Inequalities for Sums of Random Variables: a combinatorial perspective

Doctoral thesis

Matas Šileikis

Department of Mathematics and Computer Science,
Adam Mickiewicz University, Poznań

April 2012

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Andrzej Ruciński, for his priceless guidance and advice, as well as his enthusiasm he shared so kindly with me.

I thank my parents who gave me a beautiful childhood and supported my choice to take up science. And my praise goes to Asta, who was always there for me, no matter how far apart we were.

In memory of Vidmantas Kastytis Bentkus

Contents

Chapter 1

Introduction

The central object of study in this work is the probability

$$\mathbb{P}\{S_n \in I\}, \tag{1.1}$$

where $S_n = X_1 + \cdots + X_n$ is a sum of random variables possessing some degree of independence, and $I \subset \mathbb{R}$ is an interval, either bounded or unbounded. We are interested in conditions on $X_1, \cdots, X_n$ (depending, naturally, on I), which guarantee that $\mathbb{P}\{S_n \in I\}$ is small.

The thesis covers a major part of author's research carried out during his PhD studies, namely those results which have some connection to combinatorics. The work has been done in three directions: (i) obtaining probability inequalities applicable to combinatorics, (ii) applying combinatorics to obtain probability inequalities, and (iii) proving inequalities for concrete randomized combinatorial objects. The three following chapters roughly correspond to these directions and contain some results appearing in the papers of Šileikis (2009), Dzindzalieta, Juškevičius and Šileikis (2012+), and Šileikis (2012). Some statements have been proved only recently and have not yet been published. We indicate them by adding a note "(unpublished)" next to the number of the statement.

1.1 Outline

Chapter 2 concerns the concentration of Lipschitz functions in product spaces. We are mainly interested in the simplest space, the discrete cube $Q^n = \{0,1\}^n$, together with a binomial probability measure $\mathbb{P}$ such that

$$\mathbb{P}(x) = p^{|x|}(1-p)^{n-|x|} \quad \text{for every} \quad x \in Q^n.$$

The resulting probability space, which we denote by Q_p^n and call *the weighted cube*, is a basic object in probabilistic combinatorics. The natural metric on Q_n is the *Hamming distance* $d(x, x')$, which is defined as the number of coordinates at which x and x' differ. A function $f : Q^n \to \mathbb{R}$ is called *Lipschitz* if for every

$x, x' \in Q^n$ we have

$$|f(x) - f(x')| \leq d(x, x').$$

We treat f as a function of n independent Bernoulli random variables and consider the question how tightly f is concentrated around its expectation $\mathbb{E}f$. A standard way to study this is to associate with f a martingale sequence $M_0, \ldots, M_n$ such that $f = M_n$. For instance, in applications to the theory of random graphs, f is usually some parameter of a random structure, and the martingale process corresponds to the exposure of the structure in small portions, like, for example, exposure of a random graph edge by edge.

In §2.1 we compare three standard martingale inequalities which extend the three celebrated inequalities of Hoeffding (1963) for the sums of independent random variables. We also include a couple of improvements of these inequalities due to Bentkus.

In §2.2 we demonstrate how these martingale inequalities imply bounds for $\mathbb{P}\{f - \mathbb{E}f \geq x\}$. A conclusion is then made that for functions on Q_p^n the most appropriate martingale inequality is the one which takes into account the variances of the martingale differences $M_k - M_{k-1}$, $k = 1, \ldots, n$.

We finish Chapter 2 with §2.3, devoted to the folowing isoperimetric problem: given a set $A \subset Q_p^n$ of prescribed measure $\mathbb{P}(A)$, how do we minimize the set of vertices lying close to A? More formally, let us define the t-extension $A_t = \{x : d(x, A) \leq t\}$, where

$$d(x, A) = \min \{d(x, x') : x' \in A\}.$$

Then by an isoperimetric inequality we mean a lower bound on $\mathbb{P}(A_t)$. We compare isoperimetric inequalities obtained by different methods and conclude that the martingale method gives inequalities of essentially the same quality.

In Chapter 3 we obtain several optimal bounds for (1.1) when $X_1, \ldots, X_n$ are independent and distributed symmetrically around zero, while I is either $[x, \infty)$, $[x, y)$ or $\{x\}$. In other words, for a given interval I we determine

$$\sup \mathbb{P}\{S_n \in I\}. \tag{1.2}$$

To make the problem non-trivial, we impose certain boundedness conditions on X_i's. For example, when $I = [x, \infty)$, we assume that $|X_i| \in [0, 1]$ for every i. Chapter 3 is probably the most combinatorial part of the thesis, since, as it turns out, one can interpret the probability $\mathbb{P}\{S_n \in I\}$ as the normalized size of a certain family $\mathcal{F}$ of subsets of $\{1, \ldots, n\}$. Depending on the type of I, we show that $\mathcal{F}$ possesses a simple combinatorial property and then apply classical results from the combinatorial set theory to obtain optimal bounds for $|\mathcal{F}|$. The common phenomenon observed in all the bounds we obtain is that (1.2) is attained by sums of i.i.d. random variables.

In the first two chapters we consider classes of abstract random variables defined by distribution restrictions. Chapter 4 contrasts with that, since we deal there with very specific random variables arising from the basic model of random graphs. We consider the Erdős-Rényi binomial random graph $\mathbb{G}(n, p)$

on n vertices, which is obtained by including every of the $\binom{n}{2}$ possible edges independently with probability p. Let X_G be the number of copies of a fixed graph G in $\mathbb{G}(n,p)$. The random variable X_G can be written as the sum of indicators of copies of G. Each indicator has the same probability, but typically they are not independent. On the other hand, their dependence is not too strong, since, as one can easily see, a given copy of G has no common edges with most of the remaining copies.

We are interested in exponential bounds for the probability $\mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\}$, for $t > 1$ constant. The problem has asymptotic nature and the goal is to determine the order of magnitude of

$$-\log \mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\}, \quad \text{as} \quad n \to \infty. \tag{1.3}$$

Analogous problem for the lower tail $\mathbb{P}\{X_G \leq t\,\mathbb{E}X_G\}$, $t \in [0,1)$, was solved in late 1980's by Janson, Łuczak and Ruciński (1990) and Janson (1990). However, the upper tail proved to be much harder to deal with. It gave rise to a series of papers introducing new bounds for tails of sums of dependent random variables (see, e.g., a survey by Janson and Ruciński (2002)).

The best general result for the upper tail was achieved by Janson, Oleszkiewicz and Ruciński (2004) who gave upper and lower bounds differing by a factor $\log 1/p$ in the exponent. Since then this logarithmic gap has been closed for specific G's. For example, DeMarco and Kahn (2012+) have dealt with complete graphs K_r and conjectured the precise asymptotics of (1.3). We prove several results supporting their conjecture.

1.2 Preliminaries

1.2.1 Asymptotic notation

Throughout the paper we use the standard notation relating the asymptotic behaviour of two sequences of numbers (a_n) and (b_n) as $n \to \infty$ (see, e.g., Janson, Łuczak and Ruciński (1990), §2.1). We restate the definitions for completeness. Let us assume that $b_n > 0$ for n sufficiently large. Then

- $a_n = O(b_n)$, if there are constants n_0 and C such that $|a_n| \leq Cb_n$ for $n > n_0$;
- $a_n = \Omega(b_n)$, if there are constants n_0 and C such that $a_n \geq Cb_n$ (note that this implies $a_n > 0$ for large n);
- $a_n = \Theta(b_n)$, if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$;
- $a_n \asymp b_n$, if $a_n = \Theta(b_n)$;
- $a_n \sim b_n$, if $a_n/b_n \to 1$ (note that relations $\asymp$ and $\sim$ are symmetric);
- $a_n = o(b_n)$, if $a_n/b_n \to 0$ as $n \to \infty$;
- $a_n \ll b_n$ or $b_n \gg a_n$, if $a_n \geq 0$ and $a_n = o(b_n)$.

1.2.2 Probability inequalities

In the proofs below we will use a couple of standard probabilistic tools.

The one-sided Chebyshev inequality: for any random variable X with expectation $\mathbb{E}X = \mu$ and finite variance $\sigma^2 = \operatorname{Var} X$ and $x > 0$

$$\mathbb{P}\{X \geq \mu + x\} \leq \frac{\sigma^2}{\sigma^2 + x^2}. \tag{1.4}$$

Inequality (1.4) is equivalent to the following lower bound for the probability that X is not much less than its expected value (cf. Janson, Oleszkiewicz and Ruciński (2004, Lemma 3.2)):

$$\mathbb{P}\{X > \mu - x\} \geq \frac{x^2}{\sigma^2 + x^2}. \tag{1.5}$$

Proof. Noting that $\operatorname{Var}(\mu - X) = \operatorname{Var} X = \sigma^2$ and using (1.4), we get

$$\mathbb{P}\{X > \mu - x\} = 1 - \mathbb{P}\{\mu - X \geq x\} \geq 1 - \frac{\sigma^2}{\sigma^2 + x^2} = \frac{x^2}{\sigma^2 + x^2}.$$

□

Chernoff's bound (see, e.g., Janson, Łuczak and Ruciński (2000, Theorem 2.1)): if $B_n \sim \operatorname{Bi}(n, p)$ is a binomial random variable and $\varphi(\varepsilon) = (1+\varepsilon)\log(1+\varepsilon) - \varepsilon$, then

$$\mathbb{P}\{B_n \geq (1+\varepsilon)np\} \leq \exp\{-np\varphi(\varepsilon)\}, \qquad \varepsilon > 0. \tag{1.6}$$

Sometimes the following slightly weaker form is more convenient, being useful for $x > enp$:

$$\mathbb{P}\{B_n \geq x\} \leq \exp\left\{-x \log \frac{x}{enp}\right\}, \qquad x > 0. \tag{1.7}$$

As a matter of fact, inequality (1.7) can be proved directly very easily writing $m = \lceil x \rceil$ and noting that the binomial tail is at most

$$\binom{n}{m} p^m < \left(\frac{enp}{m}\right)^m = \exp\left\{-m \log \frac{m}{enp}\right\}.$$

Further write $q = 1 - p$. When x is of the same order as the variance $\operatorname{Var} B_n = npq$, it is often sufficient to use the following Bernstein-type bound (see, e.g., Hoeffding (1963, (2.13)), Janson, Łuczak and Ruciński (2000, (2.14)))

$$\mathbb{P}\{B_n \geq np + x\} \leq \exp\left\{-\frac{x^2}{2q(np + x/3)}\right\}, \qquad x > 0. \tag{1.8}$$

Bound (1.8) is a special case of Theorem 2.5 we will state in §2.1.

Let us write $Q^N = \{0,1\}^N$ for the *discrete hypercube* or, simply, the *cube*. Consider a product probability measure $\mathbb{P}_p$ on Q^N induced by a sequence of N independent biased coin tosses, each toss landing heads with probability p. More formally, if $A \subseteq Q^N$ and $|x|$ stands for the number of 1's in $x \in Q^N$, then we set

$$\mathbb{P}_p(A) = \sum_{x \in A} p^{|x|} q^{N-|x|}. \tag{1.9}$$

The measure $\mathbb{P}_p$ turns Q^N into a probability space, which we denote as Q_p^N and call the *weighted cube*.

A natural partial order on Q^N is defined by setting $x \le y$ whenever $x_i \le y_i$ for every i. We say that the event $A \subseteq Q^N$ is increasing (decreasing), if

$$y \ge x \in A \quad (y \le x \in A) \quad \text{implies} \quad y \in A.$$

The *FKG inequality* (see, e.g., Janson, Łuczak and Ruciński (1990, Theorem 2.12)), which in the presented special case is also known as *Harris' Lemma* (see, e.g., Bollobás and Riordan (2006, Lemma 2.3)), implies that monotone events are positively correlated. That is, for any two increasing (decreasing) events $A, B \subseteq Q^N$ we have

$$\mathbb{P}_p(A \cap B) \ge \mathbb{P}_p(A)\mathbb{P}_p(B), \tag{1.10}$$

and, by induction, for any increasing (decreasing) events $A_1, \ldots, A_k$

$$\mathbb{P}_p(A_1 \cap \cdots \cap A_k) \ge \prod_i \mathbb{P}_p(A_i). \tag{1.11}$$

The FKG inequality immediately implies that if A is increasing and B is decreasing, then

$$\mathbb{P}_p(A \cap B) \le \mathbb{P}_p(A)\mathbb{P}_p(B). \tag{1.12}$$

Given a graph F, consider a random subgraph F_p of F in which every edge is present independently with probability p. Such random graph is naturally associated with the weighted cube Q_p^N, where $N = e(F)$. Given an arbitrary ordering $e_1, \ldots, e_N$ of the edges of F, we assign to $x \in Q^N$ the graph with the edge set $\{e_i : x_i = 1\}$.

Chapter 2

Lipschitz functions on the weighted cube

The aim of this section is to state and prove a convenient and efficient concentration inequality for Lipschitz functions on the weighted cube Q_p^n. In §2.1 we recall a few classical inequalities for martingales together with their improvements due to Bentkus. In §2.2 we state inequalities for Lipschitz functions corresponding to the martingale inequalities from §2.1. Finally, we apply the Lipschitz function inequalities to the isoperimetric problem on the weighted cube and compare the results with the isoperimetric inequalities obtained by other methods.

The present section is mainly expository with an intention to complement certain aspects of the surveys McDiarmid (1989) and McDiarmid (1998), which we will quote frequently. The original content of the section consists of Corollaries 2.9 and 2.10 and Theorem 2.15.

2.1 Martingale inequalities

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, by a filtration we mean an increasing sequence of σ-fields

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n \subset \mathcal{F}.$$

A sequence of random variables $M_0, M_1, \ldots, M_n$ is a *martingale*, if M_i is $\mathcal{F}_i$-measurable for every $i = 0, \ldots, n$, and $\mathbb{E}[M_i \,|\, \mathcal{F}_{i-1}] = M_{i-1}$ for $i \geq 1$. For convenience, let us assume that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ so that $M_0 = \mathbb{E}M_1 = \cdots = \mathbb{E}M_n$.

Define random variables $Y_i = M_i - M_{i-1}$, $i = 1, \ldots, n$. We call $Y_1, \ldots, Y_n$ the *martingale difference sequence*. Note that we can reconstruct the martingale by setting

$$M_i = M_0 + Y_1 + \cdots + Y_i, \quad i = 0, \ldots, n.$$

Let us recall the famous Hoeffding-Azuma inequality for martingales with bounded differences.

Theorem 2.1 (McDiarmid (1989, Theorem 5.7)). *Let $a_1, \ldots, a_n, r_1, \ldots, r_n$ be real numbers. Suppose that $a_i \le Y_i \le a_i + r_i$, $i = 1, \ldots, n$. Then for $x > 0$*

$$\mathbb{P}\{M_n - M_0 \ge x\} \le \exp\left\{-\frac{2x^2}{\sum_i r_i^2}\right\}. \tag{2.1}$$

Hoeffding (1963) proved (2.1) for independent $Y_1, \ldots, Y_n$ and remarked that it can be proved for general martingales with minor changes in the proof. In the case when $a_i = -r_i/2$ for every i, Azuma (1967) proved (2.1) for a class of random variables, which includes martingales.

The bound (2.1) can be extended by noticing that it depends only on the *conditional* ranges of Y_i's. Consequently, we can assume that in the assumption $a_i \le Y_i \le a_i + r_i$ the parameter a_i is random and depends only on the past, that is $\mathcal{F}_{i-1}$. Let us make this statement precise.

Theorem 2.2 (McDiarmid (1989, Theorem 6.7)). *Suppose that for each $i = 1, \ldots, n$ there is a real number r_i and a $\mathcal{F}_{i-1}$-measurable random variable a_i such that*

$$a_i \le Y_i \le a_i + r_i. \tag{2.2}$$

Let $\sigma^2 = (\sum_{i=1}^n r_i^2)/4n$. Then for $x \ge 0$

$$\mathbb{P}\{M_n - M_0 \ge x\} \le \mathrm{e}^{-x^2/2n\sigma^2}. \tag{2.3}$$

Note that condition (2.2) is satisfied with $a_i = -r_i/2$, if $Y_1, \ldots, Y_n$ are independent two-point random variables such that

$$\mathbb{P}\{Y_i = -r_i/2\} = \mathbb{P}\{Y_i = r_i/2\} = 1/2.$$

In this case $\operatorname{Var} M_n = n\sigma^2$. Therefore, by the DeMoivre-Laplace limit theorem, the constant in the exponent of (2.3) is optimal. On the other hand, the right hand side of (2.3) is a rough estimate of the tail $\mathbb{P}\{Z \ge x/\sigma\sqrt{n}\}$, where Z is the standard normal random variable. Bentkus (2007) showed that under the assumption (2.2) we can use the normal tail as an upper bound.

Theorem 2.3 (Bentkus (2007, Theorem 1.1)). *Under the conditions of Theorem 2.2, we have*

$$\mathbb{P}\{M_n - M_0 \ge x\} \le c(1 - \Phi(x/\sigma\sqrt{n})) \le \frac{c}{\sqrt{2\pi}} \frac{\sigma\sqrt{n}}{x} \mathrm{e}^{-x^2/2n\sigma^2}, \tag{2.4}$$

where $\Phi(t) = \mathbb{P}\{Z < t\} = (2\pi)^{-1/2} \int_{-\infty}^t \mathrm{e}^{-x^2/2} \mathrm{d}t$ is the distribution function of a standard normal random variable, and $c < 8$ is an absolute constant.

The second inequality in (2.4) follows from a standard estimate of the normal tail. It shows that (2.4) is better than (2.3) for $x \ge c(2\pi)^{-1/2}\sigma\sqrt{n}$.

If we know not just the widths of the ranges of Y_i's, but also that these ranges are asymmetric with respect to zero (say, shifted to the right), we can improve the bounds above. For this, let us recall the martingale version of one of Hoeffding's inequalities.

Theorem 2.4 (McDiarmid (1998, Theorem 3.12))**.** *Suppose that $Y_1, \ldots, Y_n$ is a martingale difference sequence and for every $i = 1, \ldots, n$ there is a constant p_i such that $-p_i \le Y_i \le 1 - p_i$. Let $p = (p_1 + \cdots + p_n)/n$. Then for $x > 0$*

$$\mathbb{P}\{M_n - M_0 \ge x\} \le \exp\left\{-\frac{x^2}{2(np + x/3)}\right\}; \tag{2.5}$$

$$\mathbb{P}\{M_n - M_0 \le -x\} \le \exp\left\{-\frac{x^2}{2np}\right\}. \tag{2.6}$$

Note that Theorems 2.2 and 2.3 still apply under the conditions of Theorem 2.4, but the bounds they give correspond to a normal random variable with variance $n/4$. When $x = O(np)$ and p is small, by applying Theorem 2.4, we gain a factor of order $1/p$ in the exponent.

Let $p_0 \in [0, 1]$ be a small number and assume that n is even. Consider the conditions of Theorem 2.4, when half of the p_i's equal p_0 and the remaining half equal $1 - p_0$. In such case we get no improvement over the Hoeffding-Azuma bound even if p_0 is small, since the average of the p_i's is $1/2$. However, we know that $\operatorname{Var} M_n$ is at most $np_0(1-p_0)$, which is much less than $n/4$. To exploit the information about the variance, we need the martingale version of Bernstein's inequality (which, for binomial random variables, we stated as inequality (1.8) in §1.2.2).

Theorem 2.5 (McDiarmid (1998, (39)))**.** *Suppose that $Y_1, \ldots, Y_n$ is a martingale difference sequence. Assume that there are constants $b, \sigma_1, \ldots, \sigma_n > 0$ such that*

$$Y_i \le b \quad \text{and} \quad \mathbb{E}[Y_i^2 \,|\, \mathcal{F}_{i-1}] \le \sigma_i^2 \quad \text{for every} \quad i = 1, \ldots, n. \tag{2.7}$$

Let $\sigma^2 = (\sigma_1^2 + \cdots + \sigma_n^2)/n$. Then for $x > 0$

$$\mathbb{P}\{M_n - M_0 \ge x\} \le \exp\left\{-\frac{x^2}{2(n\sigma^2 + bx/3)}\right\}. \tag{2.8}$$

Just like Bentkus' inequality (2.4) replaces the Hoeffding-Azuma bound with a normal tail, the bound (2.8) can be replaced with a binomial tail, as was shown by Bentkus (2004).

Before proceeding to Bentkus' result, let us introduce some new notation. Given a random variable X with a *survival function* $G(x) = \mathbb{P}\{X \ge x\}$, let function G° be the log-concave hull of G, that is, the minimal function such that $G^\circ \ge G$ and the function $x \mapsto -\log G^\circ(x)$ is convex. Of course, if X is bounded from above by, say, b, then $G(x) = 0$ for $x > b$. To avoid concerns about the definition of G° in such situation, let us make an agreement that $\log 0 = -\infty$ and recall that a function $f : \mathbb{R} \to (-\infty, \infty]$ is convex whenever the set $\{(x, y) \in \mathbb{R}^2 : y \ge f(x)\}$ is convex. Define $\mathbb{P}^\circ\{X \ge x\} = G^\circ(x)$, $x \in \mathbb{R}$.

Let us further assume that $X = \alpha B_n + \beta$, where $\alpha, \beta \in \mathbb{R}, \alpha \ne 0$, and $B_n \sim \operatorname{Bi}(n, p)$ is a binomial random variable. It is known (see, e.g., Bentkus (2004)) that for such X we have $G^\circ(x) = G(x)$ whenever $G(x) = 0$ or 1 or when x is a jump point of G, while between the jump points G° is obtained by

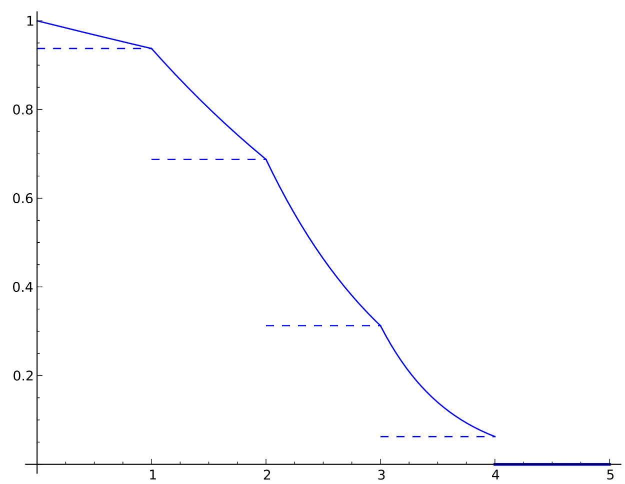

Figure 2.1: $x \to \mathbb{P}\{B_n \geq x\}$ (dashed), $x \to \mathbb{P}^\circ\{B_n \geq x\}$ (solid), $n = 4$, $p = 1/2$.

log-linear interpolation. More precisely, if $x < z < y$ and x, y are adjacent jump points of G, then

$$G^\circ(z) = G(x)^{1-\lambda} G(y)^\lambda \qquad \text{if} \quad z = (1-\lambda)x + \lambda y.$$

In particular, G° is continuous everywhere except for the point $x = \alpha n + \beta$, where G° jumps from p^n to 0, and differentiable everywhere except for the jump points. See Figure 2.1 for the graphs of $\mathbb{P}\{B_n \geq x\}$ and $\mathbb{P}^\circ\{B_n \geq x\}$, when $n = 4$ and $p = 1/2$.

Note that for the binomial tail one can obtain cruder bounds without the symbol $\mathbb{P}^\circ$ by using the following estimate:

$$\mathbb{P}^\circ\{B_n \geq x\} \leq \mathbb{P}\{B_n \geq \lfloor x \rfloor\}. \tag{2.9}$$

Theorem 2.6 (Bentkus (2004, Theorem 1.1))**.** *Suppose that martingale differences $Y_1, \ldots, Y_n$ satisfy the conditions (2.7), and $\sigma^2 = (\sigma_1^2 + \cdots + \sigma_n^2)/n$. Let*

$$S_n = S_n(b, \sigma^2) = \xi_1 + \cdots + \xi_n \tag{2.10}$$

be a sum of i.i.d. random variables such that

$$\mathbb{P}\{\xi_i = b\} = \frac{\sigma^2}{\sigma^2 + b^2}, \qquad \mathbb{P}\{\xi_i = -\sigma^2/b\} = \frac{b^2}{\sigma^2 + b^2}.$$

Then, for every $x \in \mathbb{R}$,

$$\mathbb{P}\{M_n - M_0 \geq x\} \leq c\mathbb{P}^\circ\{S_n \geq x\}. \tag{2.11}$$

where $c = \mathrm{e}^2/2$ is an absolute constant.

Remark. Simple changes in the proof of Theorem 2.6 yield (2.11) under a weaker condition

$$\mathbb{E}[Y_1^2 \,|\, \mathcal{F}_0] + \cdots + \mathbb{E}[Y_n^2 \,|\, \mathcal{F}_n] \leq n\sigma^2.$$

Remark. Note that random variables $\varepsilon_1, \ldots, \varepsilon_n$ satisfy

$$\mathbb{E}\varepsilon_i = 0, \quad \operatorname{Var} \varepsilon_i = \sigma^2, \quad \varepsilon_i \leq b.$$

Consider (2.11) as a bound in terms of n, b, and σ. Suppose that x is a jump point of the function $x \to \mathbb{P}\{S_n \geq x\}$. Then $\mathbb{P}^\circ\{S_n \geq x\} = \mathbb{P}\{S_n \geq x\}$, and therefore, by putting $M_n = S_n$, we get that (2.11) is optimal up to the constant factor c.

One can obtain more analytically manageable bounds from (2.11) by applying one's favourite bounds for the binomial tails, including those we introduced in §1.2.2. For instance, note that the right-hand side of (1.8) is a log-concave function of x. Therefore we get

$$c\mathbb{P}^\circ\{S_n \geq x\} \leq c\exp\left\{-\frac{x^2}{2(n\sigma^2 + bx/3)}\right\},$$

which shows that (2.11) essentially subsumes (2.8).

2.2 Concentration of Lipschitz functions

Let $X = (X_1, \ldots, X_n)$ be a vector of independent random variables, with X_i taking values in a measurable space A_i for each i. Suppose that a measurable function $f : A_1 \times \cdots \times A_n \to \mathbb{R}$ is *separately Lipschitz* with constants $r_1, \ldots, r_n$. That is,

$$|f(x) - f(x')| \leq r_i, \tag{2.12}$$

whenever vectors x and x' differ only in the ith coordinate. For simplicity, instead of $f(X)$ let us just write f.

We are interested in how tightly f is concentrated around its expectation $\mathbb{E}f$. A standard technique to study this is to define a filtration

$$\mathcal{F}_k = \sigma(X_1, \ldots, X_n), \qquad k = 0, \ldots, n,$$

and consider Doob's martingale

$$M_k = \mathbb{E}[f \,|\, \mathcal{F}_k], \qquad k = 0, \ldots, n. \tag{2.13}$$

Note that $M_0 = \mathbb{E}f$ and $M_n = f$. Now we can apply the martingale inequalities presented in the previous section.

As was noted by McDiarmid (1989), for every $k = 1, \ldots, n$

$$g_k(X_1, \ldots, X_{k-1}) \leq M_k \leq g_k(X_1, \ldots, X_{k-1}) + r_k,$$

where the function g_k is defined by

$$g_k(x_1, \ldots, x_{k-1}) = \inf_{y_k \in A_k} \mathbb{E}f(x_1, \ldots, x_{k-1}, y_k, X_{k+1}, \ldots, X_n),$$

so the martingale differences $Y_k = M_k - M_{k-1}$, $k = 1, \dots, n$, satisfy the condition (2.2) with $\mathcal{F}_{k-1}$-measurable random variables $a_k = g_k(X_1, \dots, X_{k-1}) - M_{k-1}$. Therefore Theorem 2.2 implies the following inequality, well known in the combinatorial community.

Corollary 2.7 (McDiarmid (1989, Lemma 1.2))**.** *For f as above and $x > 0$,*

$$\mathbb{P}\{f - \mathbb{E}f \geq x\} \leq \exp\left\{-\frac{x^2}{2n\sigma^2}\right\}, \tag{2.14}$$

where $\sigma^2 = (\sum_{i=1}^n r_i^2)/4n$.

If instead of Theorem 2.2 we apply Theorem 2.3, we get the following result, which is the essential part of Corollary 3.1 in Bentkus (2007).

Corollary (Bentkus (2007))**.** *Under the assumptions of Corollary 2.7, we have*

$$\mathbb{P}\{f - \mathbb{E}f \geq x\} \leq c(1 - \Phi(x/\sigma\sqrt{n})) \leq \frac{c}{\sqrt{2\pi}}\frac{\sigma\sqrt{n}}{x}\mathrm{e}^{-x^2/2n\sigma^2}.$$

We further consider a special setting of importance to the theory of random graphs. Let $A_1 = \cdots = A_n = \{0, 1\}$ and $X_1, \dots, X_n$ be independent Bernoulli random variables with probability p, that is $\mathbb{P}\{X_i = 1\} = p$ and $\mathbb{P}\{X_i = 0\} = q := 1 - p$. In other words, we are interested in the concentration of Lipschitz functions on the weighted cube Q_p^n introduced in §1.2.2.

As an example, consider the function $f(x) = x_1 + \cdots + x_n$, which satisfies (2.12) with $r_1 = \cdots = r_n = 1$. Then (2.14) reads as

$$\mathbb{P}\{X_1 + \cdots + X_n - np \geq x\} \leq \exp\left\{-\frac{2x^2}{n}\right\}. \tag{2.15}$$

Since $X_1 + \cdots + X_n \sim \mathrm{Bi}(n, p)$, (1.8) implies

$$\mathbb{P}\{X_1 + \cdots + X_n - np \geq x\} \leq \exp\left\{-\frac{x^2}{2q(np + x/3)}\right\}, \qquad x \geq 0, \tag{2.16}$$

which is much better than (2.15) when $x \ll n$ and p is close to 0 or 1. The heuristic reason for this is that in (2.15) we do not take into account the knowledge of the parameter p. Roughly speaking, bound (2.15) corresponds to the intuitively worst choice of p, that is, $p = 1/2$.

The argument of McDiarmid (1989) shows that Theorem 2.4 implies concentration inequalities for *monotonous* Lipschitz functions on the weighted cube. We give the proof, since it will be useful in the proof of upcoming Corollary 2.9.

Corollary 2.8 (McDiarmid (1989))**.** *Suppose that $f : Q^n \to \mathbb{R}$ satisfies* (2.12) *with $r_1 = \cdots = r_n = 1$ and $f(x) \leq f(y)$ whenever $x_i \leq y_i$ for every i. Then for $x > 0$*

$$\mathbb{P}\{f - \mathbb{E}f \geq x\} \leq \exp\left\{-\frac{x^2}{2(np + x/3)}\right\}; \tag{2.17}$$

$$\mathbb{P}\{f - \mathbb{E}f \leq -x\} \leq \exp\left\{-\frac{x^2}{2np}\right\}. \tag{2.18}$$

Proof. In view of Theorem 2.4, it is enough to show that the differences $Y_1, \ldots, Y_n$ of the martingale (2.13) satisfy

$$-p \leq Y_k \leq q = 1 - p, \qquad k = 1, \ldots, n.$$

Fix a vector $x \in \{0,1\}^{k-1}$. Conditioned on the $\mathcal{F}_{k-1}$-measurable event

$$\{X_1 = x_1, \ldots, X_{k-1} = x_{k-1}\},$$

the random variable Y_k takes two values, say, s and t, such that

$$s = \mathbb{E}f(x_1, \ldots, x_{k-1}, 0, X_{k+1}, \ldots, X_n) - \mathbb{E}f(x_1, \ldots, x_{k-1}, X_k, \ldots, X_n),$$
$$t = \mathbb{E}f(x_1, \ldots, x_{k-1}, 1, X_{k+1}, \ldots, X_n) - \mathbb{E}f(x_1, \ldots, x_{k-1}, X_k, \ldots, X_n).$$

Monotonicity of f implies that $s \leq t$, and the Lipschitz condition implies that $t - s \leq 1$. Values s and t are taken with probabilities q and p, respectively. Since $\mathbb{E}[Y_k \mid \mathcal{F}_{k-1}] = 0$, we get that $sq + tp = 0$, whence $-p \leq s, t \leq q$, as desired. □

Next we show that (2.17) holds even if f is not monotonous. Using Theorem 2.5, we obtain the following extension of (2.17).

Corollary 2.9. *Let $f : Q^n \to \mathbb{R}$ be a function satisfying condition* (2.12) *and $X = (X_1, \ldots, X_n)$ be independent Bernoulli random variables with parameter $0 < p \leq 1/2$. Let $b = q \max\{r_1, \ldots, r_n\}$ and $\sigma^2 = pq(r_1^2 + \cdots + r_n^2)/n$. Then for $x > 0$*

$$\mathbb{P}\{f - \mathbb{E}f \geq x\} \leq \exp\left\{-\frac{x^2}{2(n\sigma^2 + bx/3)}\right\}. \tag{2.19}$$

In particular, if $r_1 = \cdots = r_n = 1$, then

$$\mathbb{P}\{f - \mathbb{E}f \geq x\} \leq \exp\left\{-\frac{x^2}{2q(np + x/3)}\right\}. \tag{2.20}$$

Remark. Since Theorem 2.5 is already a standard tool within the combinatorial community, we believe that inequality (2.19) is now an "obvious corollary" to anyone who is familiar enough with the martingale method, and we are not sure who should be credited as its authors. See, e.g., §8.2 of Dubhashi and Panconesi (2009) for discussions about inequalities similar to (2.19).

Remark. For $p > 1/2$, one should interchange the roles of 0 and 1 before applying Corollary 2.9. To get a bound for the lower tail $\mathbb{P}\{f - \mathbb{E}f \leq -x\}$, one should apply Corollary 2.9 to the function $-f$. An analogue of (2.18) cannot be true, since bound (2.20) applies both to functions $f(x) = x_1 + \cdots + x_n$ and $f(x) = -x_1 - \cdots - x_n$, so it must be a bound for the heavier of the two tails. The reason for this limitation is that, informally speaking, by discarding the assumption of monotonicity we "lose the sense of orientation" in the cube.

Proof of Corollary 2.9. In view of Theorem 2.5, it is enough to show that differences $Y_1, \ldots, Y_n$ of the martingale (2.13) satisfy inequalities

$$Y_k \leq q r_k, \qquad \mathbb{E}[Y_k^2 \mid \mathcal{F}_k] \leq r_k^2 pq, \quad k = 1, \ldots, n.$$

Similarly as in the Proof of Corollary 2.8, we obtain that conditioned on the event

$$\{X_1 = x_1, \ldots, X_{k-1} = x_{k-1}\},$$

the random variable Y_k takes two values s and t such that

$$s = \mathbb{E}f(x_1, \ldots, x_{k-1}, 0, X_{k+1}, \ldots, X_n) - \mathbb{E}f(x_1, \ldots, x_{k-1}, X_k, \ldots, X_n),$$
$$t = \mathbb{E}f(x_1, \ldots, x_{k-1}, 1, X_{k+1}, \ldots, X_n) - \mathbb{E}f(x_1, \ldots, x_{k-1}, X_k, \ldots, X_n).$$

The Lipschitz condition implies that

$$|t - s| \leq r_k. \tag{2.21}$$

Values s and t are taken with probabilities q and p, respectively. However, f is not necessarily monotonous, so we have no information which one of s and t is greater. Nevertheless, since $\mathbb{E}[Y_k \,|\, \mathcal{F}_{k-1}] = 0$, we get that $sq + tp = 0$, which implies that $s = -tp/q$. Substituting this into (2.21), we obtain $|t| \leq qr_k$, and therefore $|s| \leq pr_k$. Recalling that $p \leq 1/2$, we get that $Y_k \leq \max\{s, t\} \leq qr_k$ and

$$\mathbb{E}[Y_k^2 \,|\, X_1 = x_1, \ldots, X_{k-1} = x_{k-1}] = s^2 q + t^2 p \leq r_k^2 pq,$$

as desired. □

If in the proof above we use Theorem 2.6 instead of Theorem 2.5, we obtain the following corollary.

Corollary 2.10 (Šileikis (2009))**.** *Under the conditions of Corollary 2.9,*

$$\mathbb{P}\{f - \mathbb{E}f \geq x\} \leq c\mathbb{P}^\circ\{S_n \geq x\},$$

where $S_n = S_n(b, \sigma^2)$ *is the random variable defined by* (2.10) *and* $c = \mathrm{e}^2/2$ *is an absolute constant.*

In particular, if $r_1 = \cdots = r_n = 1$*, then*

$$\mathbb{P}\{f - \mathbb{E}f \geq x\} \leq c\mathbb{P}^\circ\{B_n \geq np + x\} \leq c\mathbb{P}\{B_n \geq \lfloor np + x \rfloor\}, \tag{2.22}$$

where $B_n \sim \mathrm{Bi}(n, p)$ *is a binomial random variable.*

Remark. The second inequality in (2.22) comes from (2.9).

2.3 Applications to the isoperimetry of the cube

A natural distance on the cube $Q_n = \{0, 1\}^n$ is the Hamming distance defined by

$$d(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n|.$$

In other words, the distance between two vertices is the number of coordinates in which they differ. Given a subset $A \subset Q^n$ and $x \in Q^n$, let $d(A, x) = \min\{d(x, y) : y \in A\}$. For $t \geq 0$ define the *t-extension* of A by

$$A_t := \{x \in Q^n : d(A, x) \leq t\}.$$

Recall that $\mathbb{P}_p$ is the probability measure on Q^n induced by n independent biased toin cosses with success probability p (see (1.9)).

By an isoperimetric inequality we mean a lower bound on $\mathbb{P}_p(A_t)$ in terms of $\mathbb{P}_p(A)$ and t. Note that if $p = 1/2$, then $\mathbb{P}_p$ is just the normalized counting measure, that is, $\mathbb{P}_{1/2}(A) = |A|2^{-n}$. The best possible isoperimetric inequality in this case was obtained by Harper (1966) (see also Leader (1991) and McDiarmid (1989)). Let B_r be the ball in the Hamming metric d of radius r centered at the zero vector, i.e.,

$$B_r := \{x \in Q^n : x_1 + \cdots + x_n \le r\}, \qquad r \ge 0.$$

Note that if $A = B_r$, then $A_t = B_{r+t}$. Harper's result, in particular, implies that if $|A| = |B_r|$, $r \in \mathbb{N}$, then $|A_t| \ge |B_{r+t}|$. In other words, among sets of size $|B_r|$, the t-extension is minimized by B_r. Bollobás and Leader generalized Harper's result for down-sets $A \subset Q_p^n$ for arbitrary p. Recall that $A \subset Q^n$ is a *down-set* (*up-set*), if $x \le y$ ($x \ge y$) and $y \in A$ imply $x \in A$.

Theorem 2.11 (Bollobás and Leader (1991))**.** *Let $A \subseteq Q_p^n$ be a down-set with $\mathbb{P}_p(A) \ge \mathbb{P}_p(B_r)$, $r \in \{0, 1, \dots\}$. If $t > 0$, then $\mathbb{P}_p(A_t) \ge \mathbb{P}_p(B_{r+t})$.*

Theorem 2.11 can be reformulated as a concentration inequality. Recall that $X_1, \dots, X_n$ are i.i.d. Bernoulli random variables. Define a function $f : Q_p^n \to \mathbb{R}$ by $f(x) = d(A, x)$. Let us, as usual, write f instead of $f(X_1, \dots, X_n)$. Then the statement of Theorem 2.11 is that if $\mathbb{P}\{f = 0\} \ge \mathbb{P}\{X_1 + \cdots + X_n \le r\}$, $r = 0, 1, \dots$, then

$$\mathbb{P}\{f \le t\} \ge \mathbb{P}\{X_1 + \cdots + X_n \le r + t\}, \quad t \ge 0. \tag{2.23}$$

One could compare isoperimetric inequalities using the following test. Given a small number $\alpha \in (0, 1)$, what is the smallest t such that $\mathbb{P}_p(A) \ge \alpha$ implies $\mathbb{P}_p(A_t) \ge 1 - \alpha$? In other words, how fast a tiny set expands to occupy almost the whole cube? Let

$$t^* = t^*(n, p, \alpha) := \inf\{t > 0 : \mathbb{P}_p(A) \ge \alpha \text{ implies } \mathbb{P}_p(A_t) \ge 1 - \alpha\}. \tag{2.24}$$

The better an isoperimetric inequality, the smaller upper bound for t^* it should give.

Applying various bounds for the binomial tail, Bollobás and Leader (1991) derive several exponential isoperimetric inequalities from (2.23). We reformulate (by changing notation) the last theorem in Bollobás and Leader (1991) in such a way that it easily implies an upper bound for t^*.

Theorem 2.12 (Bollobás and Leader (1991))**.** *Let $n = 1, 2, \dots$, $p \in (0, 1/2]$, and $\alpha \in [0, 1]$ be such that*

$$q/3 \le \log 1/\alpha \le np/48.$$

If $A \subseteq Q_p^n$ is a down-set or up-set and satisfies

$$\mathbb{P}_p(A) \ge \alpha, \tag{2.25}$$

then for $t \geq \sqrt{12np \log 1/\alpha}$

$$\mathbb{P}_p(A_t) \geq 1 - \alpha. \tag{2.26}$$

Thus Theorem 2.12 implies $t^* \leq \sqrt{12np \log 1/\alpha}$.

Note that $f(x) = d(A, x)$ is a Lipschitz function and it satisfies (2.12) with $r_1 = \cdots = r_n = 1$, since changing one coordinate of x can only increase or decrease $d(A, x)$ by at most one. Therefore it is natural to try to bound $\mathbb{P}_p(A_t)$ using Corollary 2.7, the consequence of the Hoeffding-Azuma inequality. Bollobás and Leader (1991) noted that Theorem 2.12 gives much better bounds than what can be obtained from Corollary 2.7. McDiarmid (1989) pointed out that this is not because of the weakness of the martingale method, but simply because the Hoeffding-Azuma inequality is too general. Noting that $f(x) = d(A, x)$ is an increasing function (since A is a down-set), McDiarmid (1989) obtained the following result from Corollary 2.8, thus avoiding the exact isoperimetric inequality (2.23).

Theorem 2.13 (McDiarmid (1989, Proposition 7.15))**.** *Let* $A \subset Q_p^n$ *be a down-set of measure* $\mathbb{P}_p(A) = \alpha \in (0, 1)$*. If* $t \geq t_0 := \sqrt{2np \log 1/\alpha}$, $t \in \mathbb{Z}$*, then*

$$\mathbb{P}_p(A_t) \geq 1 - \exp\left\{ -\frac{(t - t_0)^2}{2(np + \frac{t - t_0}{3})} \right\}.$$

If, in addition, $t \leq t_0 + np$*, then*

$$\mathbb{P}_p(A_t) \geq 1 - \exp\left\{ -\frac{(t - t_0)^2}{3np} \right\}. \tag{2.27}$$

Remark. If $t \geq (\sqrt{3} + \sqrt{2})\sqrt{np \log 1/\alpha}$, then by (2.27) we have

$$1 - \mathbb{P}_p(A_t) \leq \exp\left\{ -\frac{(\sqrt{3np \log 1/\alpha})^2}{3np} \right\} = \alpha.$$

Thus Theorem 2.13 implies that $t^* \leq (\sqrt{3} + \sqrt{2})\sqrt{np \log 1/\alpha}$, and therefore is as good as Theorem 2.12, at least for down-sets.

So far we have assumed that $A \subseteq Q_p^n$ is a down-set. Can one obtain a good lower bound for $\mathbb{P}_p(A_t)$ without this assumption? We cannot apply Corollary 2.8, since $f(x) = d(A, x)$ is not necessarily monotonous, but we can apply Corollary 2.9. The following isoperimetric inequality for *general* sets (not just down-sets) in Q_p^n is given by Corollary 2.3.2 in Talagrand (1995).

Theorem 2.14 (Talagrand (1995))**.** *There is an absolute constant* C *such that if a* $A \subset Q_p^n$ *has measure* $\mathbb{P}_p(A) =: \alpha \in (0, 1)$ *and*

$$\sqrt{4npq \log 1/\alpha} =: t_2 \leq t \leq npq,$$

then
$$\mathbb{P}_p(A_{t-1}) \geq 1 - \exp\left\{-\frac{(t - t_2/\sqrt{2})^2}{2npq} + \frac{Ct^3}{n^2p^3q^3}\right\}. \tag{2.28}$$

Talagrand's proof avoids martingales. We conclude this section by showing that the martingale method gives a similar isoperimetric inequality as (2.28) under similar conditions.

Theorem 2.15 (Šileikis (unpublished))**.** *Suppose that set $A \subset Q_p^n$ has measure $\mathbb{P}_p(A) =: \alpha \in (0,1)$. There is a constant $C \in (0,3)$ such that if*
$$\sqrt{Cnpq \log 1/\alpha} =: t_3 \leq t \leq npq, \tag{2.29}$$
then
$$\mathbb{P}_p(A_t) \geq 1 - \exp\left\{-\frac{(t-t_3)^2}{2q(np + (t-t_3)/3)}\right\}. \tag{2.30}$$

Proof. Let X be a random element of Q_p^n distributed according to the measure $\mathbb{P}_p$. Let $f = f(X) = d(A, X)$. As we have already noticed, function f satisfies the Lipschitz property (2.12) with $r_1 = \cdots = r_n = 1$. Without loss of generality, we can assume that $p \leq 1/2$. Therefore, assuming $t \geq \mathbb{E}f$, (2.20) of Corollary 2.9 implies that
$$\begin{aligned} 1 - \mathbb{P}_p(A_{t-1}) &= \mathbb{P}\{f \geq t\} = \mathbb{P}\{f - \mathbb{E}f \geq t - \mathbb{E}f\} \\ &\leq \exp\left\{-\frac{(t-\mathbb{E}f)^2}{2q(np + (t - \mathbb{E}f)/3)}\right\}. \end{aligned} \tag{2.31}$$
The quality and the range of validity of (2.31) depend on how well one can bound the expectation $\mu := \mathbb{E}f$ from above. To conclude the proof, it suffices to show that $\mu \leq t_3$. Writing $l = \log 1/\alpha$, by (2.29) we have
$$\sqrt{Cnpql} \leq npq. \tag{2.32}$$
Applying (2.20) of Corollary 2.9 to the non-positive function $-f$, we get
$$\begin{aligned} \alpha &= \mathbb{P}\{-f = 0\} = \mathbb{P}\{-f \geq 0\} \\ &= \mathbb{P}\{-f + \mu \geq \mu\} \\ &\leq \exp\left\{-\frac{\mu^2}{2q(np + \mu/3)}\right\}. \end{aligned}$$
Therefore $\mu^2 \leq 2q(np + \mu/3)l$. Solving this quadratic inequality, we get
$$\mu \leq ql/3 + \sqrt{q^2l^2/9 + 2npql}.$$
From (2.32) we have that $l \leq npq/C$. On the other hand $q \leq 1$, therefore
$$\begin{aligned} \mu &\leq \sqrt{l}\sqrt{l}/3 + \sqrt{l^2/9 + 2npql} \\ &\leq \sqrt{\frac{npql}{9C}} + \sqrt{\frac{(npq)l}{9C} + 2npql} \\ &\leq \sqrt{Cnpql} = t_3, \end{aligned}$$

the last inequality being true for sufficiently large C. As a matter of fact, it is not hard to see, that we can take $C < 3$. $\square$

Elementary calculations show that both Theorems 2.14 and 2.15 give upper bounds for t^* of the order $\sqrt{npq \log 1/\alpha}$, so, as long as we only care about the order of the bound, the aforementioned theorems are equivalent.

Chapter 3

Sums of symmetric random variables

Let $a_1, \ldots, a_n$ be real numbers, and $\varepsilon_1, \ldots, \varepsilon_n$ be independent random variables, each taking values -1 and 1 with probabilities 1/2. The purpose of this section is to obtain several optimal bounds for the probability

$$\mathbb{P}\left\{a_1\varepsilon_1 + \cdots + a_n\varepsilon_n \in I\right\},$$

where $I \subset \mathbb{R}$ is an interval (unbounded or bounded, and, in particular, just a singleton). Depending on the properties of I, we assume appropriate boundedness conditions on a_i's. In §3.2 we consider the case when $I = [x-r, x+r)$ and the bound is depends only on r, but not on x. This is the classical Littlewood-Offord problem of the 1940's, which we reprove using a short self-contained argument. Theorem 3.5 in §3.3 provides an improvement to a Littlewood-Offord-type bound by giving a bound that depends also on x. In §3.4 we give optimal bounds for the tail probabilities, that is, when $I = [x, \infty)$ (Theorem 3.10). Finally, in §3.5 we prove that all the previous results of the present section can be extended to arbitrary symmetric random variables (not just two-point).

The section is mainly based on the paper of Dzindzalieta, Juškevičius and Šileikis (2012+) and most results here are statements from this paper: either in their original form or extended in an obvious way.

3.1 Notation and basic facts

We write $A \subseteq B$ and $A \subset B$ to denote the facts that A is a subset of B and A is a proper subset of B, respectively. We will make statements about set systems starting with some *ground set* X, in most cases X being finite. Let $\mathcal{P}(X)$ stand for the power set of X, that is, the family of all subsets of X, and let $X^{(k)}$ be the family of all subsets of X of size k. Usually we will choose $X = [n] = \{1, \ldots, n\}$, in which case $\mathcal{P}[n] = \mathcal{P}([n])$ is a disjoint union of the $n+1$

level sets $[n]^{(k)}$, $k = 0, \dots, n$. Given $A \subseteq X$, we write A^c for the complement of $A \subseteq X$, that is, $X \setminus A$. Finally, let us recall that the *symmetric difference* of two sets A and B is the set $A \triangle B := (A \cup B) \setminus (A \cap B)$.

A family $\mathcal{F} \subseteq \mathcal{P}(X)$ is called a *chain*, if every two sets $A, B \in \mathcal{F}$ are *nested*, which means that either $A \subseteq B$ or $B \subseteq A$. A family $\mathcal{F}$ is called a *Sperner family*, or an *antichain*, if no two distinct sets in $\mathcal{F}$ are nested. Simple examples of Sperner families are the level sets $[n]^{(k)}$. The reason for the terminology is a result by Sperner (1928) (see, e.g., Bollobás (1986)), which states that no Sperner family contains more elements than $[n]^{(\lfloor n/2 \rfloor)}$, the largest level set of $\mathcal{P}[n]$. Here and below $\lfloor x \rfloor$ stands for the greatest integer not exceeding x and $\lceil x \rceil$ for the least integer not less than x.

Theorem 3.1 (Sperner (1928))**.** *If $\mathcal{F}$ is a Sperner family on $[n]$, then*

$$|\mathcal{F}| \le \binom{n}{\lfloor n/2 \rfloor}.$$

A family is said to be r-*Sperner*, if it does not contain a chain of length $r+1$, that is $A_1 \subset \cdots \subset A_{r+1}$. Note that a 1-Sperner family is just a Sperner family.

3.2 The Littlewood-Offord problem

Let $a_1, \dots, a_n \in \mathbb{C}$ be complex numbers such that $|a_i| \ge 1$ for every i. Consider the 2^n sums of the form

$$\pm a_1 \pm \cdots \pm a_n.$$

Littlewood and Offord (1943) asked at most how many of these sums can lie inside a circle of a given radius r. Let us consider the simplest interesting case $r = 1$. Erdős (1945) noticed that if $a_1, \dots, a_n$ are real numbers, then by Theorem 3.1 the number of such sums is at most $\binom{n}{\lfloor n/2 \rfloor}$. Indeed, note that we can assume $a_i \ge 1$ for every i. Given $A \subseteq [n]$, write

$$s_A = \sum_{i \in A} a_i - \sum_{i \in A^c} a_i \tag{3.1}$$

and observe that for every $x \in \mathbb{R}$ the family

$$\mathcal{F} = \{A \subseteq [n] : s_A \in (x-1, x+1)\} \tag{3.2}$$

is an antichain. To see this, suppose that $A, B \in \mathcal{F}$ and $A \subset B$. Then

$$s_B - s_A = 2 \sum_{i \in B \setminus A} a_i \ge 2,$$

which is impossible.

What is more, Erdős gave a best possible bound on the number of sums falling in an interval of arbitrary width. Assuming, as above, that $a_1, \dots, a_n \ge 1$, notice that for every $r = 1, 2, \dots$ and $x \in \mathbb{R}$ the family

$$\{A \subseteq [n] : s_A \in (x-r, x+r)\}$$

is r-Sperner. Erdős generalized Theorem 3.1 by showing that the size of an r-Sperner family is at most the sum of the r largest binomial coefficients in n.

Theorem 3.2 (Erdös (1945))**.** *Let $a_1, \ldots, a_n$ be real numbers such that $|a_i| \geq 1$ for every i. For every $r = 1, \ldots, n$, the number of sums of the form $\pm a_1 \pm \cdots \pm a_n$ falling in an open interval of length $2r$ is at most the sum of the r largest binomial coefficients in n.*

Note that Theorem 3.2 is best possible, for if we choose $a_1 = \cdots = a_n = 1$, then the r most popular sums lie in, say, the interval $(1/2 - r, 1/2 + r)$.

We present an alternative proof of Theorem 3.2 due to Dzindzalieta, Juškevičius and Šileikis (2012+), which avoids considering r-Sperner systems. For this we reformulate Theorem 3.2 in probabilistic terms. Let

$$W_n = \varepsilon_1 + \cdots + \varepsilon_n$$

be the sum of independent random signs ε_i, where

$$\mathbb{P}\{\varepsilon_i = -1\} = \mathbb{P}\{\varepsilon_i = 1\} = 1/2.$$

We will refer to W_n as a *simple random walk with n steps.* Note that the sum of the r largest binomial coefficients can be written as $2^n \mathbb{P}\{W_n \in [-r, r)\}$. A moment's thought reveals that the content of Theorem 3.2 does not change if one considers half-open intervals instead of the open ones. Therefore, Theorem 3.2 is equivalent to the following result.

Theorem 3.3 (Dzindzalieta, Juškevičius and Šileikis (2012+))**.** *Let $a_1, \ldots, a_n$ be real numbers such that $|a_i| \geq 1$. For every $r = 0, 1, \ldots,$*

$$\max_{x \in \mathbb{R}} \mathbb{P}\{a_1\varepsilon_1 + \cdots + a_n\varepsilon_n \in [x - r, x + r)\} \leq \mathbb{P}\{W_n \in [-r, r)\}.$$

For purely technical reasons we have included in the statement of Theorem 3.3 the trivial cases $r = 0$ and $r > n$.

Proof of Theorem 3.3. Let us write $S_n = a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$. We can assume that $a_1 \geq a_2 \geq \ldots \geq a_n \geq 1$. Without loss of generality we can also take $a_n = 1$. This is because

$$\begin{aligned} \mathbb{P}\{S_n \in [x - r, x + r)\} &\leq \mathbb{P}\{S_n/a_n \in [x - r, x + r)/a_n\} \\ &\leq \max_{x \in \mathbb{R}} \mathbb{P}\{S_n/a_n \in [x - r, x + r)\}. \end{aligned}$$

We use induction on n. The claim is trivial for $n = 0$, so let us prove the induction step assuming $n \geq 1$. For $r = 0$ the statement is again trivial, so assuming $r \geq 1$ we get

$$\begin{aligned}
&\mathbb{P}\{S_n \in [x-r, x+r)\} \\
=&\tfrac{1}{2}\mathbb{P}\{S_{n-1} \in [x-r-1, x+r-1)\} + \tfrac{1}{2}\mathbb{P}\{S_{n-1} \in [x-r+1, x+r+1)\} \\
=&\tfrac{1}{2}\mathbb{P}\{S_{n-1} \in [x-r-1, x+r+1)\} + \tfrac{1}{2}\mathbb{P}\{S_{n-1} \in [x-r+1, x+r-1)\} \\
\leq&\tfrac{1}{2}\mathbb{P}\{W_{n-1} \in [-r-1, r+1)\} + \tfrac{1}{2}\mathbb{P}\{W_{n-1} \in [-r+1, r-1)\} \\
=&\tfrac{1}{2}\mathbb{P}\{W_{n-1} \in [-r-1, r-1)\} + \tfrac{1}{2}\mathbb{P}\{W_{n-1} \in [-r+1, r+1)\} \\
=&\mathbb{P}\{W_n \in [-r, r)\}.
\end{aligned}$$

□

The main trick is to rearrange the intervals after the second equality in such a way that we have two intervals of different lengths before applying the induction hypothesis.

After the proof of Theorem 3.3 was published in Dzindzalieta, Juškevičius and Šileikis (2012+), the author of the thesis noticed that Kleitman (1970) used a similar rearrangement idea to prove the following generalization of Theorem 3.2 for vectors in $\mathbb{R}^d$ (thus settling a conjecture of Erdös (1945)).

Theorem 3.4 (Kleitman (1970))**.** *Let $U_1, \dots, U_r$ be open subsets of the Euclidean space $\mathbb{R}^d$, each of diameter at most 2. Let $U = U_1 \cup \dots \cup U_r$. If $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$ are vectors of length at least 1, then the number of sums of the form $\pm\mathbf{a}_1 \pm \dots \pm \mathbf{a}_n$ falling in U is at most the sum of r largest binomial coefficients in n.*

Kleitman's proof easily extends to general normed spaces. For a variant of the proof of Theorem 3.4 for $r = 1$, see Bollobás (1986).

3.3 Non-uniform bounds for local concentration

In Chapter 2 we considered inequalities which imply that under certain conditions a random variable X is tightly concentrated around its expectation. The Littlewood-Offord inequalities we have seen in the previous section could be called "anti-concentration" inequalities, since they state rather an opposite fact that X is not concentrated in any sufficiently small set. Let us recall the statement of Theorem 3.3 for $r = 1$. If $a_1, \dots, a_n \geq 1$ are real numbers and $\varepsilon_1, \dots, \varepsilon_n$ are independent random signs, that is $\mathbb{P}\{\varepsilon_i = -1\} = \mathbb{P}\{\varepsilon_i = 1\} = 1/2$, then

$$\mathbb{P}\{a_1\varepsilon_1 + \dots + a_n\varepsilon_n \in [x-1, x+1)\} \leq 2^{-n}\binom{n}{\lfloor n/2 \rfloor}. \tag{3.3}$$

for every $x \in \mathbb{R}$. This bound is independent of x. Can we improve it by assuming, for example, that x is large? When n is odd, the answer is 'no', since we can attain equality in (3.3) for every $x \geq 1$ by choosing $a_1 = \dots = a_n = x$. If n is even and $x \geq 2$, the choice $a_1 = \dots = a_n = x/2$ shows that the best bound we can hope for is $2^{-n}\binom{n}{n/2-1}$, which is not a tremendous improvement.

The obvious reason why we cannot get a significant improvement for large x is that the a_i's are not restricted from above. Assuming boundedness we obtain the following estimate.

Theorem 3.5 (Dzindzalieta, Juškevičius, Šileikis (unpublished))**.** *Let $c \in (0,1]$, $x \geq 0$, $k = \lceil x \rceil$. Suppose that $c \leq a_1, \dots, a_n \leq 1$. Then*

$$\mathbb{P}\left\{a_1\varepsilon_1 + \cdots + a_n\varepsilon_n \in [x, x+2c)\right\} \leq \mathbb{P}\left\{W_n = l\right\}, \tag{3.4}$$

where

$$l = \begin{cases} k, & \text{if} \quad n+k \text{ is even,} \\ k+1, & \text{if} \quad n+k \text{ is odd.} \end{cases}$$

The bound given by Theorem 3.5 is best possible. To see this, put

$$a_1 = \cdots = a_n = a := \max\left\{c, x/l\right\}.$$

Then the sum $a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$ is just aW_n and it takes the value al with probability $\mathbb{P}\left\{W_n = l\right\}$. So it is enough to check that $al \in [x, x+2c)$. Clearly $al \geq x$ and $l \cdot x/l = x < x + 2c$, hence it remains to prove that $cl < x + 2c$. But this follows from the observation that $l \leq k+1 = \lceil x \rceil + 1 < x + 2$, whence $cl < cx + 2c \leq x + 2c$.

Let numbers $c, a_1, \dots, a_n$ be as in Theorem 3.5. Recalling the notation $s_A = \sum_{i \in A} a_i - \sum_{i \in A^c} a_i$, define, for every $x \geq 0$, a family

$$\mathcal{F}_x = \left\{A \subseteq [n] : s_A \in [x, x+2c)\right\}. \tag{3.5}$$

We say that a family of sets $\mathcal{F}$ is called *k-intersecting*, $k \in \{0, 1, \dots\}$, if for every $A, B \in \mathcal{F}$ we have $|A \cap B| \geq k$. We prove Theorem 3.5 by showing that the family $\mathcal{F}_x$ is a $\lceil x \rceil$-interseting antichain and applying the following extension of Theorem 3.1.

Theorem 3.6 (Milner (1968))**.** *If $\mathcal{F} \subseteq \mathcal{P}[n]$ is a k-intersecting antichain, then*

$$|\mathcal{F}| \leq \binom{n}{t}, \qquad t = \left\lceil \frac{n+k}{2} \right\rceil. \tag{3.6}$$

We start with an auxiliary lemma.

Lemma 3.7 (Dzindzalieta, Juškevičius and Šileikis (2012+))**.** *Let $a_1, \dots, a_n \in [0,1]$ be nonnegative numbers. Given $x \geq 0$, let $k = \lceil x \rceil$. Then the family*

$$\mathcal{F}_{\geq x} = \{A \subseteq [n] : s_A \geq x\}. \tag{3.7}$$

is k-intersecting.

Proof. Assume that $k \geq 1$, since otherwise there is nothing to prove. Suppose for contradiction that there are $A, B \in \mathcal{F}_{\geq x}$ such that $|A \cap B| \leq k - 1$. Writing $\sigma_A = \sum_{i \in A} a_i$, we have

$$\begin{aligned} s_A &= \sigma_A - \sigma_{A^c} \\ &= \sigma_{A \cap B} + \sigma_{A \cap B^c} - \sigma_{A^c \cap B} - \sigma_{A^c \cap B^c} \\ &= (\sigma_{A \cap B} - \sigma_{A^c \cap B^c}) + (\sigma_{A \cap B^c} - \sigma_{A^c \cap B}) \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} s_B &= \sigma_B - \sigma_{B^c} \\ &= \sigma_{A\cap B} + \sigma_{A^c\cap B} - \sigma_{A\cap B^c} - \sigma_{A^c\cap B^c} \\ &= (\sigma_{A\cap B} - \sigma_{A^c\cap B^c}) - (\sigma_{A\cap B^c} - \sigma_{A^c\cap B}). \end{aligned} \tag{3.9}$$

Since

$$\sigma_{A\cap B} - \sigma_{A^c\cap B^c} \le \sigma_{A\cap B} \le |A\cap B| \le k-1 < x,$$

from (3.8) and (3.9) we get

$$\min\{s_A, s_B\} = (\sigma_{A\cap B} - \sigma_{A^c\cap B^c}) - |\sigma_{A\cap B^c} - \sigma_{A^c\cap B}| < x,$$

which contradicts the fact $s_A, s_B \ge x$. □

Proof of Theorem 3.5. The left-hand side of (3.4) is $2^{-n}|\mathcal{F}_x|$, where the family $\mathcal{F}_x$ is defined by (3.5). We claim that the right-hand side of (3.4) is equal to

$$2^{-n}\binom{n}{t}, \qquad t = \left\lceil \frac{n+k}{2} \right\rceil.$$

Indeed, if $n+k \in 2\mathbb{Z}$ is even, then $t = (n+k)/2$ and

$$\mathbb{P}\{W_n = l\} = \mathbb{P}\{W_n = k\} = 2^{-n}\binom{n}{\frac{n+k}{2}};$$

whereas if $n+k$ is odd, then $t = (n+k+1)/2$ and

$$\mathbb{P}\{W_n = l\} = \mathbb{P}\{W_n = k+1\} = 2^{-n}\binom{n}{\frac{n+k+1}{2}}.$$

Thus, in view of Theorem 3.6, it is enough to show that $\mathcal{F}_x$ is a k-intersecting antichain. Since $\mathcal{F}_x$ is a subfamily of $\mathcal{F}_{\ge x}$ defined by (3.7), Lemma 3.7 implies that $\mathcal{F}_x$ is k-intersecting.

To show that $\mathcal{F}_x$ is an antichain, suppose for contradiction that there are distinct $A, B \in \mathcal{F}_x$ such that $A \subset B$. Then $s_B - s_A = 2\sum_{i\in B\setminus A} a_i \ge 2c$, which contradicts the assumption that $s_B, s_A \in [x, x+2c)$. □

Theorem 3.5 implies the following non-uniform bound for concentration at a point.

Corollary 3.8 (Dzindzalieta, Juškevičius and Šileikis (2012+))**.** *Let* $x \ge 0$, $k = \lceil x \rceil$. *Suppose that* $0 < a_1, \dots, a_n \le 1$. *Then*

$$\mathbb{P}\{a_1\varepsilon_1 + \cdots + a_n\varepsilon_n = x\} \le \begin{cases} \mathbb{P}\{W_n = k\}, & \text{if} \quad n+k \textit{ is even}, \\ \mathbb{P}\{W_n = k+1\}, & \text{if} \quad n+k \textit{ is odd}. \end{cases}$$

Proof. Let $c = \min\{a_1, \dots, a_n\} > 0$. Since

$$\mathbb{P}\{a_1\varepsilon_1 + \dots + a_n\varepsilon_n = x\} \le \mathbb{P}\{a_1\varepsilon_1 + \dots + a_n\varepsilon_n \in [x, 2c)\},$$

we are done by Theorem 3.5. □

If in Corollary 3.8 we allow some a_i's to be zero, we obtain the following bound.

Corollary 3.9 (Dzindzalieta, Juškevičius and Šileikis (2012+))**.** *If* $a_1, \dots, a_n \in [0,1]$*, then for* $x \ge 0$ *and* $k = \lceil x \rceil$ *we have*

$$\mathbb{P}\{a_1\varepsilon_1 + \dots + a_n\varepsilon_n = x\} \le \mathbb{P}\{W_m = k\}, \tag{3.10}$$

where

$$m = \begin{cases} \min\{n, k^2\}, & \text{if} \quad n+k \textit{ is even}, \\ \min\{n-1, k^2\}, & \text{if} \quad n+k \textit{ is odd}. \end{cases}$$

Proof. Write $S_n = a_1\varepsilon_1 + \dots + a_n\varepsilon_n$. Note that $\mathbb{P}\{S_n = x\} = 0$, unless at least $k = \lceil x \rceil$ of the coefficients $a_1, \dots, a_n$ are positive. Therefore, Corollary 3.8 implies

$$\mathbb{P}\{S_n = x\} \le \max_{k \le j \le n} \mathbb{P}\{W_j = k + \mathbb{I}(j,k)\},$$

where $\mathbb{I}(j,k) = \mathbb{I}\{j + k \text{ is odd}\}$. Since $k \ge 0$, we have

$$\begin{aligned} \mathbb{P}\{W_j = k\} &\ge \tfrac{1}{2}\mathbb{P}\{W_j = k\} + \tfrac{1}{2}\mathbb{P}\{W_j = k+2\} \\ &= \mathbb{P}\{W_{j+1} = k+1\}. \end{aligned}$$

Hence

$$\max_{k \le j \le n} \mathbb{P}\{W_j = k + \mathbb{I}(j,k)\} = \max_{\substack{k \le j \le n \\ k+j \text{ even}}} \mathbb{P}\{W_j = k\}.$$

To finish the proof, we show that the sequence of numbers

$$\mathbb{P}\{W_j = k\} = 2^{-j}\binom{j}{(j+k)/2}, \qquad j = k, k+2, k+4, \dots$$

is unimodal with a peak at $j = k^2$, i.e.,

$$\mathbb{P}\{W_{j-2} = k\} \le \mathbb{P}\{W_j = k\}, \quad \text{if} \quad j \le k^2,$$

and

$$\mathbb{P}\{W_{j-2} = k\} > \mathbb{P}\{W_j = k\}, \quad \text{if} \quad j > k^2.$$

Indeed, this can be shown by considering the following sequence of equivalent inequalities:

$$2^{-j+2}\binom{j-2}{(j+k)/2-1} \le 2^{-j}\binom{j}{(j+k)/2},$$

$$4\binom{j-2}{r-1} \le \binom{j}{r}, \qquad r := \frac{j+k}{2},$$

$$\frac{4(j-2)\dots(j-r)}{(r-1)!} \le \frac{(j)\dots(j-r+1)}{r!},$$
$$4r(j-r) \le j(j-1),$$
$$(j+k)(j-k) \le j^2 - j,$$
$$j \le k^2.$$

□

3.4 A bound for tails

The proof of Theorem 3.5 reveals a way to obtain the best possible bound for the tail of $a_1\varepsilon_1 + \dots + a_n\varepsilon_n$.

Theorem 3.10 (Dzindzalieta, Juškevičius and Šileikis (2012+)). *Let $x > 0$, $k = \lceil x \rceil$. If $a_1, \dots, a_n \in [0,1]$, then*

$$\mathbb{P}\{a_1\varepsilon_1 + \dots + a_n\varepsilon_n \ge x\} \le \begin{cases} \mathbb{P}\{W_n \ge k\} & \text{if } n+k \text{ is even,} \\ \mathbb{P}\{W_{n-1} \ge k\} & \text{if } n+k \text{ is odd.} \end{cases} \tag{3.11}$$

To prove Theorem 3.10, we use the optimal bound for the size of a k-intersecting family.

Theorem 3.11 (Katona (1964)). *If $k \ge 1$ and $\mathcal{F} \subseteq \mathcal{P}[n]$ is a k-intersecting family, then*

$$|\mathcal{F}| \le \begin{cases} \displaystyle\sum_{j=t}^{n} \binom{n}{j}, & \text{if } k+n = 2t, \\ \displaystyle\sum_{j=t}^{n} \binom{n}{j} + \binom{n-1}{t-1}, & \text{if } k+n = 2t-1. \end{cases} \tag{3.12}$$

Notice that if $k + n = 2t$, then

$$\sum_{j=t}^{n} \binom{n}{j} = 2^n \mathbb{P}\{W_n \ge k\}. \tag{3.13}$$

If $k + n = 2t - 1$, then using the Pascal's identity $\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1}$ we get

$$\sum_{j=t}^{n} \binom{n}{j} + \binom{n-1}{t-1} = 2 \sum_{j=t-1}^{n-1} \binom{n-1}{j} = 2^n \mathbb{P}\{W_{n-1} \ge k\}. \tag{3.14}$$

Proof of Theorem 3.10. We have

$$\mathbb{P}\{a_1\varepsilon_1 + \dots + a_n\varepsilon_n \ge x\} = 2^{-n}|\mathcal{F}_{\ge x}|.$$

By Lemma 3.7, $\mathcal{F}_{\ge x}$ is k-intersecting. Since $x > 0$, we have $k \ge 1$. Therefore (3.12), (3.13), and (3.14) imply (3.11), as desired. □

3.5 Extension to symmetric random variables

We call a random variable X *symmetric*, if X and $-X$ have the same distribution. Results from §3.3 and §3.4 can be extended to arbitrary bounded symmetric random variables via the following lemma, which slightly extends Lemma 2.1 in Dzindzalieta, Juškevičius and Šileikis (2012+). As usual, $\varepsilon_1, \ldots, \varepsilon_n$ are the independent random signs:

$$\mathbb{P}\{\varepsilon_i = -1\} = \mathbb{P}\{\varepsilon_i = 1\} = 1/2, \qquad i = 1, \ldots, n.$$

Lemma 3.12 (Dzindzalieta, Juškevičius and Šileikis (2012+))**.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a bounded measurable function. If $c \in [0, 1]$, then we have*

$$\sup_{X_1,\ldots,X_n} \mathbb{E}g(X_1, \ldots, X_n) = \sup_{a_1,\ldots,a_n} \mathbb{E}g(a_1\varepsilon_1, \ldots a_n\varepsilon_n), \tag{3.15}$$

where the supremum on the left-hand side is taken over all symmetric independent random variables $X_1, \ldots, X_n$ such that

$$c \le |X_i| \le 1,$$

and the supremum on the right-hand side is taken over all tuples $a_1, \ldots, a_n$ of numbers from $[c, 1]$.

Proof. Note that if X is a symmetric random variable and a random variable ε is independent of X and takes values ± 1, then $\varepsilon X \stackrel{\mathrm{d}}{=} X$, that is, εX has the same distribution as X. This is because for any measurable set $A \subseteq \mathbb{R}$

$$\begin{aligned}\mathbb{P}\{\varepsilon X \in A\} &= \mathbb{P}\{X \in A\}\mathbb{P}\{\varepsilon = 1\} + \mathbb{P}\{-X \in A\}\mathbb{P}\{\varepsilon = -1\} \\ &= \mathbb{P}\{X \in A\}(\mathbb{P}\{\varepsilon = 1\} + \mathbb{P}\{\varepsilon = -1\}) = \mathbb{P}\{X \in A\}.\end{aligned}$$

Observe that both suprema in (3.15) are finite, because g is bounded. Let $S = \sup_{a_1,\ldots,a_n} \mathbb{E}g(a_1\varepsilon_1, \ldots a_n\varepsilon_n)$. Clearly

$$S \le \sup_{X_1,\ldots,X_n} \mathbb{E}g(X_1, \ldots, X_n),$$

so it is enough to show the opposite inequality. Suppose $\varepsilon_1, \ldots, \varepsilon_n$ are independent of $X_1, \ldots, X_n$. Since $\varepsilon_1 X_1 \stackrel{\mathrm{d}}{=} X_1, \ldots, \varepsilon_n X_n \stackrel{\mathrm{d}}{=} X_n$, we get

$$\mathbb{E}g(X_1, \ldots, X_n) = \mathbb{E}g(X_1\varepsilon_1, \ldots, X_n\varepsilon_n).$$

If we condition on $X_1, \ldots, X_n$, then each of the random variables

$$X_1\varepsilon_1, \ldots, X_n\varepsilon_n$$

is a symmetric two-point random variable taking values in $[c, 1]$. Therefore

$$\mathbb{E}[g(X_1\varepsilon_1, \ldots, X_n\varepsilon_n) \,|\, X_1, \ldots, X_n] \le S,$$

and hence

$$\mathbb{E}g(X_1, \ldots, X_n) = \mathbb{E}\,\mathbb{E}[g(X_1\varepsilon_1, \ldots, X_n\varepsilon_n) \,|\, X_1, \ldots, X_n] \le \mathbb{E}S = S.$$

□

Letting $g(x_1, \ldots, x_n) = \mathbb{I}\{x_1 + \cdots + x_n \in [x, x+2c)\}$ be the indicator function, Theorem 3.5 and Lemma 3.12 immediately imply the following.

Theorem 3.13 (Dzindzalieta, Juškevičius, Šileikis (unpublished)). *Let $c \in (0,1]$. Suppose $S_n = X_1 + \cdots + X_n$ is a sum of independent symmetric random variables satisfying*

$$\mathbb{P}\{c \leq |X_i| \leq 1\} = 1, \qquad i = 1, \ldots, n.$$

If $x \geq 0$ and $k = \lceil x \rceil$, then

$$\mathbb{P}\{S_n \in [x, x+2c)\} \leq \begin{cases} \mathbb{P}\{W_n = k\}, & \text{if} \quad n+k \text{ is even,} \\ \mathbb{P}\{W_n = k+1\}, & \text{if} \quad n+k \text{ is odd.} \end{cases}$$

Note that Theorem 3.13 does not apply when $c = 0$, but in this case we combine Corollary 3.9 and Lemma 3.12 with the function $g(x_1, \ldots, x_n) = \mathbb{I}\{x_1 + \cdots + x_n = x\}$ to obtain the next result.

Theorem 3.14 (Dzindzalieta, Juškevičius and Šileikis (2012+)). *Suppose that $S_n = X_1 + \cdots + X_n$ is a sum of independent symmetric random variables such that*

$$\mathbb{P}\{|X_i| \leq 1\} = 1, \qquad i = 1, \ldots, n.$$

If $x \geq 0$ and $k = \lceil x \rceil$, then

$$\mathbb{P}\{S_n = x\} \leq \mathbb{P}\{W_m = k\},$$

where

$$m = \begin{cases} \min\{n, k^2\}, & \text{if} \quad n+k \text{ is even,} \\ \min\{n-1, k^2\}, & \text{if} \quad n+k \text{ is odd.} \end{cases}$$

Finally, if we set $g(x_1, \ldots, x_n) = \mathbb{I}\{x_1 + \cdots + x_n \geq x\}$, then Theorem 3.10 and Lemma 3.12 imply the following bound for tails.

Theorem 3.15 (Dzindzalieta, Juškevičius and Šileikis (2012+)). *Suppose that $S_n = X_1 + \cdots + X_n$ is a sum of independent symmetric random variables such that*

$$\mathbb{P}\{|X_i| \leq 1\} = 1, \qquad i = 1, \ldots, n.$$

If $x \geq 0$ and $k = \lceil x \rceil$, then

$$\mathbb{P}\{S_n \geq x\} \leq \begin{cases} \mathbb{P}\{W_n \geq k\} & \text{if} \quad n+k \text{ is even,} \\ \mathbb{P}\{W_{n-1} \geq k\} & \text{if} \quad n+k \text{ is odd.} \end{cases} \tag{3.16}$$

Kwapień proved (see Sztencel (1981)) that for arbitrary independent Banach space-valued symmetric random variables X_i and real numbers a_i with absolute value at most 1 we have

$$\mathbb{P}\{\|a_1 X_1 + \ldots + a_n X_n\| \geq x\} \leq 2\,\mathbb{P}\{\|X_1 + \ldots + X_n\| \geq x\}, \quad x > 0. \tag{3.17}$$

The case $n = 2$ with $X_i = \varepsilon_i$ shows that the constant 2 cannot be improved.

Theorem 3.15 improves (3.17) when X_i's are random signs. We believe that combining Theorem 3.15 with some conditioning arguments may lead to better estimates under the assumptions of Kwapień's inequality.

It is interesting to compare Theorem 3.15 with the following bound for sums of not necessarily symmetric random variables $X_1, \dots, X_n$ due to Bentkus (2001). For simplicity we state it for integers only.

Theorem (Bentkus (2001))**.** *Suppose that* $S_n = X_1 + \cdots + X_n$ *is a sum of independent (but not necessarily symmetric) random variables such that* $\mathbb{E}X_i = 0$ *and* $|X_i| \leq 1$ *for every* i*. If* $k = 0, 1, \dots,$ *then*

$$\mathbb{P}\{S_n \geq k\} \leq \begin{cases} \mathbb{P}\{W_n = k\} + 2\mathbb{P}\{W_n \geq k+1\} & \text{if } n+k \text{ is even,} \\ 2\mathbb{P}\{W_n \geq k+1\} & \text{if } n+k \text{ is odd.} \end{cases} \tag{3.18}$$

The bound (3.18) remains valid if $X_1, \dots, X_n$ are martingale differences, and for martingales (3.18) is optimal. However, finding an optimal bound for the tail of independent (but not necessarily symmetric) random variables is considered to be a very hard problem, and to our knowledge there is no conjecture what the answer should be.

Chapter 4

Subgraph counts in the random graph $\mathbb{G}(n,p)$

In this last section we study the order of magnitude of

$$-\log \mathbb{P}\left\{X_G \geq t\,\mathbb{E}X_G\right\}, \quad n \to \infty, \tag{4.1}$$

where X_G is the number of copies of a given graph G in the random graph $\mathbb{G}(n,p)$ and t is a constant. The formula for the asymptotics of (4.1) conjectured by DeMarco and Kahn is presented in §4.1.2.

We prove several partial results confirming the DeMarco-Kahn conjecture. Theorem 4.4 gives the conjectured lower bound in the regime, where it is not implied by the lower bound proved in Janson, Oleszkiewicz and Ruciński (2004). Thus what remains is the upper bound. It is partially given by Theorem 4.5 and Theorems 4.7-4.9. The former gives the upper bound in a small range of p for a large class of graphs G. What is more, we separately give an analogous result for a specific graph not in this class (see Theorem 4.6). The latter group of Theorems does the same for stars $K_{1,r}$ in various ranges of p. The proofs are provided in §4.3.

4.1 Introduction

4.1.1 Notation

We consider the Erdős-Rényi binomial random graph $\mathbb{G}(n,p)$, which is obtained by taking n labelled vertices and adding each of $\binom{n}{2}$ possible edges independently with probability p. As usual, we treat p as a function of n and study the behaviour of $\mathbb{G}(n,p)$ as n tends to infinity. Whenever we use the asymptotic notation from §1.2.1, we let the implicit constants depend on the graph in question. If these constants depend on additional parameters, we indicate that by adding subcripts, say, $a_n = O_{t,\gamma}(b_n)$.

Given a graph $G = (V(G), E(G))$, we write v_G and e_G for the numbers of vertices and edges, respectively, sometimes for typographical reasons using alternative notation $v(G)$ and $e(G)$. By a *copy* of G in another graph F we mean a subgraph of F isomorphic to G. We call the ratio e_G/v_G the *edge density* of G and define the *maximal edge density* of G as

$$m(G) := \max\{e_H/v_H : H \subseteq G, e_H > 0\}. \tag{4.2}$$

A graph is called *balanced* if $e_G/v_G = m(G)$, and *strictly balanced*, if $e_H/v_H = m(G)$ implies $H = G$. We call a subgraph $H \subseteq G$ *extreme* if it attains the maximal edge density in G, that is, $e_H/v_H = m(G)$.

We write X_G for the number of copies of G in $\mathbb{G}(n,p)$. Sometimes we consider specific sets of copies. Let D_G^e be the size of a largest collection of edge-disjoint copies of G in $\mathbb{G}(n,p)$, and let D_G^v be the corresponding count of vertex-disjoint copies. Clearly $D_G^v \leq D_G^e \leq X_G$.

To shorten notation, let us write $\Psi_H := n^{v_H} p^{e_H}$, which is roughly the expectation of X_G, and define the quantity

$$\Phi_G = \Phi_G(n,p) := \min_{H \subseteq G, e_H > 0} \Psi_H. \tag{4.3}$$

Note that our definition of Φ_G is slightly different from the one by, say, Janson, Łuczak and Ruciński (1990), who defined Φ_G as the minimum of $\mathbb{E}X_H$'s. Nevertheless, the two expressions are of the same order of magnitude.

4.1.2 History

The distribution of X_G has been studied extensively since the seminal paper of Erdős and Rényi (1960). Bollobás (1981) determined that the threshold for the property $\{X_G > 0\}$ is $p = n^{-1/m(G)}$. This means that if $p \ll n^{-1/m(G)}$, then $\mathbb{P}\{X_G > 0\} \to 0$ while if $p \gg n^{-1/m(G)}$, then $\mathbb{P}\{X_G > 0\} \to 1$. For G strictly balanced, the random variable X_G was shown to be asymptotically Poisson at the threshold in Bollobás (1981) and Karoński and Ruciński (1983). Ruciński (1988) proved that X_G is asymptotically normal as long as $p \gg n^{-1/m_G}$ and $n^2(1-p) \to \infty$.

More precise studies showed that $\mathbb{P}\{X_G = 0\}$, the probability of nonexistence, is exponentially small with respect to the expectation of the least expected subgraph. Namely, Janson, Łuczak and Ruciński (1990) showed that

$$\exp\left\{-\frac{\min_H \mathbb{E}X_H}{1-p}\right\} \leq \mathbb{P}\{X_G = 0\} \leq \exp\left\{-\Theta(\min_H \mathbb{E}X_H)\right\}, \tag{4.4}$$

where the minimum is taken over $H \subseteq G$ with $e_H > 0$. Using notation (4.3), we can rewrite (4.4) as

$$\exp\left\{-\frac{\Theta(\Phi_G)}{1-p}\right\} \leq \mathbb{P}\{X_G = 0\} \leq \exp\{-\Theta(\Phi_G)\}. \tag{4.5}$$

Janson (1990) showed that the upper bound in (4.5) also holds for the lower tail, that is, for any $\varepsilon \in (0,1]$

$$\mathbb{P}\{X_G \le (1-\varepsilon)\,\mathbb{E}X_G\} = \exp\{-\Omega_\varepsilon(\Phi_G)\}. \tag{4.6}$$

Note that the lower bound in (4.5) serves as a lower bound for the lower tail, which shows that when p is bounded away from 1, the bound (4.6) is optimal up to a constant in the exponent.

Consider a simple case when $G = K_2$, so that X_G is the number of edges, which has distribution $\mathrm{Bi}(\binom{n}{2}, p)$. Trivially, $\Phi_{K_2} = n^2p$, and a version of Chernoff's bound (1.6) gives

$$\mathbb{P}\left\{X_{K_2} \ge (1+\varepsilon)\tbinom{n}{2}p\right\} = \exp\left\{-\Omega_\varepsilon(n^2p)\right\}.$$

It is almost immediate that the same bound is valid when G is a matching, i.e., consists of several non-incident edges. This suggests a natural question: can we obtain, for every G, a bound

$$\mathbb{P}\{X_G \ge (1+\varepsilon)\,\mathbb{E}X_G\} = \exp\{-\Omega_\varepsilon(\Phi_G)\} \ ?$$

An example originating in Vu (2001) shows that this cannot be true in general. Let $G = K_3$ be a triangle, so that $\mathbb{E}X_G = \binom{n}{3}p^3$ and $\Phi_G = \min\left\{n^2p, n^3p^3\right\}$. Let $m = cnp$, where the constant $c > 0$ is such that the complete graph K_m contains at least $2\binom{n}{3}p^3$ triangles. Since $e(K_m) \asymp n^2p^2$, we get

$$\begin{aligned}\mathbb{P}\left\{X_{K_3} \ge 2\tbinom{n}{3}p^3\right\} &\ge \mathbb{P}\left\{K_m \subset \mathbb{G}(n,p)\right\}\\ &= p^{\Theta(n^2p^2)} = \exp\left\{-\Theta\left(n^2p^2\log 1/p\right)\right\}.\end{aligned} \tag{4.7}$$

However, $n^2p^2\log 1/p$ is much less than $\min\{n^2p, n^3p^3\}$, if $n^{-1}\log n \ll p \ll 1$.

In the same paper where (4.6) was proved, Janson showed that the upper tail for *disjoint* copies behaves like a sum of independent indicators. Recall that D_G^e is the maximum number of edge-disjoint copies of G, and let $\varphi(\varepsilon) = (1+\varepsilon)\log(1+\varepsilon) - \varepsilon$. Lemma 2.46 in Janson, Łuczak and Ruciński (1990) states that for every $\varepsilon > 0$

$$\mathbb{P}\{D_G^e \ge (1+\varepsilon)\,\mathbb{E}X_G\} \le \exp\{-\,\mathbb{E}X_G\varphi(\varepsilon)\} \tag{4.8}$$

$$\le \exp\left\{-\frac{\varepsilon^2}{2(1+\varepsilon/3)}\,\mathbb{E}X_G\right\}. \tag{4.9}$$

As Theorem 4.14 below shows, (4.9) is optimal, up to the constant in the exponent, at least as long as G is strictly balanced and $\mathbb{E}X_G \ll n$.

However, the upper tail of X_G proved to be a challenge. A survey by Janson and Ruciński (2002) contains a detailed overview of the methods that had been applied to the upper tail before late 2001; see also Janson, Łuczak and Ruciński (1990).

A breakthrough on the upper tail occured in early 2000's. Kim and Vu (2004) proved that

$$\mathbb{P}\left\{X_{K_3} \ge (1+\varepsilon)\tbinom{n}{3}p^3\right\} = \exp\left\{-\Omega_\varepsilon\left(n^2p^2\right)\right\}, \tag{4.10}$$

which, in view of (4.7), is optimal, at least for $p > n^{-1}\log n$, up to a factor $\log 1/p$ in the exponent. Similar bounds for $G = K_4, C_4$ in some ranges of p were obtained by Janson and Ruciński (2004). Panchenko (2004) proved a bound of essentially the same quality (losing a factor $\log\log n$ in the exponent) for all cycles. Bound (4.10) was extended to all G in the elaborate and elegant paper of Janson, Oleszkiewicz and Ruciński (2004). Let us further assume, for simplicity, that $p \geq n^{-1/m(G)}$. For smaller p the upper tail is known to be of order Φ_G (see §4.3.5).

Janson, Oleszkiewicz and Ruciński (2004) proved that for every $t > 1$

$$\mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\} = \exp\{-\Omega_t(M_G^*)\}. \tag{4.11}$$

and

$$\mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\} = p^{O_t(M_G^*)}. \tag{4.12}$$

where

$$M_G^* = \begin{cases} \Theta(\min_{H\subseteq G} \Psi_H^{1/\alpha_H^*}) & \text{if } n^{-1/m(G)} \leq p \leq n^{-1/\Delta_G}, \\ \Theta(n^2 p^{\Delta_G}) & \text{if } p \geq n^{-1/\Delta_G}, \end{cases} \tag{4.13}$$

and α_H^* is the *fractional independence number* defined as the maximum of $\sum_v \alpha_v$ over all assignments of nonnegative weights α_v to the vertices, satisfying $\alpha_u + \alpha_v \leq 1$ for every edge uv of H. Janson, Oleszkiewicz and Ruciński (2004) determined explicit asymptotics of M_G^* in a few interesting cases. For example, if G is k-regular, then $M_G^* \asymp n^2 p^k$.

The logarithms of the upper bound (4.11) and the lower bound (4.12) differ by a multiplicative factor $\log 1/p$, so the general question of determining the order of magnitude of

$$-\log \mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\} \tag{4.14}$$

remains "narrowly" open. However, some partial improvements to (4.11) exist. Janson and Ruciński (2004) proved that if $\mathbb{E}X_{C_4} \geq C\log n$ for some large $C > 0$ and $p \leq n^{-2/3-\gamma}$, then

$$\mathbb{P}\{X_{C_4} \geq 2\,\mathbb{E}X_{C_4}\} = \exp\{-\Omega_\gamma(M_{C_4}^* \log^{1/2} 1/p)\}; \tag{4.15}$$

while if $\mathbb{E}X_{K_4} \geq C\log n$ and $p \leq n^{-1/2-\gamma}$ for some constant $\gamma > 0$, then

$$\mathbb{P}\{X_{K_4} \geq 2\,\mathbb{E}X_{K_4}\} = \exp\{-\Omega_\gamma(M_{K_4}^* \log^{1/2} 1/p)\}. \tag{4.16}$$

The first results closing the logarithmic gap for a nontrivial graph were obtained by Chatterjee (2011) and, independently, by DeMarco and Kahn (2011). Chatterjee proved that for every $t > 1$ there exists a constant $C = C(t) > 0$ such that for $p > Cn^{-1}\log n$

$$\mathbb{P}\{X_{K_3} \geq t\,\mathbb{E}X_{K_3}\} = \exp\left\{-\Omega_t\left(M_{K_3}^* \log 1/p\right)\right\}. \tag{4.17}$$

However, Chatterjee's result does not cover a tiny range of interest between the threshold $p = n^{-1}$ and $p = Cn^{-1}\log n$. DeMarco and Kahn (2011) proved

(4.17) for $p \geq n^{-1} \log n$ and also showed that the upper tail in the missing range behaves similarly to the lower tail, i.e., if $n^{-1} \leq p \leq n^{-1} \log n$, then

$$\mathbb{P}\left\{X_{K_3} \geq t\,\mathbb{E}X_{K_3}\right\} = \exp\left\{-\Theta_t(\,\mathbb{E}X_{K_3})\right\}. \tag{4.18}$$

Note that (4.18) implies that neither (4.11) nor (4.12) is sharp in this regime, since $\mathbb{E}X_{K_3} \asymp n^3p^3$ and $M^*_{K_3} \asymp n^2p^2$. The results for the triangle can be combined into a single statement

$$\mathbb{P}\left\{X_{K_3} \geq t\,\mathbb{E}X_{K_3}\right\} = \exp\left\{-\Theta_t\left(\min\left\{\,\mathbb{E}X_{K_3}, M^*_{K_3}\log 1/p\right\}\right)\right\}. \tag{4.19}$$

Soon after, DeMarco and Kahn (2012+) extended (4.19) to every clique $G = K_k$:

$$\mathbb{P}\left\{X_{K_k} \geq t\,\mathbb{E}X_{K_k}\right\} = \exp\left\{-\Theta_t\left(\min\left\{\,\mathbb{E}X_{K_k}, M^*_{K_k}\log 1/p\right\}\right)\right\}. \tag{4.20}$$

DeMarco and Kahn (2012+) claim that for any graph G in a certain range of p from the threshold upwards we have

$$\mathbb{P}\left\{X_G \geq t\,\mathbb{E}X_G\right\} = \exp\left\{-O_t(\Phi_G)\right\}, \tag{4.21}$$

where, recall, $\Phi_G = \min_{H\subseteq G, e_H>0} \Psi_H$.

Let $G_0 \subseteq G$ be an extreme subgraph with the minimal number of vertices.

Let us check when the lower bound (4.21) is better than (4.12). If $\Delta_G = 1$, that is, when G is a matching, then Proposition 4.13.(iii) implies $\Phi_G = n^2p$ and $M^*_G \log 1/p \asymp n^2 p \log 1/p$ for *all* $p \geq n^{-1/m_G}$, so (4.21) is always better. Note that in this case $G_0 = K_2$.

If $\Delta_G \geq 2$, then one can show that for $p \gg n^{-1/m(G)} \log n$ we have

$$M^*_G \log 1/p \ll \Phi_G$$

(see Janson, Oleszkiewicz and Ruciński (2004, Remark 8.3)), therefore (4.21) can be better than the lower bound (4.12) only for $p = O(n^{-1/m(G)} \log n)$. By Proposition 4.13 below, in this range we have $\Phi_G \sim \Psi_{G_0}$ and

$$M^*_G \log 1/p \asymp \Psi_K^{1/\alpha^*_K} \log n, \tag{4.22}$$

where $K \subseteq G$ is an extreme subgraph minimizing the quantity v_K/α^*_K.

Thus, the maximum of the lower bounds (4.12) and (4.21) can be written as

$$\exp\left\{-O_t\left(\min\left\{\Psi_{G_0}, M^*_G \log 1/p\right\}\right)\right\}. \tag{4.23}$$

Moreover, solving inequality $\Psi_{G_0} \leq \Psi_K^{1/\alpha^*_K} \log n$ in p, we get that (for G with $\Delta_G \geq 2$) there is a constant $a = a(G) > 0$ such that the range of p where the minimum in (4.23) is Ψ_{G_0} is from the threshold $n^{-1/m(G)}$ up to $n^{-1/m(G)} \log^a n$.

DeMarco and Kahn conjecture that (4.23) gives the right description of the upper tail:

Conjecture 4.1 (DeMarco and Kahn (2012+))**.** *If G is a graph and $G_0 \subseteq G$ is its smallest extreme subgraph, then for $p \geq n^{-1/m(G)}$ and every $t > 1$*

$$\mathbb{P}\left\{X_G \geq t\,\mathbb{E}X_G\right\} = \exp\left\{-\Theta_t\left(\min\left\{\Psi_{G_0}, M^*_G \log 1/p\right\}\right)\right\}. \tag{4.24}$$

When $\Delta_G = 1$, the upper tail is rather an easy problem. Therefore we focus on G's with $\Delta_G \geq 2$. We prove the lower bound (4.21) for every G in the relevant range (see Theorems 4.3 and 4.4 below). In §4.1.3 we also state several new upper bounds.

We finish this section with an inequality by Vu (2000) that lends some support to Conjecture 4.1.

Proposition 4.2 (Vu (2000))**.** *Let G be strictly balanced (so that $G_0 = G$). If $p \geq n^{-1/m(G)}$ is so small that $\mathbb{E}X_G \leq \log n$, then*

$$\mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\} = \exp\{-\Omega_t(\mathbb{E}X_G)\}.$$

For an alternative proof of Proposition 4.2, see Janson and Ruciński (2000, Theorem 6.9). We give yet another proof of it in §4.3.2.

It is not clear if it was observed that Proposition 4.2 is optimal for any strictly balanced G until DeMarco and Kahn gave corresponding lower bounds for triangles (cf. (4.18)).

4.1.3 Results

In this section we state several results supporting Conjecture 4.1. Their proofs appear in §4.3. We start with lower bounds. Note that the following lower bound matches the upper bound given by Proposition 4.2, but is valid in a wider range of p, since $\mathbb{E}X_G \leq \log n$ is equivalent to $p \leq n^{-1/m(G)}(\log n)^{1/e(G)}$.

Theorem 4.3 (Šileikis (2012))**.** *Let G be a strictly balanced graph. Then there exists a constant $\varepsilon = \varepsilon(G) > 0$ such that for $n^{-1/m(G)} \leq p \leq n^{\varepsilon - 1/m(G)}$ and $t > 1$*

$$\mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\} = \exp\{-O_t(\Psi_G)\}. \tag{4.25}$$

Svante Janson (private communication) suggested a way to extend (4.25) to every G using the second moment method. Based on this, we prove the following.

Theorem 4.4 (Janson, Šileikis (unpublished))**.** *Let G be a graph and $G_0 \subseteq G$ be its smallest extreme subgraph. Then there exists a constant $\varepsilon = \varepsilon(G) > 0$ such that for $n^{-1/m(G)} \leq p \leq n^{\varepsilon - 1/m(G)}$ and $t > 1$*

$$\mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\} = \exp\{-O_t(\Psi_{G_0})\}.$$

Theorem 4.4 together with (4.12) gives the conjectured lower bound in (4.24).

As for the upper bound, we give several results using two different methods. The first one can be called "approximation by a disjoint subfamily" and it allows to use (4.8), the upper tail of the number of edge-disjoint copies. This will give a very short proof of Proposition 4.2.

With slightly more effort we extend Proposition 4.2 to certain graphs G, which are not strictly balanced. For this we use a bound due to Janson and Ruciński (2011) for the upper tail of rooted subgraph counts. To define the class of graphs for which our approach works, we need a few more notions.

As usual, an extreme subgraph is called *minimal* if it does not properly contain another extreme subgraph. Of course, a minimal extreme subgraph does not need to be a smallest one. A *rooted graph* (R, F) is a graph F with a fixed independent set $R \subseteq V(F)$.

Given a graph G, let $H_1, \ldots, H_m$ be the minimal extreme subgraphs of G. Note that they are vertex-disjoint. Indeed, suppose $H_i \cap H_j =: K \neq \emptyset$. Since K is a proper subgraph of H_i, by minimality we have $e_K/v_K < e_{H_i}/v_{H_i}$. This easily implies that the edge density of $H_i \cup H_j$ exceeds $m(G)$, which is a contradiction. Let $R = V(H_1 \cup \cdots \cup H_m)$ and $F = G - E(H_1 \cup \cdots \cup H_m)$. Note that R is an independent set in F, since otherwise, for some i, j, the subgraphs H_i and H_j would be joined by an edge e, yielding a subgraph $H_i \cup H_j \cup \{e\}$, denser than $m(G)$ - a contradiction.

Thus, (R, F) is a rooted graph. Its *maximal density* is defined as

$$m_R(F) = \max \left\{ \frac{e(H)}{v(H - R)} : H \subseteq F, v(H - R) > 0 \right\}.$$

It is not hard to see that every graph G satisfies $m_R(F) \leq m(G)$. Indeed, fix $H \subseteq F$ and let $K = H \cup H_1 \cup \cdots \cup H_m$. We have

$$e(K) \leq m(G)v(K) \quad \text{and} \quad e(H_i) = m(G)v(H_i) \quad \text{for every } i.$$

Therefore

$$e(H) = e(K) - \sum_{i=1}^{m} e(H_i) \leq m(G) \left(v(K) - \sum_{i=1}^{m} v(H_i) \right) = m(G)v(H - R),$$

as desired.

Theorem 4.5 (Šileikis (unpublished))**.** *Let G be a graph and (R, F) be the rooted graph obtained from G as described above. Let G_0 be a smallest extreme subgraph of G. Suppose that $m_R(F) < m(G)$. If $p \geq n^{-1/m(G)}$ is such that $\Psi_{G_0} \leq \log n$, then*

$$\mathbb{P}\left\{X_G \geq t\, \mathbb{E}X_G\right\} = \exp\left\{-\Omega_t(\Psi_{G_0})\right\}.$$

To illustrate Theorem 4.5, let G be a clique K_r, $r \geq 3$, with an extra edge attached to one of the vertices. A moment's thought tells us that $m(G) = e(K_r)/v(K_r) = (r-1)/2$. The only minimal extreme graph of G is the clique, so F is an edge rooted at one endpoint plus $r - 1$ isolated roots, so clearly $m_R(F) = 1$. Since G is not strictly balanced, Proposition 4.2 does not apply to G. However Theorem 4.5 does apply, unless $r = 3$. Nevertheless, if $r = 3$, then we can modify the proof of Theorem 4.5 to obtain the conjectured bound. We prove the following theorem in §4.3.3.

Theorem 4.6 (Šileikis (unpublished))**.** *If K_3^+ is the wisk graph, i.e., the triangle with an extra edge attached to one of the vertices, then*

$$\mathbb{P}\left\{X_{K_3^+} \geq t\, \mathbb{E}X_{K_3^+}\right\} = \exp\left\{-\Omega_t(n^3p^3)\right\}, \qquad \text{when } 1 \leq n^3p^3 \leq \log n.$$

It would be interesting to find out if such modification could be extended to cover the remaining G's.

The other group of results concerns the upper tail of stars. Further let G be $K_{1,r}$, $r \geq 2$, the r-armed star. Corollary 1.8 in Janson, Oleszkiewicz and Ruciński (2004) states that if $p \geq n^{-1/m(G)} = n^{-1-1/r}$, then

$$M_G^* \asymp \max\left\{n^{1+1/r}p, n^2p^r\right\} = \begin{cases} n^{1+1/r}p & \text{if } p \leq n^{-1/r}, \\ n^2p^r & \text{if } p \geq n^{-1/r}. \end{cases} \tag{4.26}$$

Since stars are strictly balanced, the only smallest extreme graph of G is G itself and so the conjectured optimal bound is

$$\exp\left\{-\Omega_t(\min\left\{\Psi_G, M_G^* \log 1/p\right\})\right\}. \tag{4.27}$$

Clearly $\Psi_G = n^{r+1}p^r$. It is easy to see that the minimum in (4.27) is equal to

$$\begin{array}{ll} n^{r+1}p^r & \text{if } n^{-1-1/r} \leq p \leq n^{-1-1/r}(\log n)^{1/(r-1)}, \\ n^{1+1/r}p\log 1/p & \text{if } n^{-1-1/r}(\log n)^{1/(r-1)} \leq p \leq n^{-1/r}, \\ n^2p^r \log 1/p & \text{if } p \geq n^{-1/r}. \end{array}$$

The method of approximation by a disjoint subfamily gives, together with Theorem 4.3 and (4.12), the following sharp bounds for small p.

Theorem 4.7 (Šileikis (2012) - upper bound)**.** *Let $G = K_{1,r}$, $r \geq 2$ be the r-armed star. Suppose $\gamma > 0$ and $t > 1$. If $n^{-1-1/r} \leq p \leq n^{-1-1/r+1/r^2-\gamma}$, then*

$$\mathbb{P}\left\{X_G \geq t\,\mathbb{E}X_G\right\} = \exp\left\{-\Theta_{t,\gamma}\left(\min\left\{n^{r+1}p^r, n^{1+1/r}p\log 1/p\right\}\right)\right\}.$$

In Šileikis (2012) it is shown that the factor $\log^{1/2} 1/p$ can be replaced by $\log 1/p$ in (4.15) for $p \leq n^{-4/5-\gamma}$, thus matching the lower bound. We do not include the proof of this fact here, but just mention that it is very similar to the proof of Theorem 4.7.

We use an entirely different method to obtain a sharp bound for large p, by which we mean $p \geq Cn^{-1}\log n$. For this, following DeMarco and Kahn (2011), we start with a reduction, which allows us to consider, instead of $\mathbb{G}(n,p)$, a random bipartite graph and count only stars with the center on one side of bipartition and the remaining vertices on the other. Then we apply the classical method of exponential moments to obtain the two following theorems.

Theorem 4.8 (Šileikis (unpublished))**.** *Let $G = K_{1,r}$, $r \geq 2$ and $t > 1$. Then for $p \geq n^{-1/r}$ we have*

$$\mathbb{P}\left\{X_G \geq t\,\mathbb{E}X_G\right\} = \exp\left\{-\Omega_t(n^2p^r \log 1/p)\right\}. \tag{4.28}$$

Theorem 4.9 (Šileikis (unpublished))**.** *Let $G = K_{1,r}$, $r \geq 2$ and $t > 1$. Then there is a constant $C = C(r,t) > 0$ such that for $Cn^{-1}\log n \leq p \leq n^{-1/r}$ we have*

$$\mathbb{P}\left\{X_G \geq t\,\mathbb{E}X_G\right\} = \exp\left\{-\Omega_t(n^{1+1/r}p\log 1/p)\right\}. \tag{4.29}$$

We stated the last two theorems separately for purposes of exposition. As we will see, the proof of the latter theorem is just the proof of the former one with an extra trick.

However, there is still a gap $n^{-1-1/r+1/r^2-\gamma} \leq p \leq Cn^{-1}\log n$, where a sharp upper bound has not been obtained.

4.2 Preliminaries

In this section we state some prerequisites from the theory of random graphs.

4.2.1 Ordered and rooted copies

In the introduction we considered copies of a given graph G ignoring the order of copies' vertices. Under certain circumstances it is more convenient to deal with *ordered* copies of G in F, that is, edge-preserving injections from $V(G)$ into $V(F)$. Denote their number in $\mathbb{G}(n,p)$ as Y_G. Let $\mu_G = \mu_G(p) = \mathbb{E}Y_G = (n)_{v_G}p^{e_G}$. Numbers of ordered and unordered copies are related via a simple identity $Y_G = \mathrm{aut}(G)X_G$, where $\mathrm{aut}(G)$ is the number of automorphisms of G, therefore

$$\mathbb{P}\left\{X_G \geq t\,\mathbb{E}X_G\right\} = \mathbb{P}\left\{Y_G \geq t\mu_G\right\}. \tag{4.30}$$

Recall that a rooted graph (R,G) is a graph G with a fixed independent set $R \subseteq V(G)$. When the set R is clear from the context, we write G instead of (R,G).

We say that a rooted graph (R',H) is a rooted *subgraph* of (R,G) if H is a subgraph of G and $R' = R \cap V(H)$.

Let (R,G) be a rooted graph and F a graph with a distinguished subset of vertices $W \subseteq V(F)$ such that $|W| = |R|$. Ignoring the edges spanned by W (if any), regard (W,F) as a rooted graph. A rooted subgraph of (W,F) isomorphic to (R,G) is said to be a W*-rooted copy* of (R,G). Given $W \subseteq V(K_n)$, let $X_G^R(W)$ be the number of such copies in $\mathbb{G}(n,p)$. Of course, for every W the random variable $X_G^R(W)$ has the same distribution, therefore we omit the argument W when it does not matter.

We can interpret non-rooted graphs as a special case of rooted graphs with $R = \emptyset$. Many notions extend naturally to the rooted case and the basic facts can be generalized with little effort. The edge density of a rooted graph (R,G) is defined as $e(G)/v(G-R)$ and recall that the maximal edge density is

$$m_R(G) = \max_{H\subseteq G} e(H)/v(H-R).$$

Given $H \subseteq G$, let

$$\Psi_{H,R} := n^{v(H-R)}p^{e(H)}$$

and

$$\Phi_{G,R} := \min\left\{\Psi_{H,R} : H \subseteq G, e(H) > 0\right\}.$$

Just like for non-rooted graphs, the threshold of containment of (R,G) in $\mathbb{G}(n,p)$ is $p = n^{-1/m_R(G)}$.

We will also consider ordered rooted copies, which in the literature are sometimes called extensions (see, e.g., Spencer (1990)). Suppose the roots of G are labeled $v_1,\dots,v_r$. Given a graph F with an *ordered* set of vertices $W = (w_1,\dots,w_r) \subset V(F)$, an *ordered copy* of (R,G) is an edge-preserving injection from $V(G)$ into $V(F)$ such that v_i is mapped to w_i for every i. We write $Y_G^R(W)$ for the number of W-rooted copies of (R,G) in $\mathbb{G}(n,p)$. We omit the argument W when irrelevant. We have $\mu_G^R := \mathbb{E}Y_G^R = (n-r)_{v_G - r} p_G^{e}$, where $r = |R|$.

Although $\mathbb{E}X_G^R \asymp \mathbb{E}Y_G^R$, unfortunately, $X_G^R(W)$ and $Y_G^R(W)$ are not just a multiple of each other as in the non-rooted case. Instead we have

$$X_G^R(W)\,\mathrm{aut}(R,G) = \sum_{\sigma} Y_G^R(\sigma(W)), \tag{4.31}$$

where $\mathrm{aut}(R,G)$ is the number of automorphisms of G mapping R onto R and the sum on the right-hand side is taken over all permutations σ of W.

For applications of the second moment method, we calculate the asymptotic order of $\mathrm{Var}\, Y_G^R$. The form of the answer and the idea of the proof is precisely the same as for $\mathrm{Var}\, X_G$ (cf. Janson, Łuczak and Ruciński (2000, Lemma 3.5)).

Lemma 4.10. *If (R,G) is a rooted graph, then*

$$\mathrm{Var}\, Y_G^R \asymp (1-p) \max_{H\subseteq G: e_H>0} \frac{\Psi_{G,R}^2}{\Psi_{H,R}} = (1-p)\frac{\Psi_{G,R}^2}{\Phi_{G,R}}.$$

Proof. Write $Y_G^R = \sum_{G'} \mathbb{I}_{G'}$, where $\mathbb{I}_{G'}$ is the indicator of $\{G' \subset \mathbb{G}(n,p)\}$ and the sum extends over all ordered copies of G in K_n that are rooted on, say, $\{1,\dots,|R|\} \subseteq V(K_n)$. Of course, $\mathbb{I}_{G'}$ and $\mathbb{I}_{G''}$ are independent if $E(G' \cap G'') = \emptyset$. For each $H \subseteq G$ with $V(H) \supseteq R$, the number of pairs (G', G'') intersecting in a copy of H is $\Theta(n^{v(H-R)} n^{2(v(G)-v(H))}) = \Theta(n^{2v(G-R)-v(H-R)})$. Hence

$$\begin{aligned}
\mathrm{Var}\, Y_G^R &= \sum_{G',G''} \mathbb{E}[(\mathbb{I}_{G'} - \mathbb{E}(\mathbb{I}_{G'}))(\mathbb{I}_{G''} - \mathbb{E}(\mathbb{I}_{G''}))] \\
&= \sum_{E(G')\cap E(G'')\neq\emptyset} [\mathbb{E}(\mathbb{I}_{G'}\mathbb{I}_{G''}) - \mathbb{E}(\mathbb{I}_{G'})\,\mathbb{E}(\mathbb{I}_{G''})] \\
&\asymp \sum_{\substack{R\subset H\subseteq G\\ e(H)>0}} n^{2v(G-R)-v(H-R)} (p^{2e(G)-e(H)} - p^{2e(G)}) \\
&\asymp \sum_{H\subseteq G: e(H)>0} n^{2v(G-R)-v(H-R)} p^{2e(G)-e(H)} (1-p) \\
&= (1-p) \sum_{H\subseteq G: e(H)>0} \frac{\Psi_{G,R}^2}{n^{v(H-R)} p^{e(H)}} \asymp (1-p) \max_{H\subseteq G: e(H)>0} \frac{\Psi_{G,R}^2}{\Psi_{H,R}}.
\end{aligned} \tag{4.32}$$

Note that in the fourth summation we omit, for simplicity, the unnecessary restriction $R \subset H$, since graphs H and $H \cup R$ give the same term. $\square$

Theorem 3.1 in Janson and Ruciński (2011) implies that for every $t > 1$ there is a constant $p_1 = p_1(t)$ such that if $n^{-1/m_R(G)} \leq p \leq p_1$, then

$$\mathbb{P}\left\{X_G^R \geq t\,\mathbb{E}X_G^R\right\} = \exp\left\{-\Omega_t(M_{R,G})\right\}, \tag{4.33}$$

where

$$M_{R,G} = \min\left\{(\Psi_{H,R})^{1/\alpha^*(H-R)} : H \subseteq G, e_H > 0\right\},$$

and, recall, $\alpha^*(G) = \alpha_G^*$ is the fractional independence number.

However, we will require an ordered version of inequality (4.33). Luckily, practically by making no changes to the proof of (4.33) under the same conditions we can obtain the bound of the same type:

$$\mathbb{P}\left\{Y_G^R \geq t\mu_G^R\right\} = \exp\left\{-\Omega_t(M_{R,G})\right\}. \tag{4.34}$$

As a matter of fact, if we do not care about absolute constants in the exponent, (4.34) implies (4.33) in view of (4.31).

4.2.2 Auxiliary facts

We further state a few simple auxiliary facts on subgraph counts. The advantage of using Ψ_G instead of $\mathbb{E}X_G$ is that it satisfies the log-modularity property: if G_1, G_2 are graphs, then

$$\Psi_{G_1 \cup G_2}\Psi_{G_1 \cap G_2} = \Psi_{G_1}\Psi_{G_2}. \tag{4.35}$$

Proposition 4.11. *If a graph G is strictly balanced, then the following facts hold.*

(i) *G is connected.*

(ii) *If H is a proper subgraph of G with $v_H > 0$ and $p \geq n^{-v_G/e_G}$, then*

$$\Psi_H \geq n^c, \tag{4.36}$$

where $c = c(H) = v_H - e_H v_G/e_G > 0$.

(iii) *There is a constant $b \in (0, v_G/e_G)$ such that for $p \in [n^{-v_G/e_G}, n^{-b}]$ and $H \subseteq G$ with $v_H > 0$ we have*

$$\Psi_G \leq \Psi_H. \tag{4.37}$$

Proof. (i) Suppose for contradiction that G is not connected and its connected components are $G_1, \ldots, G_m$. Since G is strictly balanced, $e(G_i) < m(G)v(G_i)$ for every i. Therefore

$$e(G) = \sum_i e(G_i) < m(G)\sum_i v(G_i) = m(G)v(G) = e(G),$$

which is false.

(ii) In order to prove (4.36), note that the condition $p \geq n^{-v_G/e_G}$ implies

$$\Psi_H = n^{v_H} p^{e_H} \geq n^{v_H - e_H v_G/e_G}.$$

Since G is strictly balanced, we have $e_G/v_G > e_H/v_H$, which is equivalent to the inequality $v_H - e_H v_G/e_G > 0$.

(iii) Inequality (4.37) follows from (ii) with $b = (v_G - \min_{H \subsetneq G} c(H))/e_G$:

$$\Psi_G = n^{v_G} p^{e_G} \leq n^{v_G - b e_G} \leq n^{c(H)} \leq \Psi_H, \qquad H \subsetneq G.$$

□

Proposition 4.12. *Let $F = G_1 \cup G_2$, where G_1 is strictly balanced. Assume that G_1 and G_2 have at least one common vertex, but do not coincide. If $p \geq n^{-1/m(G_1)}$ and $\mathbb{E}X_{G_1}, \mathbb{E}X_{G_2} = (\log n)^{O(1)}$, then*

$$\mathbb{P}\{X_F \geq 1\} = n^{-\Omega(1)}.$$

Proof. Let $H = G_1 \cap G_2$. By log-modularity (4.35), $\Psi_F = \Psi_{G_1}\Psi_{G_2}/\Psi_H$. Since H is a proper non-null subgraph of G_1, by Proposition 4.11.(ii) we have that $\Psi_H \geq n^c$ for some constant $c > 0$. Moreover, by assumption, $\Psi_{G_i} \asymp \mathbb{E}X_{G_i} \leq (\log n)^{O(1)}$. Therefore Markov's inequality implies

$$\mathbb{P}\{X_F \geq 1\} \leq \Psi_F = \frac{(\log n)^{O(1)}}{n^c} = n^{-\Omega(1)}.$$

□

Recall that that $H \subseteq G$ is extreme, if H satisfies $e_H/v_H = m(G)$. For completeness, we include the proof of the following elementary facts.

Proposition 4.13. *Suppose that G is a graph and $p \geq n^{-1/m(G)}$. Then the following statements are true.*

(i) If $G_0 \subseteq G$ is an extreme subgraph with the minimal number of vertices, then there exists a constant $\varepsilon = \varepsilon(G) > 0$ such that for $p \leq n^{\varepsilon - 1/m(G)}$ we have

$$\Phi_G = \Psi_{G_0}. \tag{4.38}$$

(ii) If $K \subseteq G$ is an exteme subgraph, which minimizes the quantity v_K/α^*_K, then there exists a constant $\varepsilon' = \varepsilon'(G) > 0$ such that for $p \leq n^{\varepsilon' - 1/m(G)}$

$$M^*_G \asymp \Psi_K^{1/\alpha^*_K}. \tag{4.39}$$

(iii) If $\Delta_G = 1$, then $m(G) = 1/2$ and for $p \geq n^{-2}$ we have $\Phi_G = n^2 p$ and $M^*_G \asymp n^2 p$.

Proof. (i) Clearly $p \geq n^{-1/m(G)}$ is equivalent to $np^{m(G)} \geq 1$. If $F \subseteq G$ is extreme, that is, $e(F)/v(F) = m(G)$, then $v(F) \geq v(G_0)$, therefore

$$\Psi_F = (np^{m(G)})^{v(F)} \geq (np^{m(G)})^{v(G_0)} = \Psi_{G_0}.$$

To deal with non-extreme F's define

$$\delta = \min \{v(F)/e(F) - 1/m(G) : F \subseteq G, 0 < e(F)/v(F) < m(G)\}.$$

Clearly $\delta > 0$. Suppose $F \subseteq G$ with $e(F) > 0$ satisfies $e(F)/v(F) < m(G)$. The definition of δ and the condition $p \geq n^{-1/m(G)}$ imply

$$\Psi_F \geq n^{v(F)-e(F)/m(G)} \geq n^{v(F)+e(F)(\delta-v(F)/e(F))} = n^{\delta e(F)} \geq n^{\delta}. \tag{4.40}$$

On the other hand, $m(G) = e(G_0)/v_{G_0}$ and $p \leq n^{\varepsilon-1/m(G)}$ imply that

$$\Psi_{G_0} \leq n^{v(G_0)-e(G_0)/m(G)+\varepsilon e(G_0)} = n^{\varepsilon e(G_0)}.$$

Therefore, choosing $\varepsilon = \delta/e(G_0)$, we get $\Psi_{G_0} \leq \Psi_F$.

(ii) By (4.13), for ε' small enough we have that $M_G^* \asymp \min_{H\subseteq G} \Psi_H^{1/\alpha_H^*}$. So it suffices to show that for any $F \subseteq G$ we have $\Psi_F^{1/\alpha_F^*} \geq \Psi_K^{1/\alpha_K^*}$. If $F \subseteq G$ is extreme, then $v(F)/\alpha_F^* \geq v(K)/\alpha_K^*$, therefore

$$\Psi_F^{1/\alpha_F^*} = (np^{m(G)})^{v(F)/\alpha_F^*} \geq (np^{m(G)})^{v(K)/\alpha_K^*} = \Psi_K^{1/\alpha_K^*}.$$

Choose $\varepsilon' \leq \delta\alpha_K^*/(\alpha_F^* e(K))$. If F is not extreme, then by (4.40), the fact that $e(K)/v(K) = m(G)$ and the assumption that $p \leq n^{\varepsilon'-1/m(G)}$ we have

$$\Psi_F^{\alpha_K^*/\alpha_F^*} \geq n^{\delta\alpha_K^*/\alpha_F^*} \geq n^{\varepsilon' e(K)} = n^{v(K)-e(K)/m(G)+\varepsilon' e(K)} \geq n^{v(K)}p^{e(K)} = \Psi_K.$$

(iii) Note that for every $H \subseteq G$, $e_H > 0$ is a matching with possibly some isolated vertices. Therefore $v_H \geq 2e_H$, whence $m(G) = e_{K_2}/v_{K_2} = 1/2$. Also

$$\Psi_H = n^2p \cdot n^{v_H-2}p^{e_H-1} \geq n^2pn^{v_H-2e_H} \geq n^2p,$$

therefore $\Phi_G = \Psi_{K_2} = n^2p$.

As for M_G^*, in view of (4.13) and the fact that $\Psi_{K_2}^{1/\alpha_{K_2}^*} = n^2p$, it suffices to show that for every $F \subseteq G$ we have $\Psi_H^{1/\alpha_H^*} \geq n^2p$, when $p \leq n^{-1}$. We have that H is a disjoint union of graphs $H_1, \ldots, H_m$, where each H_i is either an edge or a vertex. Clearly $\alpha_H^* = m$, therefore

$$\Psi_H^{1/\alpha_H^*} = (\Psi_{H_1} \ldots \Psi_{H_m})^{1/m} \geq \min_i \Psi_{H_i} = \min\left\{n^2p, n\right\} = n^2p.$$

□

4.3 Proofs

4.3.1 Lower bounds

Here we will prove Theorem 4.3, which gives the lower bound for strictly balanced G. Then we will extend it to Theorem 4.4 by a standard but somewhat tedious application of the second moment method. It is worth noting that in both proofs we will rely on the FKG inequality, which was used to prove the lower bound of the lower tail in Janson, Łuczak and Ruciński (1990).

Theorem 4.3 is a trivial corollary of the following result. Writing mG for a union of m vertex-disjoint copies of G, let

$$D_G^v = \max\{m : \mathbb{G}(n,p) \text{ contains } mG\}$$

be the size of a largest collection of vertex-disjoint copies.

Theorem 4.14 (Šileikis (2012))**.** *If G is strictly balanced and p is such that $v_G\lceil t\,\mathbb{E}X_G\rceil \le n$, then*

$$\mathbb{P}\{X_G = D_G^v = \lceil t\,\mathbb{E}X_G\rceil\} = \exp\{-O_t(\Psi_G)\}.$$

Proof of Theorem 4.3 . Let $\varepsilon \in (0, 1/e_G)$. Then assumption $p \le n^{\varepsilon - 1/m(G)}$ implies that $v_G\lceil t\,\mathbb{E}X_G\rceil = n^{1-\Omega(1)} \le n$, at least for n large enough. Therefore by the inequality $X_G \ge D_G^v$ and Theorem 4.14 we are done. □

Proof of Theorem 4.14

Let $\mathcal{G}$ be the set of all copies of G in K_n. Writing $m = \lceil t\,\mathbb{E}X_G\rceil$, consider the family

$$\mathbf{F} = \{\mathcal{S} \subset \mathcal{G} : |\mathcal{S}| = m, \text{ copies in } \mathcal{S} \text{ are vertex-disjoint}\}.$$

For every $\mathcal{S} \in \mathbf{F}$, we define events

$$A_{\mathcal{S}} = \{\mathcal{S} \text{ is the set of all copies of } G \text{ in } \mathbb{G}(n,p)\},$$

$$B_{\mathcal{S}} = \{\text{every copy } S \in \mathcal{S} \text{ appears in } \mathbb{G}(n,p)\}.$$

The events $\{A_{\mathcal{S}}\}_{\mathcal{S}\in\mathbf{F}}$ are mutually exclusive, therefore

$$\mathbb{P}\{X_G = D_G^v = m\} = \sum_{\mathcal{S}\in\mathbf{F}} \mathbb{P}\{A_{\mathcal{S}}\}. \tag{4.41}$$

Recall that $\operatorname{aut}(G)$ is the number of automorphisms of G. Then

$$|\mathbf{F}| = \frac{(n)_{v_G m}}{\operatorname{aut}(G)^m m!} \ge \frac{(n)_{v_G m}}{\operatorname{aut}(G)^m m^m}.$$

Using a standard inequality $(n)_m \ge (n/e)^m$, we get

$$|\mathbf{F}| \ge \frac{n^{v_G m}}{e^{v_G m}\operatorname{aut}(G)^m m^m} = \frac{n^{v_G m}}{m^m}\exp\{-\Theta(m)\}. \tag{4.42}$$

Note that $A_{\mathcal{S}} \subset B_{\mathcal{S}}$, therefore

$$\mathbb{P}\{A_{\mathcal{S}}\} = \mathbb{P}\{A_{\mathcal{S}}|B_{\mathcal{S}}\}\,\mathbb{P}\{B_{\mathcal{S}}\} = \mathbb{P}\{A_{\mathcal{S}}|B_{\mathcal{S}}\}\,p^{e_G m}. \tag{4.43}$$

Also, by symmetry the probability $q := \mathbb{P}\{A_{\mathcal{S}}|B_{\mathcal{S}}\}$ is independent of $\mathcal{S}$. Therefore from (4.41), (4.42), and (4.43) we infer that

$$\mathbb{P}\{X_G = D_G^v = m\} \geq \frac{(n^{v_G}p^{e_G})^m q}{m^m}\exp\{-\Theta(m)\}. \tag{4.44}$$

By the assumption that $p \geq n^{-1/m(G)}$ we have $\Psi_G = n^{v_G}p^{e_G} \asymp m/t$, so (4.44) implies

$$\mathbb{P}\{X_G = D_G^v = m\} \geq q\exp\{-\Theta_t(\Psi_G)\}. \tag{4.45}$$

Therefore it suffices to prove

$$q = \exp\{-O_t(\Psi_G)\}.$$

Fix $\mathcal{S} \in \mathbf{F}$ and let $F = \bigcup_{S\in\mathcal{S}} S$. Let $\mathbb{G}_{\mathcal{S}}(n,p)$ be the random graph $\mathbb{G}(n,p)$ conditioned on $B_{\mathcal{S}}$, i.e., the graph obtained by adding to F each of the remaining $\binom{n}{2} - e_G m$ edges with probability p, independently of others. We have

$$q = \mathbb{P}\Big(\bigcap_{G'\in\mathcal{G}\setminus\mathcal{S}}\{G' \not\subseteq \mathbb{G}_{\mathcal{S}}(n,p)\}\Big).$$

Notice that each of the events $\{G' \not\subseteq \mathbb{G}_{\mathcal{S}}(n,p)\}$ is decreasing. Therefore the FKG inequality (1.11) implies

$$q \geq \prod_{G'\in\mathcal{G}\setminus\mathcal{S}} \mathbb{P}\{G' \not\subseteq \mathbb{G}_{\mathcal{S}}(n,p)\}. \tag{4.46}$$

By Proposition 4.11.(i), G is connected, therefore F does not contain any copies of G apart from those in $\mathcal{S}$. Therefore every $G' \in \mathcal{G}\setminus\mathcal{S}$ intersects F in a proper subgraph of G. Let $N(F,H)$ be the number of copies of a graph H in F. Given $H \subsetneq G$, the number of copies $G' \in \mathcal{G}\setminus\mathcal{S}$ which intersect F in a copy of H is at most $N(F,H)(n-v_H)_{v_G-v_H} = N(mG,H)\frac{(n)_{v_G}}{(n)_{v_H}}$. The probability that such a copy G' does not exist in $\mathbb{G}_{\mathcal{S}}(n,p)$ is

$$1 - p^{e_G-e_H} \geq \exp\left\{-\frac{p^{e_G-e_H}}{1-p^{e_G-e_H}}\right\} \geq \exp\left\{-c_p\frac{p^{e_G}}{p^{e_H}}\right\},$$

where $c_p = 1/(1-p)$. Hence (4.46) implies

$$\begin{aligned} q &\geq \prod_{H\subsetneq G}\prod_{G'\cap F\cong H}\exp\left\{-c_p\frac{p^{e_G}}{p^{e_H}}\right\} \\ &\geq \prod_{H\subsetneq G}\exp\left\{-c_p N(mG,H)\cdot\frac{(n)_{v_G}p^{e_G}}{(n)_{v_H}p^{e_H}}\right\}. \end{aligned} \tag{4.47}$$

The assumption $\Psi_G \le \log^a n$ is equivalent to

$$p \le n^{-1/m(G)}(\log n)^{a/e_G}. \tag{4.48}$$

By (4.48), we have $p = o(1)$, therefore $c_p \to 1$. Also, $(n)_{v_G}p^{e_G} \asymp \Psi_G$ and $(n)_{v_H}p^{e_H} \asymp \Psi_H$. Hence (4.47) gives that

$$q \ge \exp\left\{-\Theta\left(\Psi_G \max_{H\subsetneq G} \frac{N(mG,H)}{\Psi_H}\right)\right\}. \tag{4.49}$$

To finish the proof, it is enough to show that for $H \subsetneq G$ we have $N(mG,H) = O_t(\Psi_H)$. If $H = \emptyset$ is the null graph (with zero vertices), then $N(mG,H) = 1 = \Psi_H$. Let us further assume that $v_H > 0$. By (4.48) and Proposition 4.11.(iii), the inequality $\Psi_G \le \Psi_H$ holds for n large enough, i.e., $\Psi_G = O(\Psi_H)$. If H is connected, then any copy of H counted in $N(mG,H)$ must lie entirely in one of the m copies of G. Hence

$$N(mG,H) = mN(G,H) \asymp_t \Psi_G = O_t(\Psi_H). \tag{4.50}$$

If H is not connected, then let $H_1,\dots,H_c$ be the connected components of H. Using (4.50), we get

$$N(mG,H) \le \prod_{i=1}^{c} N(mG,H_i) = \prod_{i=1}^{c} O_t(\Psi_{H_i}) = O_t\Big(\prod_{i=1}^{c}\Psi_{H_i}\Big) = O_t(\Psi_H),$$

the last equality following by log-modularity (4.35).

□

Proof of Theorem 4.4

In view of (4.30), we can consider the count Y_G of ordered copies of G. For simplicity, write $H = G_0$. Our aim is to prove that

$$\mathbb{P}\{Y_G \ge t\mu_G\} = \exp\{-O_t(\Psi_H)\}.$$

Fix a particular ordered copy of H in G (if there is more than one). Consider a rooted graph (R,F), where $R = V(H)$ and $F = G \setminus E(H)$.

Note that if we fix an ordered copy $H' \subset K_n$ of H, then for every ordered copy F' of (R,F) rooted on appropriately ordered $V(H')$, union $H' \cup F'$ is an ordered copy of G. In this case we say that H' *extends* to a copy of G.

To prove Theorem 4.4, we apply the so called *two-round exposure* method. Let $\hat{p} = 1 - \sqrt{1-p}$. It is easily checked that $\mathbb{G}(n,p)$ can be interpreted as a union of two independent graphs $\mathbb{G}(n,\hat{p})$ on the same vertex set. We refer to these two graphs as $\mathbb{G}_1$ and $\mathbb{G}_2$ and let $\mathbb{G} = \mathbb{G}_1 \cup \mathbb{G}_2$. To distinguish between parameters of $\mathbb{G}(n,p)$ and $\mathbb{G}(n,\hat{p})$, we indicate the probability in the subsript, say, $Y_{H,\hat{p}}$.

The idea of the proof is as follows: Theorem 4.14 implies that with desired probability $\mathbb{G}_1$ contains many more copies of H than expected; then we show that in $\mathbb{G}_2$ these copies extend to $t\mu_G$ copies of G.

Let Z be the number of copies of H in $\mathbb{G}_1$. Condition on the event that $Z \geq m := \lceil C_t \mu_H(\hat{p}) \rceil$. Here $C_t > 1$ is a constant to be defined later. Fix some m copies of H and call them $H'_1, \ldots, H'_m$. Let Y_i be the number of H'_i-rooted copies of F in $\mathbb{G}_2$. Note that $Y_1, \ldots, Y_m$ are identically distributed random variables with the same distribution as $Y^R_{F,\hat{p}}$.

If each copy H'_i extends to at least y copies of G in $\mathbb{G}_2$, that is, $Y_i \geq y$, then the number of copies of G in $\mathbb{G}$ is at least my. Set $y = t\mu_G(p)/m$, so that

$$\mathbb{P}\left\{Y_G \geq t\mu_G(p) | Z \geq m\right\} \geq \mathbb{P}\{Y_i \geq y, i = 1, \ldots, m\}. \tag{4.51}$$

The events $\{Y_i \geq y\}$, $i = 1, \ldots, m$, are increasing, so the FKG inequality implies

$$\mathbb{P}\left\{Y_i \geq y, i = 1, \ldots, m\right\} \geq \prod_{i=1}^{m} \mathbb{P}\left\{Y_i \geq y\right\} = \mathbb{P}\left\{Y^R_{F,\hat{p}} \geq y\right\}^m. \tag{4.52}$$

Using (4.51) and (4.52), we get

$$\begin{aligned}\mathbb{P}\left\{Y_G \geq t\mu_G(p)\right\} &\geq \mathbb{P}\left\{Z \geq m\right\} \mathbb{P}\left\{Y_G \geq t\mu_G(p) \,|\, Z \geq m\right\} \\ &\geq \mathbb{P}\{Z \geq m\} \mathbb{P}\left\{Y^R_{F,\hat{p}} \geq y\right\}^m. \end{aligned} \tag{4.53}$$

Since $m = \lceil C_t \mu_H(\hat{p}) \rceil$, $C_t > 1$, and H is strictly balanced, by Theorem 4.3 we have

$$\mathbb{P}\left\{Z \geq m\right\} = \exp\left\{-O_t(\Psi_H(\hat{p}))\right\} \tag{4.54}$$

On the other hand, it suffices to show that

$$\mathbb{P}\left\{Y^R_{F,\hat{p}} \geq y\right\} = \Omega_t(1). \tag{4.55}$$

Indeed, if (4.55) is true, then (4.53) and (4.54) imply

$$\begin{aligned}\mathbb{P}\left\{Y_G \geq t\mu_G(p)\right\} &= \exp\left\{-O_t\left(\Psi_H(\hat{p}) + m\right)\right\} \\ &= \exp\left\{-O_t(\Psi_H(p))\right\},\end{aligned}$$

where the second equality follows from the fact that $m \asymp_t \Psi_H(\hat{p}) \asymp \Psi_H(p)$. Let $\nu := \mu^R_F(\hat{p}) = (n - v_H)_{v_G - v_H} \hat{p}^{e_G - e_H}$. Recalling that $\hat{p} = 1 - \sqrt{1-p}$, by easy calculus one gets $\hat{p} \geq p/2$. Therefore

$$y = \frac{t\mu_G(p)}{C_t \mu_H(\hat{p})} = \frac{t(n)_{v_G} p^{e_G}}{C_t (n)_{v_H} \hat{p}^{e_H}} \leq \frac{2^{e_G} t}{C_t} (n - v_H)_{v_G - v_H} \hat{p}^{e_G - e_H} = \frac{2^{e_G} t}{C_t} \nu,$$

so, by choosing C_t large enough, we have $y \leq \nu/2$. Let $\sigma^2 := \operatorname{Var} Y^R_F(\hat{p})$. It is enough to prove that $\sigma^2 = O(\nu^2)$, since then the proof is completed by an application of (1.5):

$$\mathbb{P}\left\{Y^R_F \geq y\right\} \geq \mathbb{P}\left\{Y^R_F \geq \nu/2\right\} \geq \frac{\nu^2/4}{\nu^2/4 + \sigma^2} = \Omega(1).$$

Since $\hat{p}$ tends to zero, Lemma 4.10 implies that

$$\sigma^2 = \operatorname{Var} Y_F^R(\hat{p}) \asymp \frac{\Psi_{F,R}^2(\hat{p})}{\Phi_{F,R}(\hat{p})} \asymp \frac{\nu^2}{\Phi_{F,R}(\hat{p})},$$

so it is enough to show that $\Phi_{F,R}(\hat{p}) = \Omega(1)$. Since $\hat{p} \asymp p$, $\Phi_{F,R}(\hat{p}) \asymp \Phi_{F,R}(p) = \Phi_{F,R}$. We have

$$\Phi_{F,R} = \Psi_{K,R} \tag{4.56}$$

for some $K \subseteq F$ with $e(K) > 0$. Observe that $K \cup H$ is a subgraph of G and

$$\Psi_{K,R}\Psi_H = n^{v(K-R)+v(H)} p^{e(K)+e(H)} = \Psi_{K\cup H}. \tag{4.57}$$

Let $\varepsilon > 0$ be as in Proposition 4.13.(i). Then, recalling $H = G_0$,

$$\Phi_G = \Psi_H. \tag{4.58}$$

Therefore, by (4.56), (4.57) and (4.58),

$$\Phi_{F,R} = \frac{\Psi_{K,R}\Psi_H}{\Psi_H} = \frac{\Psi_{K\cup H}}{\Psi_H} \geq \frac{\Phi_G}{\Psi_H} = 1.$$

□

4.3.2 Upper bounds for small p

In the present section we give a short proof of Proposition 4.2 and prove Theorem 4.5.

Proof of Proposition 4.2. Let $\mathcal{F}_G = \{F = G_1 \cup G_2 : 0 < e(G_1 \cap G_2) < e_G\}$ be the set of all unlabelled graphs obtained by taking the union of two distinct copies of G with at least one common edge. For example, $\mathcal{F}_{K_3}$ consists of a single graph: K_4 with one edge removed. Recall that D_G^e is the maximum number of edge-disjoint copies of G in $\mathbb{G}(n,p)$. Obviously, either $X_F \geq 1$ for some $F \in \mathcal{F}_G$ or $D_G^e = X_G$, therefore

$$\mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\} \leq \mathbb{P}\{D_G^e \geq t\,\mathbb{E}X_G\} + \sum_{F\in\mathcal{F}_G} \mathbb{P}\{X_F \geq 1\}. \tag{4.59}$$

Clearly, it suffices to show that all probabilities on the right-hand side of (4.59) are bounded by $\exp\{-\Omega_t(\mathbb{E}X_G)\}$. Applying (4.9) with $\varepsilon = t-1$, we get

$$\mathbb{P}\{D_G^e \geq t\,\mathbb{E}X_G\} = \exp\{-\Omega_t(\mathbb{E}X_G)\}.$$

We bound the remaining terms in (4.59) using Proposition 4.12:

$$\begin{aligned}\mathbb{P}\{X_F \geq 1\} &= n^{-\Omega(1)} \\ &= \exp\{-\Omega(\log n)\} \\ &= \exp\{-\Omega(\mathbb{E}X_G)\}.\end{aligned}$$

□

Proof of Theorem 4.5. Let $H_1, \ldots, H_m$ be the minimal extreme subgraphs of G. Then H is one of the H_i's with the minimal number of vertices. Note that every H_i satisfies

$$\Psi_{H_i} = (np^{m(G)})^{v(H_i)} \geq (np^{m(G)})^{v(H)} = \Psi_H. \tag{4.60}$$

In view of (4.30), we can consider the count Y_G of ordered copies of G. For any $\eta > 0$ we have

$$\mathbb{P}\{Y_G \geq t\mu_G\} \leq \sum_{i=1}^{m} \mathbb{P}\{Y_{H_i} \geq (1+\eta)\mu_{H_i}\} \\ + \mathbb{P}\{Y_G \geq t\mu_G, Y_{H_i} < (1+\eta)\mu_{H_i}, i = 1, \ldots, m\}. \tag{4.61}$$

By (4.30), Proposition 4.2, and (4.60) for every i we have

$$\begin{aligned}\mathbb{P}\{Y_{H_i} \geq (1+\eta)\mu_{H_i}\} = \mathbb{P}\{X_{H_i} \geq (1+\eta)\,\mathbb{E}X_{H_i}\} &= \exp\{-\Omega_\eta(\,\mathbb{E}X_{H_i})\} \\ &= \exp\{-\Omega_\eta(\Psi_{H_i})\} \\ &= \exp\{-\Omega_\eta(\Psi_H)\}.\end{aligned}$$

Therefore it remains to bound the last probability in (4.61). For this, consider a rooted graph (R, F) with

$$R = V(H_1 \cup \cdots \cup H_m) \quad \text{and} \quad F = G \setminus E(H_1 \cup \cdots \cup H_m).$$

Let $r = |R|$ and $s = \sum_i e(H_i)$, so that $e(F) = e(G) - s$.

To inject G, we may first choose an ordered copy H_i' of each H_i, which we can do in $Y_{H_1} \cdot \ldots \cdot Y_{H_m}$ ways, and then choose a copy of F with respect to appropriately ordered set $V(H_1 \cup \cdots \cup H_m)$. Therefore

$$Y_G \leq Y_{H_1} \cdot \ldots \cdot Y_{H_m} \max_W Y_F^R(W),$$

where the maximum is taken over all r-tuples W of distinct vertices of K_n. If $Y_G \geq t\mu_G$ and $Y_{H_i} < (1+\eta)\mu_{H_i}$ for $i = 1, \ldots, m$, then

$$\exists W: \quad Y_F^R(W) \geq \frac{t\mu_G}{(1+\eta)^m \prod_i \mu_{H_i}} = (1+\eta+o(1))n^{v_G - r}p^{e_G - s}, \tag{4.62}$$

the equality being true for, say, $\eta = t^{-m-1} - 1$.

Since $\mu_F^R = \mathbb{E}Y_F^R \sim n^{v_G - r}p^{e_G - s}$, we can estimate the probability that (4.62) holds using the bound (4.34). Since there are $(n)_r$ ways to choose W, we get

$$\begin{aligned}\mathbb{P}\{Y_G \geq t\mu_G \text{ and } Y_{H_i} < (1+\eta)\mu_{H_i},\ i = 1, \ldots, m\} \\ &\leq (n)_r \mathbb{P}\left\{Y_F^R \geq (1+\eta+o(1))\mu_F^R\right\} \\ &= \exp\{r\log n - \Omega_\eta(M_{R,F})\},\end{aligned}$$

the second equality implied by (4.34). In view of the assumption $\Psi_H \leq \log n$, in order to conclude the proof it is enough to show that $M_{R,F} \gg \log n$. Recall that

$$M_{R,F} = \min\left\{\Psi_{H,R}^{1/\alpha^*(H-R)} : H \subseteq F, e(H) > 0\right\}.$$

Let us show that for every $H \subseteq F$ that $\Psi_{H,R} = n^{\Omega(1)}$. To see this, note that since $m_R(F) < m(G)$, we have $1/m(G) = 1/m_R(F) - \delta$ for some $\delta > 0$. By definition, $m_R(F) \geq e(H)/v(H-R)$, hence $1/m(G) \leq v(H-R)/e(H) - \delta$. Therefore the assumption that $p \geq n^{-1/m(G)}$ implies

$$\Psi_{H,R} = n^{v(H-R)} p^{e(H)} \geq n^{v(H-R)+e(H)[\delta - v(H-R)/e(H)]} = n^{\delta e(H)}.$$

□

4.3.3 More on small p: the whisk graph

Here we prove Theorem 4.6, which bounds, for small p, the upper tail of a specific graph K_3^+, for which Theorem 4.5 does not apply.

Proof of Theorem 4.6. Given a graph $T \subseteq K_n$ which is a union of triangles, let Z_T be the number of edges in $\mathbb{G}(n,p)$ with exactly one end in $V(T)$. Let $\mathcal{T}$ be the subgraph of $\mathbb{G}(n,p)$ formed by the union of all triangles in $\mathbb{G}(n,p)$.

We can ignore any "bad" event of probability $\exp\left\{-\Omega(n^3p^3)\right\}$. Applying Proposition 4.2 with $t = 1+\eta$, we can assume that $X_{K_3} < (1+\eta)\binom{n}{3}p^3$. In view of the assumption that $n^3p^3 \leq \log n$, by Proposition 4.12 we can assume that no two triangles intersect and no two triangles are connected by an edge (since this would create an intersection of a triangle and a whisk, the former being strictly balanced). Under such assumptions, a triangle can be extended to a whisk only if there is an edge connecting the triangle with the complement $[n] \setminus V(\mathcal{T})$. Therefore, $X_{K_3^+}$ is at most $Z_{\mathcal{T}}$, the number of edges with exactly one end in $\mathcal{T}$. It is enough to bound the probability

$$\mathbb{P}\left\{Z_{\mathcal{T}} \geq t\,\mathbb{E}X_{K_3^+} \text{ and } X_{K_3} < (1+\eta)\binom{n}{3}p^3\right\}.$$

Let $\mathbf{F}$ be the family of all possible graphs T, which can be obtained by taking a union of less than $(1+\eta)\binom{n}{3}p^3$ (not necessarily disjoint) triangles in K_n. For every $T \in \mathbf{F}$ define events

$$A_T = \{\mathcal{T} = T\}, \quad B_T = \{\mathcal{T} \supseteq T\}, \quad D_T = \left\{Z_T \geq t\,\mathbb{E}X_{K_3^+}\right\}.$$

For a given $T \in \mathbf{F}$, the number of possible cross-edges is at most $N := 3 \cdot (1+\eta)\binom{n}{3}p^3 \cdot n \sim n^4p^3/2$, so Z_T is stochastically dominated by a binomial random variable $Z \sim \mathrm{Bi}(N,p)$ with expectation $\mathbb{E}Z \sim n^4p^4/2 \sim \mathbb{E}X_{K_3^+}$. Therefore the Chernoff bound (1.6) gives

$$\begin{aligned}\mathbb{P}(D_T) \leq \mathbb{P}\left\{Z \geq t\,\mathbb{E}X_{K_3^+}\right\} &= \exp\left\{-\Omega_t\left(n^4p^4\right)\right\} \\ &= \exp\left\{-\Omega_t\left(n^3p^3\right)\right\}. \end{aligned} \tag{4.63}$$

Since $A_T \subseteq B_T$, we have

$$\mathbb{P}\{D_T, A_T\} = \mathbb{P}\{D_T, A_T | B_T\}\,\mathbb{P}\{B_T\}. \tag{4.64}$$

Consider $\mathbb{G}(n,p)$ conditioned on B_T, that is, a random graph obtained from T by adding each of the remaining $\binom{n}{2} - e(T)$ edges independently with probability p. Note that with respect to this random graph, A_T is decreasing and D_T is increasing. Therefore the FKG inequality (1.12) implies

$$\mathbb{P}\{D_T, A_T | B_T\} \le \mathbb{P}\{D_T | B_T\} \mathbb{P}\{A_T | B_T\}. \tag{4.65}$$

Note that B_T and D_T are independent, since they depend on separate sets of edges. Hence (4.64), (4.65), and the fact that $A_T \subseteq B_T$ imply

$$\begin{aligned}\mathbb{P}\{D_T, A_T\} &\le \mathbb{P}\{D_T | B_T\} \mathbb{P}\{A_T | B_T\} \mathbb{P}\{B_T\} \\ &= \mathbb{P}\{D_T\} \mathbb{P}\{A_T, B_T\} = \mathbb{P}\{D_T\} \mathbb{P}\{A_T\}.\end{aligned}$$

Hence, using (4.63), we get

$$\begin{aligned}&\mathbb{P}\left\{Z_{\mathcal{T}} \ge t\,\mathbb{E}X_{K_3^+}, X_{K_3} < (1+\eta)\binom{n}{3}p^3\right\} = \sum_{T \in \mathbf{F}} \mathbb{P}\{D_T, A_T\} \\ &\le \sum_T \mathbb{P}\{D_T\}\mathbb{P}\{A_T\} = \exp\left\{-\Omega_t(n^3p^3)\right\} \sum_T \mathbb{P}\{A_T\} \\ &\qquad = \exp\left\{-\Omega_t(n^3p^3)\right\} \mathbb{P}\left\{X_{K_3} < (1+\eta)\binom{n}{3}p^3\right\} \\ &\qquad = \exp\left\{-\Omega_t(n^3p^3)\right\},\end{aligned}$$

as desired. □

4.3.4 Upper bounds for stars

In the present section we prove Theorems 4.7-4.9, which give upper bounds for the upper tail, when G is the r-armed star $K_{1,r}$, $r \ge 2$.

Proof of Theorem 4.7

Let L be the random intersection graph, the vertices of which are the copies of G in $\mathbb{G}(n,p)$ and two vertices are connected by an edge if the corresponding copies have an edge in common. An easy graph-theoretic result, appearing implicitly in Spencer Spencer (1990) states that for any graph L

$$v_L \le \alpha_L + 2\beta_L \Delta_L, \tag{4.66}$$

where α_L is the independence number of L, β_L is the size of a largest induced matching in L, and, as usual, Δ_L is the maximum degree of L. To see that (4.66) is true, fix a maximal induced matching M and consider the set obtained by removing the neighbourhoods of the vertices in M:

$$I = V(L) \setminus U, \quad U = \bigcup_{v \in V(M)} N(v).$$

If there was an edge spanned by I, we could add it to M thus contradicting the maximality of M. Therefore I is independent and hence $|I| \le \alpha_L$. Since $|U| \le v_M \Delta_L = 2\beta_L \Delta_L$, (4.66) follows.

In our setting, $v_L = X_G$ and $\alpha_L = D_G^e$, while $\beta_L \le \sum_{F \in \mathcal{F}_G} D_F^e$. Recall that $\mathcal{F}_G$ is the set of all graphs formed by the union of two distinct edge-intersecting copies of G, as defined in the proof of Proposition 4.2 in §4.3.2.

For fixed vertices u and v of K_n, let $X_G(uv)$ be the number of copies of G in $\mathbb{G}(n,p)$ containing the edge uv. Fix a copy G' of G. If G' appears in $\mathbb{G}(n,p)$, then its degree in L is at most $\sum_{uv \in E(G')} X_G(uv)$. Hence $\Delta_L \le e_G \max_{uv} X_G(uv)$, where the maximum is taken over all edges of K_n. Clearly, the distribution of $X_G(uv)$ does not depend on uv. Therefore, when the choice of uv does not matter, we write X_G^e instead.

In view of the observations above, (4.66) implies

$$X_G \le D_G^e + \sum_{F \in \mathcal{F}_G} D_F^e \cdot 2e_G \max_{uv} X_G(uv). \tag{4.67}$$

Choose δ such that $t = 1 + \delta(1 + 2e_G|\mathcal{F}_G|)$. Then by (4.67) for every $d > 0$

$$\begin{aligned} \mathbb{P}\{X_G \ge t\, \mathbb{E}X_G\} &\le \mathbb{P}\left\{D_G^e \ge (1+\delta)\, \mathbb{E}X_G\right\} \\ &\quad + \mathbb{P}\left\{ \sum_{F \in \mathcal{F}_G} D_F^e \ge \frac{\delta|\mathcal{F}_G|\, \mathbb{E}X_G}{d} \right\} + \mathbb{P}\left\{2e_G \max_{uv} X_G(uv) \ge 2e_G d\right\} \\ &\le \mathbb{P}\{D_G^e \ge (1+\delta)\, \mathbb{E}X_G\} + \sum_{F \in \mathcal{F}_G} \mathbb{P}\left\{D_F^e \ge \frac{\delta\, \mathbb{E}X_G}{d}\right\} + \binom{n}{2} \mathbb{P}\left\{X_G^e > d\right\}. \end{aligned} \tag{4.68}$$

The conditions on p imply that $\log(1/p) \asymp \log n$. Therefore, in order to prove the theorem, it is sufficient to show that (4.68) is bounded by

$$\exp\left\{-\Omega_{\delta,\gamma}\left(\min\left\{\Psi_G, \Psi_G^{1/r} \log n\right\}\right)\right\}.$$

By inequality (4.9), the first term in (4.68) is at most

$$\exp\{-\Omega_\delta(\, \mathbb{E}X_G)\} \le \exp\{-\Omega_\delta(\min\{\Psi_G, \Psi_G^{1/r} \log n\})\}.$$

To bound the second term in (4.68), we apply the following slightly weaker but more convenient form of (4.8). Noting that $\varphi(\varepsilon) \ge (1+\varepsilon)\log\frac{1+\varepsilon}{\mathrm{e}}$ and using a substitute $x = (1+\varepsilon)\, \mathbb{E}X_G$, we get an inequality

$$\mathbb{P}\left\{D_G^e \ge x\right\} \le \exp\left\{-x \log \frac{x}{\mathrm{e}\, \mathbb{E}X_G}\right\}, \qquad \mathrm{e} = 2.71\ldots. \tag{4.69}$$

Hence, for $F \in \mathcal{F}_G$, by (4.69),

$$\mathbb{P}\left\{D_F^e \ge \frac{\delta\, \mathbb{E}X_G}{d}\right\} \le \exp\left\{-\frac{\delta\, \mathbb{E}X_G}{d} \log \frac{\delta\, \mathbb{E}X_G}{\mathrm{e}d\, \mathbb{E}X_F}\right\}.$$

Put

$$d = \max\{\log n, \Psi_G^{(r-1)/r}\} = \max\left\{\log n, n^{(r^2-1)/r} p^{r-1}\right\}.$$

Then

$$\mathbb{P}\left\{D_F^e \geq \frac{\delta\, \mathbb{E} X_G}{d}\right\} = \exp\left\{-\Omega_\delta\left(\min\left\{\frac{\Psi_G}{\log n}, \Psi_G^{1/r}\right\}\log\frac{\Psi_G}{d\Psi_F}\right)\right\}. \tag{4.70}$$

It remains to check that the logarithmic factor in (4.70) is of order $\log n$. Note that by log-modularity (4.35), $\Psi_G/\Psi_F = \Psi_H/\Psi_G$, where H is the intersection of the two copies of G that constitute F. Hence, it suffices to show that $\Psi_H/(d\Psi_G) = \Omega(n^c)$ for some $c > 0$, probably depending on γ. Consider two cases.

Case (i): $\Psi_G \leq \log^{r/(r-1)} n$. Then $d = \log n$. By Proposition 4.11.(ii), we have that Ψ_H is at least some positive power of n. Therefore

$$\frac{\Psi_H}{d\Psi_G} = \frac{\Psi_H}{\log^{O(1)} n} = \Omega(n^c), \quad c > 0.$$

Case (ii): $\Psi_G > \log^{r/(r-1)} n$. Then $d = \Psi_G^{(r-1)/r}$. Note that H is a k-armed star with $k \in \{1, \ldots, r-1\}$. Restriction $p \leq n^{-1-1/r+1/r^2-\gamma}$ implies that $np = n^{-\Omega(1)}$, therefore $\Psi_H = n^{k+1}p^k \geq n^r p^{r-1}$. Hence

$$\frac{\Psi_H}{d\Psi_G} = \frac{\Psi_H}{\Psi_G^{1+(r-1)/r}} \geq \frac{n^r p^{r-1}}{n^{2r+1-1/r} p^{2r-1}} = n^{-r-1+1/r} p^{-r} \geq n^{r\gamma}, \tag{4.71}$$

where the last inequality follows from restriction $p \leq n^{-1-1/r+1/r^2-\gamma}$.

Concerning the third term in (4.68), observe that X_G^{uv} is at most $\binom{\deg(u)}{r-1} + \binom{\deg(v)}{r-1}$, where $\deg(v)$ is the degree of vertex v in $\mathbb{G}(n,p)$. Write $\mathrm{Bi}(n,p)$ for a binomial random variable. Noting that $\deg(v) = \mathrm{Bi}(n-1,p)$, we get that

$$\begin{aligned}
\mathbb{P}\{X_G^e > d\} &\leq \mathbb{P}\left\{\max\left\{\binom{\deg(u)}{r-1}, \binom{\deg(v)}{r-1}\right\} > d\right\} \\
&\leq 2\mathbb{P}\left\{\binom{\mathrm{Bi}(n-1,p)}{r-1} > d/2\right\} \\
&\leq 2\mathbb{P}\left\{\mathrm{Bi}(n-1,p) > ((r-1)!d/2)^{1/(r-1)}\right\}.
\end{aligned}$$

Then the Chernoff bound (1.7) yields

$$\mathbb{P}\{X_K^e > d\} \leq \exp\left\{-\Omega\left(d^{1/(r-1)}\log\frac{d^{1/(r-1)}}{np}\right)\right\}. \tag{4.72}$$

The logarithmic factor in the exponent in (4.72) is $\Omega(\log n)$, since

$$\frac{d^{1/(r-1)}}{np} = \frac{n^{1+2/(r-1)}p^{1+1/(r-1)}}{np} = (n^2p)^{1/(r-1)} \geq (n^{1-1/r})^{1/(r-1)} = n^{\Omega(1)},$$

the inequality following from the assumption that p is above the threshold: $p \geq n^{-1-1/r}$. Recalling that $d = \max\left\{\log n, \Psi_G^{(r-1)/r}\right\}$, we get that the order of the exponent in (4.72) is

$$\max\left\{(\log n)^{1+\frac{1}{r-1}}, \Psi_G^{1/r}\log n\right\}.$$

We can ignore the factor $\binom{n}{2}$ in (4.68), as it contributes to the exponent only an additive term $O(\log n)$. This completes the proof, because

$$\max\{(\log n)^{1+1/(r-1)}, \Psi_G^{1/r}\log n\} \geq \min\{\Psi_G, \Psi_G^{1/r}\log n\}.$$

□

Let us further write $X = X_{K_{1,r}}$ for the number of r-armed stars, and, for simplicity, consider $\mathbb{G}(2n,p)$ instead of $\mathbb{G}(n,p)$. Given a bipartition $P = (V_1, V_2)$ of $V = \{1,\ldots,2n\}$ into sets of equal size, let X_P be the number of stars with the center in V_1 and the remaining r vertices in V_2. Observe that the distribution of X_P does not depend on the actual choice of P. In the proofs of Theorems 4.8 and 4.9 we will use the following lemma.

Lemma 4.15. *Let X be the number of r-armed stars in $\mathbb{G}(2n,p)$ and let $Z = X_P$ be the number of copies with the center in $V_1 = \{1,\ldots,n\}$ and the remaining vertices in $V_2 = \{n+1,\ldots,2n\}$. Then, for $u \in (1,t)$,*

$$\mathbb{P}\{X \geq t\,\mathbb{E}X\} \leq \frac{t-u\rho}{\rho(t-u)}\mathbb{P}\{Z \geq u\,\mathbb{E}Z\},$$

where $\rho = \rho(n)$ is a function of n such that

$$\rho \to 2^{-r-1}, \quad \text{as } n \to \infty.$$

Thus in order to get an upper bound for X, we may instead study a simpler random variable X_P. Actually, we have that $X_P = \binom{B_1}{r} + \cdots + \binom{B_n}{r}$, where $B_1,\ldots,B_n$ are independent copies of a binomial random variable $B \sim \mathrm{Bi}(n,p)$.

Lemma 4.15 and its proof follow easily from the idea by DeMarco and Kahn (2011), who showed that it is enough to study the upper tail of triangles in a random tripartite graph.

Proof of Lemma 4.15. Observe that if instead of a fixed P we choose a random bipartition $\mathcal{P}$, the distribution of $X_{\mathcal{P}}$ is remains unchanged, as long as $\mathcal{P}$ is independent of $\mathbb{G}(2n,p)$. This is because the distribution of X_P is the same for any P, and

$$\mathbb{P}\{X_{\mathcal{P}} \leq x\} = \sum_P \mathbb{P}\{X_{\mathcal{P}} \leq x \,|\, \mathcal{P} = P\}\,\mathbb{P}\{\mathcal{P} = P\}.$$

Further assume that $\mathcal{P}$ is chosen uniformly, that is, every bipartition is chosen with probability $1/\binom{2n}{n}$. For every $u \in (1, t)$ we have

$$\mathbb{P}\{X_{\mathcal{P}} \geq u\,\mathbb{E}X_{\mathcal{P}}\} \geq \mathbb{P}\{X_{\mathcal{P}} \geq u\,\mathbb{E}X_{\mathcal{P}} \,|\, X \geq t\,\mathbb{E}X\}\,\mathbb{P}\{X \geq t\,\mathbb{E}X\}\,. \tag{4.73}$$

Our aim is to bound the conditional probability away from zero. The probability that a fixed star has the center in (random) V_1 and the remaining vertices in V_2 is

$$\rho := \frac{\binom{2n-r-1}{n-1}}{\binom{2n}{n}} = \frac{n(n)^r}{(2n)_{r+1}} \to 2^{-r-1}, \quad n \to \infty.$$

Write $x = t\,\mathbb{E}X$ and $y = u\,\mathbb{E}X_{\mathcal{P}}$. Observe that $X_{\mathcal{P}}$, conditioned on X, takes values in the interval $[0, X]$. Moreover, $\mathbb{E}[X_{\mathcal{P}} \,|\, X] = \rho X$, so

$$\mathbb{E}X_{\mathcal{P}} = \mathbb{E}\,\mathbb{E}[X_{\mathcal{P}} \,|\, X] = \rho\,\mathbb{E}X. \tag{4.74}$$

By Markov's inequality

$$\mathbb{P}\{X - X_{\mathcal{P}} > X - y \,|\, X\} \leq \frac{\mathbb{E}[X - X_{\mathcal{P}} \,|\, X]}{X - y} = \frac{X - \rho X}{X - y} = \frac{1-\rho}{1 - y/X},$$

so

$$\mathbb{P}\{X_{\mathcal{P}} \geq y \,|\, X\} = 1 - \mathbb{P}\{X - X_{\mathcal{P}} > X - y \,|\, X\} \geq 1 - \frac{1-\rho}{1-y/X} =: f(X).$$

Since f is an increasing function,

$$\mathbb{P}\{X_{\mathcal{P}} \geq y \,|\, X \geq x\} \geq f(x) = \frac{x - y - x + \rho x}{x - y} = \frac{\rho x - y}{x - y}.$$

As $x = t\,\mathbb{E}X$ and, by (4.74), $y = u\rho\,\mathbb{E}X$, plugging in and simplifying yields

$$\mathbb{P}\{X_{\mathcal{P}} \geq y \,|\, X \geq x\} \geq \frac{\rho(t-u)}{t - u\rho}. \tag{4.75}$$

□

Proof of Theorem 4.8

In view of Lemma 4.15, we can consider Z instead of X. For each vertex $v \in V_1$, let random variable B_v be the number of neighbours of v in V_2. Then

$$Z = \sum_{v \in V_1} \binom{B_v}{r} \leq \frac{1}{r!} \sum_{v \in V_1} B_v^r =: \frac{1}{r!} S_n.$$

Let $l = \ln 1/p$. We have $\mathbb{E}Z = n\binom{n}{r}p^r \sim n^{r+1}p^r/r!$, therefore it suffices to prove

$$\mathbb{P}\left\{S_n \geq tn^{r+1}p^r\right\} \leq \exp\left\{-\Omega_t(n^2p^rl)\right\}, \qquad t > 1. \tag{4.76}$$

If $h > 0$, Markov's inequality implies

$$\mathbb{P}\left\{S_n \geq tn^{r+1}p^r\right\} = \mathbb{P}\left\{\mathrm{e}^{hS_n} \geq \mathrm{e}^{htn^{r+1}p^r}\right\} \leq \mathbb{E}\mathrm{e}^{hS_n} \exp\{-htn^{r+1}p^r\}. \quad (4.77)$$

Since $S_n = \sum_{v \in V_1} B_v^r$ is the sum of n independent random variables, we get

$$\mathbb{E}\mathrm{e}^{hS_n} = \left(\mathbb{E}\exp\{hB^r\}\right)^n, \quad (4.78)$$

where $B \sim \mathrm{Bi}(n,p)$. To find a good choice of h, let us consider a lower bound of the right-hand side of (4.77). Since $\mathbb{P}\{B = n\} = p^n = e^{-nl}$, we have

$$\mathbb{E}\mathrm{e}^{hS_n} \exp\{-htn^{r+1}p^r\} \geq \exp\left\{hn^{r+1} - n^2 l - htn^{r+1}p^r\right\}.$$

So, in order to have a chance that the right-hand side of (4.77) is exponentially small, we need to choose h of order l/n^{r-1}. Let $\delta, \varepsilon > 0$ be small constants depending on r and t only, which we will choose later. Set $h = \varepsilon l/n^{r-1}$. Writing $q = 1 - p$, we have

$$\begin{aligned}\mathbb{E}\exp\{hB^r\} &\leq \exp\{h[(1+\delta)np]^r\} + \sum_{k=\lceil(1+\delta)np\rceil}^{n} \exp\{hk^r\}\mathbb{P}\{B = k\} \\ &= \exp\{(1+\delta)^r \varepsilon np^r l\} + \sum_k E(k), \quad (4.79)\end{aligned}$$

where $E(k) = \exp\{\varepsilon lk^r/n^{r-1}\}\mathbb{P}\{B = k\}$. A major part of the remaining proof is to obtain a uniform, that is, independent of k bound on $E(k)$. Substituting $k = (1+d)np$, we rewrite

$$E(k) = \exp\left\{\varepsilon np^r l(1+d)^r\right\}\mathbb{P}\left\{B = (1+d)np\right\}. \quad (4.80)$$

Applying the Chernoff bound (1.6) for the probabilities $\mathbb{P}\{B = (1+d)np\}$, we obtain

$$E(k) \leq \exp\left\{\varepsilon np^r l(1+d)^r - np\varphi(d)\right\} = \exp\left\{npf(d)\right\}, \quad (4.81)$$

where $f(d) := \varepsilon p^{r-1} l(1+d)^r - \varphi(d)$. Note that d takes values in the interval $[\delta, q/p]$. Straighforward calculation gives that

$$\begin{aligned}f'(d) &= r\varepsilon p^{r-1} l(1+d)^{r-1} - \ln(1+d), \\ f''(d) &= r(r-1)\varepsilon p^{r-1} l(1+d)^{r-2} - \frac{1}{d+1}.\end{aligned}$$

Since $r(r-1)\varepsilon p^{r-1} l > 0$, function f'' is increasing on $[0, \infty)$ and therefore f' is convex. Simple calculus shows that $p^{r-1} l \leq 1/\mathrm{e}(r-1)$. Therefore

$$f'(\delta) \leq \frac{r}{\mathrm{e}(r-1)}\varepsilon(1+\delta)^{r-1} - \ln(1+\delta).$$

On the other hand

$$f'(q/p) = r\varepsilon p^{r-1} l(1/p)^{r-1} - \log 1/p = l(r\varepsilon - 1).$$

Recalling that $\delta = \delta(r,t)$, let us choose $\varepsilon = \varepsilon(r,t)$ small enough so that $f'(\delta), f'(q/p) \le 0$. Since f' is convex, we get that $f'(d) \le 0$ for every $d \in [\delta, q/p]$, whence f is decreasing, and so $f(d) \le f(\delta)$. Then (4.81) implies

$$E(k) \le \exp\{npf(\delta)\} = \exp\{(1+\delta)^r \varepsilon np^r l - np\varphi(\delta)\}.$$

Since the sum in (4.79) has at most $n-1$ terms, we get

$$\begin{aligned}\mathbb{E}\exp\{hB^r\} &\le \exp\{(1+\delta)^r \varepsilon np^r l\}(1 + n\exp\{-np\varphi(\delta)\}) \\ &=: \exp\{(1+\delta)^r \varepsilon np^r l\} F_\delta(n,p). \end{aligned} \tag{4.82}$$

Recalling that $h = \varepsilon l / n^{r-1}$, by (4.77), (4.78) and (4.82), we have

$$\begin{aligned}\mathbb{P}\left\{S_n \ge tn^{r+1}p^r\right\} &\le \exp\left\{(1+\delta)^r \varepsilon n^2 p^r l - htn^{r+1}p^r\right\} F_\delta(n,p)^n. \\ &= \exp\left\{-\left[t - (1+\delta)^r\right]\varepsilon n^2 p^r l\right\} F_\delta(n,p)^n.\end{aligned}$$

Note that the assumption $p \ge n^{-1/r}$ implies $F_\delta(n,p)^n = 1 + o(1)$. Choose $\delta > 0$ such that $t > (1+\delta)^r$. Since $\varepsilon = \varepsilon(r,t)$, we have

$$\mathbb{P}\left\{S_n \ge tn^{r+1}p^r\right\} \le \exp\left\{-\Omega_t(n^2 p^r l)\right\},$$

as desired.

$\square$

Proof of Theorem 4.9

By the argument at the beginning of the proof of Theorem 4.8, we can consider S_n instead of X. Assumptions on p imply that $\log 1/p \asymp \log n$. Therefore, writing $M := n^{1+1/r}p$, we aim to show

$$\mathbb{P}\{S_n \ge tM^r\} \le \exp\{-cM\log n\}, \tag{4.83}$$

where $S_n = \sum_{v \in V_1} B_v^r$ and B_v's are independent copies of a random variable $B \sim \mathrm{Bi}(n,p)$.

The reason why the proof of Theorem 4.8 did not work for smaller p was that in the exponential moment of B^r the contribution of the largest values of B was too large. However, we can exclude the event that some B_v is large as follows. By the Chernoff bound (1.7) and the assumption $p \ge Cn^{-1}\log n$,

$$\mathbb{P}\left\{\max_v B_v > M\right\} \le n\exp\left\{-M\log\left(\frac{M}{enp}\right)\right\} = \exp\left\{-\Omega_r\left(M\log n\right)\right\}.$$

Therefore, instead of S_n we can further consider the tail of $\tilde{S}_n = \sum_{v \in V_1} \tilde{B}_v$, where $\tilde{B}_v$ is B_v, conditioned on the event $\{B_v \le M\}$. Clearly $\tilde{B}_v$'s are independent copies of a random variable $\tilde{B}$ such that for $m \in \mathbb{Z} \cap [0, M]$ we have

$$\mathbb{P}\left\{\tilde{B} = m\right\} = \frac{1}{p_M}\mathbb{P}\{B = m\},$$

where $p_M = \mathbb{P}\{B \le M\}$. Then

$$\mathbb{P}\{S_n \ge tM^r\} \le \exp\{-\Omega_r(M \log n)\} + p_M^n \mathbb{P}\left\{\tilde{S}_n \ge tM^r\right\}. \tag{4.84}$$

Similarly as in the proof of Theorem 4.9, for $h > 0$ we have

$$\mathbb{P}\left\{\tilde{S}_n \ge tM^r\right\} \le \left(\mathbb{E}\exp\left\{h\tilde{B}^r\right\}\right)^n \exp\{-htM^r\}. \tag{4.85}$$

Assume, for a moment, that M is an integer. We have

$$-\log \mathbb{P}\left\{\tilde{B}_v = M\right\} \asymp M \log(M/enp) \asymp M \log n.$$

Therefore the right-hand side of (4.85) is at least

$$\exp\{hnM^r - \Theta(nM \log n) - htM^r\}.$$

This suggests that we should take h of order $M^{1-r} \log n$. Let $\delta, \varepsilon > 0$ be small constants depending on r and t, which we will choose later. Set $h = \varepsilon M^{1-r} \log n$. Writing $q = 1 - p$, we have

$$\begin{aligned} p_M \exp\{h\tilde{B}_v^r\} &\le \exp\{h[(1+\delta)np]^r\} + \sum_{k=\lceil(1+\delta)np\rceil}^{\lfloor M \rfloor} \exp\{hk^r\}\mathbb{P}\{B = k\} \\ &= \exp\{(1+\delta)^r \varepsilon n^{1/r} p \log n\} + \sum_k E(k), \end{aligned} \tag{4.86}$$

where $E(k) = \exp\{\varepsilon M^{1-r} k^r \log n\}\mathbb{P}\{B = k\}$. Substituting $k = (1+d)np$, and using the Chernoff bound (1.6), we have

$$E(k) \le \exp\left\{\varepsilon(1+d)^r (np)^r M^{1-r} \log n - np\varphi(d)\right\} = \exp\{npf(d)\}, \tag{4.87}$$

where $f(d) := \varepsilon(1+d)^r n^{1/r-1} \log n - \varphi(d)$. Analysis of f is very similar to that in the proof of Theorem 4.8. Note that d ranges in $[\delta, D]$, where $D = M/(np) - 1 = n^{1/r} - 1$. We have

$$\begin{aligned} f'(d) &= r\varepsilon(1+d)^{r-1} n^{1/r-1} \log n - \log(1+d), \\ f''(d) &= r(r-1)\varepsilon(1+d)^{r-2} n^{1/r-1} \log n - \frac{1}{d+1}. \end{aligned}$$

It is easy to see that $n^{1/r-1} \log n \le r/(r-1)$, therefore

$$f'(\delta) \le r^2/(r-1)\varepsilon(1+\delta)^{r-1} - \log(1+\delta). \tag{4.88}$$

On the other hand,

$$f'(D) = f'(n^{1/r} - 1) = (r\varepsilon - 1/r) \log n.$$

So, by choosing ε small enough (depending on r and δ), we make sure that $f'(\delta), f'(D) \le 0$. Since f'' is increasing on $[0, \infty)$, f' is convex and hence nonpositive on $[\delta, D]$. Consequently f decreases on $[\delta, D]$ and therefore

$$f(d) \le f(\delta), \qquad d \in [\delta, D]. \tag{4.89}$$

Notice that the sum in (4.86) has less than n terms. By (4.87) and (4.89), we get that (4.86) is at most

$$\exp\left\{\Big((1+\delta)^r \varepsilon n^{1/r} p \log n\Big)\right\}(1 + n\exp\{-np\varphi(\delta)\}).$$

Therefore

$$p_M^n \mathbb{P}\left\{\tilde{S}_n \ge tM^r\right\} \le \exp\{-cM\log n\}(1 + \exp\{\log n - np\varphi(\delta)\})^n, \tag{4.90}$$

where $c = \varepsilon(t - (1+\delta)^r)$. Let us choose δ so that $(1+\delta)^r < t$. To see that the second factor on the right-hand side of (4.90) can be ignored, let, say, $C = 3/\varphi(\delta)$. Then the hypothesis $p \ge Cn^{-1}\log n$ implies

$$(1 + \exp\{\log n - np\varphi(\delta)\})^n \le (1 + \exp\{-2\log n\})^n \le \mathrm{e}^{1/n} \to 1.$$

We conclude the proof by combining (4.84) and (4.90).

□

4.3.5 Below the threshold

The upper tail is usually studied under an assumption that $p \ge n^{-1/m(G)}$. Although it is not such an interesting thing if $p < n^{-1/m(G)}$, but it is not entirely trivial. The following proposition is a kind of folklore.

Proposition 4.16. *If $p \le n^{-1/m(G)}$, which is equivalent to $\Phi_G \le 1$, then we have*

$$\mathbb{P}\{X_G \ge t\,\mathbb{E}X_G\} \asymp \Phi_G. \tag{4.91}$$

Sketch of Proof. (i) If $t\,\mathbb{E}X_G \le 1$, then the left hand side is just

$$\mathbb{P}\{X_G \ge 1\} = 1 - \mathbb{P}\{X_G = 0\},$$

therefore (4.91) easily follows from (4.5) and the fact that $1 - e^{-x} \asymp x$ as $x \to 0$.

(ii) If $t\,\mathbb{E}X_G > 1$, then the upper bound is easy, because for every $H \subseteq G$ Markov's inequality implies

$$\mathbb{P}\{X_G \ge t\,\mathbb{E}X_G\} \le \mathbb{P}\{X_H \ge 1\} \le \mathbb{E}X_H.$$

However, for the lower bound the easiest way we know would be to proceed in the same fashion as in the proof of Theorem 4.4. Here we give just a sketch of the proof. Let $H \subseteq G$ be such that $\Phi_G = \Psi_H$. Consider the two-round exposure, i.e., treat $\mathbb{G}(n,p)$ as a union $\mathbb{G}_1 \cup \mathbb{G}_2$, where $\mathbb{G}_1$ and $\mathbb{G}_2$ are independent $\mathbb{G}(n,p')$'s with appropriate p'. The probability $\mathbb{P}\{X_G \geq t\,\mathbb{E}X_G\}$ is at least the probability that (a) in $\mathbb{G}_1$ there is at least one copy of H and (b) in $\mathbb{G}_2$ this copy extends to $t\,\mathbb{E}X_G$ copies of G. As we have seen in the case (i), the event (a) happens with probability of order $\Phi_H = \Psi_H = \Phi_G$. Assuming $p \leq cn^{-1/m(G)}$ for some sufficiently small constant c, and using the second moment method, one can show that the event (b) happens with probability $\Omega(1)$. □

Bibliography

Azuma, K.: 1967, Weighted sums of certain dependent random variables, *Tôhoku Math. J. (2)* **19**, 357–367.

Bentkus, V.: 2001, An inequality for large deviation probabilities of sums of bounded i. i. d. random variables., *Lith. Math. J.* **41**(2), 112–119.

Bentkus, V.: 2004, On Hoeffding's inequalities, *Ann. Probab.* **32**(2), 1650–1673.

Bentkus, V.: 2007, On measure concentration for separately Lipschitz functions in product spaces, *Israel J. Math.* **158**, 1–17.

Bollobás, B.: 1981, Threshold functions for small subgraphs, *Math. Proc. Cambridge Philos. Soc.* **90**(2), 197–206.

Bollobás, B.: 1986, *Combinatorics. Set systems, hypergraphs, families of vectors and combinatorial probability.*, Cambridge University Press.

Bollobás, B. and Leader, I.: 1991, Isoperimetric inequalities and fractional set systems, *J. Combin. Theory Ser. A* **56**(1), 63–74.

Bollobás, B. and Riordan, O.: 2006, *Percolation*, Cambridge University Press.

Chatterjee, S.: 2011, The missing log in large deviations for triangle counts, *Random Struct. Algorithms* . doi: 10.1002/rsa.20381.

DeMarco, B. and Kahn, J.: 2011, Upper tails for triangles, *Random Struct. Algorithms* .
URL: *http://dx.doi.org/10.1002/rsa.20382*

DeMarco, B. and Kahn, J.: 2012+, Upper tails for cliques. Preprint available from `http://arxiv.org/pdf/1111.6687v1`.

Dubhashi, D. P. and Panconesi, A.: 2009, *Concentration of measure for the analysis of randomized algorithms*, Cambridge Univ. Press.

Dzindzalieta, D., Juškevičius, T. and Šileikis, M.: 2012+, Optimal probability inequalities for random walks related to problems in extremal combinatorics. To appear in *SIAM J. Discrete Math.* Preprint available from http://arxiv.org/pdf/1111.6332.

Erdös, P.: 1945, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51**, 898–902.

Erdős, P. and Rényi, A.: 1960, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**, 17–61.

Harper, L. H.: 1966, Optimal numberings and isoperimetric problems on graphs, *J. Combinatorial Theory* **1**, 385–393.

Hoeffding, W.: 1963, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58**, 13–30.

Janson, S.: 1990, Poisson approximation for large deviations, *Random Struct. Algorithms* **1**(2), 221–229.
URL: *http://dx.doi.org/10.1002/rsa.3240010209*

Janson, S., Łuczak, T. and Ruciński, A.: 1990, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, *Random graphs '87 (Poznań, 1987)*, Wiley, Chichester, pp. 73–87.

Janson, S., Łuczak, T. and Ruciński, A.: 2000, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York.

Janson, S., Oleszkiewicz, K. and Ruciński, A.: 2004, Upper tails for subgraph counts in random graphs, *Israel J. Math.* **142**, 61–92.

Janson, S. and Ruciński, A.: 2000, The deletion method for upper tail estimates. Preprint, available from `http://www.math.uu.se/ svante`.

Janson, S. and Ruciński, A.: 2002, The infamous upper tail, *Random Struct. Algorithms* **20**(3), 317–342.
URL: *http://dx.doi.org/10.1002/rsa.10031*

Janson, S. and Ruciński, A.: 2004, The deletion method for upper tail estimates, *Combinatorica* **24**(4), 615–640.
URL: *http://dx.doi.org/10.1007/s00493-004-0038-3*

Janson, S. and Ruciński, A.: 2011, Upper tails for counting objects in randomly induced subhypergraphs and rooted random graphs, *Ark. Mat.* **49**(1), 79–96.
URL: *http://dx.doi.org/10.1007/s11512-009-0117-1*

Karoński, M. and Ruciński, A.: 1983, On the number of strictly balanced subgraphs of a random graphs., *Graph Theory, Proceedings, Lagów, 1981, eds. M. Borowiecki, J. W. Kennedy & M.M. Systo*, Vol. 1018 of *Lecture Notes in Math.*, Springer, Berlin, pp. 79–83.

Katona, G.: 1964, Intersection theorems for systems of finite sets., *Acta Math. Acad. Sci. Hung.* **15**, 329–337.

Kim, J. H. and Vu, V. H.: 2004, Divide and conquer martingales and the number of triangles in a random graph, *Random Structures Algorithms* **24**(2), 166–174.
URL: *http://dx.doi.org/10.1002/rsa.10113*

Kleitman, D. J.: 1970, On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors, *Advances in Math.* **5**, 155–157 (1970).

Leader, I.: 1991, Discrete isoperimetric inequalities, *Probabilistic combinatorics and its applications (San Francisco, CA, 1991)*, Vol. 44 of *Proc. Sympos. Appl. Math.*, Amer. Math. Soc., Providence, RI, pp. 57–80.

Littlewood, J. E. and Offord, A. C.: 1943, On the number of real roots of a random algebraic equation. III, *Rec. Math. [Mat. Sbornik] N.S.* **12(54)**, 277–286.

McDiarmid, C.: 1989, On the method of bounded differences, *Surveys in combinatorics, 1989 (Norwich, 1989)*, Vol. 141 of *London Math. Soc. Lecture Note Ser.*, Cambridge Univ. Press, Cambridge, pp. 148–188.

McDiarmid, C.: 1998, Concentration, *Probabilistic methods for algorithmic discrete mathematics*, Vol. 16 of *Algorithms Combin.*, Springer, Berlin, pp. 195–248.

Milner, E. C.: 1968, A combinatorial theorem on systems of sets, *J. London Math. Soc.* **43**, 204–206.

Panchenko, D.: 2004, Deviation inequality for monotonic Boolean functions with application to the number of k-cycles in a random graph, *Random Structures Algorithms* **24**(1), 65–74.
URL: *http://dx.doi.org/10.1002/rsa.10105*

Ruciński, A.: 1988, When are small subgraphs of a random graph normally distributed?, *Probab. Theory Related Fields* **78**(1), 1–10.
URL: *http://dx.doi.org/10.1007/BF00718031*

Šileikis, M.: 2009, An application of large deviation inequalities to the graph isoperimetry, *Matematika ir matematinis modeliavimas* **5**, 6–10.

Šileikis, M.: 2012, On the upper tail of counts of strictly balanced subgraphs, *Electr. J. Comb.* **19**, P4.

Spencer, J.: 1990, Counting extensions, *J. Combin. Theory Ser. A* **55**(2), 247–255.
URL: *http://dx.doi.org/10.1016/0097-3165(90)90070-D*

Sperner, E.: 1928, Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27**(1), 544–548.
URL: *http://dx.doi.org/10.1007/BF01171114*

Sztencel, R.: 1981, On boundedness and convergence of some Banach space valued random series., *Probab. Math. Stat.* **2**, 83–88.

Talagrand, M.: 1995, Concentration of measure and isoperimetric inequalities in product spaces, *Inst. Hautes Études Sci. Publ. Math.* (81), 73–205.

Vu, V. H.: 2000, On the concentration of multi-variate polynomials with small expectation, *Random Struct. Algorithms* **16**(4), 344–363.

Vu, V. H.: 2001, A large deviation result on the number of small subgraphs of a random graph, *Combin. Probab. Comput.* **10**(1), 79–94.