline{\Sigma}_t^{(r+1)}\right), \tag{5.3.2}$$

where

$$\overline{\Sigma}_t^{(r+1)} = \sum_{\substack{u\in\mathbf{S}\\ h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} c(u,t,h)\frac{\lambda_u}{\lambda_t}\Gamma_u^{(r)}\prod_{i=1}^{k}\frac{g_{t_i}}{\Gamma_{t_i}^{(r)}}. \tag{5.3.3}$$

Then the current series on the right side give rise to updated series on the left side.

Given $f \in \mathcal{F}$, for each $t \in \mathbf{S}$, we approximate $\gamma(f,t)$ by

$$\overline{\gamma}_t(f) := \Gamma_t\left(n,p,\widetilde{\mathbf{g}}\right), \text{ with } \widetilde{\mathbf{g}} = (\widetilde{g}_1,\ldots,\widetilde{g}_s) = \left(\frac{f(1)}{\lambda_1},\ldots,\frac{f(s)}{\lambda_s}\right), \tag{5.3.4}$$

which is in turn approximated, for some suitable r, by

$$\overline{\gamma}_t^{(r)}(f) := \Gamma_t^{(r)}\left(n,p,\widetilde{\mathbf{g}}\right). \tag{5.3.5}$$

Then, given numerical values of n and p, $\overline{\gamma}_t(\cdot)$ maps $f \in \mathcal{F}$ to numbers, whereas Γ_t is a power series in $\widetilde{\mathbf{g}}$.

Since $f \in \mathcal{F}$ and we take p to be a function of n such that $x = x(n,p) = O(n^{-\varepsilon})$ by (5.1.4), we have that for given n, p and f satisfying these constraints, there is a unique value of $\overline{\gamma}_t(f)$ determined from the series Γ_t and (5.3.4), as long as n is large enough. One way to see this is to consider an initial approximation $\overline{\gamma}_t^{(0)}(f) = 1$ for each $\overline{\gamma}_t(f)$, and then, iterating the approximations using (5.3.2), with g_t set to $f(t)/\lambda_t$ in (5.3.3), the current values of $\overline{\gamma}_t^{(r)}(f)$ on the right side giving rise to updated values on the left side. We will show that this in a sense determines a contractive mapping which has a fixed point near the initial approximate solution.

In order to do this, we start with some necessary results, following [98] closely. First rewrite (5.3.2) as

$$\Gamma_t = 1 + w_0(t) - \sum_{\substack{u \in \mathbf{S} \\ h \in \mathcal{F} \\ (u,h) \neq (t,\mathbf{0})}} w(u,t,h)\Gamma_u \prod_{i=1}^{k} \frac{1}{\Gamma_{t_i}}, \tag{5.3.6}$$

with

$$w_0(t) = \frac{1}{c(t,t,\mathbf{0})} - 1, \quad w(u,t,h) = \frac{\lambda_u c(u,t,h)}{\lambda_t c(t,t,\mathbf{0})} \prod_{i=1}^{k} g_{t_i}. \tag{5.3.7}$$

In view of (5.3.4), here (5.3.6) defines $\overline{\gamma}_t$ as a power series in the ws, which, if substituted appr-operately as power series in n, p and $\mathbf{g}$ using (5.3.7), results in the same series as defined in (5.3.1). Let $\widetilde{w}(u,t,h)$ denote the value of $w(u,t,h)$ obtained if we replace g_{t_i} by $\widetilde{g}_{t_i}$ in (5.3.7), and similarly, set $\widetilde{w}_0(t) = w_0(t)$.

Next is a reformulation of Lemma 2.4 in [98] for the present setting.

Lemma 5.6. *Suppose that* $0 \leqslant \widetilde{g}_t = \widetilde{g}_t(n) = O\left(n^{o(1)}\right)$*, with* $\widetilde{g}_t = O(1)$ *for all* $t \in \mathbf{S}_1$*. Then* $\widetilde{w}_0(t) = O(p)$ *and* $\widetilde{w}(u,t,h) = O(x)$ *for each term in* (5.3.6) *uniformly.*

Proof. Since k is bounded in (5.3.7) and $c(t,t,\mathbf{0}) = 1 + O(p) = 1 + o(1)$, then $\widetilde{w}_0(t) = O(p)$, and

$$\widetilde{w}(u,t,h) = O\left(\frac{\lambda_u c(u,t,h)}{\lambda_t}\left(\max_i \widetilde{g}_{t_i}\right)^k\right). \tag{5.3.8}$$

- Firstly, if $h = \mathbf{0}$, then $k = 0$, and $u \succ t$ by the condition in the summation. So $\widetilde{w}(u,t,h) = O(x)$.
- Secondly, suppose that $h \neq \mathbf{0}$ and $u = t$.
  - If $h = \delta_{t^*}$ (recall that t^* is the type of the single-edge cluster), then $c(u,t,h) = O(p)$ and $t^* \in \mathbf{S}_1$. So, using the hypothesis of this lemma, the maximum in (5.3.8) is $O(1)$, and thus $\widetilde{w}(u,t,h) = O(p) = O(x)$.
  - In all other cases, if $c(u,t,h) \neq 0$, then $c(t,t,h) = O(p^2)$ since $s_H = h$ implies $|H| \geqslant 2$. By (5.3.8), again $\widetilde{w}(u,t,h) = O(x)$.
- Lastly, suppose that $h \neq \mathbf{0}$ and $u \succ t$.
  - If $t_i \in \mathbf{S}_1$ for all $i \in [k]$, then the maximum in (5.3.8) is $O(1)$, Since $\lambda_u/\lambda_t = O(x)$, then $\widetilde{w}(u,t,h) = O(x)$.
  - If there is $t_i \in \mathbf{S}_2$ for some $i \in [k]$. Since $J \setminus Q \subseteq H \subseteq J$, so H contains only sub-clusters of cluster J of type $u \in \mathbf{S}$, then we must have $t_i = u$, $h = \delta_u$, and hence $Q \subseteq J$ and $|Q \cap H| \geqslant 1$, and so $c(u,t,h) = O(p) = O(x)$. Since the maximum in (5.3.8) is $O(n^{o(1)})$, the bound obtained is $O(x^2 n^{o(1)})$, and the result follows in this case.

This completes the proof. □

Recall that $\overline{\gamma}_t(f)$ defined by (5.3.4) is a function of n, p and f. We estimate it by a reformulation of Lemma 2.5 in [98] with the proof essentially the same.

Lemma 5.7. *For $f \in \mathcal{F}_{\mathbf{S}}$ and p satisfying (5.1.3), the series definition of $\overline{\gamma}_t(f)$ in (5.3.4) converges absolutely for n sufficiently large, and*

$$\overline{\gamma}_t(f) = 1 + O(x). \tag{5.3.9}$$

Proof. For any $t \in \mathbf{S}_2$, it follows from the definition of $\widetilde{g}_t$, the upper bounds (5.2.19) and (5.2.18) on $f(t)$, that $\widetilde{g}_t = O(n^{o(1)})$. On the other hand, if $t \in \mathbf{S}_1$, then $0 \leqslant \widetilde{g}_t \leqslant 3$. Thus the conditions of Lemma 5.6 are satisfied.

For polynomials or formal power series P and $\widehat{P}$, denote by P^+ the formal power series obtained by replacing all coefficients of P by their absolute values, and write $P \leqslant \widehat{P}$ if the coefficient of any monomial in P is no greater than the corresponding coefficient in $\widehat{P}$. We will use the obvious fact that if P^+ is absolutely convergent (for a particular assignment of the indeterminates) then so is P.

With (5.3.6) in mind, define the power series Γ^* for each $t \in \mathbf{S}$ by

$$\Gamma^* = 1 + w_0^+ + \sum_{\substack{u \in \mathbf{S} \\ h \in \mathcal{F} \\ (u,h) \neq (t,\mathbf{0})}} w(u,t,h)^+ \Gamma_u^* \prod_{i=1}^{k} \frac{1}{2-\Gamma_{t_i}^*}, \tag{5.3.10}$$

which by induction has a unique solution in formal power series with constant terms all 1. Then

$$\frac{1}{2-\Gamma_{t_i}^*} = \sum_{j \geqslant 0} (\Gamma_{t_i}^* - 1)^j$$

and so by induction, all coefficients of Γ_t^* are nonnegative for each $t \in \mathbf{S}$. Thus

$$\frac{1}{2-\Gamma_{t_i}^*} \geqslant \sum_{j \geqslant 0} (1 - \Gamma_{t_i}^*)^j = \frac{1}{\Gamma_{t_i}^*}$$

and, again by induction, comparing (5.3.6) with (5.3.10) gives

$$\Gamma_t^+ \leqslant \Gamma_t^* \tag{5.3.11}$$

for each $t \in \mathbf{S}$.

Now consider summing the terms of $\Gamma_t^*(n, p, \widetilde{\mathbf{g}})$ for p and f as in the lemma, when n is sufficiently large. It is immediate from the proof of Lemma 5.6 that $w(u,t,h)^+ = O(x)$ and $w_0^+ = O(p) = O(x)$. It is now straightforward to verify from (5.3.10), by a sequence of successive approximations beginning with $\Gamma_t^* \approx 1$ for all t, that

$$\Gamma_t^*(n, p, \widetilde{\mathbf{g}}) = 1 + O(x). \tag{5.3.12}$$

The lemma now follows since from (5.3.11), and the fact that the constant terms in all Γs and Γ^*s are all 1, $(\Gamma_t - 1)^+ \leqslant \Gamma_t^* - 1$. $\square$

Next is a reformulation of Proposition 2.2 in [98].

Claim 5.8. *Uniformly for all $t \in \mathbf{S}$ and $f \in \mathcal{F}_{\mathbf{S}}$, we have*

$$\gamma(f,t) = 1 + O(\phi_t x) \tag{5.3.13}$$

and

$$\frac{\theta(f,\delta_t)}{|t|\mathbf{P}(f)} = O\left(\frac{\phi_t \lambda_{\mathcal{L}}}{\lambda_t}\right), \tag{5.3.14}$$

where $\phi_t = n^{o(1)}$ for $t = t^*$ and $\phi_t = 1$ otherwise.

Proof. We will prove that for some sufficiently large constants C and C' (not depending on f, t, n, or p), and some function $1 \leqslant \phi^* = \phi^*(n) = n^{o(1)}$ such that, for n large enough and all relevant f and t,

$$\frac{\theta(f,\delta_t)}{|t|\mathbf{P}(f)} \leqslant C\phi_t \frac{\lambda_{\mathcal{L}}}{\lambda_t}, \tag{5.3.15}$$

and

$$|\gamma(f,t) - 1| \leqslant C'\phi_t x \leqslant 1/2, \tag{5.3.16}$$

with $\phi_t = \phi^*$ for $t = t^*$ and $\phi_t = 1$ otherwise. We use an induction on $f \in \mathcal{F}$ in the lexicographic order, that is, $g < f$ if and only if $g \neq f$ and g has a smaller value than f in the first entry at which they differ.

(i) The bound (5.3.15) *involving* θ:

First recall the bound on $\theta(f,\delta_t)$ in (5.2.7). Since the number of clusters of the complete r-graph $K_{n,r}$ that are isomorphic to a given cluster L is $O(n^{\nu_L})$, and the number of types of unavoidable clusters is by definition bounded for p satisfying (5.1.3), we have

$$\begin{aligned}\frac{\theta(f,\delta_t)}{|t|\mathbf{P}(f)} &= O(1) \max_{\substack{\tau(L)\in\mathcal{U} \\ \tau(Q)=t,\ Q\subseteq L \\ L\setminus Q\subseteq H\subseteq L}} n^{\nu_L-\nu_Q}\left(\frac{p}{1-p}\right)^{|H|}\frac{\mathbf{P}(f-s_H)}{\mathbf{P}(f)} \\ &= O(1) \max_{\substack{\tau(L)\in\mathcal{U} \\ \tau(Q)=t,\ Q\subseteq L \\ L\setminus Q\subseteq H\subseteq L}} O\left(\frac{\lambda_{\tau(L)}}{\lambda_t}\right) p^{|Q\cap H|}\frac{\mathbf{P}(f-s_H)}{\mathbf{P}(f)},\end{aligned} \tag{5.3.17}$$

where the second equality is by noting

$$|H| \geqslant |L| - |Q|, \quad \lambda_t = O\left(n^{\nu_Q}p^{|Q|}\right), \text{ and } \lambda_{\tau(L)} = O\left(n^{\nu_L}p^{|L|}\right).$$

In the case $f = \mathbf{0}$, we may assume $s_H = \mathbf{0}$, since otherwise, $\mathcal{C}_{f-s_H}$ is empty. Since $\tau(L) \notin \mathbf{S}$, we have (5.3.15) for sufficiently large C.

For the case $\mathbf{0} \neq f \in \mathcal{F}_{\mathbf{S}}$, we may suppose the claim has been shown when f is replaced by any $g < f$. Denoting a general term in the maximum in (5.3.17) by M, since $\tau(L) \in \mathcal{U} \subseteq \mathcal{L}$, it suffices to show that $M = O(\lambda_{\tau(L)}/\lambda_t)$, or $M = O(n^{o(1)}\lambda_{\tau(L)}/\lambda_t)$ in the case of the $\mathcal{R}_0^*$-clustering, that is, when

$t = t^\star$. Applying the same telescoping technique in (5.2.11), we have the ratio of probabilities

$$\frac{\mathbf{P}(f - s_H)}{\mathbf{P}(f)} = \prod_{i=1}^{k} \frac{f_i(t_i)+1}{\lambda_{t_i}\gamma(f_i,t_i)}, \tag{5.3.18}$$

where $s_H = \sum_{i=1}^{k} \delta_{t_i}$ and $f_i = f - \sum_{j=1}^{i} \delta_{t_j}$ for all $i \in [k]$. By definition, an unavoidable cluster has size at most $R(R-1)$, where R is the size of the largest small cluster. Hence, the number of terms in the product (5.3.18) is at most $R(R-1)$.

Note also that each f_i occurs before f in lexicographic order, and (5.3.16) inductively implies $1/2 \leqslant \gamma(f_j - \delta_{u_j}, \delta_{u_j}) \leqslant 3/2$ for all $j \geqslant 1$. Suppose firstly that, in (5.3.18), $t_i \in \mathbf{S}_1$ for all i. Then by (5.2.18), $f_j(t_i)/\lambda_{t_i} \leqslant 3$ for all i, so we have that the product in (5.3.18) is $O(1)$ and $M = O\left(\lambda_{\tau(L)}/\lambda_t\right)$, as required.

If there is some $j' \in [k]$ in (5.3.18), for which $t_{j'} \in \mathbf{S}_2$, then $\lambda_{t_{j'}} = n^{o(1)}$. There are two subcases to consider.

- Firstly, if $H \cap Q \neq \emptyset$, then $p^{|H \cap Q|} n^{o(1)} \leqslant p n^{o(1)} = o(1)$ and hence $M = O(\lambda_{\tau(L)}/\lambda_t)$ as required.
- The second subcase is $H \cap Q = \emptyset$. Then H contains a cluster Q' of type $t_{j'}$, disjoint from Q. It follows that there is a sequence $Q_1, \ldots, Q_\ell$ of elementary clusters, each nontrivially intersecting the next, with $Q_1 \cap Q' \neq \emptyset$, $Q_\ell \cap Q \neq \emptyset$, and $Q_\ell \neq Q$. Note that here we only claim that $Q_1, \ldots, Q_\ell$ are elementary clusters in H, and H may not necessarily be the union $Q' \cup \left(\cup_{i=1}^{\ell} Q_i\right)$. We will consider two subsubcases of this second case.
  - Suppose firstly that $Q \not\subseteq Q_\ell$, and so $Q'' := Q' \cup \left(\cup_{i=1}^{\ell} Q_i\right)$ is a cluster satisfying $Q' \subset Q'' \subset L$, where the inclusions are proper and $\tau(Q') = t_{j'}$. It follows that $\lambda_{\tau(Q'')} = O(\lambda_{t_{j'}} x) = O(n^{o(1)} x)$ since $t_{j'} \in \mathbf{S}_2$. Thus $\tau(Q'') \in \mathcal{L}$, and hence by the definition (5.2.16) of $\lambda_{\mathcal{L}}$, we have $\lambda_{\tau(Q'')} \leqslant \lambda_{\mathcal{L}}$. Similarly, $\lambda_{\tau(L)} = O(x \lambda_{\tau(Q'')}) = O(x \lambda_{\mathcal{L}})$, and $M = O\left(x \lambda_{\tau(L)} n^{o(1)}/\lambda_t\right) = O\left(\lambda_{\tau(L)}/\lambda_t\right)$ as required.
  - For the other subsubcase $Q \subseteq Q_\ell$, recall that $Q_\ell \neq Q$. As Q_ℓ is elementary, it follows that this is for the $\mathcal{R}_0^*$-clustering, and Q must be a single edge of type t^*. Hence $M = O\left(\lambda_{\tau(L)} n^{o(1)}/\lambda_t\right)$ in this case, as required. We note that in fact the bound can be strengthened to $O\left(\lambda_{\tau(L)}/\lambda_t\right)$ unless $Q_\ell = L$, $\ell = 1$ and $j = 1$, and looking back at the above argument, we may use the maximum of $f(t')/\lambda_{t'}$ for $t' \in \mathbf{S}$ in place of $n^{o(1)}$, which is always at most $\log n$.

(ii) The bound (5.3.16) :

Here we may assume by induction that (5.3.16) holds with f replaced by any $g < f$, and that (5.3.15) holds as shown above. Since $c(t,t,\mathbf{0}) = 1 + O(p) = 1 + O(x)$, for the first inequality in (5.3.16), it suffices to show that Σ in (5.2.6) is of order $O(n^{o(1)} x)$. Since $\mathbf{S}$ is fixed, there is a bounded number of terms in the sum, and each may be written as

$$c(u,t,h) \frac{\lambda_u}{\lambda_t} \gamma(f - h, u) \frac{\mathbf{P}(f - s_H)}{\mathbf{P}(f)}, \tag{5.3.19}$$

as also in (5.2.10). Similar to the argument above, we have $\mathbf{P}\left(f - s_H\right)/\mathbf{P}\left(f\right) = O(n^{o(1)})$ for all cases. So we only need to show that the product of the remaining factors in (5.3.19) is $O(n^{o(1)} x)$.

Let $\mathcal{F}_1$ denote the set of $h \in \mathcal{F}_{\mathbf{S}}$ for which there are $t, u \in \mathbf{S}$ such that $c(u,t,h) \neq 0$. Note that the cardinality of $\mathcal{F}_1$ is bounded.

Inside the inductive step, we use a second level of induction on t, going from greatest to smallest in the relation '$\prec$'. Assume first that t is maximal. Since $u \in \mathbf{S}$, it is necessary that $u = t$ and $h \neq \mathbf{0}$ for such a term to be included in Σ, because $(u, h) \neq (t, \mathbf{0})$. Then $\gamma(f-h, t) \leqslant 3/2$ by (5.3.16) inductively. Furthermore, since $H \neq \emptyset$ in (5.2.2), we have $c(t, t, h) = O(p) = O(x)$, which gives the first inequality in (5.3.16).

Suppose next that t is not maximal. A term (5.3.19) with $u = t$ and $h \neq \mathbf{0}$ is $O(n^{o(1)}x)$ for reasons as above. For the case $u \neq t$ and $h \in \mathcal{F}_1$, clearly $c(u, t, h) = O(1)$. If $c(u, t, h) \neq 0$, then $t \prec u$, $\lambda_u/\lambda_t = O(x)$, and $\gamma(f-h, u) \leqslant 3/2$. So (5.3.19) is $O(n^{o(1)}x)$ and we conclude that $\Sigma = O(n^{o(1)}x)$ in (5.2.6), establishing the first inequality.

Moreover, in view of the bound (5.1.4) on x, for n large enough, we have (5.3.16) in full.

This completes the inductive step, and (5.3.15) and (5.3.16) imply the lemma. $\square$

Next we bound the approximation error for the successive iterations. The iterative scheme can be seen as a contraction mapping that updates $\{\Gamma_t^{(r)}\}_t$ and yeilds more accurate approximation of $\{\Gamma_t\}_t$ that corresponds to the fixed point. We use a similar analysis to that in [68] by Liebenau and Wormald by combining ingredients from Proposition 2.3 and Proposition 2.6 in [98].

Claim 5.9. *If for any integer $r \geqslant 0$, we have*

$$|\overline{\gamma}_t^{(r)}(f) - \gamma(f, t)| = \xi$$

for some $\xi > 0$ uniformly for all $f \in \mathcal{F}_{\mathbf{S}}$ and $t \in \mathbf{S}$, then we have that $\Gamma_t^{(r+1)}$ defined by (r1) and (r2) satisfies

$$|\overline{\gamma}_t^{(r+1)}(f) - \gamma(f, t)| = O\left(\tfrac{x+\phi_t\lambda_{\mathcal{L}}}{\lambda_t} + x\xi\right) \tag{5.3.20}$$

uniformly for all $f \in \mathcal{F}_{\mathbf{S}}$ and $t \in \mathbf{S}$.

Proof. First recall the defining equation of the correction factor

$$\gamma(f, t) = \tfrac{1}{c(t,t,\mathbf{0})}\left(1 - \Sigma(f, t)\right) - \tfrac{1}{c(t,t,\mathbf{0})} \cdot \tfrac{\theta(f,\delta_t)}{|t|\mathbf{P}(f)}, \tag{5.3.21}$$

and its estimate

$$\overline{\gamma}_t^{(r+1)}(f) = \tfrac{1}{c(t,t,\mathbf{0})}\left(1 - \overline{\Sigma}_t^{(r+1)}(f)\right),$$

where

$$\overline{\Sigma}_t^{(r+1)}(f) = \sum_{\substack{u\in\mathbf{S}\\ h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} c(u, t, h)\tfrac{\lambda_u}{\lambda_t}\overline{\gamma}_u^{(r)}(f)\prod_{i=1}^{k}\tfrac{\widetilde{g}_{t_i}}{\overline{\gamma}_{t_i}^{(r)}(f)}, \tag{5.3.22}$$

with $h = \sum_{i=1}^{k}\delta_{t_i}$, $f_i = f - \sum_{j=1}^{i}\delta_{t_j}$, and $\widetilde{g}_{t_i}$ defined in (5.3.4).

Noting $c(t, t, \mathbf{0}) = 1 + O(p) = 1 + o(1)$, and in view of the bound (5.3.14) on the θ factor in (5.3.21), in order to prove (5.3.20), it suffices to show that for all $t \in \mathbf{S}$ and $f \in \mathcal{F}_{\mathbf{S}}$,

$$\left|\Sigma(f,t) - \overline{\Sigma}_t^{(r+1)}(f)\right| = O\left(\frac{x+\phi_t\lambda_{\mathcal{L}}}{\lambda_t} + x\xi\right). \tag{5.3.23}$$

We will in fact show that the terms with $f-h \in \mathcal{F}$ appearing in both summations $\Sigma(f,t)$ and $\overline{\Sigma}_t^{(r+1)}(f)$ and the extra summands with $f-h \notin \mathcal{F}$ in $\overline{\Sigma}_t^{(r+1)}(f)$ are all negligible. This would then yield (5.3.23). Specifically, we prove that

$$\sum_{\substack{u\in\mathbf{S}\\ h,f-h\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} c(u,t,h)\frac{\lambda_u}{\lambda_t}\left|\gamma(f-h,u)\prod_{i=1}^{k}\frac{f_i(t_i)+1}{\lambda_{t_i}\gamma(f_i,t_i)} - \overline{\gamma}_u^{(r)}(f)\prod_{i=1}^{k}\frac{\widetilde{g}_{t_i}}{\overline{\gamma}_{t_i}^{(r)}(f)}\right| = O\left(\frac{x+\phi_t\lambda_{\mathcal{L}}}{\lambda_t} + x\xi\right) \tag{5.3.24}$$

and

$$\sum_{\substack{u\in\mathbf{S}\\ h\in\mathcal{F},f-h\notin\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} c(u,t,h)\frac{\lambda_u}{\lambda_t}\overline{\gamma}_u^{(r)}(f)\prod_{i=1}^{k}\frac{\widetilde{g}_{t_i}}{\overline{\gamma}_{t_i}^{(r)}(f)} = O\left(\frac{x+\phi_t\lambda_{\mathcal{L}}}{\lambda_t}\right). \tag{5.3.25}$$

If $f=\mathbf{0}$, then we have $h=\mathbf{0}$ in (5.3.24) since $f-h\in\mathcal{F}$, and the terms in (5.3.25) with $h\neq\mathbf{0}$ are 0, because $\widetilde{g}_{t_i}=0$ for all i by (5.3.4). If $h=\mathbf{0}$, then the value of k in (5.3.24) is 0, and the products in (5.3.24) is empty, and is equal to 1. Hence we have that

$$\sum_{\substack{u\in\mathbf{S}\\ h,f-\mathbf{0}\in\mathcal{F}\\ (u,h)\neq(t,\mathbf{0})}} c(u,t,h)\frac{\lambda_u}{\lambda_t}\left|\gamma(f-\mathbf{0},u) - \overline{\gamma}_u^{(r)}(f)\right| = \sum_{\substack{u\in\mathbf{S},u\neq t\\ f\in\mathcal{F}}} c(u,t,h)\frac{\lambda_u}{\lambda_t}\left|\gamma(f,u) - \overline{\gamma}_u^{(r)}(f)\right| = O\left(x\xi\right).$$

In this case, the summation in (5.3.25) is empty, since $f-\mathbf{0}\in\mathcal{F}_{\mathbf{S}}$.

It remains to consider $f\neq\mathbf{0}$ and $h\neq\mathbf{0}$. First, notice that if some $t_i=u$ in (5.3.24), then it must be that $k=1$, $h=\delta_{t_1}=\delta_u$, and $f_1=f-h=f-\delta_u$. Thus by noting $g_u=f(u)/\lambda_u=(f_1(u)+1)/\lambda_u$, we have

$$c(u,t,h)\frac{\lambda_u}{\lambda_t}\left|\gamma(f-h,u)\frac{f_1(t_1)+1}{\lambda_{t_1}\gamma(f_1,t_1)} - \overline{\gamma}_u^{(r)}(f)\frac{\widetilde{g}_{t_1}}{\overline{\gamma}_{t_1}^{(r)}(f)}\right| = 0,$$

since $\gamma(f-h,u)=\gamma(f_1,t_1)$, and similarly, $\overline{\gamma}_{t_1}^{(r)}(f)=\overline{\gamma}_u^{(r)}(f)$. For (5.3.25), if $f-h=f-\delta_u\notin\mathcal{F}$, then $f(u)<1$, thus $\widetilde{g}_{t_1}=\widetilde{g}_u=0$, and (5.3.25) holds.

So henceforth whenever $k\geqslant 1$, we assume that $t_i\prec u$ for all i. We treat (5.3.25) first. Since $f-h\notin\mathcal{F}$, so that we have $f(t_{i'})<h(t_{i'})$ for some $t_{i'}$ in (5.3.25), then $f(t_{i'})=O(1)$ and so $\widetilde{g}_{t_{i'}}=O(1/\lambda_{t_{i'}})$. The contribution of such a term in (5.3.25) is $O\left(\lambda_u/\lambda_t\lambda_{t_{i'}}\right)$, which in the case $t_{i'}\prec u$ is $O(x/\lambda_t)$.

Now we consider (5.3.24), and estimate the difference using

$$(x+\Delta_x)(y+\Delta_y)-xy = O\left(|y\Delta_x|+|x\Delta_y|\right),$$

which holds provided that $\Delta_x=O(x)$ or $\Delta_y=O(y)$. Note that $\lambda_{t_i}\to\infty$ as $n\to\infty$ for all $i\in[k]$ in (5.3.24), because we have $t_i\prec u$ for all i, and if any of these were bounded, then it would imply $\lambda_u=O(x)=O\left(n^{-\varepsilon}\right)$ and so $u\notin\mathbf{S}$. The ratios $(f_i(t_i)+1)/\lambda_{t_i}$ in (5.3.24) is therefore $O(1)$ for $f\in\mathcal{F}_{\mathbf{S}}$. We have from Claim 5.8 that $\gamma(f,t)=1+o(1)$ uniformly, and from Claim 5.7 that $\overline{\gamma}(f,t)=1+o(1)$

uniformly. This implies in particular that two products in (5.3.24) are in all cases $O(1)$.

Since $|\overline{\gamma}_t^{(r)}(f) - \gamma(f,t)| = \xi$ uniformly for all $f \in \mathcal{F}_{\mathbf{S}}$ and $t \in \mathbf{S}$, then we have for (u,h) as in the scope of the summation in (5.3.24),

$$c(u,t,h)\frac{\lambda_u}{\lambda_t}\left|\overline{\gamma}_u^{(r)}(f-h) - \gamma(f-h,u)\right| = O\left(x\xi\right),$$

and, for factors appearing in the product in (5.3.24) with $t_i \prec u$,

$$\left|\overline{\gamma}_{t_i}^{(r)}(f_i) - \gamma(f_i,t_i)\right|\frac{\lambda_u}{\lambda_t} = O\left(x\xi\right).$$

First note that for the replacement of $f_i(t_i)+1$ by $f_i(t_i)$ when evaluating $\widetilde{g}_{t_i}$, we have

$$c(u,t,h)\frac{\lambda_u}{\lambda_t\lambda_{t_i}} = O\left(\frac{x}{\lambda_t}\right). \tag{5.3.26}$$

We next show that for all u in (5.3.24),

$$c(u,t,h)\frac{\lambda_u}{\lambda_t}\left|\overline{\gamma}_u^{(r)}(f-h_0) - \overline{\gamma}_u^{(r)}(f)\right| = O\left(\frac{x}{\lambda_t}\right)$$

for any fixed h_0 with bounded entries and $f - h_0 \in \mathcal{F}_{\mathbf{S}}$, similarly we also have

$$c(u,t,h)\frac{\lambda_u}{\lambda_t}\left|\overline{\gamma}_{t_i}^{(r)}(f_i) - \overline{\gamma}_{t_i}^{(r)}(f)\right| = O\left(\frac{x}{\lambda_t}\right)$$

for all $i \in [k]$. We can assume $h_0 \neq \mathbf{0}$. By Lemma 5.7, equation (5.3.6) can be expanded in increasing powers of the ws, which are $O(x)$ under the substitution $g_v = f(v)/\lambda_v$ by Lemma 5.6. By (5.1.4), we may ignore terms whose total degree in ws is larger than some fixed value, and substitute $f(t_i)/\lambda_{t_i}$ and $(f(t_i) - h(t_i))/\lambda_{t_i}$ for g_{t_i} in the definition of $w(u,t,h)$ at (5.3.7), and subtract the two resulting expressions term by term. Since the entries of h_0 are bounded, the dominating terms for the difference of substitution $f(v)/\lambda_v$ and $(f(v) - h(v))/\lambda_v$ for g_v is bounded by $O\left(x/\lambda_t\right)$ similar to (5.3.26).

Hence combining above, we have (5.3.24) holds, since the summations in (5.3.24) contain a bounded number of terms, in particular, the size of $\mathbf{S}$ is bounded, $h = \sum_{i=1}^{k}\delta_{t_i}$ and k is bounded. This completes the proof. □

Now we are ready to bound the approximation error by a reformulation of Corollary 2.7 in [98].

Theorem 5.10. *For all $t \in \mathbf{S}$, there are power series Γ_t in n, p, $\mathbf{g}$, such that for all fixed $\varepsilon > 0$, truncations $\Gamma_{t,\varepsilon}$ of the series Γ_t, to a finite number of terms, such that we have the following.*

(b1) For all $p = n^{-\chi-\varepsilon'+o(1)}$ with $\varepsilon' \geqslant \varepsilon$, we have uniformly for all $f \in \mathcal{F}_{\mathbf{S}}$,

$$\left|\overline{\gamma}_{t,\varepsilon}(f) - \gamma(f,t)\right| = O\left(\frac{x+n^{o(1)}\lambda_{\mathcal{L}}}{\lambda_t}\right), \tag{5.3.27}$$

where $\overline{\gamma}_{t,\varepsilon}(f) := \Gamma_{t,\varepsilon}\left(n,p,\widetilde{\mathbf{g}}\right)$ with $\widetilde{\mathbf{g}}$ defined by (5.3.4).

(b2) For $\mathbf{i} \neq \mathbf{0}$, we have $[\mathbf{g}^{\mathbf{i}}]\Gamma_{t,\varepsilon} = O(x)$, for p satisfying (5.1.3), as $n \to \infty$, where $\mathbf{g}^{\mathbf{i}}$ denotes $g_1^{i_1}g_2^{i_2}\cdots g_s^{i_s}$.

(b3) For each $\mathbf{i}$, the coefficient $[\mathbf{g}^{\mathbf{i}}]\Gamma_{t,\varepsilon}$ is a multiple of $\prod_{v\in\mathbf{S}} p^{\mu_v i_v}$.

Note that we have for all $t \in \mathbf{S}$ and all $f \in \mathcal{F}_{\mathbf{S}}$, combining (5.3.27), and (5.3.13) in Claim 5.8, we have

$$|\overline{\gamma}_{t,\varepsilon}(f) - 1| \leqslant |\overline{\gamma}_{t,\varepsilon}(f) - \gamma(f,t)| + |\gamma(f,t) - 1| = O\left(n^{o(1)}x\right). \tag{5.3.28}$$

Proof of Theorem 5.10. We first fix some $\varepsilon > 0$ and show the existence of series; this gets relaxed later to show the series Γ_t are independent of the choice of ε. For (b1), combining Lemma 5.7 and Lemma 5.8, we have the bound on the initial approximation error

$$|\gamma(f,t) - 1| = |\gamma(f,t) - \overline{\gamma}_t^{(0)}(f)| = O\left(n^{o(1)}x\right),$$

holds uniformly for all $t \in \mathbf{S}$ and $f \in \mathcal{F}_{\mathbf{S}}$. Then for any fixed constant $r \geqslant 0$, by iterating successively r times using (r1) and (r2) to obtain $\Gamma_t^{(r)}$, we get from Corollary 5.9 that

$$|\overline{\gamma}_t^{(r)}(f) - \gamma(f,t)| = O\left(\tfrac{x+\phi_t\lambda_{\mathcal{L}}}{\lambda_t} + x^{r+1}\right). \tag{5.3.29}$$

In view of bound on $\lambda_{\mathcal{L}}$ in (5.2.16) and the constraint on x in (5.1.4), there exists an integer constant $r_t = r_t(\varepsilon) \geqslant 0$ such that

$$x^{r_t} = O\left(\tfrac{\phi_t\lambda_{\mathcal{L}}}{\lambda_t}\right).$$

Then for all $r \geqslant r_t$, we have

$$|\overline{\gamma}_t^{(r)}(f) - \gamma(f,t)| = O\left(\tfrac{x+\phi_t\lambda_{\mathcal{L}}}{\lambda_t}\right). \tag{5.3.30}$$

Hence we set $\Gamma_{t,\varepsilon}$ equal to the truncation of $\Gamma_t^{(r_t)}$ to those terms whose value, with $\mathbf{g}$ set equal to be all 1, is not of order $o\left(1/\lambda_t\right)$ and obtain (5.3.27). Also note that using $\Gamma_{t,\varepsilon}^{(r)}$ for any $r > r_t$ would define the same $\Gamma_{t,\varepsilon}$.

We now have shown for the case of some fixed $\varepsilon > 0$. We next claim that (5.3.27) is also valid for $\varepsilon' > \varepsilon$. Then p is smaller for the case of ε', and the recursive definition of $\Gamma_{t,\varepsilon'}$ is the same as for ε except that the definition of $\mathbf{S}$ may be different. No new types would be added to $\mathbf{S}$, and any terms in the summation in (5.3.3) corresponding to types t that are in $\mathbf{S}$ for ε, and not in $\mathbf{S}$ for ε', are now omitted. These terms are of the order $c(u,t,h)\lambda_u/\lambda_t$ times a finite product of g_i, for some $u \notin \mathbf{S}$. Since all g_i are substituted with values $n^{o(1)}$, the claim holds.

If some $\varepsilon' < \varepsilon$ is considered, then some new types may enter $\mathbf{S}$ in $\Gamma_{t,\varepsilon'}$ and are not included in $\Gamma_{t,\varepsilon}$. Also, the appropriate value of r_t may be larger for ε' than for ε, but as noted above, truncating with the larger value of r gives the same function $\Gamma_{t,\varepsilon}$, so the extra terms generated serves as higher-order corrections and cannot include any of the same monomials as appearing in $\Gamma_{t,\varepsilon}$.

For (b2), the coefficients of any non-constant monomial $\mathbf{g}^{\mathbf{i}}$ in Γ_t, as it arises recursively from (5.3.2) and (5.3.3) are $O(x)$. So we have $[\mathbf{g}^{\mathbf{i}}]\Gamma_t = O(x)$, for $\mathbf{i} \neq \mathbf{0}$ and p satisfying (5.1.3), as $n \to \infty$.

For (b3), since $\Gamma_{t,\varepsilon}$ is set to be some truncation of $\Gamma_t^{(r_t)}$ that is obtained recursively, and in each recursive step, every new product $\prod_{i=1}^k g_{t_i}$ that is introduced in (5.3.3) is accompanied by the factor $\lambda_u c(u,t,h)/\lambda_t$. Therefore it suffices to show this factor possess the claimed divisibility, by noting that the expansions of $1/c(t,t,\mathbf{0})$ and $1/\Gamma_{t_i}$ in (5.3.3) do not affect as their terms have nonnegative exponents. By its definition (5.2.1), each term of $c(u,t,h)$ is associated with a cluster of J of type u, a cluster Q of type t, and pairwise edge-disjoint clusters $J_1, \dots, J_k$ of types $t_1, \dots, t_k$, where the v-th entry of $\mathbf{i}$ satisfies $i_v = |\{j : t_j = v\}|$ for $v \in \mathbf{S}$. Hence the term $[\mathbf{g}^{\mathbf{i}}]\Gamma_{t,\varepsilon}$ is a multiple of

$$p^{|Q \cap (\bigcup_{i=1}^k J_i)|+\mu_u-\mu_t} = p^{\mu_u+|Q\cap(\bigcup_{i=1}^k J_i)|-|Q|} = p^{\mu_u-|Q\setminus(\bigcup_{i=1}^k J_i)|},$$

and therefore is divisible by $p^{\sum_{i=1}^k |J_i|}$. □

5.4 Non-existence of subhypergraphs in $H_r(n,p)$

We prove Theorem 5.1.6 in this section, that is to obtain the asymptotic probability that the binomial random hypergraph $H_r(n,p)$ is $\mathcal{R}$-free (that is having no subhypergraph that is isomorphic to some hypergraph in $\mathcal{R}$). By partitioning the probability space according to the cluster counts in random hypergraph $H_r(n,p)$, the reciprocal of $\mathcal{R}$-free probability is approximated by a summation of ratios of the probability of "typical" cluster counts to the probability of $\mathcal{R}$-free. This further gets reduced to estimate the ratio of the "adjacent" probabilities of having "perturbed" cluster counts by using

$$\frac{\mathbf{P}(f+\delta_t)}{\mathbf{P}(f)} = \frac{\lambda_t}{f(t)+1}\gamma(f,t), \tag{5.4.1}$$

where the correction factor $\gamma(f,t)$ is defined by (5.2.4), and will be approximated by $\overline{\gamma}_{t,\varepsilon}(f)$ using Theorem 5.10. In this section, we only work with the $\mathcal{R}$-clustering with one exception (Lemma 5.15).

Let $X_{\mathcal{L}}$ be the number of all large clusters occurring in $H_r(n,p)$. For $t \in \mathbf{S}$, let X_t count clusters of type t occurring in $H_r(n,p)$, and define X to be the number of subhypergraphs occurring in $H_r(n,p)$ that are isomorphic to some hypergraph in $\mathcal{R}$. Then we have

$$\mathbf{P}\left(X=0\right) = \mathbf{P}\left(H_r(n,p) \text{ is } \mathcal{R}\text{-free}\right) = \mathbf{P}\left(\left\{X_{\mathcal{L}}=0\right\} \cap \bigcap_{t\in\mathbf{S}}\{X_t=0\}\right). \tag{5.4.2}$$

In view of the bounds on the expected numbers of small clusters (5.2.15), we may fix a linear ordering on $\mathbf{S} = [s] = \{1, 2, \dots, s\}$ in decreasing order of $\nu_t - \mu_t(\chi+\varepsilon)$, and break ties arbitrarily when needed. Then by (5.2.15), we have for all $t < s$,

$$\lambda_{t+1} < n^{o(1)}\lambda_t. \tag{5.4.3}$$

Recall that $\mathcal{F}$ is the set of all non-negative integer cluster counts f defined on the set of small clusters $\mathbf{S}$, and $\mathcal{F}_{\mathbf{S}}$ contains restricted cluster counts. Also note that $\mathcal{C}_f$ contains r-graphs with no large clusters. The next lemma allows us to restrict our consideration to cluster counts in $\mathcal{F}_{\mathbf{S}}$ only.

Lemma 5.11. *We have*

$$\mathbf{P}\left(\left\{X_{\mathcal{L}} \neq 0\right\} \cup \left\{\exists t \in \mathbf{S} : X_t > m_t\right\}\right) = o(1), \tag{5.4.4}$$

and moreover,

$$\mathbf{P}\left(X=0\right)^{-1} = (1+o(1)) \sum_{f\in\mathcal{F}_{\mathbf{S}}} \frac{\mathbf{P}(\mathcal{C}_f)}{\mathbf{P}(X=0)}. \tag{5.4.5}$$

Proof. Since every large cluster contains an unavoidable cluster, and the number of types of unavoidable clusters is by its definition bounded. Let $X_{\mathcal{U}}$ be the number of all unavoidable clusters occurring in $H_r(n,p)$. Then by Markov's inequality, we have

$$\mathbf{P}\left(X_{\mathcal{L}} \neq 0\right) \leqslant \mathbf{P}\left(X_{\mathcal{U}} \geqslant 1\right) \leqslant \mathbf{E}\left[X_{\mathcal{U}}\right] = O\left(\lambda_{\mathcal{L}}\right) = o(1), \tag{5.4.6}$$

where we use $\lambda_{\mathcal{L}} = \sup_{t \notin \mathbf{S}} \lambda_t = O\left(n^{-\varepsilon_{\mathcal{L}}}\right)$ from (5.2.16),

The expected number of sets of j edge-disjoint clusters of type $t \in \mathbf{S}$ occurring in $H_r(n,p)$ is at most

$$\binom{|\mathfrak{C}_t|}{j} p^{\mu_t j} \leqslant \left(\frac{e|\mathfrak{C}_t|p^{\mu_t}}{j}\right)^j = \left(\frac{e\lambda_t}{j}\right)^j,$$

where we use the formula for λ_t in (5.2.3).

Recall that $m_t = 3\lambda_t = \Theta(n^{c+o(1)})$ for each $t \in \mathbf{S}_1$ with some $c = c(t) \geqslant 0$, and $m_t = \lambda_t \log n$ for each $t \in \mathbf{S}_2$. Let Y_t count the sets of $\lfloor m_t \rfloor + 1$ disjoint clusters of type t. By taking $j = \lfloor m_t \rfloor + 1$, we have for all $t \in \mathbf{S}_1$,

$$\mathbf{E}\left[Y_t\right] \leqslant \left(\frac{e}{3}\right)^{m_t} = O\left(e^{-\Omega(n^c)}\right);$$

similarly, we have for $t \in \mathbf{S}_2$,

$$\mathbf{E}\left[Y_t\right] \leqslant \left(\frac{e}{\log n}\right)^{\lceil \lambda_t \log n \rceil} = o(1).$$

Consequently by Markov's inequality, we have, for all $t \in \mathbf{S}$, that

$$\mathbf{P}\left(X_t > m_t\right) = \mathbf{P}\left(Y_t \geqslant 1\right) \leqslant \mathbf{E}\left[Y_t\right] = o(1).$$

Hence the union bound over $\mathbf{S}$ yields $\mathbf{P}\left(\left\{\exists t \in \mathbf{S} : X_t > m_t\right\}\right) = o(1)$. This, combining with (5.4.6), gives (5.4.4) by the union bound.

Now we are ready to show (5.4.5) by noting that

$$\mathbf{P}\left(X_{\mathcal{L}} \neq 0\right) + \sum_{f \in (\mathcal{F} \setminus \mathcal{F}_{\mathbf{S}})} \mathbf{P}\left(C_f\right) + \sum_{f \in \mathcal{F}_{\mathbf{S}}} \mathbf{P}\left(C_f\right) = 1. \tag{5.4.7}$$

Diving $\mathbf{P}\left(X = 0\right)$ on both sides of (5.4.7) gives

$$\mathbf{P}\left(X = 0\right)^{-1} = \frac{\mathbf{P}(X_{\mathcal{L}} \neq 0)}{\mathbf{P}(X=0)} + \frac{\mathbf{P}\left(\left\{\exists t \in \mathbf{S} : X_t > m_t\right\} \cap \left\{X_{\mathcal{L}} = 0\right\}\right)}{\mathbf{P}\left(X = 0\right)} + \sum_{f \in \mathcal{F}_{\mathbf{S}}} \frac{\mathbf{P}(f)}{\mathbf{P}(X=0)}.$$

Hence, by (5.4.4), we obtain (5.4.5). □

For all $t \in (\mathbf{S} \cup \{0\})$, let $\mathcal{P}(\mathbf{g}_t)$ be the ring of multivariate polynomials in $\mathbf{g}_t = (g_1, \ldots, g_t)$ such that

(1) all coefficients are polynomials in n, p and n^{-1}, and

(2) the coefficient of $g_1^{i_1} \cdots g_t^{i_t}$ is divisible by $\prod_{j=1}^{t} p^{\mu_j i_j}$.

Note that $\mathcal{P}(\mathbf{g}_0)$ is a polynomial in n, p and n^{-1}, and does not contain any indeterminate g_i.

For all $t \in [s]$, we say that the cluster counts $(j_1, \ldots, j_t)$ for types $[t]$ is *t-amenable* if $j_i \in [0, m_i]$ for all $i \in [t]$. It is useful to define the scaled cluster count for each cluster type $t \in [s]$ by

$$\zeta_t(j) = \zeta_t(j_1, \ldots, j_{t-1}, j) = \left(\frac{j_1}{\lambda_1}, \ldots, \frac{j_{t-1}}{\lambda_{t-1}}, \frac{j}{\lambda_t}\right). \tag{5.4.8}$$

Note that $\zeta_t(j)$ depends on $j_1, j_2, \ldots, j_{t-1}$, but these are suppressed from the notation for simplicity. For a t-amenable cluster count, each entry of $\zeta_t(j)$ is at most $\log n$ in view of the constraint on m_i in (5.2.19).

For $t \in [s]$, define the event

$$\mathcal{J}_t(j_1, \ldots, j_t) := \bigcap_{i \in [t]} \{X_i = j_i\} \cap \bigcap_{u=t+1}^{s} \{X_u \leqslant m_u\} \cap \{X_{\mathcal{L}} = 0\}, \tag{5.4.9}$$

similarly when $t = 0$, define

$$\mathcal{J}_0 := \{X_{\mathcal{L}} = 0\} \cap \bigcap_{u=1}^{s} \{X_u \leqslant m_u\}. \tag{5.4.10}$$

Then we have

$$\mathbf{P}\left(\mathcal{J}_s(j_1, \ldots, j_s)\right) = \mathbf{P}\left(\mathcal{C}_{j_1, \ldots, j_s}\right)$$

for all $(j_1, \ldots, j_s) \in \mathcal{F}$, and in particular,

$$\mathbf{P}\left(\mathcal{J}_s(0, \ldots, 0)\right) = \mathbf{P}\left(X = 0\right).$$

We will approximate recursively, for each $t \in \mathbf{S}$, the conditional probability of not having any small cluster of type u for all $u > t$, given the cluster counts of type v for all $v \leqslant t$. This recursive approximation method is similar to [98, Section 3] and includes some simplified intermediate steps.

Theorem 5.12. *For $t \in [s]$ and all t-amenable $(j_1, \ldots, j_t)$, we have that*

$$\mathbf{P}\left(\sum_{u=t+1}^{s} X_u = 0 \,\middle|\, \mathcal{J}_t(j_1, \ldots, j_t)\right) = \exp\left(-P_{t,\varepsilon}(\zeta(j_t)) + o(1)\right), \tag{5.4.11}$$

and

$$\mathbf{P}\left(\sum_{u=1}^{s} X_u = 0 \,\middle|\, \mathcal{J}_0\right) = \exp\left(-P_{0,\varepsilon} + o(1)\right), \tag{5.4.12}$$

for some polynomials $\{P_{t,\varepsilon}\}_{t \in (\mathbf{S} \cup \{0\})}$ such that the following hold.

(i) $P_{t,\varepsilon} \in \mathcal{P}(\mathbf{g}_t)$ for $t \in (\mathbf{S} \cup \{0\})$; in particular, $P_{0,\varepsilon}$ is a polynomial in n, p, n^{-1}.

(ii) The constant coefficient of $P_{t,\varepsilon}$, that is, $P_{t,\varepsilon}(0, \ldots, 0)$, equals $\left(1 + O\left(n^{o(1)}x\right)\right) \sum_{u=t+1}^{s} \lambda_u$, and the other coefficients are $O\left(n^{o(1)}xP_{t,\varepsilon}(0, \ldots, 0)\right)$. The implicit constants in these bounds are are independent of ε.

(iii) The error $o(1)$ is uniform over all t-amenable $(j_1, \ldots, j_t)$.

The proof of this theorem is in the next subsection. The above theorem gives $\mathbf{P}(X=0)$ as a direct corollary.

Corollary 5.13. *There exists a polynomial $P_{0,\varepsilon}$ with the properties claimed in Theorem 5.12 such that*

$$\mathbf{P}(X=0) = \exp\left(-P_{0,\varepsilon} + o(1)\right). \tag{5.4.13}$$

Proof. In view of (5.4.2), to prove (5.4.13), it suffices to show that

$$\mathbf{P}\left(\left\{\sum_{t=1}^{s} X_t = 0\right\} \cap \left\{X_{\mathcal{L}} = 0\right\}\right) = \exp\left(-P_{0,\varepsilon} + o(1)\right). \tag{5.4.14}$$

Since $\left\{\sum_{t=1}^{s} X_t = 0\right\} \subseteq \bigcap_{u=1}^{s} \left\{X_u \leqslant m_u\right\}$, we have

$$\frac{\mathbf{P}\left(\left\{\sum_{t=1}^{s} X_t = 0\right\} \cap \left\{X_{\mathcal{L}} = 0\right\}\right)}{\mathbf{P}\left(\bigcap_{u=1}^{s} \left\{X_u \leqslant m_u\right\} \cap \left\{X_{\mathcal{L}} = 0\right\}\right)} = \mathbf{P}\left(\sum_{t=1}^{s} X_t = 0 \;\middle|\; \left\{X_{\mathcal{L}} = 0\right\} \cap \bigcap_{u=1}^{s} \left\{X_u \leqslant m_u\right\}\right). \tag{5.4.15}$$

Plugging the defining equation (5.4.10) of $\mathcal{J}_0$ into (5.4.12) gives that

$$\mathbf{P}\left(\sum_{u=1}^{s} X_u = 0 \;\middle|\; \left\{X_{\mathcal{L}} = 0\right\} \cap \bigcap_{u=1}^{s} \left\{X_u \leqslant m_u\right\}\right) = \exp\left(-P_{0,\varepsilon} + o(1)\right). \tag{5.4.16}$$

By (5.4.4) in Lemma 5.11, we have

$$\mathbf{P}\left(\left\{X_{\mathcal{L}} = 0\right\} \cap \bigcap_{u=1}^{s} \left\{X_u \leqslant m_u\right\}\right) = 1 - \mathbf{P}\left(\left\{X_{\mathcal{L}} \neq 0\right\} \cup \left\{\exists t \in \mathbf{S} : X_t > m_t\right\}\right) = 1 - o(1). \tag{5.4.17}$$

Plugging estimates (5.4.16) and (5.4.17) into (5.4.15) gives (5.4.14), which completes the proof. $\square$

The rest of this section is devoted to the proof of Theorem 5.12 and Theorem 5.1.6. We first introduce some notation. Given an s-amenable $(j_1, j_2, \ldots, j_s)$, for each $t \in [s]$ and $j \in \mathbb{N}$, we define the cluster count $f_{t,j}$ on $\mathbf{S}$ by

$$f_{t,j}(u) = \begin{cases} j_u & \text{if } u < t, \\ j & \text{if } u = t, \\ 0 & \text{if } u > t. \end{cases} \tag{5.4.18}$$

Note that $f_{t,j}(\cdot)$ depends on $j_1, j_2, \ldots, j_{t-1}$, but these parameters are suppressed from the notation for simplicity.

For all $t \in \mathbf{S}$, we derive a recursive formula for the reciprocal of the conditional probability of not having any small cluster of type u for all $u \geqslant t$, given j_u clusters of type u for all $u < t$. Specifically, by the law of total probability,

$$\begin{aligned}
\frac{\mathbf{P}(\mathcal{J}_{t-1}(j_1,\ldots,j_{t-1}))}{\mathbf{P}(\mathcal{J}_s(j_1,\ldots,j_{t-1},0,\ldots,0))} &= \sum_{j_t=0}^{\lfloor m_t \rfloor} \frac{\mathbf{P}(\mathcal{J}_t(j_1,\ldots,j_{t-1},j_t))}{\mathbf{P}(\mathcal{J}_s(j_1,\ldots,j_{t-1},j_t,0,\ldots,0))} \frac{\mathbf{P}(\mathcal{J}_s(j_1,\ldots,j_{t-1},j_t,\ldots,0))}{\mathbf{P}(\mathcal{J}_s(j_1,\ldots,j_{t-1},0,\ldots,0))} \\
&= \sum_{j_t=0}^{\lfloor m_t \rfloor} \frac{\mathbf{P}(\mathcal{J}_t(j_1,\ldots,j_{t-1},j_t))}{\mathbf{P}(\mathcal{J}_s(j_1,\ldots,j_{t-1},j_t,0,\ldots,0))} \frac{\lambda_t^{j_t}}{j_t!} \prod_{j=0}^{j_t-1} \gamma(f_{t,j},t), \qquad (5.4.19)
\end{aligned}$$

where in the second equality, we use a telescoping product of ratios of adjacent probabilities (5.4.1).

Combining Claim 5.8 and Theorem 5.10 yields that for all $t \in \mathbf{S}$ and $f \in \mathcal{F}_{\mathbf{S}}$,

$$\gamma(f,t) = \overline{\gamma}_{t,\varepsilon}(f) + O\left(\frac{x+n^{o(1)}\lambda_{\mathcal{L}}}{\lambda_t}\right) = \overline{\gamma}_{t,\varepsilon}(f)\left(1 + O\left(\frac{n^{-\varepsilon'}}{\lambda_t}\right)\right),$$

with $\varepsilon' = \min(\varepsilon, \varepsilon_{\mathcal{L}})$. Recalling the definition of $\mathcal{J}_t$ in (5.4.9), we have that

$$\begin{aligned}
\frac{\mathbf{P}(\mathcal{J}_{t-1}(j_1,\ldots,j_{t-1}))}{\mathbf{P}(\mathcal{J}_s(j_1,\ldots,j_{t-1},0,\ldots,0))} &= \frac{\mathbf{P}\left(\bigcap_{i\in[t-1]}\{X_i = j_i\} \cap \bigcap_{u=t}^{s}\{X_u \leqslant m_u\} \cap \{X_{\mathcal{L}} = 0\}\right)}{\mathbf{P}\left(\bigcap_{i\in[t-1]}\{X_i = j_i\} \bigcap_{i=t}^{s}\{X_i = 0\} \cap \{X_{\mathcal{L}} = 0\}\right)} \\
&= \mathbf{P}\left(\sum_{u\geqslant t} X_u = 0 \,\middle|\, \mathcal{J}_{t-1}(j_1,\ldots,j_{t-1})\right)^{-1},
\end{aligned}$$

and consequently, it follows from (5.4.19) that

$$\begin{aligned}
&\mathbf{P}\left(\sum_{u\geqslant t} X_u = 0 \,\middle|\, \mathcal{J}_{t-1}(j_1,\ldots,j_{t-1})\right)^{-1} \\
&= \sum_{j_t=0}^{\lfloor m_t \rfloor} \mathbf{P}\left(\sum_{u>t} X_u = 0 \,\middle|\, \mathcal{J}_t(j_1,\ldots,j_{t-1},j_t)\right)^{-1} \frac{\lambda_t^{j_t}}{j_t!} \prod_{j=0}^{j_t-1} \overline{\gamma}_{t,\varepsilon}(f_{t,j}) \left(1 + O\left(\frac{n^{-\varepsilon'}}{\lambda_t}\right)\right) \\
&= (1+o(1)) \sum_{j_t=0}^{\lfloor m_t \rfloor} \mathbf{P}\left(\sum_{u>t} X_u = 0 \,\middle|\, \mathcal{J}_t(j_1,\ldots,j_{t-1},j_t)\right)^{-1} \frac{\lambda_t^{j_t}}{j_t!} \prod_{j=0}^{j_t-1} \overline{\gamma}_{t,\varepsilon}(f_{t,j}), \qquad (5.4.20)
\end{aligned}$$

where the last equality is by noting that $m_t \leqslant \lambda_t \log n$ for all $t \in \mathbf{S}$, and hence,

$$\left(1 + O\left(\frac{n^{-\varepsilon'}}{\lambda_t}\right)\right)^{m_t} = 1 + o(1).$$

We will use an inductive argument beginning with $t = s$, and then proceed through decreasing values of t. It finishes with the case $t = 0$, that is (5.4.12). The initial step of the induction argument, when $t = s$, is trivial, and we set $P_{s,\varepsilon} = 0$. Now we assume that (5.4.11) holds for some particular value of t and consider the case of $t-1$.

It is useful to define

$$T_t(j) = \exp\left(P_{t,\varepsilon}\left(\zeta_t(j)\right)\right) \frac{\lambda_t^j}{j!} \prod_{i=0}^{j-1} \overline{\gamma}_{t,\varepsilon}(f_{t,i}). \qquad (5.4.21)$$

Then by the induction hypothesis (5.4.11), we have

$$T_t(j_t) \sim \mathbf{P}\left(\sum_{u>t} X_u = 0 \,\middle|\, \mathcal{J}_t(j_1, \ldots, j_{t-1}, j_t)\right)^{-1} \frac{\lambda_t^{j_t}}{j_t!} \prod_{j=0}^{j_t-1} \overline{\gamma}_{t,\varepsilon}(f_{t,j}).$$

In view of (5.4.20), the summation of $T_t(j)$ approximates the conditional probability:

$$\mathbf{P}\left(\sum_{u \geqslant t} X_u = 0 \,\middle|\, \mathcal{J}_t(j_1, \ldots, j_{t-1})\right)^{-1} = (1 + o(1)) \sum_{j=0}^{\lfloor m_t \rfloor} T_t(j). \tag{5.4.22}$$

To prove the theorem, it suffices to show that the above summation on the right-hand side of (5.4.22) admits the desired properties, that is,

$$\sum_{j=0}^{\lfloor m_t \rfloor} T_t(j) = \exp\left(-P_{t-1,\varepsilon}(\zeta_{t-1}(j_{t-1})) + o(1)\right), \tag{5.4.23}$$

for some $P_{t-1,\varepsilon}$ with the properties (i), (ii), and (iii) claimed in theorem statement.

We introduce a constant that will be used to determine the truncation point for various expansions. Specifically, define

$$\ell_0 = \ell_0(t, \varepsilon) = 2\left(\tfrac{\nu_t - \chi\mu_t - \varepsilon\mu_t}{\varepsilon} + 2\right) = \tfrac{2}{\varepsilon}(\nu_t - \chi\mu_t) - 2\mu_t + 4. \tag{5.4.24}$$

By noting $\nu_t/\mu_t \geqslant \chi + \varepsilon$ for all $t \in \mathbf{S}$, by the definition (5.2.14) of $\mathbf{S}$, we have $\ell_0 > 0$, and

$$x^{\ell_0/2} = O\left(\frac{x^2}{n^{\nu_t - \chi\mu_t - \varepsilon\mu_t}}\right) = O\left(\frac{n^{o(1)} x^2}{\lambda_t}\right).$$

Recall from Theorem 5.10 that $\Gamma_{t,\varepsilon}$ is a polynomial in $\mathbf{g}_t$ with all coefficients $O(x)$, and for $f \in \mathcal{F}_{\mathbf{S}}$, we have $\overline{\gamma}_{t,\varepsilon}(f) = \Gamma_{t,\varepsilon}(n, p, \widetilde{\mathbf{g}}) = 1 + O(x)$ with

$$\widetilde{\mathbf{g}} = (\widetilde{g}_1, \ldots, \widetilde{g}_s) = \left(\frac{f(1)}{\lambda_1}, \ldots, \frac{f(s)}{\lambda_s}\right),$$

Hence, defining

$$\widehat{\zeta}_{t-1} = \widehat{\zeta}_{t-1}(j_1, \ldots, j_{t-1}) = \left(\tfrac{j_1}{\lambda_1}, \ldots, \tfrac{j_{t-1}}{\lambda_{t-1}}\right), \tag{5.4.25}$$

and expanding the logarithm gives

$$\begin{aligned} \log \overline{\gamma}_{t,\varepsilon}(f_{t,i}) &= \sum_{k=1}^{\ell_0} \frac{(-1)^{k-1}}{k} \left(\overline{\gamma}_{t,\varepsilon}(f_{t,i}) - 1\right)^k + O\left(\left(\overline{\gamma}_{t,\varepsilon}(f_{t,i}) - 1\right)^{\ell_0+1}\right) \\ &= \sum_{v=0}^{d'_{\max}} R_v^{(1)}(\widehat{\zeta}_{t-1}) \left(\frac{i}{\lambda_t}\right)^v + O\left(\frac{n^{o(1)} x}{\lambda_t}\right), \end{aligned} \tag{5.4.26}$$

where

- $\{R_v^{(1)}\}$ are polynomials in $\mathbf{g}_{t-1}$, and it is easily checked that $R_v^{(1)} \in \mathcal{P}(\mathbf{g}_{t-1})$ has all coefficients $O\left(n^{o(1)} x\right)$ for all $0 \leqslant v \leqslant d'_{\max}$;

- the bound on truncation error is established by noting that $|\overline{\gamma}_{t,\varepsilon}(f_{t,i})-1| = O\left(n^{o(1)}x\right)$ by (5.3.28), and in view of the definition of ℓ_0 in (5.4.24).
- and $d'_{\max} = \ell_0 \cdot \deg_{g_t}(\Gamma_{t,\varepsilon})$ with $\deg_{g_t}(\Gamma_{t,\varepsilon})$ denoting the degree of $\Gamma_{t,\varepsilon}$ in g_t.

Adapting the derivation of [98, Eq (3.19)], for all $j \leqslant m_t$, we have

$$\begin{aligned}\sum_{i=0}^{j-1} \log \overline{\gamma}_{t,\varepsilon}(f_{t,i}) &= \sum_{v=0}^{d'_{\max}} R_v^{(1)}(\widehat{\zeta}_{t-1}) \sum_{i=0}^{j-1} \left(\frac{i}{\lambda_t}\right)^v + O\left(\frac{jn^{o(1)}x}{\lambda_t}\right) \\ &= \sum_{v=0}^{d'_{\max}} \frac{R_v^{(1)}(\widehat{\zeta}_{t-1})}{v+1} \cdot \frac{j^{v+1}}{\lambda_t^v} + \sum_{v=0}^{d'_{\max}} R_v^{(1)}(\widehat{\zeta}_{t-1}) O\left(\left(\frac{j}{\lambda_t}\right)^v\right) + O\left(\frac{jn^{o(1)}x}{\lambda_t}\right) \\ &= O\left(n^{o(1)}x\right) + \sum_{v=0}^{d'_{\max}} \frac{R_v^{(1)}(\widehat{\zeta}_{t-1})}{v+1} \frac{j^{v+1}}{\lambda_t^v}, \end{aligned} \tag{5.4.27}$$

by noting that for all $(t-1)$-amenable $(j_1, \ldots, j_{t-1})$, all terms in $R_v^{(1)}(\widehat{\zeta}_{t-1})$ are $O\left(n^{o(1)}x\right)$ since all coefficients are $O(x)$.

Let $d_{\max}$ denote the degree, $\deg_{g_t}(P_{t,\varepsilon}(\mathbf{g}_t))$, of $P_{t,\varepsilon}(\mathbf{g}_t)$ in g_t. By the properties claimed in induction hypothesis, it is useful to write

$$P_{t,\varepsilon}(\mathbf{g}_t) = \sum_{v=0}^{d_{\max}} R_v^{(2)}(\widehat{\zeta}_{t-1}) \left(\frac{j}{\lambda_t}\right)^v, \tag{5.4.28}$$

where $R_v^{(2)} = [g_t^v] P_{t,\varepsilon}(\mathbf{g}_t)$ for $0 \leqslant v \leqslant d_{\max}$.

Claim 5.14. *Let $d_t = \max(d_{\max} - 1, d'_{\max})$. For all $t \in \mathbf{S}$ and t-amenable $(j_1, \ldots, j_{t-1}, j)$, we have*

$$T_t(j) = \frac{\lambda_t^j}{j!} \exp\left(R_0(\widehat{\zeta}_{t-1}) + \sum_{v=0}^{d_t} R_{v+1}(\widehat{\zeta}_{t-1}) \frac{j^{v+1}}{\lambda_t^v} + O\left(n^{o(1)}x\right)\right), \tag{5.4.29}$$

where

(i) For all $1 \leqslant v \leqslant d_t$, we have $R_v \in \mathcal{P}(\mathbf{g}_{t-1})$ with all coefficients $O\left(n^{o(1)}x\right)$.

(ii) We have

$$R_0(\widehat{\zeta}_{t-1}) = P_{t,\varepsilon}(\zeta(0)), \tag{5.4.30}$$

and for $0 \leqslant v \leqslant d_t$,

$$R_{v+1}(\widehat{\zeta}_{t-1}) = \frac{R_v^{(2)}(\widehat{\zeta}_{t-1})}{\lambda_t} + \frac{1}{v+1} R_v^{(1)}(\widehat{\zeta}_{t-1}) + O\left(\frac{n^{o(1)}x}{\lambda_t}\right). \tag{5.4.31}$$

Proof. Using (5.4.27) and (5.4.28), we rewrite the defining equation (5.4.21) of $T_t(j)$ as

$$\begin{aligned} T_t(j) &= \frac{\lambda_t^j}{j!} \exp\left(P_{t,\varepsilon}(\zeta_t(j)) + \sum_{i=0}^{j-1} \log \overline{\gamma}_{t,\varepsilon}(f_{t,i})\right) \\ &= \frac{\lambda_t^j}{j!} \exp\left(R_0^{(2)}(\widehat{\zeta}_{t-1}) + \sum_{v=0}^{d_{\max}-1} \frac{R_{v+1}^{(2)}(\widehat{\zeta}_{t-1})}{\lambda_t} \frac{j^{v+1}}{\lambda_t^v} + \sum_{v=0}^{d'_{\max}} \frac{R_v^{(1)}(\widehat{\zeta}_{t-1})}{v+1} \frac{j^{v+1}}{\lambda_t^v} + O\left(n^{o(1)}x\right)\right). \end{aligned} \tag{5.4.32}$$

Now we set

$$R_0(\widehat{\zeta}_{t-1}) = R_0^{(2)}(\widehat{\zeta}_{t-1}),$$

and claim that for $1 \leqslant v \leqslant d_{\max}$, there exists $R_v^{(3)} \in \mathcal{P}(\mathbf{g}_{t-1})$ having all coefficients $O(x)$ such that

$$\frac{R_v^{(2)}}{\lambda_t} = R_v^{(3)} + O\left(\frac{n^{o(1)}x}{\lambda_t}\right). \tag{5.4.33}$$

Since $R_v^{(2)} = [g_t^v]P_{t,\varepsilon}(\mathbf{g}_t)$, we have $R_v^{(2)}$ does not contain g_t, and $R_v^{(2)} \in \mathcal{P}(\mathbf{g}_{t-1})$. In view of λ_t in (5.2.3), and noting that $P_{t,\varepsilon} \in \mathcal{P}(\mathbf{g}_t)$ by hypothesis (i), we have the exponent of p in any term of $R_v^{(2)}/\lambda_t$ is greater than zero because the coefficient of $g_1^{i_1}\cdots g_{t-1}^{i_{t-1}}g_t^v$ of $P_{t,\varepsilon}$ is divisible by $p^{\mu_t v}\prod_{j=1}^{t-1}p^{\mu_j i_j}$, which follows from the claimed properties of $\mathcal{P}(\mathbf{g}_t)$.

Moreover, by hypothesis (ii), we have that for all $(i_1, i_2, \ldots, i_{t-1}) \neq \mathbf{0}$ and $v \geqslant 1$,

$$\frac{[g_1^{i_1}\cdots g_{t-1}^{i_{t-1}}]R_v^{(2)}}{\lambda_t} = O\left(n^{o(1)}x\sum_{u=t+1}^{s}\frac{\lambda_u}{\lambda_t}\right) = O\left(n^{o(1)}x\right),$$

where the last step is by inequality (5.4.3). Therefore we conclude (5.4.33) by noting the coefficient of $g_1^{i_1}\cdots g_{t-1}^{i_{t-1}}$ in $R_v^{(2)}$ is divisible by $\prod_{j=1}^{t-1}p^{\mu_j i_j}$.

Therefore, we set

$$R_{v+1} = R_{v+1}^{(3)} + \frac{1}{v+1}R_v^{(1)} \tag{5.4.34}$$

for $0 \leqslant v \leqslant d_t$. Then in view of (5.4.26) and (5.4.33), we have $R_v(\widehat{\zeta}_{t-1}) \in \mathcal{P}(\mathbf{g}_{t-1})$ with all coefficients $O\left(n^{o(1)}x\right)$. This gives part (i).

We have (5.4.31) in part (ii) by (5.4.33) and (5.4.34). It is easy to see (5.4.30) by (5.4.32) since $R_0^{(2)}(\widehat{\zeta}_{t-1}) = [g_t^0]P_{t,\varepsilon}(\mathbf{g}_t)$. This completes the proof. □

Next, we approximate the maximum term in the summation in (5.4.23) by estimating the point where the ratio of consecutive terms is close to 1. Note that the following lemma holds for single edge cluster $t^\star$ as well, not only for $t \in \mathbf{S}$, and this will be utilised later in next section.

Lemma 5.15. *For $t \in \mathbf{S}$ and t-amenable $(j_1, \ldots, j_{t-1}, j)$, let $T_t(j)$ be defined by (5.4.21); for $t = t^\star$, let*

$$T_{t^\star}(j) := \frac{\lambda_{t^\star}^j}{j!}\prod_{i=0}^{j-1}\overline{\gamma}_{t^\star,\varepsilon}(i\delta_{t^\star}). \tag{5.4.35}$$

Then for all $t \in \mathbf{S} \cup \{t^\star\}$, there exists $j^\star = j^\star(t) = (1+o(1))\lambda_t$ that is defined as the unique solution of a certain equation, such that the following holds.

(i) For sufficiently large n, the ratio of consecutive terms $T_t(j)/T_t(j-1)$ increases for $j < j^\star$, and decreases for $j > j^*$.

(ii) For $|k - j^*| = O\left(\sqrt{j^*}\log j^*\right)$, we have

$$T_t(k) = \exp\left(-\frac{(k-j^\star)^2}{2j^\star} + o(1)\right)T_t(\widetilde{j}).$$

Proof of Lemma 5.15. We have, from the definition of $T_t(\cdot)$ in (5.4.21), that

$$\frac{T_t(j)}{T_t(j-1)} = \frac{\lambda_t}{j} \exp\left(P_{t,\varepsilon}\left(\zeta_t(j)\right) - P_{t,\varepsilon}\left(\zeta_t\left(j-1\right)\right) + \log \overline{\gamma}_{t,\varepsilon}(f_{t,j-1})\right). \tag{5.4.36}$$

Recall that our expression $P_{t,\varepsilon}(\mathbf{g}_t)$ in (5.4.28). Using this, and then the expansion of the logarithm of the correction factor in (5.4.26), by noting its bound (5.3.27), we have

$$\begin{aligned}\frac{T_t(j)}{T_t(j-1)} &= \frac{\lambda_t}{j} \exp\left(\sum_{v=1}^{d_{\max}} \frac{R_v^{(2)}(\widehat{\zeta}_{t-1})}{\lambda_t} \frac{j^v-(j-1)^v}{\lambda_t^{v-1}} + \log \overline{\gamma}_{t,\varepsilon}(f_{t,j-1})\right) \\ &= \frac{\lambda_t}{j} \exp\left(\sum_{v=0}^{d_{\max}} \frac{(v+1)R_v^{(2)}(\widehat{\zeta}_{t-1})}{\lambda_t}\left(\frac{j}{\lambda_t}\right)^v + \sum_{v=0}^{d'_{\max}} R_v^{(1)}(\widehat{\zeta}_{t-1})\left(\frac{j-1}{\lambda_t}\right)^v + O\left(\frac{n^{o(1)}x}{\lambda_t}\right)\right).\end{aligned} \tag{5.4.37}$$

By the discussion after (5.4.26), we have that for all $v \geqslant 1$, the coefficients of all terms in $R_v^{(1)}$ are $O(x)$. Recall that $\widehat{\zeta}_{t-1}$ is defined by (5.4.25), then for t-amenable $(j_1, \ldots, j_{t-1}, j)$, we have

$$R_v^{(1)}(\widehat{\zeta}_{t-1})\left(\frac{j-1}{\lambda_t}\right)^v = R_v^{(1)}(\widehat{\zeta}_{t-1})\left(\frac{j}{\lambda_t}\right)^v + O\left(\frac{n^{o(1)}x}{\lambda_t}\right). \tag{5.4.38}$$

In view of (5.4.29), define q as a function in $\mathbf{g}_t = (g_1, \ldots, g_t)$ by

$$q(\mathbf{g}_t) = \exp\left(\sum_{v=0}^{d_t}(v+1)R_{v+1}(\mathbf{g}_{t-1})g_t^v\right). \tag{5.4.39}$$

Combining (5.4.38) with Claim 5.14 (ii), the ratio estimate (5.4.37) becomes

$$\frac{T_t(j)}{T_t(j-1)} = \left(1 + O\left(\frac{n^{o(1)}x}{\lambda_t}\right)\right) \frac{\lambda_t}{j} q(\zeta_t(j)), \tag{5.4.40}$$

where the scaled cluster count $\zeta_t(j)$ is defined in (5.4.8), and

$$q(\zeta_t(j)) = \exp\left(\sum_{v=0}^{d_t}(v+1)R_{v+1}(\widehat{\zeta}_{t-1})\left(\frac{j}{\lambda_t}\right)^v\right), \tag{5.4.41}$$

by noting (5.4.39).

To approximate j that maximize $T_t(j)$, we use an estimate, $j^\star$, by determining the ratio $T_t(j)/T_t(j-1)$ is close to 1 asymptotically. There are two steps.

- Step 1: We define $j^\star$ as the solution of a certain equation.
- Step 2: After expanding the defining equation, we use an iterative scheme to approximate $j^\star$.

Step 1: In view of the ratio estimate (5.4.40), since the coefficients of all R_v in (5.4.41) are $O\left(n^{o(1)}x\right)$ by Claim 5.14 (i), the terms in the summation in exponent of (5.4.41) are all $O\left(n^{o(1)}x\right)$ for all t-amenable $(j_1, \ldots, j_{t-1}, j)$, recalling $\widehat{\zeta}_{t-1}$ is defined in (5.4.25). Therefore we conclude $q(\zeta_t(j)) \sim 1$ for all t-amenable $(j_1, \ldots, j_{t-1}, j)$.

Note that the partial derivative of $q(\mathbf{g}_t)$ with respect to g_t is

$$\exp\left(\sum_{v=0}^{d_t}(v+1)R_{v+1}(\mathbf{g}_{t-1})g_t^v\right)\sum_{v=1}^{d_t}(v+1)R_{v+1}(\mathbf{g}_{t-1})g_t^{v-1} = O\left(n^{o(1)}x\right), \tag{5.4.42}$$

when $\mathbf{g}_{t-1} = \widehat{\zeta}_{t-1}$ and $g_t = g_t(j) = j/\lambda_t$ for $0 \leqslant j \leqslant 3\lambda_t$. Hence there exists a unique $j^\star \in \mathbb{R}$ such that

$$q(\zeta_t(j^\star)) = \frac{j^\star}{\lambda_t}, \tag{5.4.43}$$

or equivalently,

$$\frac{j^\star}{\lambda_t} = \exp\left(\sum_{v=0}^{d_t}(v+1)R_{v+1}(\widehat{\zeta}_{t-1})\left(\frac{j^\star}{\lambda_t}\right)^v\right). \tag{5.4.44}$$

Note that $g_t(j^\star) = j^\star/\lambda_t = 1 + o(1)$.

Step 2: Recalling its definition in (5.4.39), we expand the multivariate function $q(\mathbf{g}_t)$ in (5.4.43) around $(0, \dots, 0)$, and approximate $j^\star$ by λ_t with a multiplicative polynomial factor in $\widehat{\zeta}_{t-1}$.

Since the terms in the summation on the right-hand side of (5.4.41) are all $O\left(n^{o(1)}x\right)$ for $j^\star \sim \lambda_t$ and $\widehat{\zeta}_{t-1}$ under consideration, expanding the exponential function $q(\zeta_t(j))$ defined by (5.4.41), and truncating at ℓ_0 defined by (5.4.24), yield

$$\begin{aligned} q(\zeta_t(j)) &= 1 + \sum_{k=1}^{\ell_0}\frac{1}{k!}\left(\sum_{v=0}^{d_t}(v+1)R_{v+1}(\widehat{\zeta}_{t-1})\left(\frac{j}{\lambda_t}\right)^v\right)^k + O\left(\left(n^{o(1)}x\right)^{\ell_0+1}\right) \\ &= \widetilde{q}(g_1, \dots, g_t) + O\left(\frac{n^{o(1)}x}{\lambda_t}\right), \end{aligned}$$

where for $0 \leqslant v \leqslant \ell_0 d_t$, there exist $\widehat{R}_v \in \mathcal{P}(\mathbf{g}_{t-1})$ with all coefficients $O\left(n^{o(1)}x\right)$ such that

$$\widetilde{q}(g_1, \dots, g_t) = 1 + \sum_{v=0}^{\ell_0 d_t}\widehat{R}_v(\widehat{\zeta}_{t-1})g_t^v. \tag{5.4.45}$$

Recalling that $q(\zeta_t(j)) \sim 1$, the ratio (5.4.40) at $j^\star$ becomes

$$\frac{T_t(j^\star)}{T_t(j^\star{-}1)} = \left(1 + O\left(\frac{n^{o(1)}x}{\lambda_t}\right)\right)\frac{\lambda_t}{j^\star}\widetilde{q}\left(\zeta_t(j^\star)\right).$$

Note that the partial derivative of $\widetilde{q}$ with respect to g_t is also $O(n^{o(1)}x)$ when $\mathbf{g}_{t-1} = \widehat{\zeta}_{t-1}$ and $0 \leqslant g_t \leqslant 3$, as $g_t = j/\lambda_t$ and $0 \leqslant j \leqslant 3\lambda_t$.

Therefore for n sufficiently large, beginning with $\beta_0 = 1$, a fixed-point iteration by repeated substitutions

$$\beta_\ell = \widetilde{q}(g_1, \dots, g_{t-1}, \beta_{\ell-1}), \tag{5.4.46}$$

yields polynomials $\beta_\ell = \beta_\ell(\widehat{\zeta}_{t-1}) \in \mathcal{P}(\mathbf{g}_{t-1})$ that approximate its fixed point $\beta = \widetilde{q}(g_1, \dots, g_{t-1}, \beta) = 1 + o(1)$. In each iteration, the approximation error gets multiplied by $O(n^{o(1)}x)$, hence, for ℓ sufficiently large, setting $\beta^\star = \beta_\ell(\widehat{\zeta}_{t-1})$ gives

$$j^\star = \beta^\star\lambda_t + O\left(n^{o(1)}x\right) = (1 + o(1))\lambda_t. \tag{5.4.47}$$

Note that $\beta^\star \in \mathcal{P}(\mathbf{g}_{t-1})$ and has all coefficients $O\left(n^{o(1)}x\right)$.

Now we show the asymptotic monotonicity claimed in (i), that is, the ratio of consecutive terms $T_t(j)/T_t(j-1)$ increases for $j < j^\star$, and decreases for $j > j^*$ for large n. First, we consider the terms

$T_t(j)$ with $j > j^\star$. Let

$$F(j) = \lambda_t q(\zeta_t(j)) - j. \tag{5.4.48}$$

By the defining equation of $j^\star$ in (5.4.43), and the bound (5.4.42) on the partial derivative of q, we have, from (5.4.48), that

$$F(j^\star) = 0, \text{ and } \quad \frac{\mathrm{d}}{\mathrm{d}j}F(j) = -1 + \frac{\partial}{\partial g_t}q(\zeta(j)) = -1 + O\left(n^{o(1)}x\right) \text{ for } 0 \leqslant j \leqslant m_t. \tag{5.4.49}$$

In view of the ratios of consecutive terms (5.4.40), also recalling that $q(\zeta_t(j)) \sim 1$, if for some large constant $C > 0$, we have

$$\frac{\lambda_t}{j}q(\zeta_t(j)) - 1 < -C\frac{n^{o(1)}x}{\lambda_t},$$

or equivalently,

$$F(j) < -Cn^{o(1)}x\frac{j}{\lambda_t}, \tag{5.4.50}$$

then we have $T_t(j) < T_t(j-1)$. It then suffices to show (5.4.50) for $j > j^\star$. For all integers j such that $j^\star < j < m_t$, by noting the bound on $j^\star$ in (5.4.47), we have

$$j - j^\star > 2Cn^{o(1)}x\frac{j^\star}{\lambda_t} = o(1).$$

By the mean value theorem, in view of derivative bound (5.4.49), we have, for $j^\star < j < m_t$, there exists $\psi \in [j^*, j]$ such that

$$F(j) = (j - j^*)F'(\psi) < -(2 + o(1))Cn^{o(1)}x\frac{j^\star}{\lambda_t},$$

which satisfies (5.4.50). Therefore we conclude that $T_t(j) < T_t(j-1)$. The lower side follows a similar analysis.

What remains is to show (ii), that is to approximate the ratios to the maximum term in (5.4.23) for terms with indices that are near it, specifically, we estimate

$$Q_t(k) := \log\left(\frac{T_t(k)}{T_t(\widetilde{j})}\right) \tag{5.4.51}$$

for $|k - j^*| = O\left(\sqrt{j^*}\log j^*\right)$.

Taking the logarithm of the ratio estimate in the right-hand side of (5.4.40) yields

$$\begin{aligned}\log\left(\frac{\lambda_t}{j}q(\zeta_t(j))\right) &= \log\left(\frac{\lambda_t}{j}\right) + \sum_{v=0}^{d_t}(v+1)R_{v+1}(\widehat{\zeta}_{t-1})\left(\frac{j}{\lambda_t}\right)^v \\ &= \log\left(\frac{\lambda_t}{j^\star}\right) - \log\left(1 + \frac{j-j^\star}{j^\star}\right) + \sum_{v=0}^{d_t}(v+1)R_{v+1}(\widehat{\zeta}_{t-1})\left(\frac{j^\star}{\lambda_t}\right)^v\left(1 + \frac{j-j^\star}{j^\star}\right)^v,\end{aligned} \tag{5.4.52}$$

where the second equality is by noting $j^\star \neq 0$, and rewriting

$$j = j^\star\left(1 + \tfrac{j-j^\star}{j^\star}\right).$$

We also have, from the defining equation of $j^\star$ in (5.4.44), that

$$\log\left(\tfrac{j^\star}{\lambda_t}\right) = \sum_{v=0}^{d_t}(v+1)R_{v+1}(\widehat{\zeta}_{t-1})\left(\tfrac{j^\star}{\lambda_t}\right)^v.$$

Plugging this in (5.4.52) gives

$$\log\left(\tfrac{\lambda_t}{j}q(\zeta_t(j))\right) = -\tfrac{j-j^\star}{j^\star} + O\left(\left(\tfrac{j-j^\star}{j^\star}\right)^2\right) + \sum_{v=1}^{d_t}(v+1)R_{v+1}(\widehat{\zeta}_{t-1})\left(\tfrac{j^\star}{\lambda_t}\right)^v O\left(\tfrac{j-j^\star}{j^\star}\right), \qquad (5.4.53)$$

where we expand the logarithm for j such that $|j - j^\star| = o(j^\star)$. Note that the terms in the summation on the right-hand side of above equation are all $O(x(j-j^\star)/j^\star)$.

Let $\widetilde{j} := \lfloor j^* \rfloor$. Recalling the definition of $Q_t(k)$ in (5.4.51), and using the ratio estimate (5.4.40), we obtain

$$Q_t(k) = \sum_{j=\widetilde{j}+1}^{k}\log\left(\frac{T_t(j)}{T_t(j-1)}\right) = O\left(|k-j^*|\frac{n^{o(1)}x}{\lambda_t}\right) + \sum_{j=\widetilde{j}+1}^{k}\log\left(\tfrac{\lambda_t}{j}q(\zeta_t(j))\right).$$

Since $j^\star = (1+o(1))\lambda_t$ by (5.4.47), for $k = j^* + O\left(\sqrt{j^*}\log j^*\right)$, in view of (5.4.53), we have, from (5.4.53), that

$$Q_t(k) = -\tfrac{(k-j^\star)^2}{2j^\star} + O\left(n^{o(1)}x\log^2 j^* + \frac{\log^3 j^*}{\sqrt{j^*}}\right) = -\tfrac{(k-j^\star)^2}{2j^\star} + o(1), \qquad (5.4.54)$$

which completes the proof. □

5.4.1 Proof of Theorem 5.12

Now we are ready to deal with the summation (5.4.23). First we consider the case that $t \in \mathbf{S}_2$. The other case, of $t \in \mathbf{S}_1$, is much more involved, and will be treated separately.

Case 1: $t \in \mathbf{S}_2$

First recall from (5.2.15) that $\lambda_t = \Theta(n^{o(1)})$ and $m_t = \lambda_t \log n$ for $t \in \mathbf{S}_2$. Let $P_{s,\varepsilon} = 0$. For all $t \in \mathbf{S}_2 \setminus \{s\}$, we inductively define $P_{t-1,\varepsilon} = P_{t-1,\varepsilon}(\widehat{\zeta}_{t-1})$ by

$$P_{t-1,\varepsilon} = P_{t,\varepsilon}(\zeta_t(0)) + \lambda_t.$$

Claim 5.16. *For all $t \in \mathbf{S}_2$, we have*

$$\mathbf{P}\left(\sum_{u\geqslant t}X_u = 0 \,\middle|\, \mathcal{J}_{t-1}(j_1,\ldots,j_{t-1})\right)^{-1} = \exp\left(\sum_{u\geqslant t}\lambda_u + o(1)\right). \qquad (5.4.55)$$

Proof. For $T_t(j)$ defined by (5.4.21), with $P_{t,\varepsilon}$ defined by (5.4.55), in view of Claim 5.14, for $t \in \mathbf{S}_2$, we have

$$\sum_{j=0}^{\lfloor m_t \rfloor} T_t(j) = \sum_{j=0}^{\lfloor \lambda_t \log n \rfloor} \frac{\lambda_t^j}{j!} \exp\left(R_0(\widehat{\zeta}_{t-1}) + \sum_{v=0}^{d_t} R_{v+1}(\widehat{\zeta}_{t-1}) \frac{j^{v+1}}{\lambda_t^v} + O\left(n^{o(1)} x\right)\right)$$
$$= \left(1 + O\left(n^{o(1)}x\right)\right) \exp\left(R_0(\widehat{\zeta}_{t-1})\right) \sum_{j=0}^{\lfloor \lambda_t \log n \rfloor} \frac{\lambda_t^j}{j!} \exp\left(\sum_{v=0}^{d_t} R_{v+1}(\widehat{\zeta}_{t-1}) \frac{j^{v+1}}{\lambda_t^v}\right). \tag{5.4.56}$$

Since $j_u/\lambda_u \leqslant \log n$ for all $u \leqslant t$, and R_v has all coefficients $O\left(n^{o(1)}x\right)$ by Claim 5.14 (i), noting the constraint on x in (5.1.4), we have

$$R_{v+1}(\widehat{\zeta}_{t-1}) \frac{j^{v+1}}{\lambda_t^v} = O\left(n^{o(1)}x\right) = o(1),$$

for all $0 \leqslant v \leqslant d_t$.

Let the random variable Y satisfy $Y \overset{d}{\sim} \mathrm{Poi}\,(\lambda_t)$. Then

$$\sum_{j \geqslant \lfloor \lambda_t \log n \rfloor + 1} \frac{\lambda_t^j}{j!} = e^{\lambda_t} \mathbf{P}\left(Y \geqslant \lfloor \lambda_t \log n \rfloor + 1\right).$$

If $\lambda_t \log n = \omega(1)$, we have, for some $\phi(n) \to 0$, that

$$\sum_{j \geqslant \lfloor \lambda_t \log n \rfloor + 1} \frac{\lambda_t^j}{j!} = e^{\lambda_t} \mathbf{P}\left(Y - \lambda_t \geqslant (1 + \phi(n))\lambda_t \log n\right)$$
$$\leqslant \exp\left(\lambda_t - \frac{(1+\phi(n))\lambda_t^2 \log^2 n}{2(\lambda_t + \lambda_t \log n)}\right) = \exp\left(-(1 + \phi(n))\lambda_t \log n/2\right) = o(1),$$

where the inequality is by Poisson concentration (see, for example, combining [58, Remark 2.6] and [58, Theorem 2.1, Eq. (2.5), (2.6)]). For $\lambda_t \log n = O(1)$, we have, by Markov's inequality, that

$$\sum_{j \geqslant \lfloor \lambda_t \log n \rfloor + 1} \frac{\lambda_t^j}{j!} \leqslant \frac{\mathbf{E}[Y]}{\lambda_t \log n} e^{\lambda_t} = \frac{1+O(\lambda_t)}{\log n} = o(1),$$

since $\mathbf{E}\left[Y\right] = \lambda_t$. Hence we conclude

$$\sum_{j=0}^{\lfloor \lambda_t \log n \rfloor} \frac{\lambda_t^j}{j!} = e^{\lambda_t} - o(1) = e^{\lambda_t + o(1)},$$

where the last equality is by noting that $\lambda_t = \Theta(n^{o(1)})$ for $t \in \mathbf{S}_2$.

From (5.4.56), by noting (5.4.30), we have that

$$\sum_{j=0}^{\lfloor m_t \rfloor} T_t(j) = (1 + o(1)) \exp\left(R_0(\widehat{\zeta}_{t-1})\right) \sum_{j=0}^{\lfloor \lambda_t \log n \rfloor} \frac{\lambda_t^j}{j!}$$
$$= (1 + o(1)) \exp\left(P_{t,\varepsilon}(\zeta_t(0))\right) e^{\lambda_t + o(1)} = \exp\left(P_{t,\varepsilon}(\zeta_t(0)) + \lambda_t + o(1)\right).$$

It follows from (5.4.22) and induction hypothesis (5.4.11) that

$$\mathbf{P}\left(\sum_{u\geqslant t} X_u = 0 \,\middle|\, \mathcal{J}_{t-1}(j_1,\dots,j_{t-1})\right)^{-1} = \exp\left(P_{t,\varepsilon}(\zeta_t(0)) + \lambda_t + o\left(1\right)\right).$$

Note that λ_t is a polynomial in n and p. Setting $P_{t-1,\varepsilon} = P_{t,\varepsilon}(\zeta_t(0)) + \lambda_t$ gives the inductive hypotheses (i) and (ii); and the uniformity in (iii) implies that (iii) holds with t replaced by $t-1$. □

Case 2: $t \in \mathbf{S}_1$

In this case, we have $\lambda_t \to \infty$ and $m_t = 3\lambda_t$. To handle the summation in the left-hand side of (5.4.23) for $t \in \mathbf{S}_1$, it is useful to rewrite the non-error terms in the exponential factor of $T_t(j)$ in (5.4.29) as follows:

$$R_0(\widehat{\zeta}_{t-1}) + \sum_{v=0}^{d_t} R_{v+1}(\widehat{\zeta}_{t-1})\frac{j^{v+1}}{\lambda_t^v} = \widetilde{F}_t\left(\frac{j^\star}{\lambda_t}, \frac{j}{j^\star}\right), \tag{5.4.57}$$

where $j^\star$ is defined by (5.4.43), and we introduce a bivariate polynomial

$$\widetilde{F}_t(\theta,\eta) := R_0(\widehat{\zeta}_{t-1}) + j^\star \sum_{k=1}^{d_t+1} R_k(\widehat{\zeta}_{t-1})\theta^{k-1}\eta^k.$$

For $0 \leqslant \ell_1 \leqslant d_t$ and $0 \leqslant \ell_2 \leqslant d_t+1$ with $\ell_1+\ell_2 \geqslant 1$, we have

$$\frac{\partial^{\ell_1+\ell_2}}{\partial\theta^{\ell_1}\partial\eta^{\ell_2}}\widetilde{F}_t(\theta,\eta) = j^\star \sum_{k=\max(\ell_1+1,\ell_2)}^{d_t+1} R_k(\widehat{\zeta}_{t-1})[k-1]_{\ell_1}[k]_{\ell_2}\theta^{k-\ell_1-1}\eta^{k-\ell_2}. \tag{5.4.58}$$

We use the Taylor series at $(\beta^\star, 1)$ to obtain

$$\widetilde{F}_t(\theta,\eta) = \widetilde{F}_t(\beta^\star,1) + \sum_{\substack{0\leqslant\ell_1\leqslant d_t\\ 0\leqslant\ell_2\leqslant d_t+1\\ \ell_1+\ell_2\geqslant 1}} \frac{1}{\ell_1!\ell_2!}\frac{\partial^{\ell_1+\ell_2}\widetilde{F}_t(\beta^\star,1)}{\partial\theta^{\ell_1}\partial\eta^{\ell_2}}(\theta-\beta^\star)^{\ell_1}(\eta-1)^{\ell_2}. \tag{5.4.59}$$

In view of (5.4.47), we have $j^\star = O\left(\lambda_t\right)$, and therefore,

$$j^\star\left(\frac{j^\star}{\lambda_t} - \beta^\star\right) = O\left(n^{o(1)}x\right).$$

Hence by recalling from Claim 5.14 (i) that all coefficients of each R_v are $O\left(n^{o(1)}x\right)$, and $j^\star = (1+o(1))\lambda_t$, we have, by (5.4.58), that for all $\ell_1 \geqslant 1$,

$$\frac{\partial^{\ell_1+\ell_2}\widetilde{F}_t(\beta^\star,1)}{\partial\theta^{\ell_1}\partial\eta^{\ell_2}}\left(\frac{j^\star}{\lambda_t} - \beta^\star\right)^{\ell_1} = O\left(n^{o(1)}x\right).$$

Then we have, by (5.4.59) and then (5.4.58), that

$$
\begin{aligned}
&\widetilde{F}_t\left(\frac{j^\star}{\lambda_t}, \frac{j}{j^\star}\right) \\
&= \widetilde{F}_t(\beta^\star, 1) + \sum_{\ell_2=1}^{d_t+1} \frac{1}{\ell_2!} \frac{\partial^{\ell_2} \widetilde{F}_t(\beta^\star,1)}{\partial \eta^{\ell_2}} \left(\frac{j}{j^\star} - 1\right)^{\ell_2} + \sum_{\substack{\ell_1 \geqslant 1 \\ 0 \leqslant \ell_2 \leqslant d_t+1}} \frac{1}{\ell_1! \ell_2!} \frac{\partial^{\ell_1+\ell_2} \widetilde{F}_t(\beta^\star,1)}{\partial \theta^{\ell_1} \partial \eta^{\ell_2}} \left(\frac{j^\star}{\lambda_t} - \beta^\star\right)^{\ell_1} \left(\frac{j}{j^\star} - 1\right)^{\ell_2} \\
&= \widetilde{F}_t(\beta^\star, 1) + \sum_{\ell_2=1}^{d_t+1} j^\star \sum_{k=\ell_2}^{d_t+1} R_k(\widehat{\zeta}_{t-1})(\beta^\star)^{k-1} \frac{[k]_{\ell_2}}{\ell_2!} \left(\frac{j}{j^\star} - 1\right)^{\ell_2} + O\left(n^{o(1)} x \left(\frac{j}{j^\star} - 1\right)^{d_t+1}\right),
\end{aligned}
$$

where

$$
\widetilde{F}_t(\beta^\star, 1) = R_0(\widehat{\zeta}_{t-1}) + j^\star \sum_{k=1}^{d_t+1} R_k(\widehat{\zeta}_{t-1})(\beta^\star)^{k-1}. \tag{5.4.60}
$$

Combining above expansion and (5.4.57) gives

$$
\begin{aligned}
&R_0(\widehat{\zeta}_{t-1}) + \sum_{v=0}^{d_t} R_{v+1}(\widehat{\zeta}_{t-1}) \frac{j^{v+1}}{\lambda_t^v} \\
&\quad = \widetilde{F}_t(\beta^\star, 1) + \sum_{v=1}^{d_t+1} \left(\sum_{k=v}^{d_t+1} \binom{k}{v} R_k(\widehat{\zeta}_{t-1})(\beta^\star)^{k-1}\right) \frac{(j-j^\star)^v}{(j^\star)^{v-1}} + O\left(n^{o(1)} x\right).
\end{aligned} \tag{5.4.61}
$$

Using this, we rewrite $T_t(j)$ in (5.4.29) as

$$
\begin{aligned}
T_t(j) &= \frac{\lambda_t^j}{j!} \exp\left(\widetilde{F}_t(\beta^\star, 1) + \sum_{v=1}^{d_t+1} \left(\sum_{k=v}^{d_t+1} \binom{k}{v} R_k(\widehat{\zeta}_{t-1})(\beta^\star)^{k-1}\right) \frac{(j-j^\star)^v}{(j^\star)^{v-1}} + O\left(n^{o(1)} x\right)\right) \\
&= (1+o(1)) \frac{\lambda_t^j}{j!} \exp\left(\widetilde{R}_0(\widehat{\zeta}_{t-1}) + \widetilde{R}_1(\widehat{\zeta}_{t-1}) j + \sum_{v=1}^{d_t} \widetilde{R}_{v+1}(\widehat{\zeta}_{t-1}) \frac{(j-j^\star)^{v+1}}{(j^\star)^v}\right),
\end{aligned} \tag{5.4.62}
$$

where we seperate the the constant and linear term in the exponent, and by plugging in $\widetilde{F}_t(\beta^\star, 1)$ from (5.4.60), we have

$$
\widetilde{R}_0(\widehat{\zeta}_{t-1}) = R_0(\widehat{\zeta}_{t-1}) - j^\star \sum_{k=2}^{d_t+1} (k-1) R_k(\widehat{\zeta}_{t-1})(\beta^\star)^{k-1} \tag{5.4.63}
$$

$$
\widetilde{R}_1(\widehat{\zeta}_{t-1}) = \sum_{k=1}^{d_t+1} k R_k(\widehat{\zeta}_{t-1})(\beta^\star)^{k-1}, \tag{5.4.64}
$$

and for $v \in [d_t]$,

$$
\widetilde{R}_{v+1}(\widehat{\zeta}_{t-1}) = \sum_{k=v+1}^{d_t+1} \binom{k}{v+1} R_k(\widehat{\zeta}_{t-1})(\beta^\star)^{k-1}.
$$

Recalling that $\beta^\star$ and all R_v are in $\mathcal{P}(\mathbf{g}_{t-1})$ with all coefficients $O\left(n^{o(1)} x\right)$, we also have $\widetilde{R}_v \in \mathcal{P}(\mathbf{g}_{t-1})$, and has all coefficients $O(n^{o(1)} x)$ for all $1 \leqslant v \leqslant d_t$.

Taking the derivative with respect to j on both sides of (5.4.57), after expanding $\widetilde{F}_t$ with respect to the second variable around 1 only, with the first variable setting to be $j^\star/\lambda_t$, we have

$$\begin{aligned}
&\sum_{v=0}^{d_t}(v+1)R_{v+1}(\widehat{\zeta}_{t-1})\left(\frac{j}{\lambda_t}\right)^v \\
&=\frac{\mathrm{d}}{\mathrm{d}j}\left(\widetilde{F}_t\left(\frac{j^\star}{\lambda_t},1\right)+\sum_{\ell_2=1}^{d_t+1}\frac{1}{\ell_2!}\frac{\partial^{\ell_2}\widetilde{F}_t(j^\star/\lambda_t,1)}{\partial\eta^{\ell_2}}\left(\frac{j}{j^\star}-1\right)^{\ell_2}\right) \\
&=\sum_{k=1}^{d_t+1}kR_k(\widehat{\zeta}_{t-1})\left(\frac{j^\star}{\lambda_t}\right)^{k-1}+\sum_{\ell_2=2}^{d_t+1}\frac{1}{\ell_2!}\sum_{k=\ell_2}^{d_t+1}R_k(\widehat{\zeta}_{t-1})[k]_{\ell_2}\left(\frac{j^\star}{\lambda_t}\right)^{k-1}\ell_2\left(\frac{j}{j^\star}-1\right)^{\ell_2-1},
\end{aligned}$$

where the first summand in the last line is when $\ell_2=1$. In view of $q(\zeta_t(j))$ defined in (5.4.41), and $q(\zeta_t(j^\star))$ in (5.4.43), evaluating above equation at $j=j^\star$ yields

$$\begin{aligned}
q(\zeta_t(j^\star)) &= \exp\left(\sum_{k=1}^{d_t+1}kR_k(\widehat{\zeta}_{t-1})\left(\frac{j^\star}{\lambda_t}\right)^{k-1}\right) \\
&= \exp\left(\sum_{k=1}^{d_t+1}kR_k(\widehat{\zeta}_{t-1})(\beta^\star)^{k-1}+\sum_{k=1}^{d_t+1}kR_k(\widehat{\zeta}_{t-1})\left(\frac{j^\star}{\lambda_t}-\beta^\star\right)\sum_{i=0}^{k-2}\left(\frac{j^\star}{\lambda_t}\right)^i(\beta^\star)^{n-1-i}\right) \\
&= \exp\left(\widetilde{R}_1(\widehat{\zeta}_{t-1})+O\left(\frac{n^{o(1)}x}{\lambda_t}\right)\right),
\end{aligned}$$

where the last step is by recalling $\widetilde{R}_1$ defined in (5.4.64), and by noting $\beta^\star=1+o(1)$, $j^\star=(1+o(1))\lambda_t$, and

$$\frac{j^\star}{\lambda_t}-\beta^\star=O\left(\frac{n^{o(1)}x}{\lambda_t}\right),$$

by (5.4.47). Recalling that $j^\star$ satisfies $q(\zeta_t(j^\star))=j^\star/\lambda_t$ by (5.4.43), it follows that

$$j^\star=\exp\left(\widetilde{R}_1(\widehat{\zeta}_{t-1})+O\left(\frac{n^{o(1)}x}{\lambda_t}\right)\right)\lambda_t. \tag{5.4.65}$$

Using this, we rewrite (5.4.62) by separating the constant and linear term from the exponent, and obtain, for $0\leqslant j\leqslant m_t$, that

$$\begin{aligned}
T_t(j) &= (1+o(1))\exp\left(\widetilde{R}_0(\widehat{\zeta}_{t-1})\right)\frac{(j^\star)^j}{j!}\left(1+O\left(\frac{n^{o(1)}x}{\lambda_t}\right)\right)^j\exp\left(\sum_{v=1}^{d_t}\widetilde{R}_{v+1}(\widehat{\zeta}_{t-1})\frac{(j-j^\star)^{v+1}}{(j^\star)^v}\right) \\
&= \exp\left(\widetilde{R}_0(\widehat{\zeta}_{t-1})+o(1)\right)\frac{(j^\star)^j}{j!}\exp\left(\sum_{v=1}^{d_t}\widetilde{R}_{v+1}(\widehat{\zeta}_{t-1})\frac{(j-j^\star)^{v+1}}{(j^\star)^v}\right). \qquad (5.4.66)
\end{aligned}$$

Define the set of integers

$$J_t=\left\{0\leqslant j\leqslant\lfloor 3\lambda_t\rfloor : |j-j^\star|\leqslant\sqrt{j^\star}\log j^\star\right\}.$$

Recalling that $Q_t(\cdot)$ is defined as the logarithm of the ratio in (5.4.51), we may divide the summation into two parts

$$\sum_{j=0}^{\lfloor m_t \rfloor} T_t(j) = \sum_{j\in J_t} T_t(j) + T_{j^\star} \sum_{0\leqslant j\leqslant \lfloor m_t\rfloor : j\notin J_t} \exp\left(Q_t(j)\right). \tag{5.4.67}$$

We will show that the first term on the right-hand side is the main contribution.

By the asymptotic monotonicity claimed in Lemma 5.15 (ii), we use T_j at the end points of set J_t as upper bounds for terms T_j such that $j \notin J_t$, and obtain

$$\sum_{0\leqslant j\leqslant \lfloor m_t\rfloor : j\notin J_t} \exp\left(Q_t(j)\right) \leqslant 3\lambda_t \exp\left(-\frac{j^\star \log^2 j^\star}{2j^\star}(1+o(1))\right) = O\left(\lambda_t^{-\log \lambda_t}\right),$$

and therefore, by noting $\lambda_t = \Theta(n^c)$ with some $c > 0$ for $t \in \mathbf{S}_1$, we have, for some constant $c' > 0$, that

$$\sum_{j=0}^{\lfloor m_t \rfloor} T_t(j) = \sum_{j\in J_t} T_t(j) + O\left(n^{-c'\log n}\right) T_{j^\star} = (1+o(1)) \sum_{j\in J_t} T_t(j). \tag{5.4.68}$$

Hence we have the summation for $j \in J_t$ is the main contribution.

For $j \in J_t$ and $v \geqslant 1$, we have

$$\frac{(j-j^\star)^{v+1}}{(j^\star)^v} = O\left(\frac{(j-j^\star)^2}{j^\star}\right) = O\left(\log^2 j^\star\right).$$

For $1 \leqslant v \leqslant d_t$, we have $\widetilde{R}_{v+1}(\widehat{\zeta}_{t-1}) = O\left(n^{o(1)}x\right)$, and therefore

$$\begin{aligned}\sum_{j\in J_t} T_t(j) &= \exp\left(\widetilde{R}_0(\widehat{\zeta}_{t-1}) + o(1)\right) \sum_{j\in J_t} \frac{(j^\star)^j}{j!} \exp\left(O\left(n^{o(1)} x \log^2 j^\star\right)\right) \\ &= \exp\left(\widetilde{R}_0(\widehat{\zeta}_{t-1}) + o(1)\right) \sum_{j\in J_t} \frac{(j^\star)^j}{j!}.\end{aligned}$$

Let random variable Y' follow Poisson distribution

$$Y' \overset{d}{\sim} \mathrm{Poi}\left(j^*\right).$$

Then

$$\sum_{j\in J_t} T_t(j) = \exp\left(\widetilde{R}_0(\widehat{\zeta}_{t-1}) + j^\star + o(1)\right) \left(1 - \sum_{j\geqslant 0 : j\notin J_t} \mathbf{P}\left(Y' = j\right)\right). \tag{5.4.69}$$

By Poisson concentration (see, for example, combining [58, Remark 2.6] and [58, Theorem 2.1, Eq. (2.5), (2.6)]), and noting $\lambda_t = \Theta(n^c)$, we have

$$\mathbf{P}\left(|Y' - j^*| \geqslant \sqrt{j^*}\log j^*\right) \leqslant 2\exp\left(-\frac{j^* \log^2 j^*}{2(j^* + \sqrt{j^*}\log j^*)}\right) \leqslant \exp\left(-(1+o(1))\log^2 \lambda_t / 2\right) = o(1). \tag{5.4.70}$$

Combining (5.4.68), (5.4.69) and (5.4.70) yields

$$\sum_{j=0}^{\lfloor m_t \rfloor} T_t(j) = (1+o(1)) \sum_{j \in J_t} T_t(j) = \exp\left(\widetilde{R}_0(\widehat{\zeta}_{t-1}) + j^\star + o(1)\right). \tag{5.4.71}$$

Hence by noting the defining equation (5.4.63) of $\widetilde{R}_0(\widehat{\zeta}_{t-1})$, we set

$$\begin{aligned} P_{t-1,\varepsilon} &= R_0(\widehat{\zeta}_{t-1}) + \beta^\star \lambda_t \left(1 - \sum_{k=2}^{d_t+1} (k-1) R_k(\widehat{\zeta}_{t-1}) (\beta^\star)^{k-1}\right) \\ &= P_{t,\varepsilon}(\zeta(0)) + \lambda_t \left(\beta^\star - \sum_{k=2}^{d_t+1} (k-1) R_k(\widehat{\zeta}_{t-1}) (\beta^\star)^{k}\right), \end{aligned} \tag{5.4.72}$$

where we use (5.4.30) in the last step. Then we have, by (5.4.22), that

$$\mathbf{P}\left(\sum_{u \geqslant t} X_u = 0 \,\middle|\, \mathcal{J}_{t-1}(j_1, \ldots, j_{t-1})\right)^{-1} = \exp\left(P_{t-1,\varepsilon}(\widehat{\zeta}_{t-1}) + o\,(1)\right).$$

In view of (5.4.72), we have part (i) by recalling that $P_{t,\varepsilon}(\zeta(0)), \beta^\star$ and $R_k(\widehat{\zeta}_{t-1})$ are all in $\mathcal{P}(\mathbf{g}_{t-1})$. For part (ii), since the induction hypothesis applies to the first term of (5.4.72), it suffices to focus on the second term of (5.4.72). Recalling the discussion after (5.4.47), we have that $\beta^\star \in \mathcal{P}(\mathbf{g}_{t-1})$ and has all coefficients $O\left(n^{o(1)}x\right)$. By Claim 5.14 (i), the coefficients of all R_v are also $O\left(n^{o(1)}x\right)$. So the constant coefficient of $\beta^\star - \sum_{k=2}^{d_t+1}(k-1)R_k(\widehat{\zeta}_{t-1})(\beta^\star)^k$ is $1 + O\left(n^{o(1)}x\right)$, and all other coefficients are $O\left(n^{o(1)}x\right)$ by noting $\beta^\star \sim 1$.

Since each step is valid for all t-amenable $(j_1, \ldots, j_{t-1}, j_t)$, the expansions and derivations are, inductively, uniform over all appropriate $\widehat{\zeta}_{t-1}$, and $o(1)$ error does not depend on $j_1, \ldots, j_t$, therefore, the uniformity in part (iii) holds with t replaced by $t-1$. The inductive step is now fully established and this completes the proof of Theorem 5.12.

5.4.2 Proof of Theorem 5.1

We proceed under the assumption on p in (5.2.12), that is, $p = n^{-\chi-\varepsilon+o(1)}$, where ε is fixed. This constraint will be relaxed at the end of this section. By Corollary 5.13, we have

$$\mathbf{P}\,(X=0) = \exp\left(-P_{0,\varepsilon} + o(1)\right), \tag{5.4.73}$$

where $P_{0,\varepsilon}$ possesses the properties claimed in Theorem 5.12, in particular,

$$P_{0,\varepsilon} = \sum_{\ell=1}^{M_\varepsilon} c_{\varepsilon,\ell} n^{i_{\varepsilon,\ell}} p^{j_{\varepsilon,\ell}}, \quad \text{and} \quad P_{0,\varepsilon} = \left(1 + O\left(n^{o(1)}x\right)\right) \sum_{t \in \mathbf{S}} \lambda_t, \tag{5.4.74}$$

where $c_{\varepsilon,\ell}, i_{\varepsilon,\ell}$ and $j_{\varepsilon,\ell}$ are constants depending on ε.

We first show the positivity of i_ℓ, j_ℓ for all terms $n^{i_\ell}p^{j_\ell}$ in P_0, and moreover, it does not depend on ε. We consider the following cases:

- If $j_\ell < 0$, the term $n^{i_\ell}p^{j_\ell}$ cannot be included in P_0; otherwise, when p gets very small such that it does not satisfy (5.2.17), including term $n^{i_\ell}p^{j_\ell}$ does not give the correct asymptotic probability 1;

- If $j_\ell = 0$:
  - if $i_\ell \geqslant 0$, the term $n^{i_\ell} p^{j_\ell} = n^{i_\ell}$ cannot be included in P_0, for the same reason for the case $j_\ell < 0$;
  - if $i_\ell < 0$, the term $n^{i_\ell} p^{j_\ell} = n^{i_\ell} = o(1)$, therefore is dropped from P_0, and absorbed in error.
- If $j_\ell > 0$:
  - if $i_\ell \leqslant 0$, the term $n^{i_\ell} p^{j_\ell} = o(1)$, therefore is dropped from P_0, and absorbed in error;
  - if $i_\ell > 0$, the term $n^{i_\ell} p^{j_\ell}$ is included in P_0.

Next we show that different choices of ε for $P_{0,\varepsilon}$ simply determine different places to truncate a series that is independent of ε. Specifically, we follow the derivation of [98, Eq. (3.31)] and prove the following.

Lemma 5.17. *There exists a power series $P_0 = \sum_{\ell \geqslant 1} c_\ell n^{i_\ell} p^{j_\ell}$, where P_0 is independent of ε, and satisfies for any $\varepsilon > 0$,*

(i) $P_{0,\varepsilon}$ is a truncation of P_0 to a finite number of terms,

(ii) different orderings of types that are valid all lead to the same terms in $P_{0,\varepsilon}$.

Proof. We first deal with part (ii) and consider different orderings of types for some fixed ε. Note that the ordering of types in $\mathbf{S}$ determines the order of the inductive arguments concerning the conditional probabilities, and a given linear ordering leads to a unique set of terms in $P_{0,\varepsilon}$. Recall that the set of small types $\mathbf{S}$ admits a unique partial ordering π_0 such that (5.4.3) holds, and we first show that various linear extensions of π_0 all result in the same terms in $P_{t,\varepsilon}$.

For fixed ε and any two linear extensions π, π' of π_0, we have two corresponding polynomials $P_{0,\varepsilon}$ and $P'_{0,\varepsilon}$, respectively, and

$$\mathbf{P}\left(\sum_{u=1}^{s} X_u = 0 \,\middle|\, \mathcal{J}_0 \right) = \exp\left(-P_{0,\varepsilon} + o(1)\right) = \exp\left(-P'_{0,\varepsilon} + o(1)\right). \tag{5.4.75}$$

For any valid choice of $p = n^{-\chi-\varepsilon+o(1)}$, in view of (5.4.74), the terms in $P_{t,\varepsilon}$ and $P'_{t,\varepsilon}$ are

$$c_{\varepsilon,\ell} n^{i_{\varepsilon,\ell} - (\chi+\varepsilon) j_{\varepsilon,\ell} + o(1)} \quad \text{and} \quad c'_{\varepsilon,\ell} n^{i'_{\varepsilon,\ell} - (\chi+\varepsilon) j'_{\varepsilon,\ell} + o(1)},$$

respectively. Note a term $c_{\varepsilon,\ell} n^{i_{\varepsilon,\ell} - (\chi+\varepsilon) j_{\varepsilon,\ell} + o(1)}$ is $o(1)$ if $i_{\varepsilon,\ell} < (\chi+\varepsilon) j_{\varepsilon,\ell}$, and therefore, $P_{t,\varepsilon}$ only contain terms $c_{\varepsilon,\ell} n^{i_{\varepsilon,\ell}} p^{j_{\varepsilon,\ell}}$ with $i_{\varepsilon,\ell} \geqslant (\chi+\varepsilon) j_{\varepsilon,\ell}$. Similarly for $P'_{t,\varepsilon}$.

We have, from (5.4.75), that for all $p = n^{-\chi-\varepsilon+o(1)}$,

$$\operatorname{diff}(n,p) := P_{0,\varepsilon} - P'_{0,\varepsilon} = o(1), \tag{5.4.76}$$

which remains true if we choose $p_0 = C_0 n^{-\chi-\varepsilon}$ with any $C_0 > 0$, for which

$$\widetilde{\operatorname{diff}}(n, C_0) := \operatorname{diff}(n, C_0 n^{-\chi-\varepsilon}) = \sum_{\ell=1}^{M_\varepsilon} C_0^{j_\ell} a_\ell n^{i_\ell - (\chi+\varepsilon) j_\ell},$$

for some a_ℓ, i_ℓ, j_ℓ such that $i_\ell - (\chi+\varepsilon)j_\ell \geqslant 0$ for all ℓ. Then by (5.4.76), we conclude that $\widetilde{\mathrm{diff}}(n, C_0) = 0$ and two polynomials are equal.

Next we consider different choices of ε and prove part (i). To do so, it suffices to show that for all $0 < \varepsilon < \varepsilon'$, the polynomial $P_{0,\varepsilon'}$ is a truncation of $P_{0,\varepsilon}$; specifically, $i_{\varepsilon,\ell} = i_{\varepsilon',\ell}$, $j_{\varepsilon,\ell} = j_{\varepsilon',\ell}$, $c_{\varepsilon,\ell} = c_{\varepsilon',\ell}$ for all $\ell \in [M'_\varepsilon]$.

For any ε, the sets of types will be denoted explicitly by $\mathbf{S}(\varepsilon)$, $\mathbf{S}_1(\varepsilon)$ and $\mathbf{S}_2(\varepsilon)$. Recall the bound on λ_t in (5.2.15) and the definitions of $\mathbf{S}$, $\mathbf{S}_1$, and $\mathbf{S}_2$ in (5.2.14), (t1), and (t2) respectively, a type can move from $\mathbf{S}_1(\varepsilon)$ to $\mathbf{S}_2(\varepsilon')$, or from $\mathbf{S}_2(\varepsilon)$ to a large type for ε'. Therefore, we have $\mathbf{S}(\varepsilon') \subseteq \mathbf{S}(\varepsilon)$ for all $\varepsilon' > \varepsilon$.

As ε increases to ε', in view of the definition of $\mathbf{S}_2$ in (t2), and its dependence on ε, there is a finite number of values of ε such that the corresponding $\mathbf{S}_2$ are non-empty and are different. (The types in $\mathbf{S}_1$ are also different.) We list these special values in increasing order $\varepsilon_1 < \varepsilon_2 < \cdots$. We consider arbitrarily two consecutive $\varepsilon_i < \varepsilon_{i+1}$, and let ε be in open interval $(\varepsilon_i, \varepsilon_{i+1})$. We claim that

(c1) $P_{t,\varepsilon}$ is a truncation of P_{t,ε_i} for $t \in \mathbf{S}(\varepsilon)$;

(c2) $P_{t,\varepsilon_{i+1}}$ is a truncation of $P_{t,\varepsilon}$ for $t \in \mathbf{S}(\varepsilon_{i+1})$.

Then part (i) follows from these two claims.

For both claims, we need to show that the difference between the corresponding series, lies only in the terms that are absorbed by the error terms $o(1)$. We will analyse the inductive derivations of $P_{0,\varepsilon}$ for different ε. Since $\mathbf{S}(\varepsilon') \subseteq \mathbf{S}(\varepsilon)$ for all $\varepsilon' > \varepsilon$, and different valid orderings of types all lead to the same terms in the final formula, we assume, for any case, that the types in $\mathbf{S}(\varepsilon')$ have the same ordering for ε' as they do for ε for any $\varepsilon' > \varepsilon$.

- **Claim (c1)**:

  Recalling that $\{\varepsilon_i\}_i$ is the set of special values that types in $\mathbf{S}_2$ get changed. Since $\varepsilon_i < \varepsilon < \varepsilon_{i+1}$, we have ε is not a special value, and therefore $\mathbf{S}_2(\varepsilon) = \emptyset$. Hence we have $\mathbf{S}(\varepsilon) = \mathbf{S}_1(\varepsilon) = \mathbf{S}_1(\varepsilon_i)$. To show $P_{t,\varepsilon}$ is a truncation of P_{t,ε_i}, we consider the recursive derivation of P_{t,ε_i} and $P_{t,\varepsilon}$ in Theorem 5.12.

  - **For type $t' \in \mathbf{S}(\varepsilon_i) \setminus \mathbf{S}(\varepsilon)$**:

    We follow the described derivation of P_{t',ε_i} for ε_i. Note that $t' \notin \mathbf{S}(\varepsilon)$, we may simply set $P_{t',\varepsilon} = 0$, and it is easy to verify that $P_{t',\varepsilon_i} = o(1)$ when evaluated at the value of p occurring in the argument for ε. Therefore, we have the difference between $P_{t',\varepsilon}$ and $P_{t',\varepsilon'}$ is $o(1)$.

  - **For type $t \in \mathbf{S}(\varepsilon)$**:

    The rest of the derivation involves the same set of types with the same ordering, and they are all in $\mathbf{S}_1$. The argument for ε is the same as the argument for ε_i except that whenever an expansion occurs, different terms may be truncated. In particular, in view of the definition of truncation point ℓ_0 in (5.4.24), since $\varepsilon > \varepsilon_i$, we need less terms when expanding $\log \overline{\gamma}_{t,\varepsilon}(f_{t,i})$ in (5.4.26) to achieve larger error

$$O\left(\frac{n^{o(1)}x}{\lambda_t}\right) = O\left(n^{-\nu_t+\mu_t\chi+(\mu_t-1)\varepsilon+o(1)}\right),$$

which is sufficient for the derivation for ε. Suppose instead that we retain all terms that are significant for either ε_i. The extra retained terms are dominated by the error term, and therefore all fall into the error terms during all the remaining inductive steps in the argument involving series expansions. In particular, the ratio estimate (5.4.40) remains true for ε because of the assertion about the truncations in Theorem 5.10, as $\Gamma_{t,\varepsilon}$ and Γ_{t,ε_i} are both obtained by truncating series Γ_t with Γ_{t,ε_i} possibly including more terms as $\varepsilon_i < \varepsilon$. Then the series for ε_i in the exponent of $T_t(j)$ in (5.4.29) contains extra terms in comparison with that for ε.

Since the ratio estimate (and those following it) holds with extra terms, the iterative scheme described before (5.4.47) leads to estimates of $\beta^\star$ and $j^\star$ with extra error terms. By noting that the recursive defining formula (5.4.72) depends on $\beta^\star$ and $R_k(\widehat{\zeta}_{t-1})$ in (5.4.29), we conclude that $P_{t',\varepsilon}$ and $P_{t',\varepsilon'}$ differ by $o(1)$.

- **Claim (c2)**:

  Next, we will show $P_{t,\varepsilon_{i+1}}$ is a truncation of $P_{t,\varepsilon}$. Since $\varepsilon_{i+1} > \varepsilon$, then $\mathbf{S}(\varepsilon_{i+1}) \subseteq \mathbf{S}(\varepsilon)$. Moreover, if $t \in \mathbf{S}(\varepsilon)$, we have $t \in \mathbf{S}(\varepsilon_{i+1})$ by noting that ε_{i+1} is one of the special values. The reason is that for any $t \in \mathbf{S}(\varepsilon)$, it is possible that $t \in \mathbf{S}_2(\varepsilon_{i+1})$, but it cannot be large for ε_{i+1}, since ε_{i+1} is the next special value by definition. Therefore we conclude $\mathbf{S}(\varepsilon) = \mathbf{S}(\varepsilon_{i+1})$.

  It then remains to show that $P_{t,\varepsilon_{i+1}}$ equals $P_{t,\varepsilon}$ except for any terms of $P_{t,\varepsilon}$ that are $o(1)$ for ε_{i+1}. There may be a type t is in $\mathbf{S}_1(\varepsilon)$ but in $\mathbf{S}_2(\varepsilon_{i+1})$, that is $t \in (\mathbf{S}_2(\varepsilon_{i+1}) \cap \mathbf{S}_1(\varepsilon))$. For any such type, we claim that it is a maximal types in $\mathbf{S}_1(\varepsilon)$; otherwise, they cannot be in $\mathbf{S}_2(\varepsilon_{i+1})$. Therefore, its contribution to $P_{t,\varepsilon}$ is $\lambda_t + o(1)$. When p is taken in the appropriate range for ε_{i+1}, as $t \in \mathbf{S}_2(\varepsilon_{i+1})$, then we have its contribution to $P_{t,\varepsilon_{i+1}}$ is $\lambda_t = n^{o(1)}$. Therefore, the difference between $P_{t,\varepsilon}$ and $P_{t,\varepsilon_{i+1}}$ is $o(1)$.

  The rest of the argument for this case only involves $t \in \mathbf{S}_1(\varepsilon_{i+1})$, so is similar to the argument above.

This completes the proof of Lemma 5.17. □

By Lemma 5.17, the series $P_{0,\varepsilon}$ in (5.4.74) is a truncation of the power series P_0 to a finite number of terms. Since there is a bounded number of terms in (5.1.6) that are $o(1)$ for a given ε, we have now established (5.1.6) for this power series and for $p = n^{-\chi-\varepsilon+o(1)}$, or equivalently, we have

$$g(n,p) := \log \mathbf{P}\left(X = 0\right) - \sum_{\ell=1}^{M_\varepsilon} c_\ell n^{i_\ell} p^{j_\ell}$$

satisfies $g(n,p) \to 0$ as $n \to \infty$ for all p of the form $p = n^{-\chi-\varepsilon+o(1)}$ with some fixed $\varepsilon > 0$.

Next we need to show that this holds uniformly for $p = O\left(n^{-\chi-\varepsilon}\right)$ with any $\varepsilon > 0$. To do so, we will use the following lemma from [98].

Lemma 5.18. *[98, Lemma 1.4] For a closed interval $[a,b]$, suppose that $g(n,p)$ is a function such that $g(n,p) \to 0$ as $n \to \infty$ for all p of the form $p = n^{-\kappa+o(1)}$ when $\kappa \in [a,b]$ is fixed. Then $g(n,p) \to 0$ uniformly for all $p(n)$ satisfying $p(n) = n^{-\kappa(n)}$ with $\kappa(n) \in [a,b]$ for all n.*

By setting $a = \chi + \varepsilon$ and $b = r - \varepsilon''$ in above lemma, where ε'' is in (5.2.17), we have $g(n,p) = o(1)$ uniformly for all p such that $n^{-r+\varepsilon''} \leqslant p \leqslant n^{-\chi-\varepsilon}$. The remaining case of smaller p, that is $p \leqslant n^{-r+\varepsilon''}$, is discussed in the justification of assumption (5.2.17). The proof of Theorem 5.1.6 is complete.

5.5 Non-existence of subhypergraphs in $H_r(n,m)$

We will show that the $H_r(n,p)$ case can be extended to get a similar result for $H_r(n,m)$. The argument here is more or less the same as that in Section 4 of [98]. For convenience, the asymptotics are expressed in terms of n and the parameter $d = m/N$, where $N := \binom{n}{r}$. We employ the $H_r(n,p)$ case inside the proof, for a value of p that is close to d, with some adjustment that maximises the probability of having m edges with $X = 0$.

Let Y denote the number of edges of a hypergraph. The probability that $X = 0$ in $H_r(n,m)$ is exactly $\mathbf{P}(X = 0 \mid Y = m)$ in $H_r(n,p)$. For the rest of the section, we estimate this quantity, with all probabilities referring to $H_r(n,p)$ for some p that will be expressed using d. By Bayes' theorem, for all $0 < p < 1$,

$$\mathbf{P}(X = 0 \mid Y = m) = \mathbf{P}(Y = m \mid X = 0)\,\frac{\mathbf{P}(X=0)}{\mathbf{P}(Y=m)}. \tag{5.5.1}$$

The value of p we will use, which is specified below, is asymptotic to d and lies in the range required for the $H_r(n,p)$ case of Theorem 5.1 with the same restrictions on ε, which determines the set of small clusters $\mathbf{S}$. Then Theorem 5.1 estimates $\mathbf{P}(X = 0)$ in $H_r(n,p)$, and the remaining task is to compute the other two factors on the right hand side of (5.5.1).

We compute $\mathbf{P}(Y = m \mid X = 0)$ by considering the $\mathcal{R}^*$-clustering in $H_r(n,p)$. Recall that this is obtained by adding to $\mathbf{S}$ the type t^* (denoted by 0) of maximal cluster corresponding to a single hyperedge. We will first obtain an estimate of p using the approximation of the correction factor for the single edge cluster.

Let $\widehat{\mathbf{g}}(j) = (\widehat{g}_0(j), \ldots, \widehat{g}_s(j))$ be defined by

$$\widehat{g}_0(j) = j/\lambda_0$$

and $\widehat{g}_i(j) = 0$ for $i \geqslant 1$. Considering the power series $\Gamma_{0,\varepsilon}(n,p,\mathbf{g})$ evaluated at $\mathbf{g} = \widehat{\mathbf{g}}(j)$, we have, by Theorem 5.10 (b1), that for $j/\lambda_0 \leqslant 3$,

$$\Gamma_{0,\varepsilon}(n,p,\widehat{\mathbf{g}}(j)) = \gamma(j\delta_0, 0) + O\left(\frac{x+n^{o(1)}\lambda_{\mathcal{L}}}{\lambda_0}\right), \tag{5.5.2}$$

by recalling that $\overline{\gamma}_{0,\varepsilon}(i\delta_0) := \Gamma_{0,\varepsilon}(n,p,\widehat{\mathbf{g}}(i))$. Also define $\widetilde{\mathbf{g}}(m)$ by

$$\widetilde{g}_0(m) = m/\lambda_0$$

and $\widetilde{g}_i(m) = 0$ for $i \geqslant 1$. Then by Theorem 5.10 (b2), in view of the differences between $\widehat{g}_0(j)$ and $\widetilde{g}_0(m)$, we also have

$$\Gamma_{0,\varepsilon}\big(n,p,\widehat{\mathbf{g}}(j)\big) = \Gamma_{0,\varepsilon}(n,p,\widetilde{\mathbf{g}}(m)) + O\left(\frac{x}{\lambda_0}(m-j)\right). \tag{5.5.3}$$

For brevity, we write

$$\overline{\gamma} = \Gamma_{0,\varepsilon}(n,p,\widetilde{\mathbf{g}}(m)). \tag{5.5.4}$$

Combining (5.5.2) and (5.5.3) gives

$$\gamma(j\delta_0, 0) = \overline{\gamma} + O\left(\frac{n^{o(1)}\lambda_{\mathcal{L}}}{\lambda_0} + \frac{x}{\lambda_0}(m-j)\right). \tag{5.5.5}$$

We will use $\overline{\gamma}$ as an estimate of $\gamma(j\delta_0, 0)$ for j in the range under consideration.

The probability that $H_r(n,p)$ has no copies of any hypergraph in $\mathcal{R}$ and m' edges will be maximised, given p, at $m' \approx m$ provided that

$$\frac{\mathbf{P}((m+1)\delta_0)}{\mathbf{P}(m\delta_0)} \approx 1.$$

In view of the definition of correction factor $\gamma(f,t)$ in (5.2.4), and noting $\lambda_0 = pN$, this is equivalent to

$$\gamma(m\delta_0, 0) = \frac{m}{\lambda_0} \cdot \frac{\mathbf{P}((m+1)\delta_0)}{\mathbf{P}(m\delta_0)} \approx \frac{m}{\lambda_0} = \frac{d}{p}.$$

Consequently, we define p by

$$p = d/\overline{\gamma}. \tag{5.5.6}$$

One can interpret $d/\overline{\gamma}$ as the "maximum likelihood estimation" of the parameter p given m edges with no copies of any hypergraph in $\mathcal{R}$.

By Theorem 5.10 (b1), we have

$$\left|\Gamma_{0,\varepsilon}(n,p,\widetilde{\mathbf{g}}(m)) - \gamma(\widetilde{f}, 0)\right| = O\left(\frac{x+n^{o(1)}\lambda_{\mathcal{L}}}{\lambda_0}\right),$$

where the cluster count $\widetilde{f} := (m, 0, \ldots, 0)$. We also have $\gamma(\widetilde{f}, 0) = 1 + O\left(n^{o(1)}x\right)$ by Claim 5.8. Therefore,

$$\overline{\gamma} = \Gamma_{0,\varepsilon}(n,p,\widetilde{\mathbf{g}}(m)) = 1 + O\left(n^{o(1)}x + \frac{x+n^{o(1)}\lambda_{\mathcal{L}}}{\lambda_0}\right),$$

and moreover, by (5.5.6),

$$p = d\left(1 + O\left(n^{o(1)}x + \frac{x+n^{o(1)}\lambda_{\mathcal{L}}}{\lambda_0}\right)\right). \tag{5.5.7}$$

By (5.5.4), $\overline{\gamma}$ is a function of n and p, in the form of some power series. The estimate of p, in view of (5.5.6) and (5.5.7), and noting the constraint on x in (5.1.4), be can obtained via repeated substitutions in the form of a power series in n and d as

$$p = d\widetilde{J}_1\left(1 + o\left(\frac{1}{\lambda_0}\right)\right), \tag{5.5.8}$$

where $\widetilde{J}_1$ is the truncation of some power series J_1 in n and d to significant terms that is of a finite number, and J_1 is independent of ε.

Now we are ready to approximate $\mathbf{P}(Y = m \mid X = 0)$.

Claim 5.19.

$$\mathbf{P}(Y = m \mid X = 0) \sim \frac{1}{\sqrt{2\pi m}}. \tag{5.5.9}$$

Proof. The number Y of edges in $H_r(n,p)$ is distributed as $\mathrm{Bin}(N,p)$ where $N = \binom{n}{r}$, with mean

$\lambda_0 = Np \sim m$. Therefore, by the concentration of binomial variables [58, Theorem 2.1], we have

$$\mathbf{P}\left(|Y-m| \geqslant 3m\right) \leqslant \exp\left(-c_0 m\right),$$

for some constant $c_0 > 0$. From the $H_r(n,p)$ case of Theorem 5.1, we have $\mathbf{P}\left(X=0\right)$ in $H_r(n,p)$ is $e^{-\Theta(\lambda_1)}$ with $\lambda_1 = o(\lambda_0)$. It follows that

$$\mathbf{P}\left(X=0\right) = \sum_{j\geqslant 0} \mathbf{P}\left(X=0, Y=j\right) = \sum_{j\geqslant 0} \mathbf{P}\left(j\delta_0\right) = (1+o(1)) \sum_{j=0}^{3m} \mathbf{P}\left(j\delta_0\right). \tag{5.5.10}$$

Also by (5.5.5), we have

$$\frac{\mathbf{P}((j+1)\delta_0)}{\mathbf{P}(j\delta_0)} = \frac{\lambda_0}{j+1}\gamma(j\delta_0, 0) = \frac{m}{j+1}\left(1 + O\left(\frac{n^{o(1)}x}{\lambda_0}\right) + O\left(\frac{x(m-j)}{\lambda_0}\right)\right), \tag{5.5.11}$$

where the last step is by noting $\lambda_0 = pN$ and $\overline{\gamma}pN = dN = m$. Hence (5.5.10) gives

$$\frac{\mathbf{P}(X=0)}{\mathbf{P}(Y=m, X=0)} \sim \sum_{j=0}^{3m} \frac{\mathbf{P}(j\delta_0)}{\mathbf{P}(m\delta_0)} = \frac{\mathbf{P}(0\delta_0)}{\mathbf{P}(m\delta_0)} \sum_{j=0}^{3m} \frac{\mathbf{P}(j\delta_0)}{\mathbf{P}(0\delta_0)}. \tag{5.5.12}$$

Recall the definition of $T_{t^\star}(\cdot)$ in (5.4.35), we have

$$T_0(j) = \frac{\lambda_0^j}{j!}\prod_{i=0}^{j-1}\overline{\gamma}_{0,\varepsilon}(i\delta_0) = \frac{m^j}{j!}\prod_{i=0}^{j-1}\left(1 + O\left(\frac{n^{o(1)}x}{\lambda_0}\right) + O\left(\frac{x(m-j)}{\lambda_0}\right)\right), \tag{5.5.13}$$

with the second equality is by noting (5.5.5). Therefore, rewriting (5.5.12) using the correction factor defined by (5.2.4), gives that

$$\frac{\mathbf{P}(X=0)}{\mathbf{P}(Y=m, X=0)} \sim \frac{1}{T_0(m)} \sum_{j=0}^{3m} T_0(j). \tag{5.5.14}$$

We now can deduce, from (5.5.13), that

$$\frac{T_0(m)}{T_0(m-1)} = 1 + O\left(\frac{n^{o(1)}x}{\lambda_0}\right).$$

From Lemma 5.15 and its proof, we have that $j^\star = m + o(1)$, and

$$T_0(j) = \exp\left(-\frac{(j-m)^2}{2m} + o(1)\right) T_0(m).$$

for $j = m + O\left(\sqrt{m}\log m\right)$. Thus, in view of (5.5.14),

$$\frac{\mathbf{P}(X=0)}{\mathbf{P}(Y=m, X=0)} \sim \sum_{|j-m| = O\left(\sqrt{m}\log m\right)} \exp\left(-\frac{(j-m)^2}{2m}\right) \sim \sqrt{2\pi m}.$$

Taking the reciprocal completes the proof. $\square$

Now we have

$$\mathbf{P}\left(X=0 \mid Y=m\right) \sim \frac{1}{\sqrt{2\pi m}} \frac{\mathbf{P}(X=0)}{\mathbf{P}(Y=m)}, \tag{5.5.15}$$

and it remains to estimate probabilities $\mathbf{P}\left(X=0\right)$ and $\mathbf{P}\left(Y=m\right)$. We substitute p using (5.5.8) into the polynomial obtained by truncating the power series for $\log \mathbf{P}\left(X=0\right)$ from the $H_r(n,p)$ case of Theorem 5.1, at an appropriate level, to obtain

$$\log \mathbf{P}\left(X=0\right) = \widetilde{J}_2 + o(1), \tag{5.5.16}$$

where $\widetilde{J}_2$ is a truncation of a power series J_2 in n and d, with J_2 independent of ε.

The remaining factor $\mathbf{P}\left(Y=m\right)$ is simply the binomial probability that can be estimated using Stirling's formula. Note that $m = dN = o\left(N\right)$ since $d \sim p$ and $p = O\left(n^{-\varepsilon}\right)$, then by Stirling's formula,

$$\begin{aligned}
\mathbf{P}\left(Y=m\right) = \mathbf{P}\left(Y=dN\right) &= \binom{N}{dN} p^{dN}(1-p)^{(1-d)N} \\
&\sim \frac{1}{\sqrt{2\pi dN}} \frac{N^N}{(dN)^{dN}((1-d)N)^{(1-d)N}} p^{dN}(1-p)^{(1-d)N} \qquad (5.5.17)\\
&= \frac{1}{\sqrt{2\pi m}} \left(\frac{p}{d}\right)^{dN} \left(\frac{1-p}{1-d}\right)^{(1-d)N} \\
&\sim \frac{1}{\sqrt{2\pi m}} \exp\left(dN \log \widetilde{J}_1 + (1-d)N \log\left(1 - \frac{d(\widetilde{J}_1-1)}{1-d}\right)\right), \qquad (5.5.18)
\end{aligned}$$

with factor $1/\sqrt{2\pi m}$ cancelling with $\mathbf{P}\left(Y=m \mid X=0\right)$ obtained above in Claim 5.19.

Then plugging (5.5.16) and (5.5.18) in (5.5.15) yeilds

$$\mathbf{P}\left(X=0 \mid Y=m\right) \sim \exp\left(\widetilde{J}_2 - d\binom{n}{r}\log\left(\widetilde{J}_1\right) - \binom{n}{r}(1-d)\log\left(1-\frac{d(\widetilde{J}_1-1)}{1-d}\right)\right). \tag{5.5.19}$$

Further expanding gives a power series in n and d, by noting $\widetilde{J}_1 - 1 = O\left(n^{o(1)}x + \lambda_0^{-1}\right)$ from (5.5.7). The positivity of the exponents i_ℓ and j_ℓ follows by arguing as in the proof of the $H_r(n,p)$ case.

5.6 Computation of the asymptotic probability that $H_3(n,p)$ is linear

To obtain the probability of a random hyprgraph being linear, recall that the set $\mathcal{H}_0$ contains 'forbidden' hypergraphs with hyperedge pair intersecting in more than one vertex; therefore, we set $\mathcal{R} = \mathcal{H}_0$ and let $\mathcal{R}^*$ be the union of $\mathcal{R}$ and the single edge cluster for this and the next section.

The goal of this section is to compute the terms in Theorem 5.1 explicitly to prove Theorem 5.3. To obtain an explicit formula, we need to first obtain estimates of the correction factors, and these need to be determined to a required accuracy, along with the quantities $c(u,t,h)$ appearing in the iterative approximation (5.3.1).

5.6.1 Clusters and correction factors

For 3-graphs, the only forbidden hypergraph is on four vertices with two hyperedges sharing two vertices and we have $\chi = 1$ according to (5.1.1). For $p = O\left(n^{-7/5-\varepsilon}\right)$, we have, by (5.1.4), that $x = np = O\left(n^{-2/5-\varepsilon}\right)$.

When $n^7p^5 \to 0$ and $n^6p^4 \to \infty$, simple computations reveal that $\mathbf{S} = [8]$ and $\mathbf{S}$ consists of the eight clusters listed in Figure 5.1. Note that for smaller p, we may have less types. All other cluster types have expected number tending to 0 asymptotically, and are therefore are not in $\mathbf{S}$. Note that the types $\{4,5,6,7,8\}$ are all maximal, and therefore not comparable in the poset ordering $\prec$ on $\mathbf{S}$. Nevertheless, we have that $\lambda_{t+1} < n^{o(1)}\lambda_t$ for all $t < 8$.

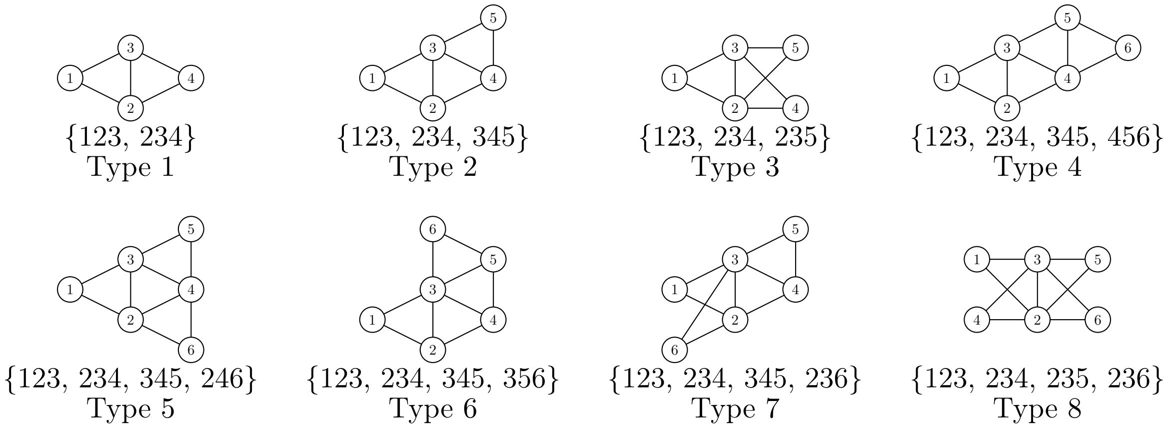


Figure 5.1: Types of possible small clusters for $\mathcal{H}_0$-clustering in $H_3(n,p)$ when $p = o\left(n^{-7/5}\right)$. For each type, the hyperedges are listed below each diagram.

We first calculate λ_t for $t \in \mathbf{S}$, as listed in Table 5.1.

$\lambda_1 = \dfrac{[n]_4p^2}{4}$	$\lambda_2 = \dfrac{[n]_5p^3}{2}$	$\lambda_3 = \dfrac{[n]_5p^3}{12}$	$\lambda_4 = \dfrac{[n]_6p^4}{2}$
$\lambda_5 = \dfrac{[n]_6p^4}{6}$	$\lambda_6 = \dfrac{[n]_6p^4}{2}$	$\lambda_7 = \dfrac{[n]_6p^4}{2}$	$\lambda_8 = \dfrac{[n]_6p^4}{48}$

Table 5.1: Expected numbers of small clusters.

Our next task is to find the polynomial $\Gamma_{t,\varepsilon}$ defined in Theorem 5.10, iteratively for all $t \in \mathbf{S}$. In order to do this, we first list all non-negligible terms in summation (5.3.3) for computing Γ_t in Table 5.2, where the "cofactor" column of Table 5.2 shows the non-negligible part of

$$\frac{\lambda_u}{\lambda_t}\Gamma_u \prod_{i=1}^{k} \frac{g_{t_i}}{\Gamma_{t_i}}.$$

The calculations get simplified by omitting terms in (5.3.3) that have coefficients of variables g_{t_i} being $O(n^{-\varepsilon_0}/\lambda_t)$ for some $\varepsilon_0 > 0$, and therefore, h is quite restricted. This is because in when evaluating $\overline{\gamma}_{t,\varepsilon}(f)$, each g_{t_i} is assigned a value that is $n^{o(1)}$, and hence the dropped terms are subsumed into the error term in (5.3.27), recalling that $\overline{\gamma}_t(f) = 1 + O(x)$ by Lemma 5.7. In particular,

- We omit any $O(p)$ term of the expansion of $1/c(t,t,\mathbf{0})$ in (5.3.2), since for all $t \in \mathbf{S}$, we have $p\lambda_t = O(p\lambda_1) = O(n^4p^3)$. Hence, we can treat $c(1,1,\mathbf{0})$ as 1; this also applies for $c(t,t,\mathbf{0})$ for all other $t \in \mathbf{S}$.

- We omit terms inside the summation in (5.3.2) with factor $c(t,t,h)$ for any non-zero $h \in \mathcal{F}$ and for all t, since $c(t,t,h) = O(p)$.

u	t	h	$c(u,t,h)$	cofactor
2	1	$\mathbf{0}$	2	$2np\Gamma_2$
3	1	$\mathbf{0}$	3	$\frac{1}{3}np\Gamma_3$
4	1	$\mathbf{0}$	1	$2n^2p^2$
4	1	δ_1	2	$2n^2p^2\Gamma_4 \cdot g_1\Gamma_1^{-1}$
5	1	$\mathbf{0}$	3	$\frac{2}{3}n^2p^2$
6	1	$\mathbf{0}$	1	$2n^2p^2$
6	1	δ_1	2	$2n^2p^2\Gamma_6 \cdot g_1\Gamma_1^{-1}$
7	1	$\mathbf{0}$	2	$2n^2p^2$
7	1	δ_1	2	$2n^2p^2\Gamma_7 \cdot g_1\Gamma_1^{-1}$
8	1	δ_1	6	$\frac{1}{12}n^2p^2 \cdot g_1\Gamma_1^{-1}$
4	2	$\mathbf{0}$	2	np
5	2	$\mathbf{0}$	3	$\frac{1}{3}np$
6	2	$\mathbf{0}$	2	np
7	2	$\mathbf{0}$	2	np
7	3	$\mathbf{0}$	1	$6np$
8	3	$\mathbf{0}$	4	$\frac{1}{4}np$

Table 5.2: Contribution to the iteration (5.3.3) for the $\mathcal{H}_0$-clustering.

- We omit contributions from $t \geqslant 4$, since all such t are maximal in $\mathbf{S}$, and $c(u,t,h) = 0$ unless $t \prec u$ or $u = t$.
- Any other factors which appear to be missing have simply been replaced by 1.

Note that the above omissions and simplifications are not necessary, one could include all relevant terms and factors in iterations and obtain the same series in the end with the required accuracy.

Using the above values, we obtain the iterative formulae

$$\Gamma_{1,\varepsilon}^{(r+1)} = 1 - 4np\Gamma_{2,\varepsilon}^{(r)} - np\Gamma_{3,\varepsilon}^{(r)} - 10n^2p^2 - 4n^2p^2\frac{\Gamma_{4,\varepsilon}^{(r)}}{\Gamma_1^{(r)}}g_1 - 4n^2p^2\frac{\Gamma_{6,\varepsilon}^{(r)}}{\Gamma_1^{(r)}}g_1 - 4n^2p^2\frac{\Gamma_{7,\varepsilon}^{(r)}}{\Gamma_1^{(r)}}g_1 - \frac{1}{2}n^2p^2\frac{\Gamma_{8,\varepsilon}^{(r)}}{\Gamma_1^{(r)}}g_1,$$

$$\Gamma_{2,\varepsilon} = 1 - 7np, \tag{5.6.1}$$

$$\Gamma_{3,\varepsilon} = 1 - 7np, \tag{5.6.2}$$

$$\Gamma_{t,\varepsilon} = 1 \quad (t \geqslant 4). \tag{5.6.3}$$

Substitution of $\{\Gamma_{t,\varepsilon}\}_{t\geqslant 2}$ to $\Gamma_{1,\varepsilon}$ gives that

$$\Gamma_{1,\varepsilon}^{(1)} = 1 - 5np(1-7np) - 10n^2p^2 - \frac{25}{2}n^2p^2g_1 = 1 - 5np + 25n^2p^2 - \frac{25}{2}n^2p^2g_1.$$

Further iterations give $\Gamma_{1,\varepsilon}^{(2)} = \Gamma_{1,\varepsilon}^{(1)}$, and therefore we have

$$\Gamma_{1,\varepsilon} = 1 - 5np + 25n^2p^2 - \frac{25}{2}n^2p^2g_1. \tag{5.6.4}$$

We will evaluate the expressions given in Section 5.4 for p such that $n^7p^5 \to 0$ and $n^6p^4 \to \infty$, so that $\mathbf{S}_1 = [8]$ and $\mathbf{S}_2 = \emptyset$. The ultimate result will then be valid for all values of p such that $n^7p^5 \to 0$ by Lemma 5.17. We also fix ε in the range $0 < \varepsilon < 7/5 - 1 = 2/5$. With p and ε in these ranges, the

$\Gamma_{t,\varepsilon}$ are given by the expressions (5.6.1)–(5.6.4).

5.6.2 Conditional probabilities

The recursive approximation of conditional probabilities for $t \leqslant 8$ starts with setting $P_{8,\varepsilon} = 0$ in (5.4.11). The next step is to determine $P_{7,\varepsilon}$. Now we have $j^* = \lambda_7$ and we conclude that $\mathbf{P}\left(X_8 = 0 \,\middle|\, \mathcal{J}_7(j_1,\ldots,j_7)\right)^{-1} \sim e^{\lambda_8}$ and then $P_{7,\varepsilon} = \lambda_8$. Similarly, one can show that

$$\mathbf{P}\left(\sum_{u>t} X_u = 0 \,\middle|\, \mathcal{J}_t(j_1,\ldots,j_t)\right)^{-1} \sim \exp\left(\sum_{u=t+1}^{8} \lambda_u\right)$$

for $t \in \{6,5,4,3\}$. In particular we have $\mathbf{P}\left(\sum_{u>3} X_u = 0 \,\middle|\, \mathcal{J}_3(j_1,j_2,j_3)\right)^{-1} \sim \exp\left(P_{3,\varepsilon}\right)$ with $P_{3,\varepsilon} = \sum_{u=4}^{8} \lambda_u$.

5.6.2.1 Summation for type 3 and 2

For type 3, plugging the (5.6.2) into the summation (5.4.21) gives

$$\begin{aligned}\sum_{j=0}^{\lfloor m_3 \rfloor} \frac{\lambda_3^j}{j!} \exp\left(\sum_{i=4}^{8} \lambda_i\right)(1-7np)^j &= \exp\left(\sum_{i=4}^{8} \lambda_i\right) \sum_{j=0}^{\lfloor m_t \rfloor} \frac{\lambda_3^j}{j!} \exp\left(j \log\left(1-7np\right)\right) \\ &= \exp\left(\sum_{i=4}^{8} \lambda_i\right) \sum_{j=0}^{\lfloor m_t \rfloor} \frac{\lambda_3^j}{j!} \exp\left(-7npj + O\left(n^2p^2j\right)\right),\end{aligned}$$

where the last step is by expanding the logarithm and noting $np = o(1)$. Since $0 \leqslant j \leqslant 3\lambda_3$, we have $n^2p^2j = O\left(n^7p^5\right) = o\left(1\right)$, and

$$\begin{aligned}\sum_{j=0}^{\lfloor m_3 \rfloor} \frac{\lambda_3^j}{j!} \exp\left(\sum_{i=4}^{8} \lambda_i\right)(1-7np)^j &= \exp\left(\sum_{i=4}^{8} \lambda_i + o\left(1\right)\right) \sum_{j=0}^{\lfloor m_t \rfloor} \frac{(e^{-7np}\lambda_3)^j}{j!} \\ &= \exp\left(\sum_{i=4}^{8} \lambda_i + e^{-7np}\lambda_3 + o\left(1\right)\right) = \exp\left(\sum_{i=4}^{8} \lambda_i + (1-7np)\lambda_3 + o\left(1\right)\right),\end{aligned}$$

by noting $n^2p^2\lambda_3 = o(1)$. Thus we have $\mathbf{P}\left(\sum_{u>2} X_u = 0 \,\middle|\, \mathcal{J}_2(j_1,j_2)\right)^{-1} \sim \exp\left(P_{2,\varepsilon}\right)$ with $P_{2,\varepsilon} = (1-7np)\lambda_3 + \sum_{u=4}^{8} \lambda_u$.

Similarly for type 2, we obtain

$$\sum_{j=0}^{\lfloor m_2 \rfloor} \frac{\lambda_2^j}{j!} \exp\left(\sum_{i=4}^{8} \lambda_i + (1-7np)\lambda_3\right)(1-7np)^j = \exp\left(\sum_{i=4}^{8} \lambda_i + (1-7np)(\lambda_3+\lambda_2) + o\left(1\right)\right),$$

and $\mathbf{P}\left(\sum_{u>1} X_u = 0 \,\middle|\, \mathcal{J}_1(j_1)\right)^{-1} \sim \exp\left(P_{1,\varepsilon}\right)$ with $P_{1,\varepsilon} = (1-7np)(\lambda_3+\lambda_2) + \sum_{u=4}^{8} \lambda_u$.

5.6.2.2 Summation for type 1

For type 1, note that g_1 does not appear in $P_{1,\varepsilon}$, we have

$$\sum_{j=0}^{\lfloor m_1 \rfloor} T_1(j) = \sum_{j=0}^{\lfloor m_1 \rfloor} \frac{\lambda_1^j}{j!} \exp\left(P_{1,\varepsilon}\right) \prod_{i=0}^{j-1} \left(1 - 5np + 25n^2p^2 - \frac{25}{2} n^2p^2 g_1\right) \tag{5.6.5}$$

$$= \exp\left(P_{1,\varepsilon}\right) \sum_{j=0}^{\lfloor m_1 \rfloor} \frac{\lambda_1^j}{j!} \exp\left(\sum_{i=0}^{j-1} \log\left(1 - 5np + 25n^2p^2 - \frac{25}{2} n^2p^2 \frac{i}{\lambda_1}\right)\right). \tag{5.6.6}$$

For $0 \leqslant i \leqslant 3\lambda_1$, expanding logarithm gives that

$$\log\left(1 - 5np + 25n^2p^2 - \frac{25}{2} n^2p^2 \frac{i}{\lambda_1}\right) = -5np + \frac{25}{2} n^2p^2 - \frac{25}{2} n^2p^2 \frac{i}{\lambda_1} + O\left(n^3p^3\right).$$

Then we obtain

$$T_1(j) = \frac{\lambda_1^j}{j!} \exp\left(P_{1,\varepsilon} + \left(-5np + \frac{25}{2} n^2p^2\right) j - \frac{25}{4} n^2p^2 \frac{j^2}{\lambda_1} + o\left(1\right)\right). \tag{5.6.7}$$

We also have the ratio from (5.6.5),

$$\frac{T_1(j)}{T_1(j-1)} = \left(1 - 5np + 25n^2p^2 - \frac{25}{2} n^2p^2 \frac{j}{\lambda_1}\right) \frac{\lambda_1}{j}.$$

This leads to the equation for the maximum term

$$\frac{j^\star}{\lambda_1} = 1 - 5np + 25n^2p^2 - \frac{25}{2} n^2p^2 \frac{j^\star}{\lambda_1}.$$

Iterations give that

$$j^\star = \left(1 - 5np + \frac{25}{2} n^2p^2\right) \lambda_1 + o\left(1\right). \tag{5.6.8}$$

In view of (5.6.7), we rewrite

$$P_{1,\varepsilon} + \left(-5np + \frac{25}{2} n^2p^2\right) j - \frac{25}{4} n^2p^2 \frac{j^2}{\lambda_1} = \widetilde{R}_0 + \widetilde{R}_1 j + \widetilde{R}_2 \frac{(j-j^\star)^2}{j^\star}, \tag{5.6.9}$$

where expanding (5.6.9) and comparing with (5.6.7) give

$$\widetilde{R}_0 + \widetilde{R}_2 j^\star + (\widetilde{R}_1 - 2\widetilde{R}_2) j + \widetilde{R}_2 \frac{j^2}{j^\star} = P_{1,\varepsilon} + \left(-5np + \frac{25}{2} n^2p^2\right) j - \frac{25}{4} n^2p^2 \frac{j^2}{\lambda_1}.$$

Recursively solving, in view of (5.6.8), yields

$$\widetilde{R}_2 = -\frac{25}{4} n^2p^2 \left(1 - 5np + \frac{25}{2} n^2p^2\right),$$

$$\widetilde{R}_1 = -5np + \frac{25}{2} n^2p^2 + 2\widetilde{R}_2 = -5np + \frac{25}{2} n^2p^2 \left(5np - \frac{25}{2} n^2p^2\right), \tag{5.6.10}$$

$$\widetilde{R}_0 = \frac{25}{4} n^2p^2 \left(1 - 5np + \frac{25}{2} n^2p^2\right)^2 \lambda_1 + P_{1,\varepsilon}. \tag{5.6.11}$$

Here we can also verify (5.4.65) by (5.6.8) and (5.6.10), that is,

$$\exp\left(\widetilde{R}_1(\widehat{\zeta}_{t-1})\right)\lambda_1 = \left(1-5np+\frac{25}{2}n^2p^2\right)\lambda_1 + o\,(1)\,.$$

Then we have the summation for type 1, following (5.4.71),

$$\sum_{j=0}^{\lfloor m_1 \rfloor} T_1(j) = \exp\left(\widetilde{R}_0 + j^\star + o\,(1)\right)$$
$$= \exp\left(P_{1,\varepsilon} + \frac{25}{4}n^2p^2\left(1-5np+\frac{25}{2}n^2p^2\right)^2\lambda_1 + \left(1-5np+\frac{25}{2}n^2p^2\right)\lambda_1 + o\,(1)\right) = \exp\left(P_{0,\varepsilon}+o\,(1)\right),$$

where

$$P_{0,\varepsilon} = P_{1,\varepsilon} + \left(1-5np+\frac{75}{4}n^2p^2\right)\lambda_1.$$

Plugging in the expectations in Table 5.1 yields the final asymptotic formula

$$\begin{aligned}
\mathbf{P}\,(X=0) &= \exp\left(-\sum_{i=4}^{8}\lambda_i - (1-7np)\lambda_3 - (1-7np)\lambda_2 - \left(1-5np+\frac{75}{4}n^2p^2\right)\lambda_1\right)\\
&= \exp\left(-\frac{81}{48}n^6p^4 - (1-7np)\frac{n^5p^3}{12} - (1-7np)\frac{n^5p^3}{2} - \left(1-5np+\frac{75}{4}n^2p^2\right)\frac{[n]_4p^2}{4}\right)\\
&= \exp\left(-\frac{1}{4}n^4p^2 + \frac{3}{2}n^3p^2 + \frac{2}{3}n^5p^3 - \frac{55}{24}n^6p^4\right).
\end{aligned}$$

We for now have the case when $p = O\left(n^{-7/11-\varepsilon}\right)$. To relax this to $p = o(n^{-7/11})$, we only need to note that, from this conclusion, all other terms in the series in Theorem 5.1.7 must have $i_\ell/j_\ell \leqslant 7/5$. Such terms tend to zero for $p = o(n^{-7/5})$, and this completes the proof of Theorem 5.3.

5.7 Computation of the asymptotic probability that $H_3(n,m)$ is linear

Here we extend the previous case to give the probability that $H_3(n,m)$ is linear by finding the asymptotics of $\mathbf{P}(X=0)$ and $\mathbf{P}(Y=m)$. This requires first obtaining the asymptotic expansion of $p = d/\overline{\gamma}$. Recall by (5.5.4), we have $\overline{\gamma} = \overline{\gamma}_{0,\varepsilon}(n,p,\widetilde{\mathbf{g}})$ with $\widetilde{g}_0 = m/\lambda_0$ and $\widetilde{g}_i = 0$ for $i \geqslant 1$.

Recall that we need $\mathcal{R}^*$-clustering with single hyperedge clusters of type 0. To calculate $\overline{\gamma}_0(n,p,\widetilde{\mathbf{g}})$, under the same assumption that $p = O(n^{-7/5-\varepsilon})$, we extend Table 5.2 by considering Table 5.3 for this new clustering.

We may also ignore and simplify certain terms in the product of $c(u,t,h)$ with its cofactor in (5.3.3). In particular,

- Note that $\lambda_0 = p\binom{n}{3}$, and in view of (5.5.5) and (5.5.6),

  $$\widetilde{g}_0 = m/\lambda_0 = d/p = \overline{\gamma} = \Gamma_{0,\varepsilon}(n,p,\widetilde{\mathbf{g}}) = \gamma(j\delta_0,0) + o(\lambda_0^{-1}).$$

  The factors $\widetilde{g}_0/\Gamma_0$ can be replaced by 1 in (5.3.3).

- We omit all terms of the form $c(u,0,\delta_u)$ for $u>0$, since $\Gamma_u = \Gamma_{t_i}$ and get cancelled in (5.3.3) and the factor g_u is 0 during evaluation.

u	t	h	$c(u,t,h)$	cofactor
0	0	δ_0	p	Γ_0
1	0	δ_0	2	$\frac{3}{2}np\Gamma_1 \cdot g_0\Gamma_0^{-1}$
2	0	$2\delta_0$	1	$3n^2p^2\Gamma_2 g_0^2\Gamma_0^{-2}$
5	0	$3\delta_0$	1	$n^3p^3g_0^3\Gamma_0^{-3}$
2	1	δ_0	2	$2np\Gamma_2 \cdot g_0\Gamma_0^{-1}$
3	1	δ_0	3	$\frac{1}{3}np\Gamma_3 \cdot g_0\Gamma_0^{-1}$
4	1	$2\delta_0$	1	$2n^2p^2\Gamma_4 \cdot g_0^2\Gamma_0^{-2}$
5	1	$2\delta_0$	3	$\frac{2}{3}n^2p^2\Gamma_5 \cdot g_0^2\Gamma_0^{-2}$
6	1	$2\delta_0$	1	$2n^2p^2\Gamma_6 \cdot g_0^2\Gamma_0^{-2}$
7	1	$2\delta_0$	2	$2n^2p^2\Gamma_7 \cdot g_0^2\Gamma_0^{-2}$
4	2	δ_0	2	$np \cdot g_0\Gamma_0^{-1}$
5	2	δ_0	3	$\frac{1}{3}np \cdot g_0\Gamma_0^{-1}$
6	2	δ_0	2	$np \cdot g_0\Gamma_0^{-1}$
7	2	δ_0	2	$np \cdot g_0\Gamma_0^{-1}$
7	3	δ_0	1	$6np \cdot g_0\Gamma_0^{-1}$
8	3	δ_0	4	$\frac{1}{4}np \cdot g_0\Gamma_0^{-1}$

Table 5.3: The contribution to the iteration (5.3.3) for the $\mathcal{H}_0^\star$-clustering.

Then, we obtain the iterative formulae, by noting from (5.2.1), that $c(0,0,\mathbf{0}) = 1-p$ and $c(1,1,\mathbf{0}) = (1-p)^2$,

$$\begin{aligned}
\Gamma_{0,\varepsilon}^{(r+1)} &= \frac{1}{1-p}\left(1 - p\Gamma_{0,\varepsilon}^{(r)} - 3np\frac{\Gamma_{1,\varepsilon}^{(r)}}{\Gamma_{0,\varepsilon}^{(r)}}g_0 - 3n^2p^2\frac{\Gamma_{2,\varepsilon}^{(r)}}{(\Gamma_{0,\varepsilon}^{(r)})^2}g_0^2 - n^3p^3\frac{1}{(\Gamma_{0,\varepsilon}^{(r)})^3}g_0^3\right)\\
\Gamma_{1,\varepsilon} &= \frac{1}{(1-p)^2}\left(1 - 4np\Gamma_{2,\varepsilon} - np\Gamma_{3,\varepsilon} - 10n^2p^2\right)\\
\Gamma_{2,\varepsilon} &= 1 - 7np\\
\Gamma_{3,\varepsilon} &= 1 - 7np\\
\Gamma_{t,\varepsilon} &= 1 \quad (t \geqslant 4).
\end{aligned}$$

Substitution of $\{\Gamma_{t,\varepsilon}\}_{t\geqslant 2}$ to $\Gamma_{1,\varepsilon}$, and further expansions and simplifications give that

$$\Gamma_{1,\varepsilon} = \frac{1}{(1-p)^2}(1 - 5np(1-7np) - 10n^2p^2) = (1 - 5np + 25n^2p^2)(1 + p + p^2 + p^3 + p^4)^2,$$

and similarly we have for type 0,

$$\Gamma_{0,\varepsilon}^{(r+1)} = (1 + p + p^2 + p^3)\left(1 - p\Gamma_{0,\varepsilon}^{(r)} - 3np\Gamma_{1,\varepsilon} - 3n^2p^2\Gamma_{2,\varepsilon} - n^3p^3\right).$$

Iterations lead to

$$\Gamma_{0,\varepsilon} = 1 - 3np + 12n^2p^2 - 55n^3p^3 + o\left(n^3p^3\right).$$

Then repeated substituting $p = d/\overline{\gamma}$ into itself yields

$$p = \frac{d}{\overline{\gamma}} = d\left(1 + 3dn + o\left(dn\right)\right).$$

Define $\varepsilon_0 = p/d - 1$ for brevity. Then by substituting p, we have from (5.1.8) that

$$\begin{aligned}\log \mathbf{P}\left(X = 0\right) &= \widetilde{J}_2 + o(1) \\ &= -\frac{1}{4}n^4d^2(1+\varepsilon_0)^2 + \frac{3}{2}n^3d^2(1+\varepsilon_0)^2 + \frac{2}{3}n^5d^3(1+\varepsilon_0)^3 - \frac{55}{24}n^6d^4(1+\varepsilon_0)^4 + o(1).\end{aligned}$$

By noting (5.5.8), plugging into (5.5.19) gives $\mathbf{P}\left(X = 0\right)$ in $H_3(m, n)$,

$$\begin{aligned}\log \mathbf{P}\left(X = 0 \mid Y = m\right) &= \widetilde{J}_2 - d\binom{n}{3}\log\left(1+\varepsilon_0\right) - \binom{n}{3}(1-d)\log\left(1 - \frac{d\varepsilon_0}{1-d}\right) + o\left(1\right) \\ &= \widetilde{J}_2 + \binom{n}{3}\left(\frac{d\varepsilon_0^2}{2}(1+d) - \frac{d\varepsilon_0^3}{3}\right) + o\left(1\right).\end{aligned}$$

Calculations and truncations yield

$$\mathbf{P}\left(H_3(n,m) \in \mathcal{L}_3(n)\right) = \exp\left(-\frac{1}{4}n^4d^2 - \frac{1}{12}n^5d^3 - \frac{1}{24}n^6d^4 + \frac{3}{2}n^3d^2 + o\left(1\right)\right).$$

For the same reasons as in the $H_3(n,p)$ case, the validity extends to all $d = o(n^{-7/5})$. This completes the proof of Theorem 5.4.

5.8 Concluding remarks

In this chapter, we use the perturbation method to give the asymptotic probability of linearity of random hypergraphs by avoiding certain forbidden hypergraphs. The same method also gives the probability of non-existence of some other subhypergraphs in random hypergraphs by considering the corresponding clustering. In this chapter, we focus on the case when r is some fixed integer, it may be possible to extend our method and to consider the case when $r = r(n)$ by considering the corresponding clusterings therein.

5.9 Appendix: Computational details for Section 5.7

We include Wolfram Mathematica code used in Section 5.7 here.

```
ga2 = 1 - 7 (y*z)
ga1 = (1 - 5 y*z +
    25 (y*z)^2)*(1 + p*z + (p*z)^2 + (p*z)^3 + (p*z)^4)^2
ga0i0 = (1 + (p*z) + (p*z)^2 + (p*z)^3 + (p*z)^4)*(1 - (p*z) -
    3*(y*z)*ga1 - 3 (y*z)^2*ga2 - (y*z)^3)
ga0i1 = (1 + (p*z) + (p*z)^2 + (p*z)^3 + (p*z)^4)*(1 - (p*z)*ga0i0 -
    3*(y*z)*ga1 - 3 (y*z)^2*ga2 - (y*z)^3)
ga0i2 = (1 + (p*z) + (p*z)^2 + (p*z)^3 + (p*z)^4)*(1 - (p*z)*ga0i1 -
    3*(y*z)*ga1 - 3 (y*z)^2*ga2 - (y*z)^3)
```

```
ga0trun = Expand[ga0i2] /. z^b_ /; b >= 6 -> 0
ga = With[{z = 1, y = n*p*a, p = a*p}, Evaluate[ga0trun]]

p0 = Expand[
  d*a*Expand[1 + (1 - ga) + (1 - ga)^2 + (1 - ga)^3] /.
   a^b_ /; b >= 6 -> 0]
p1 = Expand[
  d*a*Expand[
     With[{p = p0/a},
      Evaluate[1 + (1 - ga) + (1 - ga)^2 + (1 - ga)^3]]] /.
   a^b_ /; b >= 6 -> 0]
p2 = Expand[
  d*a*Expand[
     With[{p = p1/a},
      Evaluate[1 + (1 - ga) + (1 - ga)^2 + (1 - ga)^3]]] /.
   a^b_ /; b >= 6 -> 0]
p3 = Expand[
  d*a*Expand[
     With[{p = p2/a},
      Evaluate[1 + (1 - ga) + (1 - ga)^2 + (1 - ga)^3]]] /.
   a^b_ /; b >= 6 -> 0]
eps = With[{a = 1}, Evaluate[Expand[p3/d - 1]]]
first = -n^4 p3^2/4 + 3 n^3 p3^2/2 + 2 n^5 p3^3/3 - 55 n^6 p3^4/24
second = n (n - 1) (n - 2)/6*(d*(1 + d)*eps^2/2 - d*eps^3/3)
With[{a = 1}, Evaluate[Expand[first + second]]] /.
  d^b_ /; b > 5 -> 0 /. n^b_ /; b > 8 -> 0
```

Chapter 6

Concluding remarks and future work

This thesis presents several examples of obtaining accurate asymptotics of the probability of non-existence of small substructures in random objects via the consideration of clusters and cumulants. We believe these are some tips of the iceberg. More interesting and fundamental results are to be investigated.

6.1 Non-existence of subhypergraphs and hypergraph independence polynomials

Now we point out some connections between the probability of non-existence of small subhypergraphs and the hypergraph independence polynomials.

An independent set in a hypergraph H is a subset of vertices $U \subseteq V(H)$ that contains no edge in $E(H)$. Let $\mathcal{I}(H)$ denote the set of all independent sets of hypergraphs H, and $p \in (0,1)$. Then the independence polynomial of the hypergraph H with parameter p is

$$\mathbf{I}_H(p) = \sum_{U \in \mathcal{I}(H)} p^{|U|}.$$

It is natural to consider the hypergraph cluster expansion following the derivation in Section 2.2 to obtain an expansion of $\log \mathbf{I}_H(p)$.

Let F be a hypergraph and let n be a positive integer. We define a hypergraph H_F such that its vertex set $V(H_F) = E(K_n) = \binom{[n]}{r}$ is the edge set of the complete r-uniform hypergraph with vertex set $[n] = \{1, \dots, n\}$, and let $E(H_F)$ be the collection of the edge sets of all copies of F in the complete r-uniform hypergraph $K_{n,r}$ defined by (2.3.1).

Then the probability of the non-existence of subhypergraphs F in a random binomial hypergraph is

$$\begin{aligned} \mathbf{P}\left(H_r(n,p) \text{ is } F\text{-free}\right) = \sum_{U \in \mathcal{I}(H_F)} p^{|U|}(1-p)^{\binom{n}{r}-|U|} &= (1-p)^{\binom{n}{r}} \sum_{U \in \mathcal{I}(H_F)} \left(\tfrac{p}{1-p}\right)^{|U|} \\ &= (1-p)^{\binom{n}{r}} \mathbf{I}_{H_F}\left(\tfrac{p}{1-p}\right), \end{aligned} \tag{6.1.1}$$

and therefore,

$$\log \mathbf{P}\left(H_r(n,p) \text{ is } F\text{-free}\right) = \binom{n}{r} \log(1-p) + \log \mathbf{I}_{H_F}\left(\tfrac{p}{1-p}\right). \tag{6.1.2}$$

In a hypergraph with maximum degree Δ, each vertex appears in at most Δ edges. Improving upon a recent result by Galvin, McKinley, Perkins, Sarantis and Tetali in [35], Bencs and Buys [8] show that the optimal zero-free disc around 0 for the graph independence polynomials obtained by Shearer [89], that is $p_{\text{Shearer}}(\Delta)$ defined by (1.3.15), is also the zero-free region for the hypergraph independence polynomials. This essentially characterises the absolute convergence of the hypergraph cluster expansion for hypergraphs with maximum degree Δ.

For the non-existence of triangle in $\mathcal{G}(n,p)$, we have $F = K_3$, and therefore the maximum degree of hypergraph H_F is

$$\Delta(H_{K_3}) = |\{e \in E(H_{K_3}) : v \in e\}| = n - 2,$$

and hence,

$$p_{\text{Shearer}}(\Delta) \lesssim \frac{1}{e(\Delta-1)} = \frac{1}{e(n-3)} = \Theta\left(n^{-1}\right).$$

In comparison, the assumption by Stark and Wormald [98, Theorem 1.1] and also by Mousset, Noever, Panagiotou and Samotij [77, Corollary 13] is

$$p = O\left(n^{-1/2-\varepsilon}\right)$$

for some $\varepsilon > 0$. It is unclear whether $p_{\text{Shearer}}(\Delta)$ is necessary for using cluster expansion in (6.1.2).

6.2 Open problems

We have shown that the truncation of the cluster expansion series gives the asymptotic linearity of binomial random hypergraphs. The analysis of the truncation utilised the cumulant series by by Mousset, Noever, Panagiotou and Samotij [77], whose derivation exploits the correlation among random variables and relies heavily on FKG inequality. It would be interesting to investigate whether this is necessary for the truncation. Alternative ways of handling truncation that are commonly used include establishing the absolute convergence of the series via the Koteckỳ-Preiss criterion [61], for example, see Section 2.5 or [40, 59], etc.

Question 1. *What is the proper criterion for the absolute convergence of cluster expansion* (2.2.1) *and* (6.1.2)*? In particular, for the probability of the non-existence of subhypergraphs in* $H_r(n,p)$*, recalling the definition of extension value* x *in* (5.1.2)*, is* $x = o(1)$ *enough?*

The local dependence is assumed to use the cluster expansion in this thesis due to the assumption in the binomial random graph $H_r(n,p)$ setting. It is not clear whether the framework can be adapted to the settings with weak long-range dependencies, such as the study of Markov chains, etc.

It also would be interesting to investigate whether the cluster expansion series also gives the probability of the non-existence of subhypergraphs in random hypergraphs with given number of edges $H_r(n,m)$. We could also consider random graphs with given degree sequence, or the even simpler random regular graphs $\mathcal{G}(n,d)$. In these cases, all graph-dependent indicators are dependent, and the only valid dependency graph for them is the complete graph. We may need to modify the method by incorporating the notion of weak dependence, for instance, mixing coefficients, etc, see, for example, [48, 49], or to consider variants of "weighted dependency graph" introduced in [25].

Question 2. *How to design proper dependency graphs to study the probability of the non-existence of subhypergraphs in $H_r(n, m)$ or subgraphs in random regular graphs $\mathcal{G}(n, d)$?*

6.3 Future work

Here we list some feasible future work.

(fw1) **Eulerian orientation count of random regular graphs.**

The complex martingale method [46] gives asymptotic enumeration of Eulerian orientations of random regular graphs for the dense case. Together with Brendan McKay, the preliminary computations give an interesting correction multiplicative factor to a naive formula. This formula is consistent with simulations for the sparse range, for which the switching method will be used. It remains to write a rigorous proof. We also conjecture that the formula holds for the middle range.

(fw2) **Complex cumulants and asymptotic enumeration.**

The complex martingale method [46] can be used to obtain accurate asymptotic enumeration of orientations of a graph as a function of the out-degree sequence. This would extend results in [45].

(fw3) **More accurate asymptotic probability of non-existence of small subhypergraphs.**

We used the perturbation method to study the probability of the non-existence of small subhypergraphs in random hypergraphs, and we can modify our approach to obtain more accurate asymptotic formulae to any desired power of n^{-1}. This would be analogous to the results in Theorem 2.18.

(fw4) **Maximum likelihood estimation of edge probability using small cluster counts.**

By keeping track of the numbers of clusters, we obtain approximations of the conditional probabilities of avoiding certain sets of clusters given the counts for smaller ones and the non-existence of even larger ones. This can be uilized, from a Bayesian perspective, to obtain the maximum likelihood estimation of edge probability using cluster counts.

(fw5) **Asymptotic linearity of binomial random hypergraphs via cluster expansion, II. Convergent series.**

There is another formulation to relate the probability of a random hypergaph $H_r(n, p)$ being linear and the independence polynomial. We define a graph G such that its vertex set contains all hyperedges of the complete r-graph, and there are edges between hyperedges with overlap greater than 1, that is, $V(G) = \binom{n}{r}$, and

$$E(G) = \left\{ \{v_1, v_2\} \in \binom{V(G)}{2} : |v_1 \cap v_2| \geqslant 1 \right\}.$$

Then the maximum degree of G is

$$\Delta = \sum_{i=2}^{r-1} \binom{r}{i} \binom{n-r}{r-i} > \binom{r}{2} \binom{n-r}{r-2},$$

and therefore, if $r = \Theta(1)$, then

$$p_{\text{Shearer}}(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{\Delta^{\Delta}} < \frac{1}{e(\Delta-1)} \lesssim \frac{1}{e\binom{r}{2}\binom{n-r}{r-2}} = \Theta\left(n^{2-r}\right).$$

In comparison, Theorem 2.7 requires $p = o\left(n^{2-r}\right)$. Hence, this approach gives stronger results for even less restricted p.

Bibliography

[1] Abdelmalek Abdesselam, Aldo Procacci, and Benedetto Scoppola. Clustering bounds on n-point correlations for unbounded spin systems. *Journal of Statistical Physics*, 136(3):405–452, 2009.

[2] Noga Alon. The number of spanning trees in regular graphs. *Random Structures & Algorithms*, 1(2):175–181, 1990.

[3] Noga Alon and Joel Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[4] Richard Arratia, Larry Goldstein, and Louis Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *The Annals of Probability*, pages 9–25, 1989.

[5] Richard Arratia, Larry Goldstein, and Louis Gordon. Poisson approximation and the Chen-Stein method. *Statistical Science*, pages 403–424, 1990.

[6] Andrew D Barbour, Lars Holst, and Svante Janson. *Poisson approximation*, volume 2. The Clarendon Press Oxford University Press, 1992.

[7] Rodney J Baxter. F model on a triangular lattice. *Journal of Mathematical Physics*, 10(7):1211–1216, 1969.

[8] Ferenc Bencs and Pjotr Buys. Optimal zero-free regions for the independence polynomial of bounded degree hypergraphs. *arXiv preprint arXiv:2306.00122*, 2023.

[9] Olivier Bernardi. On the spanning trees of the hypercube and other products of graphs. *the electronic journal of combinatorics*, 19(4):P51, 2012.

[10] Béla Bollobás. The distribution of the maximum degree of a random graph. *Discrete Mathematics*, 32(2):201–203, 1980.

[11] Béla Bollobás. Degree sequences of random graphs. *Discrete Mathematics*, 33(1):1–19, 1981.

[12] Béla Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001.

[13] Richard C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

[14] Robert M Burton, Marc Goulet, and Ronald Meester. On 1-dependent processes and k-block factors. *The Annals of Probability*, 21(4):2157–2168, 1993.

[15] Olivier Catoni. Laplace transform estimates and deviation inequalities. In *Annales de l'IHP Probabilités et statistiques*, volume 39, pages 1–26, 2003.

[16] Fraser Daly. Compound Poisson approximation with association or negative association via Stein's method. *Electronic Communications in Probability*, 18, 2013.

[17] Roland L Dobrushin. Estimates of semi-invariants for the Ising model at low temperatures. *Translations of the American Mathematical Society-Series 2*, 177:59–82, 1996.

[18] Fengming Dong, Khee-Meng Koh, and Kee L Teo. *Chromatic polynomials and chromaticity of graphs*. World Scientific, 2005.

[19] Hanna Döring and Peter Eichelsbacher. Moderate deviations via cumulants. *Journal of Theoretical Probability*, 26:360–385, 2013.

[20] Artūras Dubickas. An estimate for the probability of dependent events. *Statistics & probability letters*, 78(17):2839–2843, 2008.

[21] Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II*, pages 609–627. Colloq. Math. Soc. János Bolyai, Vol. 10. North-Holland, Amsterdam, 1975.

[22] Paul Erdős and Joel Spencer. Lopsided Lovász local lemma and Latin transversals. *Discrete Applied Mathematics*, 30(151-154):10–1016, 1991.

[23] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[24] Michael Falk, Jürg Hüsler, and Rolf-Dieter Reiss. *Laws of small numbers: extremes and rare events*. Springer Science & Business Media, 2010.

[25] Valentin Féray. Weighted dependency graphs. *Electronic Journal of Probability*, 23, 2018.

[26] Valentin Féray, Pierre-Loïc Méliot, and Ashkan Nikeghbali. Mod-ϕ convergence: Normality zones and precise deviations. In *Springer Briefs in Probability and Mathematical Statistics*. Springer, 2016.

[27] Roberto Fernández and Aldo Procacci. Cluster expansion for abstract polymer models. new bounds from an old approach. *Communications in Mathematical Physics*, 274(1):123–140, 2007.

[28] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press, 1928.

[29] Cees M Fortuin, Pieter W Kasteleyn, and Jean Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

[30] Sacha Friedli and Yvan Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017.

[31] Janes Galambos. Variants of the graph dependent model in extreme value theory. *Communications in Statistics-Theory and Methods*, 17(7):2211–2221, 1988.

[32] Janos Galambos. On the distribution of the maximum of random variables. *The Annals of Mathematical Statistics*, pages 516–521, 1972.

[33] Janos Galambos and Italo Simonelli. *Bonferroni-type inequalities with applications.* Springer, 1996.

[34] David Galvin and Jeff Kahn. On phase transition in the hard-core model on $\mathbb{Z}^d$. *Combin. Probab. Comput.*, 13(2):137–164, 2004.

[35] David Galvin, Gwen McKinley, Will Perkins, Michail Sarantis, and Prasad Tetali. On the zeroes of hypergraph independence polynomials. *Combinatorics, Probability and Computing*, page 1–20, 2023.

[36] ML Glasser and FY Wu. On the entropy of spanning trees on a large triangular lattice. *The Ramanujan Journal*, 10(2):205–214, 2005.

[37] Boris Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.

[38] I. S Gradshteyn and I. M Ryzhik. *Table of Integrals, Series, and Products.* Elsevier Science & Technology, Kent, 1980.

[39] Theodore E Harris. A lower bound for the critical probability in a certain percolation process. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 56, pages 13–20. Cambridge University Press, 1960.

[40] Tyler Helmuth, Will Perkins, and Guus Regts. Algorithmic Pirogov–Sinai theory. *Probability Theory and Related Fields*, 176(3):851–895, 2020.

[41] Wassily Hoeffding and Herbert Robbins. The central limit theorem for dependent random variables. *Duke Mathematical Journal*, 15(3):773–780, 1948.

[42] Jürg Hüsler. Asymptotic approximation of crossing probabilities of random sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 63:257–270, 1983.

[43] Jürg Hüsler. Extreme values of non-stationary random sequences. *Journal of applied probability*, 23(4):937–950, 1986.

[44] Mikhail Isaev and Kseniia Isaeva. Asymptotic enumeration of Eulerian orientations for graphs with strong mixing properties. *Diskretn. Anal. Issled. Oper.*, 20(6):40–58, 101, 2013.

[45] Mikhail Isaev, Tejas Iyer, and Brendan D McKay. Asymptotic enumeration of orientations of a graph as a function of the out-degree sequence. *Electronic Journal of Combinatorics*, 27(1):P1–26, 2020.

[46] Mikhail Isaev and Brendan D McKay. Complex martingales and asymptotic enumeration. *Random Structures & Algorithms*, 52(4):617–661, 2018.

[47] Mikhail Isaev, Brendan D McKay, and Rui-Ray Zhang. Cumulant expansion for counting eulerian orientations. *arXiv preprint arXiv:2309.15473*, 2023.

[48] Mikhail Isaev, Igor Rodionov, R-R Zhang, and Maksim Zhukovskii. Extreme value theory for triangular arrays of dependent random variables. *Russian Mathematical Surveys*, 75(5):968, 2020.

[49] Mikhail Isaev, Igor Rodionov, Rui-Ray Zhang, and Maksim Zhukovskii. Extremal independence in discrete random systems. *arXiv preprint arXiv:2105.04917*, 2021.

[50] Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

[51] Grigorii Ivanovich Ivchenko. On the asymptotic behavior of degrees of vertices in a random graph. *Theory of Probability & Its Applications*, 18(1):188–195, 1973.

[52] François Jaeger, Dirk L Vertigan, and Dominic JA Welsh. On the computational complexity of the jones and tutte polynomials. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 108, pages 35–53. Cambridge University Press, 1990.

[53] Adam Jakubowski and Natalia Soja-Kukieła. Managing local dependencies in asymptotic theory for maxima of stationary random fields. *Extremes*, 22(2):293–315, 2019.

[54] Svante Janson. Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. *The Annals of Probability*, 16(1):305–312, 1988.

[55] Svante Janson. Poisson approximation for large deviations. *Random Structures & Algorithms*, 1(2):221–229, 1990.

[56] Svante Janson. New versions of Suen's correlation inequality. *Random Structures and Algorithms*, 13(3-4):467–483, 1998.

[57] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In *Random graphs*, volume 87, pages 73–87, 1990.

[58] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs*, volume 45. John Wiley & Sons, 2000.

[59] Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. *SIAM Journal on Computing*, 49(4):681–710, 2020.

[60] Kumar Joag-Dev and Frank Proschan. Negative association of random variables with applications. *The Annals of Statistics*, pages 286–295, 1983.

[61] Roman Kotecký and David Preiss. Cluster expansion for abstract polymer models. *Communications in Mathematical Physics*, 103(3):491–498, 1986.

[62] M. R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin, 1983.

[63] M Ross Leadbetter. On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 28(4):289–303, 1974.

[64] M Ross Leadbetter. Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 65:291–306, 1983.

[65] M Ross Leadbetter and Holger Rootzén. On extreme values in stationary random fields. *Stochastic Processes and Related Topics: In Memory of Stamatis Cambanis 1943–1995*, pages 275–285, 1998.

[66] Elliott H Lieb. Residual entropy of square ice. *Physical Review*, 162(1):162, 1967.

[67] Elliott H Lieb and Fa Yueh Wu. Two-dimensional ferroelectric models. In *Phase Transitions and Critical Phenomena*, volume 1, pages 331–490. Academic Press, 1972.

[68] Anita Liebenau and Nick Wormald. Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph. *Journal of the European Mathematical Society*, 2023.

[69] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[70] RM Loynes. Extreme values in uniformly mixing stationary stochastic processes. *The Annals of Mathematical Statistics*, 36(3):993–999, 1965.

[71] Linyuan Lu and Laszlo A Székely. A new asymptotic enumeration technique: the Lovász local lemma. *arXiv preprint arXiv:0905.3983*, 2009.

[72] Brendan D McKay. Applications of a technique for labelled enumeration. *Congressus Numerantium*, 40:207–221, 1983.

[73] Brendan D McKay. The asymptotic numbers of regular tournaments, Eulerian digraphs and Eulerian oriented graphs. *Combinatorica*, 10(4):367–377, 1990.

[74] Brendan D McKay and Fang Tian. Asymptotic enumeration of linear hypergraphs with given number of vertices and edges. *Advances in Applied Mathematics*, 115:102000, 2020.

[75] Milena Mihail and Peter Winkler. On the number of Eulerian orientations of a graph. *Algorithmica*, 16(4):402–414, 1996.

[76] John W Moon. *Counting Labelled Trees*. Canadian mathematical monographs. 1970.

[77] Frank Mousset, Andreas Noever, Konstantinos Panagiotou, and Wojciech Samotij. On the probability of nonexistence in binomial subsets. *The Annals of Probability*, 48(1):493–525, 2020.

[78] Saralees Nadarajah and Kosto Mitov. Asymptotics of maxima of discrete random variables. *Extremes*, 5(3):287, 2002.

[79] Gordon Frank Newell. Asymptotic extremes for m-dependent random variables. *The Annals of Mathematical Statistics*, 35(3):1322–1325, 1964.

[80] GL O'Brien. The maximum term of uniformly mixing stationary processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 30(1):57–63, 1974.

[81] Linus Pauling. The structure and entropy of ice and of other crystals with some randomness of atomic arrangement. *Journal of the American Chemical Society*, 57(12):2680–2684, 1935.

[82] Oliver Penrose. Convergence of fugacity expansions for classical systems. *Statistical mechanics: foundations and applications*, page 101, 1967.

[83] L Pereira and H Ferreira. Limiting crossing probabilities of random fields. *Journal of applied probability*, 43(3):884–891, 2006.

[84] Igor Rodionov and Maksim Zhukovskii. The distribution of the maximum number of common neighbors in the random graph. *European Journal of Combinatorics*, 107:103602, 2023.

[85] Anders Rosengren. On the number of spanning trees for the 3D simple cubic lattice. *Journal of Physics A: Mathematical and General*, 20(14):L923, 1987.

[86] Andrzej Ruciński. When are small subgraphs of a random graph normally distributed? *Probability Theory and Related Fields*, 78(1):1–10, 1988.

[87] Alexander Schrijver. Bounds on the number of Eulerian orientations. *Combinatorica*, 3(3):375–380, 1983.

[88] Alexander D Scott and Alan D Sokal. The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *Journal of Statistical Physics*, 118(5-6):1151–1261, 2005.

[89] James B. Shearer. On a problem of Spencer. *Combinatorica*, 5(3):241–245, 1985.

[90] Matas Šileikis and Lutz Warnke. Counting extensions revisited. *Random Structures & Algorithms*, 61(1):3–30, 2022.

[91] N. Soja-Kukieła. On maxima of stationary fields. *J. Appl. Probab.*, 56(4):1217–1230, 2019.

[92] Alan D Sokal. Bounds on the complex zeros of (di)chromatic polynomials and Potts-model partition functions. *Combinatorics, Probability and Computing*, 10(1):41–77, 2001.

[93] Terry P. Speed. Cumulants and partition lattices. *Austral. J. Statist.*, 25(2):378–388, 1983.

[94] Joel Spencer. Random regular tournaments. *Periodica Mathematica Hungarica*, 5(2):105 – 120, 1974.

[95] Joel Spencer. Counting extensions. *Journal of Combinatorial Theory, Series A*, 55(2):247–255, 1990.

[96] Richard P Stanley. Enumerative combinatorics. vol. 2, volume 62 of. *Cambridge Studies in Advanced Mathematics*, 1999.

[97] Richard P Stanley. Enumerative combinatorics volume 1 second edition. *Cambridge studies in advanced mathematics*, 2011.

[98] Dudley Stark and Nick Wormald. The probability of non-existence of a subgraph in a moderately sparse random graph. *Combinatorics, Probability and Computing*, 27(4):672–715, 2018.

[99] WC Stephen Suen. A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Structures & Algorithms*, 1(2):231–242, 1990.

[100] Peter P Valkó and Joseph Abate. Comparison of sequence accelerators for the gaver method of numerical Laplace transform inversion. *Computers & Mathematics with Applications*, 48(3-4):629–636, 2004.

[101] Geoffrey S Watson. Extreme values in samples from m-dependent stationary stochastic processes. *The Annals of Mathematical Statistics*, 25(4):798–800, 1954.

[102] Dominic JA Welsh. The computational complexity of some classical problems from statistical physics, 1990.

[103] Nicholas C Wormald. The perturbation method and triangle-free random graphs. *Random Structures & Algorithms*, 9(1-2):253–270, 1996.

[104] Rui-Ray Zhang. Asymptotic linearity of binomial random hypergraphs via cluster expansion under graph-dependence. *Advances in Applied Mathematics*, 139:102378, 2022.

[105] Rui-Ray Zhang. When Janson meets McDiarmid: Bounded difference inequalities under graph-dependence. *Statistics & Probability Letters*, 181:109272, 2022.