The lower tail: Poisson approximation revisited

Svante Janson* and Lutz Warnke†

May 27, 2014

Abstract

The well-known "Janson's inequality" gives Poisson-like upper bounds for the lower tail probability $\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X)$ when X is the sum of dependent indicator random variables of a special form. We show that, for large deviations, this inequality is optimal whenever X is approximately Poisson, i.e., when the dependencies are weak. We also present correlation-based approaches that, in certain symmetric applications, yield related conclusions when X is no longer close to Poisson. As an illustration we, e.g., consider subgraph counts in random graphs, and obtain new lower tail estimates, extending earlier work (for the special case $\varepsilon = 1$) of Janson, Łuczak and Ruciński.

1 Introduction

In probabilistic combinatorics and related areas it often is important to estimate the probability that a sum X of dependent indicator random variables is small or zero (to, e.g., show that few or none of a collection of events occurs). Moreover, it frequently is desirable that these probabilities are exponentially small (to, e.g., make union bound arguments amenable). In this paper we focus on such sharp estimates for the *lower tail* $\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X)$, where X is of a form that is commonly used in, e.g., applications of the probabilistic method or random graph theory, see [1, 16]. More precisely, the underlying probability space is the random subset $\Gamma_{\mathbf{p}} \subseteq \Gamma$, with $|\Gamma| = N$ and $\mathbf{p} = (p_i)_{i\in\Gamma}$, where each $i \in \Gamma$ is included, independently, with probability p_i. Given a family $\big(Q(\alpha)\big)_{\alpha\in\mathcal{X}}$ of subsets of Γ (often $\mathcal{X} \subseteq 2^\Gamma$ and $Q(\alpha) = \alpha$ is convenient) we define $I_\alpha = \mathbb{1}_{\{Q(\alpha)\subseteq\Gamma_{\mathbf{p}}\}}$, so that

$$X = \sum_{\alpha\in\mathcal{X}} I_\alpha \tag{1}$$

counts the number of sets $Q(\alpha)$ that are entirely contained in $\Gamma_{\mathbf{p}}$. We write $\alpha \sim \beta$ if $Q(\alpha) \cap Q(\beta) \neq \emptyset$ and $\alpha \neq \beta$, which intuitively means that there are 'dependencies' between I_α and I_β. Let

$$\mu = \mathbb{E}X = \sum_{\alpha\in\mathcal{X}} \mathbb{E}I_\alpha, \qquad \Pi = \max_{\alpha\in\mathcal{X}} \mathbb{E}I_\alpha,$$
$$\Lambda = \mu + \sum_{(\alpha,\beta)\in\mathcal{X}\times\mathcal{X}:\alpha\sim\beta} \mathbb{E}I_\alpha I_\beta = (1+\delta)\mu.$$

(We write $\mu(X)$, $\Pi(X)$, $\Lambda(X)$ and $\delta(X)$ in case of ambiguity.) Note that δ measures how dependent the indicators I_α are (with $\delta = 0$ in the case of independent summands), and that $\operatorname{Var} X \leqslant \Lambda$ holds. In [13] the first author proved the following lower tail analogue (often called *Janson's inequality*, see, e.g., [1]) of the Bernstein and Chernoff bounds for sums of independent indicators (the case $\delta = 0$): with $\varphi(x) = (1+x)\log(1+x) - x$, for all $\varepsilon \in [0,1]$ we have

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \leqslant \exp\big\{-\varphi(-\varepsilon)\mu/(1+\delta)\big\} = \exp\big\{-\varphi(-\varepsilon)\mu^2/\Lambda\big\}, \tag{2}$$

*Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden. E-mail: svante.janson@math.uu.se. Partly supported by the Knut and Alice Wallenberg Foundation.

†Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK. E-mail: L.Warnke@dpmms.cam.ac.uk.

where $\varphi(-1) = 1$, $\varepsilon^2/2 \leqslant \varphi(-\varepsilon) \leqslant \varepsilon^2$ and $\varphi(-\varepsilon) = \varepsilon^2/2 + O(\varepsilon^3)$ for $\varepsilon \in [0,1]$. As discussed in [13, 16, 1], inequality (2) is quite attractive because it (i) yields Poisson-like tail estimates in the weakly dependent case $\delta = O(1)$, (ii) usually corresponds to a (one-sided) exponential version of Chebyshev's inequality, and (iii) often qualitatively matches the tail behaviour suggested by the central limit theorem. For example, it is well-known (and not hard to check) that $\Lambda = \Theta(\mathrm{Var}\, X)$ if $\widehat{p} = \max\{\Pi, \max_i p_i\}$ is bounded away from one, that $\widehat{p} \to 0$ implies $\Lambda \sim \mathrm{Var}\, X$, and that $\delta, \Pi \to 0$ implies $\Lambda \sim \mu \sim \mathrm{Var}\, X$.

The inequality (2) is nowadays a widely used tool in probabilistic combinatorics (see, e.g., [1, 16] and the references therein), which makes it important to understand how 'sharp' it is, i.e., whether the exponential rate of decay given by (2) is best possible. For sums of independent Bernoulli random variables we have $\delta = 0$ and (2) coincides with the Chernoff bounds, where the exponent is well-known to be best possible if $\max_i p_i = o(1)$. However, it is doubtful whether such examples are of any significance for concrete applications with $\delta > 0$. Fortunately, whenever $\Pi < 1$, Harris' inequality [12] gives, as noted in [15],

$$\mathbb{P}(X = 0) \geqslant \prod_{\alpha \in \mathcal{X}} (1 - \mathbb{E} I_\alpha) \geqslant \exp\bigl\{-\mu/(1-\Pi)\bigr\}. \tag{3}$$

The point is that (2) and (3) yield $\log \mathbb{P}(X = 0) \sim -\mu$ whenever $\delta, \Pi \to 0$. This raises the intriguing question whether the exponent of (2) is also sharp for other choices of ε, in particular when $\varepsilon \to 0$ (which, of course, is also an interesting problem in concentration of measure).

1.1 Main result

In this paper we prove that "Janson's inequality" (2) is close to best possible in many situations of interest. Our first result shows that, for large deviations, the rate of decay of (2) is optimal for *any* random variable X of type (1) that is approximately Poisson, i.e., whenever $\delta, \Pi \to 0$ (see [13]).

Theorem 1. *With notations as above, if* $\varepsilon \in [0,1]$, $\max\{\Pi, \mathbb{1}_{\{\varepsilon<1\}}\delta\} \leqslant 2^{-14}$ *and* $\varepsilon^2\mu \geqslant \mathbb{1}_{\{\varepsilon<1\}}$*, then*

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant \exp\bigl\{-(1+\xi)\varphi(-\varepsilon)\mu\bigr\}, \tag{4}$$

with $\xi = 135 \max\{\Pi^{1/8}, \mathbb{1}_{\{\varepsilon<1\}}\delta^{1/8}, \mathbb{1}_{\{\varepsilon<1\}}(\varepsilon^2\mu)^{-1/4}\}$.

With $\varphi(-1) = 1$ in mind, note that (4) qualitatively extends the lower bound (3) resulting from Harris' inequality [12] to general ε. Here the condition $\varepsilon^2\mu = \Omega(1)$ is natural in the context of exponentially small probabilities since $(1+\xi)\varphi(-\varepsilon) = \Theta(\varepsilon^2)$. As discussed, our favourite range is when $\delta, \Pi \to 0$. For large deviations, i.e., when $\varepsilon^2\mu \to \infty$ holds, (2) and (4) then yield

$$\log \mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \sim -\varphi(-\varepsilon)\mu.$$

In words, Theorem 1 determines the *large deviation rate function* $\log \mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X)$ up to second order error terms, closing a gap that was left open by the first author nearly 25 years ago. Indeed, Theorem 2 in [13] gives a lower bound, but it is at best off from the upper bound (2) by a (multiplicative) constant factor in the exponent, and even this holds only for a more restricted range of the parameters. Furthermore, Theorem 1 with $\delta = 0$ also implies the optimality of the Chernoff bounds mentioned above.

Our second result yields a related conclusion when $\delta = O(1)$ and Π is bounded away from one. More precisely, in this 'weakly dependent' case Theorem 2 shows that the decay of the inequality (2) is best possible up to constant factors in the exponent.

Theorem 2. *With notations as above, if* $\varepsilon \in [0,1]$, $\Pi < 1$ *and* $\varepsilon^2\mu \geqslant \mathbb{1}_{\{\varepsilon<1/50\}}(1+\delta)^{-1/2}$*, then*

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant \exp\bigl\{-K\varphi(-\varepsilon)\mu(1+\delta^*)\bigr\} \geqslant \exp\bigl\{-K\varepsilon^2\mu(1+\delta^*)\bigr\}, \tag{5}$$

with $K = 5000/(1-\Pi)^5$ *and* $\delta^* = \mathbb{1}_{\{\varepsilon<1/50\}}\delta$.

A key feature of (5) is that it holds for *any* $\Pi < 1$ (and that the dependence of K on Π is explicit). Note that usually $K = \Theta(1)$. Whenever $\delta = O(1)$, inequalities (2) and (5) then yield

$$\log \mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) = -\Theta(\varepsilon^2\mu),$$

where the implicit constants differ by a factor of at most $2K(1+\delta)^2 = O(1)$. This subsumes the folklore fact that Chernoff bounds (where $\delta = 0$) are sharp up to constants in the exponent if $\max_i p_i$ is bounded away from one. While the numerical value of K is often immaterial, better constant factors can typically be obtained, if desired, by reworking the proof (optimizing certain parameters to the situation at hand).

The proofs of Theorem 1 and 2 hinge on Hölder's inequality and several estimates of the Laplace transform (which in turn are based on correlation inequalities), see Section 2. In fact, an inspection of the proofs reveals that Theorem 1 and 2 (as well as (3), Theorem 6 and Lemma 7) remain valid for the more general correlation conditions (and setup) stated by Riordan and Warnke [23]. It would be interesting to know whether similar results also hold under the weaker dependency assumptions of Suen's inequality [28, 14].

1.2 Main example

From an applications point of view it is important to also understand the sharpness of (2) in the case $\delta = \Omega(1)$, i.e., when X is no longer close to Poisson. In Section 3 we present correlation-inequality based bootstrapping approaches which often allow us to deal with this remaining 'strongly dependent' case. The punchline seems to be that, in the presence of certain symmetries, the inequality (2) is oftentimes best possible up to constant factors in the exponent.

In this paper our main example is the number of small subgraphs in the binomial random graph $G_{n,p}$, which is a classical topic in random graph theory (see, e.g., [10, 3, 24]). It frequently serves as a test-bed for new probabilistic estimates (see, e.g., [2, 15, 27, 21, 18, 17, 7]), and we shall use it to demonstrate the applicability of our bootstrapping approaches. In fact, we consider the more general random hypergraph $G^{(k)}_{n,p}$, with $k \geqslant 2$, where each of the $\binom{n}{k}$ edges of the complete k-uniform hypergraph $K^{(k)}_n$ is included, independently, with probability p. Given a k-uniform hypergraph H, or briefly *k-graph*, we define $X_H = X_H(n,p)$ as the number of copies of H in $G^{(k)}_{n,p}$, where by a *copy* we mean, as usual, a subgraph isomorphic to H. Furthermore, we write $e_H = |E(H)|$ and $v_H = |V(H)|$ for the number of edges and vertices of H, respectively. Theorem 3 shows that the lower tail of the distribution of X_H is governed by Φ_H, i.e., the expected number of copies of the 'least expected' subgraph of H. This exponential rate of decay is consistent with normal approximation heuristics since $\Phi_H = \Theta\big((1-p)(\mathbb{E}X_H)^2/\operatorname{Var} X_H\big)$, see Lemma 3.5 in [16].

Theorem 3. *Let H be a k-graph with $e_H \geqslant 1$. Define $\Phi_H = \Phi_H(n,p) = \min\{\mathbb{E}X_J : J \subseteq H, e_J \geqslant 1\}$. There are positive constants c, C, D and n_0, all depending only on H, such that for all $n \geqslant n_0$, $p \in [0,1)$ and $\varepsilon \in [0,1]$ satisfying $\varepsilon^2\Phi_H \geqslant \mathbb{1}_{\{\varepsilon<1\}}D$ we have*

$$\exp\big\{-(1-p)^{-5}C\varepsilon^2\Phi_H\big\} \leqslant \mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \leqslant \exp\big\{-c\varepsilon^2\Phi_H\big\}. \tag{6}$$

The upper bound of (6) follows from (2) via standard calculations (see, e.g., [16] or Lemma 22), and so the real content of this theorem is the 'matching' lower bound. A key feature of Theorem 3 is that ε is *not* fixed, but may depend on n. In the context of exponentially decaying probabilities, note that the $\varepsilon^2\Phi_H = \Omega(1)$ condition is natural (unless $p \approx 1$). In applications p is typically bounded away from one (in fact, $p = o(1)$ is often standard), in which case (6) yields

$$\log \mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) = -\Theta(\varepsilon^2\Phi_H), \tag{7}$$

determining the large deviation rate function of X_H up to constants factors. For the special case $\varepsilon = 1$ (and $k = 2$) this was established more than 25 years ago by Janson, Łuczak and Ruciński [15], and for $\varepsilon \geqslant \varepsilon_0$ an analogous statement is nowadays easily deduced from (2) and (3), see also (73). By contrast, the case $\varepsilon \to 0$ seems to have eluded further attention, and Theorem 3 rectifies this (surprising) gap in the literature.

Although not our primary focus, in certain ranges our proof techniques are strong enough to establish the finer behaviour of the large deviation rate function. In particular, for the case in which there is only one subgraph $G \subseteq H$ with $\mathbb{E}X_G = \Theta(\Phi_H)$ we have two results that determine the leading constant in (7). More precisely, Theorem 4 applies if there is only one copy of G in H (which includes the case $G = H$), and Theorem 5 applies if G is an edge (in which case there are e_H copies of G in H). To state these results, for any given k-graph H we set

$$m_k(H) = \mathbb{1}_{\{e_H \geqslant 2\}} \max_{J \subseteq H, e_J \geqslant 2} \frac{e_J - 1}{v_J - k} + \mathbb{1}_{\{e_H = 1\}} \frac{1}{k}. \tag{8}$$

In addition, we define $\mathrm{ex}(n,H)$ as the maximum number of edges in an H-free k-graph with n vertices. It is well-known (see, e.g., [20]) that $\pi_H = \lim_{n\to\infty} \mathrm{ex}(n,H)/\binom{n}{k}$ exists, with $\pi_H \in [0,1)$, and that for graphs (i.e., $k=2$) we have $\pi_H = 1 - 1/(\chi(H)-1)$, where $\chi(H)$ is the chromatic number of H.

Theorem 4. *Let $G \subseteq H$ be k-graphs with $e_G \geqslant 1$. Assume that there is exactly one copy of G in H, and that $p = p(n) = o(1)$ is such that $\mathbb{E}X_G = o(\mathbb{E}X_J)$ for all $G \neq J \subseteq H$ with $e_J \geqslant 1$. If $\varepsilon = \varepsilon(n) \in (0,1]$ satisfies $\varepsilon^2 \mathbb{E}X_G \geqslant \mathbb{1}_{\{\varepsilon<1\}}\omega\big(1 + \mathbb{1}_{\{G\neq H, e_G \geqslant 2\}} \log(1/\varepsilon)\big)$, then we have*

$$\log \mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \sim -\varphi(-\varepsilon)\mathbb{E}X_G. \tag{9}$$

Theorem 5. *Let H be a k-graph with $e_H \geqslant 1$. If $p = p(n) = o(1)$ and $\varepsilon = \varepsilon(n) \in [0,1]$ satisfy $p = \omega(n^{-1/m_k(H)})$ and $\varepsilon^2\binom{n}{k}p = \omega(1)$, then we have*

$$\log \mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \sim \begin{cases} -\varphi(-\varepsilon)\binom{n}{k}p/e_H^2, & \text{if } \varepsilon = o(1), \\ -\varphi(-\varepsilon)\binom{n}{k}p(1-\pi_H), & \text{if } \varepsilon = 1 - o(1). \end{cases} \tag{10}$$

Here our main contributions are the tight lower bound of (9), and the case $\varepsilon = o(1)$ of (10). Theorem 4 is a natural extension of earlier work of Janson, Łuczak and Ruciński [15] for the special case $\varepsilon = 1$ (and $k = 2$). Theorem 5 partially solves an open problem of [15], but in the relevant case $\varepsilon = 1$ inequality (10) is a fairly simple consequence of the recent 'hypergraph container' results of Saxton and Thomason [25], see also Lemma 23. With $\varphi(-\varepsilon) = \Theta(\varepsilon^2)$ in mind the conditions involving ε^2 are natural in both results – up to the logarithmic term in case of Theorem 4, which seems to be an artefact of our proof (we leave its removal as an open problem, see Section 3.2). The form of the exponent in Theorem 5 differs in an intriguing way for $\varepsilon = o(1)$ and $\varepsilon = 1 - o(1)$. In particular, (10) provides a natural example where the inequality (2) does *not* always give the correct constants in the exponent when $\delta = \omega(1)$: in the case $\varepsilon = 1 - o(1)$, the 'extremal' structural properties of H-free graphs come into play. We leave it as an open problem to determine the finer behaviour of the exponent (i.e., with explicit constants) in the 'intermediate' range $\varepsilon = \Theta(1)$. This seems of particular interest since Theorem 4 and 5 nearly cover all edge probabilities p for *balanced* k-graphs with $e_H \geqslant 2$ and $m_k(H) = (e_H - 1)/(v_H - k)$, where $G = H$ for $p = o(n^{-1/m_k(H)})$; for $k = 2$ (when this class usually is called *2-balanced*) this class includes, e.g., trees, cycles, complete graphs, complete r-partite graphs $K_{t,\dots,t}$ and the d-dimensional cube.

Finally, Theorems 3–5 compare favourable with related work for the *upper tail* probability $\mathbb{P}(X_H \geqslant (1+\varepsilon)\mathbb{E}X_H)$, where the case $\varepsilon = \Theta(1)$ has been extensively studied for $k = 2$, see, e.g., [27, 29, 17, 5, 8, 26, 6] and the references therein. Indeed, for most graphs H the order of magnitude of the large deviation rate function $\log \mathbb{P}(X_H \geqslant (1+\varepsilon)\mathbb{E}X_H)$ is only known up to logarithmic factors when $\varepsilon = \Theta(1)$, whereas Theorem 3 determines $\log \mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H)$ up to constant factors, even when $\varepsilon = \varepsilon(n) \to 0$. For triangles the finer behaviour of $\log \mathbb{P}(X_{K_3} \geqslant (1+\varepsilon)\mathbb{E}X_{K_3})$ has very recently been determined for $\varepsilon = \Theta(1)$ and $n^{-1/42+o(1)} \leqslant p = o(1)$, see [22]. By contrast, for all balanced k-graphs H (which for $k = 2$ includes $H = K_3$) Theorems 4–5 apply for essentially all $p = o(1)$ of interest, excluding only $p = \Theta(n^{-1/m_k(H)})$. However, the key conceptual difference is that Theorem 4 includes the case $\varepsilon = \varepsilon(n) \to 0$.

The rest of the paper is organized as follows. First, in Section 2, we prove Theorem 1 and 2. Next, in Section 3, we present several bootstrapping approaches that yield lower bounds for the lower tail, which are subsequently illustrated in Section 4. Namely, in Section 4.1 we apply them to the number of arithmetic progressions in random subsets of the integers, and in Section 4.2 we apply them to subgraph counts in random hypergraphs and prove Theorems 3–5.

2 Lower bounds for the lower tail

In this section we prove Theorem 1 and 2, i.e., establish lower bounds for the lower tail. Since our core argument breaks down when ε is very close to one, en route to Theorem 1 we establish the following (slightly sharper) complementary estimates.

Theorem 6. *Let $X = \sum_{\alpha\in\mathcal{X}} I_\alpha$, $\mu = \mathbb{E}X$, Π and δ be defined as in Section 1. If $e(1-\varepsilon)\varepsilon^2\mu \geqslant 1$ and $0 \leqslant \varepsilon \leqslant 1 - 4\max\{\Pi^{1/4}, \delta^{1/4}\}$, then*

$$\mathbb{P}(X < (1-\varepsilon)\mathbb{E}X) \geqslant \exp\big\{-(1+\xi)\varphi(-\varepsilon)\mu\big\}, \tag{11}$$

with $\xi = 135\max\{\Pi^{1/4}, \delta^{1/4}, [e(1-\varepsilon)\varepsilon^2\mu]^{-1/2}\}$.

Lemma 7. *Let* $X = \sum_{\alpha\in\mathcal{X}} I_\alpha$*,* $\mu = \mathbb{E}X$ *and* Π *be defined as in Section 1. If* $1-e^{-1} \leqslant \varepsilon \leqslant 1$ *and* $\Pi < 1$*, then*

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant \mathbb{P}(X=0) \geqslant \exp\bigl\{-(1+\zeta)\varphi(-\varepsilon)\mu\bigr\}, \tag{12}$$

with $\zeta = 10\max\{\sqrt{1-\varepsilon}, \Pi/(1-\Pi)\}$.

While Lemma 7 follows from (3) via calculus (see Lemma 11), the remaining proofs are not a mere refinement of [13], but contain several new ideas and ingredients. This includes integrating the logarithmic derivative of the Laplace transform over the interval $[r,t]$ instead of the usual $[0,t]$ (see the proof of Lemma 9), using Hölder's inequality with parameter $p \to 1$ instead of the Cauchy–Schwarz inequality (see Section 2.2), and a careful treatment of second order error terms (see, e.g., Lemma 8 and 14).

2.1 Preliminaries

We first collect some basic estimates of the Laplace transform of X as defined in Section 1.

Lemma 8. *For all* $s \geqslant 0$ *satisfying* $\lambda = \Pi(1-e^{-s}) < 1$ *we have*

$$\mathbb{E}e^{-sX} \geqslant \exp\left\{-\mu(1-e^{-s}) - \frac{\mu\Pi(1-e^{-s})^2}{2(1-\lambda)}\right\}. \tag{13}$$

Proof. The FKG inequality [11] (or Harris's inequality [12]) yields

$$\mathbb{E}e^{-sX} = \mathbb{E}\prod_{\alpha\in\mathcal{X}} e^{-sI_\alpha} \geqslant \prod_{\alpha\in\mathcal{X}} \mathbb{E}e^{-sI_\alpha} = \prod_{\alpha\in\mathcal{X}} \bigl(1 - \mathbb{E}I_\alpha(1-e^{-s})\bigr).$$

Now, for $x \in [0,1)$ we have

$$\log(1-x) = -\sum_{j\geqslant 1}\frac{x^j}{j} \geqslant -x - \frac{x^2}{2(1-x)}, \tag{14}$$

and (13) follows since $\mathbb{E}I_\alpha \leqslant \Pi$ and $\mu = \sum_{\alpha\in\mathcal{X}} \mathbb{E}I_\alpha$. □

Lemma 9. *For all* $t \geqslant r \geqslant 0$ *we have*

$$\frac{\mathbb{E}e^{-rX}}{\mathbb{E}e^{-tX}} \geqslant \exp\left\{\frac{\mu}{1+\delta}\left(e^{-(1+\delta)r} - e^{-(1+\delta)t}\right)\right\}. \tag{15}$$

Proof. Let $\Psi(x) = \mathbb{E}e^{-xX}$. The proof of Lemma 1 in [13] establishes $-\frac{d}{dx}\log\Psi(x) \geqslant \mu e^{-(1+\delta)x}$ for $x \geqslant 0$ (see also [23]). Hence

$$\begin{aligned}\log\left(\frac{\mathbb{E}e^{-rX}}{\mathbb{E}e^{-tX}}\right) &= -\log\Psi(t) + \log\Psi(r) = \int_r^t \left(-\frac{d}{dx}\log\Psi(x)\right)dx \\ &\geqslant \int_r^t \mu e^{-(1+\delta)x}dx = \frac{\mu}{1+\delta}\left(e^{-(1+\delta)r} - e^{-(1+\delta)t}\right),\end{aligned}$$

and (15) follows. □

Next, we state some technical estimates of $\varphi(-\varepsilon) = (1-\varepsilon)\log(1-\varepsilon) + \varepsilon$ for later reference (these can safely be skipped on first reading). Following standard conventions, for $k \in \{1,2\}$ we have $0\log^k(0) = \lim_{\varepsilon\nearrow 1}(1-\varepsilon)\log^k(1-\varepsilon) = 0$, so that $\varphi(-1) = 1$.

Lemma 10. *For all* $\varepsilon \in [0,1]$ *we have*

$$\max\bigl\{(1-\varepsilon)\log^2(1-\varepsilon), \varepsilon^2\bigr\} \leqslant 2\varphi(-\varepsilon) \leqslant \min\bigl\{\log^2(1-\varepsilon), 2\varepsilon^2\bigr\}. \tag{16}$$

Lemma 11. *For all* $1-e^{-1} \leqslant \varepsilon \leqslant 1$ *we have*

$$\varphi(-\varepsilon) \leqslant 1 \leqslant (1+5\sqrt{1-\varepsilon})\varphi(-\varepsilon). \tag{17}$$

Lemma 12. *For all $\varepsilon \in [0,1]$ and $A \in [0,\infty)$ we have, with $\gamma = A-1$,*

$$\varphi(-A\varepsilon) \leqslant \begin{cases} (1+A\varepsilon)A^2\varphi(-\varepsilon), & \text{if } A\varepsilon \leqslant 1, \\ (1+\sqrt{\gamma})\varphi(-\varepsilon), & \text{if } 0 \leqslant 3\sqrt{\gamma} \leqslant 1-\varepsilon. \end{cases} \tag{18}$$

The elementary proofs of Lemma 10–12 are deferred to Appendix A.

2.2 Proof strategy

We start with a general lower bound for $\mathbb{P}(X < (1-\varepsilon)\mathbb{E}X)$. If $p, q \in (1,\infty)$ satisfy $1/p + 1/q = 1$, then Hölder's inequality implies

$$\mathbb{E}(e^{-sX}\mathbb{1}_{\{X<(1-\varepsilon)\mathbb{E}X\}}) \leqslant (\mathbb{E}e^{-psX})^{1/p}\mathbb{P}(X < (1-\varepsilon)\mathbb{E}X)^{1/q}.$$

Noting that $q = q/p + 1 = 1/(p-1) + 1$, we infer

$$\begin{aligned} \mathbb{P}(X < (1-\varepsilon)\mathbb{E}X) &\geqslant \left(\frac{\mathbb{E}(e^{-sX}\mathbb{1}_{\{X<(1-\varepsilon)\mathbb{E}X\}})}{(\mathbb{E}e^{-psX})^{1/p}}\right)^q \\ &= \left(\frac{\mathbb{E}(e^{-sX}\mathbb{1}_{\{X<(1-\varepsilon)\mathbb{E}X\}})}{\mathbb{E}e^{-sX}}\right)^{\frac{p}{p-1}} \cdot \left(\frac{\mathbb{E}e^{-sX}}{\mathbb{E}e^{-psX}}\right)^{\frac{1}{p-1}} \mathbb{E}e^{-sX}. \end{aligned} \tag{19}$$

In the following we heuristically outline how we estimate $\mathbb{P}(X < (1-\varepsilon)\mathbb{E}X)$ when $\delta, \Pi \to 0$ and $\varepsilon < 1$ (to be precise, ε bounded away from one). The idea is to first consider $p > 1$ and $s > z = -\log(1-\varepsilon)$, and then let $p \to 1$ and $s \to z$. Since $\Pi \to 0$, using Lemma 8 we have

$$\mathbb{E}e^{-sX} \geqslant \exp\Big\{-\mu\big(1 - e^{-s} + o(1)\big)\Big\}. \tag{20}$$

So, using Lemma 9 together with $\delta \to 0$, we expect that (replacing the difference quotient by the derivative), as $p \to 1$,

$$\left(\frac{\mathbb{E}e^{-sX}}{\mathbb{E}e^{-psX}}\right)^{\frac{1}{p-1}} \geqslant \exp\left\{\mu s\left(\frac{e^{-(1+\delta)s} - e^{-(1+\delta)ps}}{(1+\delta)(p-1)s}\right)\right\} = \exp\Big\{\mu\big(se^{-s} + o(1)\big)\Big\}. \tag{21}$$

The point is that $1 - e^{-s} - se^{-s} \to \varphi(-\varepsilon)$ as $s \to z$. So, if (20) and (21) essentially determine the right hand side of (19), then our previous considerations suggest

$$\mathbb{P}(X < (1-\varepsilon)\mathbb{E}X) \geqslant \exp\Big\{-\mu\big(\varphi(-\varepsilon) + o(1)\big)\Big\}.$$

Luckily, our later calculations confirm that (for suitable choices of p and s) we can indeed essentially ignore the first term on the right hand side of (19) for large deviations, i.e., when $\varepsilon^2\mu \to \infty$ holds.

2.3 Proofs of Theorem 2 and 6

Assume that $\varepsilon, \tau \in (0,1)$ and $\sigma \in (0,\infty)$. Let

$$p = 1 + \sigma \quad \text{and} \quad q = 1 + 1/\sigma, \tag{22}$$

so that $p, q \in (1,\infty)$ and $1/p + 1/q = 1$. Furthermore, let

$$z = -\log(1-\varepsilon) \quad \text{and} \quad s = pz. \tag{23}$$

With (19) in mind, the following two lemmas are at the heart of our argument.

Lemma 13. *With definitions as above, if* $\Pi(1-e^{-s}) \leqslant 1/2$, *then*

$$\left(\frac{\mathbb{E}e^{-sX}}{\mathbb{E}e^{-psX}}\right)^{\frac{1}{p-1}} \mathbb{E}e^{-sX} \geqslant e^{-(1+\eta)\varphi(-\varepsilon)\mu}, \tag{24}$$

with $\eta = 2p^2(\sigma + p\delta + \Pi) + 2p\sigma$.

Proof. Since $f(x) = -e^{-x}$ satisfies $f'(x) = e^{-x}$, the mean value theorem implies that there is $\zeta \in [1,p]$ such that

$$\frac{e^{-(1+\delta)s} - e^{-(1+\delta)ps}}{(1+\delta)(p-1)s} = e^{-(1+\delta)\zeta s} \geqslant e^{-(1+\delta)ps}. \tag{25}$$

Furthermore, since $g(x) = e^{-x}$ satisfies $g'(x) = -e^{-x}$ and $g''(x) = e^{-x} \geqslant 0$, using Taylor's theorem with remainder, we obtain

$$e^{-(1+\delta)ps} \geqslant e^{-s} - \big((1+\delta)p - 1\big)se^{-s}. \tag{26}$$

Note that $(1+\delta)p - 1 = \sigma + p\delta$. Furthermore, since $s = -p\log(1-\varepsilon)$, Bernoulli's inequality yields

$$(1-e^{-s})^2 = (1-(1-\varepsilon)^p)^2 \leqslant p^2\varepsilon^2. \tag{27}$$

So, by combining Lemmas 8 and 9 with (25)–(27), using $\Pi(1-e^{-s}) \leqslant 1/2$, it follows that

$$\begin{aligned}\left(\frac{\mathbb{E}e^{-sX}}{\mathbb{E}e^{-psX}}\right)^{\frac{1}{p-1}} \mathbb{E}e^{-sX} &\geqslant \exp\left\{\frac{\mu s\big(e^{-(1+\delta)s} - e^{-(1+\delta)ps}\big)}{(1+\delta)(p-1)s} - \mu(1-e^{-s}) - \mu\Pi(1-e^{-s})^2\right\} \\ &\geqslant \exp\left\{-\mu\Big(1 - e^{-s} - se^{-s} + \big(\sigma + p\delta\big)s^2e^{-s} + \Pi p^2\varepsilon^2\Big)\right\}.\end{aligned}$$

Let $g(x) = 1 - e^{-x} - xe^{-x}$, and note that $g(z) = \varphi(-\varepsilon)$. Furthermore, for $z \leqslant x \leqslant s$ we have $g'(x) = xe^{-x} \leqslant se^{-z}$. So, using Taylor's theorem with remainder, we deduce that

$$1 - e^{-s} - se^{-s} \leqslant \varphi(-\varepsilon) + (s-z)se^{-z}.$$

Consequently, since $s = pz \geqslant z$, we obtain

$$\left(\frac{\mathbb{E}e^{-sX}}{\mathbb{E}e^{-psX}}\right)^{\frac{1}{p-1}} \mathbb{E}e^{-sX} \geqslant \exp\Big\{-\varphi(-\varepsilon)\mu - \big(z^2e^{-z}\eta_1 + \varepsilon^2\eta_2\big)\mu\Big\},$$

where $\eta_1 = p^2(\sigma + p\delta) + p\sigma$ and $\eta_2 = p^2\Pi$. Finally, recalling $z = -\log(1-\varepsilon)$, the point is that Lemma 10 yields $\max\{z^2e^{-z}, \varepsilon^2\} \leqslant 2\varphi(-\varepsilon)$, yielding the result with $\eta = 2\eta_1 + 2\eta_2$. □

Lemma 14. *With definitions as above, if* $\lambda = \Pi(1-e^{-s}) < 1$ *and* $(1-\tau)\sigma^2(1-\varepsilon)^p \geqslant p^2\Pi/(1-\lambda) + \delta/(1+\delta)$, *then*

$$\left(\frac{\mathbb{E}(e^{-sX}\mathbb{1}_{\{X<(1-\varepsilon)\mathbb{E}X\}})}{\mathbb{E}e^{-sX}}\right)^{\frac{p}{p-1}} \geqslant \exp\left\{-\left(\frac{4p}{\tau\sigma^3(1-\varepsilon)^p\varepsilon^4\mu^2}\right)\varphi(-\varepsilon)\mu\right\}. \tag{28}$$

Proof. As $p = 1 + \sigma$, we write

$$\left(\frac{\mathbb{E}(e^{-sX}\mathbb{1}_{\{X<(1-\varepsilon)\mu\}})}{\mathbb{E}e^{-sX}}\right)^{\frac{p}{p-1}} = \left(1 - \frac{\mathbb{E}(e^{-sX}\mathbb{1}_{\{X\geqslant(1-\varepsilon)\mu\}})}{\mathbb{E}e^{-sX}}\right)^{\frac{p}{\sigma}}. \tag{29}$$

Let $t = z/(1+\delta)$. Recalling $\varphi(-\varepsilon) = (1-\varepsilon)\log(1-\varepsilon) + \varepsilon$, note that

$$t(1-\varepsilon)\mu - \frac{\mu}{1+\delta}\left(1 - e^{-(1+\delta)t}\right) = -\frac{\varphi(-\varepsilon)\mu}{1+\delta}.$$

So, using $t \leqslant s$ and Lemma 9 (with $r = 0$), it follows that

$$\mathbb{E}(e^{-sX}\mathbb{1}_{\{X\geqslant(1-\varepsilon)\mu\}}) \leqslant e^{-(s-t)(1-\varepsilon)\mu} \cdot \mathbb{E}e^{-tX} \leqslant \exp\left\{-s(1-\varepsilon)\mu - \frac{\varphi(-\varepsilon)\mu}{1+\delta}\right\}. \tag{30}$$

Set $h(x) = (1-\varepsilon)x - (1-e^{-x})$, and note that $h(z) = -\varphi(-\varepsilon)$ and $h'(z) = 0$. Furthermore, for $x \leqslant s$ we have $h''(x) = e^{-x} \geqslant e^{-s}$. So, using Taylor's theorem with remainder, we obtain

$$(1-\varepsilon)s - (1-e^{-s}) \geqslant -\varphi(-\varepsilon) + (s-z)^2 e^{-s}/2. \tag{31}$$

Recalling $p = 1+\sigma$, $s = pz$ and $\lambda = \Pi(1-e^{-s})$, by combining Lemma 8 with (30), (31) and $(1-e^{-s})^2 \leqslant s^2$, we infer

$$\begin{aligned}\frac{\mathbb{E}(e^{-sX}\mathbb{1}_{\{X \geqslant (1-\varepsilon)\mu\}})}{\mathbb{E}e^{-sX}} &\leqslant \exp\left\{-\mu\left((1-\varepsilon)s - (1-e^{-s}) + \frac{\varphi(-\varepsilon)}{1+\delta} - \frac{\Pi s^2}{2(1-\lambda)}\right)\right\} \\ &\leqslant \exp\left\{-\mu\left(\frac{\sigma^2(1-\varepsilon)^p z^2}{2} - \frac{\Pi p^2 z^2}{2(1-\lambda)} - \frac{\delta\varphi(-\varepsilon)}{1+\delta}\right)\right\}.\end{aligned}$$

Since Lemma 10 gives $\varphi(-\varepsilon) \leqslant \log^2(1-\varepsilon)/2 = z^2/2$, we have, by assumption,

$$\frac{\mathbb{E}(e^{-sX}\mathbb{1}_{\{X \geqslant (1-\varepsilon)\mu\}})}{\mathbb{E}e^{-sX}} \leqslant \exp\Big\{-\tau\sigma^2(1-\varepsilon)^p z^2\mu/2\Big\}. \tag{32}$$

Now, inserting (32) into (29), using the fact that $e^{-x} + e^{-1/x} \leqslant 1$ for $x > 0$ (as in the proof of Theorem 2 in [13]), we obtain

$$\left(\frac{\mathbb{E}(e^{-sX}\mathbb{1}_{\{X < (1-\varepsilon)\mu\}})}{\mathbb{E}e^{-sX}}\right)^{\frac{p}{p-1}} \geqslant \exp\left\{-\frac{2p}{\tau\sigma^3(1-\varepsilon)^p z^2\mu}\right\}.$$

Finally, recalling $z = -\log(1-\varepsilon)$, Lemma 10 yields $z^2 \geqslant \varepsilon^2$ and $1 \leqslant 2\varphi(-\varepsilon)/\varepsilon^2$. □

Combining (19) with Lemma 13 and 14, the proofs of Theorem 2 and 6 reduce to defining suitable parameters σ and τ (our choices are somewhat ad-hoc, and yield fairly transparent error-terms).

Proof of Theorem 6. With foresight, let $\tau = 5/8$ and

$$\sigma = \max\big\{\Pi^{1/4}, \delta^{1/4}, [e(1-\varepsilon)\varepsilon^2\mu]^{-1/2}\big\}. \tag{33}$$

Note that the assumption $0 \leqslant \varepsilon \leqslant 1 - 4\max\{\Pi^{1/4}, \delta^{1/4}\}$ implies $\max\{\Pi, \delta\} \leqslant 4^{-4}$, so that $\lambda = \Pi(1-e^{-s}) \leqslant \Pi \leqslant 1/5$. Hence, using $e(1-\varepsilon)\varepsilon^2\mu \geqslant 1$, we see that $\sigma \leqslant 1$ and thus $p \leqslant 2$. Consequently, by (33), we have

$$\sigma^4(1-\varepsilon)^p\varepsilon^4\mu^2 \geqslant \sigma^4(1-\varepsilon)^2\varepsilon^4\mu^2 \geqslant e^{-2} \tag{34}$$

and $\sigma^2 \geqslant \max\{\Pi^{1/2}, \delta^{1/2}\}$. In addition, by assumption, we have $(1-\varepsilon)^p \geqslant (1-\varepsilon)^2 \geqslant 16\max\{\Pi^{1/2}, \delta^{1/2}\}$. Since $16(1-\tau) = 6$ and $p^2/(1-\lambda) \leqslant 5$, it follows that

$$(1-\tau)\sigma^2(1-\varepsilon)^p \geqslant 6\max\{\Pi, \delta\} \geqslant p^2\Pi/(1-\lambda) + \delta/(1+\delta).$$

Now, combining (19) with Lemmas 13–14 and (34), we obtain

$$\mathbb{P}(X < (1-\varepsilon)\mu) \geqslant e^{-(1+\kappa)\varphi(-\varepsilon)\mu},$$

with $\kappa = 2p^2(\sigma + p\delta + \Pi) + 2p\sigma + 4e^2\tau^{-1}p\sigma$. Finally, using $\sigma \geqslant \sigma^4 \geqslant \max\{\delta, \Pi\}$, $p \leqslant 2$ and $\tau = 5/8$, we see that $\kappa \leqslant 135\sigma$. □

Proof of Theorem 2. Let $\tau = (1-\Pi)/5$, so that, by assumption, $\tau \in (0, 1/5]$. The proof distinguishes two cases, which eventually establish (5) by noting that Lemma 10 gives $\varphi(-\varepsilon) \leqslant \varepsilon^2$.

First, we assume $0 \leqslant \varepsilon < \tau^2/2$. Note that then, by assumption, we have $0 < \varepsilon < 1/50$ and $\delta = \delta^*$. Let $p = 2/\tau$ and $\sigma = p - 1$. Analogous to (27) we have $1 - e^{-s} = 1 - (1-\varepsilon)^p \leqslant p\varepsilon$, so that $\Pi \leqslant 1$ implies

$$\lambda = \Pi(1-e^{-s}) \leqslant \Pi p\varepsilon \leqslant \tau,$$

which in particular yields $\lambda \leqslant 1/2$, with room to spare. Next observe that, since $\sigma/p = 1 - 1/p$ and $\max\{2/p, p\varepsilon, \lambda\} = \tau$, by the definition of τ we have

$$\begin{aligned}\frac{(1-\tau)\sigma^2(1-\varepsilon)^p(1-\lambda)}{p^2} - \frac{1}{p^2} &\geqslant (1-\tau)(1-2/p)(1-p\varepsilon)(1-\lambda) - \tau^2/4 \\ &\geqslant (1-\tau)^4 - \tau^2/4 \geqslant 1 - 5\tau = \Pi,\end{aligned}$$

which in turn readily yields $(1-\tau)\sigma^2(1-\varepsilon)^p \geqslant p^2\Pi/(1-\lambda) + \delta/(1+\delta)$. Similarly, using $\sigma \geqslant p/2 = \tau^{-1}$ and $\tau \leqslant 1/2$ we obtain

$$\tau\sigma^3(1-\varepsilon)^p \geqslant \tau^{-2}(1-\tau) \geqslant \tau^{-2}/2.$$

Since $\varepsilon^4\mu^2 \geqslant (1+\delta)^{-1}$ by assumption, analogously to the proof of Theorem 6, using (19) together with Lemmas 13–14, we obtain

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mu) \geqslant \mathbb{P}(X < (1-\varepsilon)\mu) \geqslant e^{-(1+\kappa)\varphi(-\varepsilon)\mu},$$

with $\kappa = 2p^2(\sigma + p\delta + \Pi) + 2p\sigma + 8\tau^2 p(1+\delta)$. Now, using $\max\{\Pi, \tau\} \leqslant 1$ and $\sigma \leqslant p = 2/\tau = 10/(1-\Pi)$, a short calculation shows that, say,

$$1 + \kappa \leqslant 17 + 2p^3 + 4p^2 + (2p^3 + 16)\delta \leqslant 2500(1+\delta)/(1-\Pi)^3.$$

Finally, we assume $\tau^2/2 \leqslant \varepsilon \leqslant 1$. Using the lower bound (3) resulting from Harris' inequality [12], it follows that

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mu) \geqslant \mathbb{P}(X = 0) \geqslant e^{-\mu/(1-\Pi)}. \tag{35}$$

The point is that, by assumption, we have $2/\varepsilon^2 \leqslant 8/\tau^4 = 5000/(1-\Pi)^4$, so that Lemma 10 implies $1 \leqslant 5000\varphi(-\varepsilon)/(1-\Pi)^4$. □

2.4 Proofs of Theorem 1 and Lemma 7

The remaining proofs of Theorem 1 and Lemma 7 are straightforward.

Proof of Lemma 7. Note that, by assumption, $5\sqrt{1-\varepsilon} \leqslant 5e^{-1/2} \leqslant 4$. So, using Lemma 11, we infer

$$1/(1-\Pi) \leqslant (1 + 5\sqrt{1-\varepsilon})\big(1 + \Pi/(1-\Pi)\big)\varphi(-\varepsilon) \leqslant (1+\zeta)\varphi(-\varepsilon),$$

with $\zeta = 10\max\{\sqrt{1-\varepsilon}, \Pi/(1-\Pi)\}$. Now an application of (3), analogous to (35), completes the proof. □

Proof of Theorem 1. Note that, using the assumption,

$$\eta = \max\{4\Pi^{1/4}, \mathbb{1}_{\{\varepsilon<1\}}4\delta^{1/4}, \mathbb{1}_{\{\varepsilon<1\}}e^{-1}(\varepsilon^2\mu)^{-1/2}\}$$

satisfies $\eta \in [0, e^{-1}]$. If $1 - \eta \leqslant \varepsilon \leqslant 1$, then $\varepsilon \geqslant 1 - e^{-1}$ and $1 - \varepsilon \leqslant \eta$, so that Lemma 7 implies (4). If $0 \leqslant \varepsilon < 1 - \eta$, then $e(1-\varepsilon)\varepsilon^2\mu \geqslant e\eta\varepsilon^2\mu \geqslant (\varepsilon^2\mu)^{1/2} \geqslant 1$ and $\varepsilon \leqslant 1 - 4\max\{\Pi^{1/4}, \delta^{1/4}\}$, so that Theorem 6 establishes (4). □

3 Bootstrapping lower bounds for the lower tail

As discussed, Theorem 1 and 2 only give reasonable lower bounds for the lower tail if $\delta = O(1)$, i.e., as long as the dependencies are 'weak'. In this section we present a bootstrapping strategy, which often allows us to deal with the remaining case, where $\delta = \Omega(1)$ holds.

In order to establish a competent lower bound on the lower tail, we usually need to (approximately) identify the most likely way to obtain $X \leqslant (1-\varepsilon)\mathbb{E}X$. At first glance it seems that this would require fairly detailed information about the random variable X, where $\mu = \mathbb{E}X$. However, in the general setting of this paper, we discovered that, perhaps surprisingly, we can *systematically* guess suitable (nearly) 'extremal' events by only inspecting the form of the variance $\operatorname{Var} X \leqslant \Lambda = \Lambda(X)$. Indeed, assume that there is a random variable Y, of the same type as (1), satisfying

$$\Lambda = \Theta(\mu^2/\mathbb{E}Y) \quad \text{and} \quad \delta(Y) = O(1). \tag{36}$$

For example, if X_H counts the number of copies of a given graph H in $G_{n,p}$, then (36) holds for $X = X_H$ with $Y = X_G$, where $G \subseteq H$ is a suitable subgraph (see [15, 16] or Lemma 22). Defining $\mathcal{E}$ as the event that $Y \leqslant (1-\varepsilon)\mathbb{E}Y$ holds, our starting point is the basic inequality

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant \mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X \mid \mathcal{E})\mathbb{P}(\mathcal{E}). \tag{37}$$

Assuming that Theorem 1 or 2 applies to Y, using (36) there are constants $c_1, c_2 > 0$ such that

$$\mathbb{P}(\mathcal{E}) \geqslant e^{-c_1\varphi(-\varepsilon)\mathbb{E}Y} \geqslant e^{-c_2\varphi(-\varepsilon)\mu^2/\Lambda}. \tag{38}$$

Hence it remains to estimate $\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X \mid \mathcal{E})$ from below. It turns out that if X and Y are suitably related (as in the subgraphs example), then under fairly mild conditions we can prove that $\mathbb{E}(X \mid \mathcal{E})$ is quite a bit smaller than $(1-\varepsilon)\mathbb{E}X$. In other words, by conditioning on $\mathcal{E}$ we intuitively 'convert' the *rare* event $X \leqslant (1-\varepsilon)\mathbb{E}X$ into a *typical* one (this subtle conditioning idea is at the heart of our approach). With this in mind it seems plausible that we have, say,

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X \mid \mathcal{E}) = \Omega(1), \tag{39}$$

although $\geqslant e^{-c_3\varphi(-\varepsilon)\mu^2/\Lambda}$ suffices for our purposes. Note that for the special case $\varepsilon = 1$ this inequality is immediate in the subgraphs example (where $X_G = 0$ implies $X_H = 0$). Finally, by combining (37)–(39) we obtain

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) = \Omega(e^{-c_2\varphi(-\varepsilon)\mu^2/\Lambda}), \tag{40}$$

which qualitatively matches the upper bound of (2), as desired.

To implement this proof strategy, we need to be able to verify that (39) holds (or a related inequality). Here the main technical challenge is that, after conditioning on $\mathcal{E}$, the $i \in \Gamma$ are no longer added independently to $\Gamma_{\mathbf{p}}$. In Sections 3.1–3.3 we present three approaches that, in symmetric situations, allow us to *routinely* overcome this difficulty (each of them hinges on an event that is similar to $\mathcal{E}$). Since we are interested in large deviations (with exponentially small probabilities), here $(\varepsilon\mu)^2 = \Omega(\Lambda)$ is a natural condition in view of (2), (40) and the fact $\varphi(-\varepsilon) = \Theta(\varepsilon^2)$.

3.1 Binomial random subset

The first approach is motivated by the following simple observation: if $|\Gamma_{\mathbf{p}}| = 0$, then *deterministically* $X = 0$. Indeed, this yields

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant \mathbb{P}(X = 0) \geqslant \mathbb{P}(|\Gamma_{\mathbf{p}}| = 0),$$

which for $\varepsilon = \Theta(1)$ may give a fair lower bound. The next theorem, for the case of equal p_i, is based on the following heuristic extension of this observation: if $|\Gamma_{\mathbf{p}}|$ is 'too small', then we expect that X is *typically* also 'too small'. As we shall see, the crux is that conditioning on $|\Gamma_{\mathbf{p}}| \leqslant (1-\varepsilon)\mathbb{E}|\Gamma_{\mathbf{p}}|$ decreases the expected value of X, which intuitively increases the probability that $X \leqslant (1-\varepsilon)\mathbb{E}X$ occurs. Note that $\mathbb{E}(X \mid |\Gamma_{\mathbf{p}}| = 0) = 0$ confirms this phenomenon in the special case $\varepsilon = 1$.

Theorem 15. *Let $X = \sum_{\alpha \in \mathcal{X}} I_\alpha$, $\mu = \mathbb{E}X$ and Λ be defined as in Section 1. Suppose that $\mathbf{p} = (p, \ldots, p) \in [0,1]^N$ and $\min_{\alpha \in \mathcal{X}} |Q(\alpha)| \geqslant 2$. For all $\varepsilon \in (0,1]$ satisfying $(\varepsilon\mu)^2 \geqslant \mathbb{1}_{\{\varepsilon<1\}}\Lambda$, with $c = 1/2 + \mathbb{1}_{\{\varepsilon=1\}}1/2$,*

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant c\mathbb{P}(|\Gamma_{\mathbf{p}}| \leqslant (1-\varepsilon)\mathbb{E}|\Gamma_{\mathbf{p}}|). \tag{41}$$

In the proof of Theorem 15 we use the following one-sided version of Chebyshev's inequality (see, e.g., Theorem A.17 in [9]).

Claim 16. *If $\operatorname{Var} Z \leqslant v$, then $\mathbb{P}(Z \geqslant \mathbb{E}Z + t) \leqslant v/(v+t^2)$ for all $t > 0$.*

Proof of Theorem 15. Given $0 \leqslant j \leqslant N$, we write $\mathbb{P}(\cdot \mid |\Gamma_{\mathbf{p}}| = j) = \mathbb{P}_j(\cdot)$ for brevity. Note that for $m = (1-\varepsilon)Np = (1-\varepsilon)\mathbb{E}|\Gamma_{\mathbf{p}}|$ we have

$$\begin{aligned}\mathbb{P}(X \leqslant (1-\varepsilon)\mu) &\geqslant \sum_{0 \leqslant j \leqslant m} \mathbb{P}_j(X \leqslant (1-\varepsilon)\mu)\mathbb{P}(|\Gamma_{\mathbf{p}}| = j)\\ &\geqslant \mathbb{P}(|\Gamma_{\mathbf{p}}| \leqslant m) \min_{0 \leqslant j \leqslant m} \mathbb{P}_j(X \leqslant (1-\varepsilon)\mu).\end{aligned} \tag{42}$$

Since $\mathbb{P}_0(X \leqslant (1-\varepsilon)\mu) \geqslant \mathbb{P}_0(X=0) = 1$, we henceforth may assume $m \geqslant 1$. Consequently $\varepsilon < 1$ and $p > 0$ hold, so that $\mu \geqslant \min_{\alpha \in \mathcal{X}} \mathbb{E} I_\alpha \geqslant p^N > 0$.

In the following we estimate the conditional expected value and variance of X. Given $0 \leqslant j \leqslant m$, we write $\mathbb{E}(\cdot \mid |\Gamma_{\mathbf{p}}| = j) = \mathbb{E}_j(\cdot)$ and $\mathrm{Var}(\cdot \mid |\Gamma_{\mathbf{p}}| = j) = \mathrm{Var}_j(\cdot)$ for brevity. Let $\Gamma_j \subseteq \Gamma$ with $|\Gamma_j| = j$ be chosen uniformly at random. Since $\mathbf{p} = (p, \ldots, p)$, it follows that $\Gamma_{\mathbf{p}}$ conditioned on $|\Gamma_{\mathbf{p}}| = j$ has the same distribution as Γ_j. As $|Q(\alpha)| \geqslant 2$ and $j \leqslant m \leqslant N$, using $I_\alpha = \mathbb{1}_{\{Q(\alpha) \subseteq \Gamma_{\mathbf{p}}\}}$ we infer

$$\begin{aligned} \mathbb{E}_j(I_\alpha) = \mathbb{1}_{\{|Q(\alpha)| \leqslant j\}} \frac{\binom{N-|Q(\alpha)|}{j-|Q(\alpha)|}}{\binom{N}{j}} &= \mathbb{1}_{\{|Q(\alpha)| \leqslant j\}} \prod_{0 \leqslant i < |\alpha|} \frac{j-i}{N-i} \\ &\leqslant \left(\frac{j}{N}\right)^{|Q(\alpha)|} \leqslant (1-\varepsilon)^{|Q(\alpha)|} p^{|Q(\alpha)|} \leqslant (1-\varepsilon)^2 \mathbb{E} I_\alpha. \end{aligned} \tag{43}$$

Since $I_\alpha I_\beta = \mathbb{1}_{\{Q(\alpha) \cup Q(\beta) \subseteq \Gamma_{\mathbf{p}}\}}$, we analogously obtain $\mathbb{E}_j(I_\alpha I_\beta) \leqslant (1-\varepsilon)^2 \mathbb{E}(I_\alpha I_\beta)$. Furthermore, if $Q(\alpha) \cap Q(\beta) = \emptyset$ and $|Q(\alpha)| + |Q(\beta)| \leqslant j$, then a similar calculation shows that

$$\mathbb{E}_j(I_\alpha \mid I_\beta = 1) = \frac{\binom{N-|Q(\beta)|-|Q(\alpha)|}{j-|Q(\beta)|-|Q(\alpha)|}}{\binom{N-|Q(\beta)|}{j-|Q(\beta)|}} = \prod_{0 \leqslant i < |Q(\alpha)|} \frac{j - |Q(\beta)| - i}{N - |Q(\beta)| - i} \leqslant \mathbb{E}_j(I_\alpha).$$

If $|Q(\alpha) \cup Q(\beta)| > j$ then, trivially, $\mathbb{E}_j(I_\alpha I_\beta) = 0$. It follows that $Q(\alpha) \cap Q(\beta) = \emptyset$ implies $\mathbb{E}_j(I_\alpha I_\beta) - \mathbb{E}_j(I_\alpha)\mathbb{E}_j(I_\beta) \leqslant 0$. Combining our findings, we deduce that

$$\max_{0 \leqslant j \leqslant m} \mathbb{E}_j(X) \leqslant (1-\varepsilon)^2 \mu \quad \text{and} \quad \max_{0 \leqslant j \leqslant m} \mathrm{Var}_j(X) \leqslant (1-\varepsilon)^2 \Lambda. \tag{44}$$

Finally, using (44) and the one-sided Chebyshev's inequality (Claim 16) we infer that for every $0 \leqslant j \leqslant m$ we have

$$\mathbb{P}_j(X > (1-\varepsilon)\mu) \leqslant \mathbb{P}_j(X \geqslant \mathbb{E}_j(X) + (1-\varepsilon)\varepsilon\mu) \leqslant \Lambda/(\Lambda + (\varepsilon\mu)^2),$$

which together with $(\varepsilon\mu)^2 \geqslant \Lambda$ and (42) establishes (41). □

The proof shows that (41) holds with c replaced by $1 - \mathbb{1}_{\{\varepsilon<1,\mu>0\}} \Lambda/(\Lambda + (\varepsilon\mu)^2)$, and that the left hand side of (41) can be strengthened to $\mathbb{P}(X < (1-\varepsilon)\mathbb{E}X)$ whenever $\varepsilon \in (0,1)$ and $\mu > 0$ (we henceforth omit analogous remarks).

In applications where constant factors in the exponent are important, the following variant of Theorem 15 usually gives better results when $\varepsilon \to 0$ and $L = (\varepsilon\mu)^2/\Lambda \to \infty$ (by setting $\tau = 6\max\{\varepsilon, L^{-1/2}\}$; see Lemma 12 with $A = (1+\tau)/k$).

Theorem 17. *Let $X = \sum_{\alpha \in \mathcal{X}} I_\alpha$, $\mu = \mathbb{E}X$ and Λ be defined as in Section 1. Suppose that $\mathbf{p} = (p, \ldots, p) \in [0,1]^N$ and $\min_{\alpha \in \mathcal{X}} |Q(\alpha)| \geqslant k \geqslant 1$. For all $\varepsilon, \tau \in (0,1]$ satisfying $\tau \geqslant \mathbb{1}_{\{k>1\}} 6\varepsilon$ and $(\varepsilon\mu)^2 \geqslant 4\tau^{-2}\Lambda$, with $c = 1/2$,*

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant c\mathbb{P}(|\Gamma_{\mathbf{p}}| \leqslant (1-(1+\tau)\varepsilon/k)\mathbb{E}|\Gamma_{\mathbf{p}}|). \tag{45}$$

Proof. Let $\lambda = (1+\tau)\varepsilon/k$ and $m = (1-\lambda)\mathbb{E}|\Gamma_{\mathbf{p}}|$. As (45) is trivial otherwise, we henceforth assume $\mathbb{P}(|\Gamma_{\mathbf{p}}| \leqslant m) > 0$, which implies $m \geqslant 0$. Now, (42) carries over mutatis mutandis, and, with similar reasoning as in the proof of Theorem 15, we may henceforth assume $\min\{m, p, \mu\} > 0$. Furthermore, as $\min_{\alpha \in \mathcal{X}} |Q(\alpha)| \geqslant k$, the calculations leading to (44) imply

$$\max_{0 \leqslant j \leqslant m} \mathbb{E}_j(X) \leqslant (1-\lambda)^k \mu \quad \text{and} \quad \max_{0 \leqslant j \leqslant m} \mathrm{Var}_j(X) \leqslant \Lambda. \tag{46}$$

If $k = 1$, then $(1-\varepsilon) - (1-\lambda)^k = \lambda - \varepsilon = \tau\varepsilon$, and we now establish a similar bound for $k > 1$. Note that $\lambda k = (1+\tau)\varepsilon \leqslant 2\varepsilon \leqslant \tau/3 < 1$ and

$$(1-\lambda)^k \leqslant e^{-\lambda k} \leqslant 1 - \lambda k + \sum_{j \geqslant 2} \frac{(\lambda k)^j}{j!} \leqslant 1 - \lambda k + \frac{(\lambda k)^2}{2(1-\lambda k)}.$$

Recalling $\lambda k = (1+\tau)\varepsilon$, $\varepsilon \leqslant \tau/6$ and $\tau \leqslant 1$, a short calculation shows that

$$(1-\varepsilon) - (1-\lambda)^k \geqslant \tau\varepsilon\left(1 - \frac{(1+\tau)^2\varepsilon}{2\tau(1-(1+\tau)\varepsilon)}\right) \geqslant \tau\varepsilon/2.$$

Consequently, using (46) and the one-sided Chebyshev's inequality (Claim 16), we infer that for every $0 \leqslant j \leqslant m$ we have

$$\mathbb{P}_j(X > (1-\varepsilon)\mu) \leqslant \mathbb{P}_j(X \geqslant \mathbb{E}_j(X) + \tau\varepsilon\mu/2) \leqslant \Lambda/(\Lambda + \tau^2(\varepsilon\mu)^2/4),$$

which together with $(\varepsilon\mu)^2 \geqslant 4\tau^{-2}\Lambda$ and (42) establishes (45). $\square$

3.2 Symmetric decomposition

In general, the conditional expected value of X is difficult to compute (as we do not have explicit formulas as in (43)). Our second approach shows that we can overcome this obstacle using a *symmetric decomposition* of X. As an illustration, we again consider the number of copies of H in $G_{n,p}$. Clearly, for every $G \subseteq H$ we have $\mathbb{P}(X_H = 0) \geqslant \mathbb{P}(X_G = 0)$. The basic idea is now that, by counting the number of H-copies extending each copy of G, we ought to be able to argue as follows: if X_G is 'too small', then the (conditional) expected value of X_H is also 'too small'. To avoid clutter, we henceforth use the abbreviation

$$I_{\alpha\setminus\beta} = \mathbb{1}_{\{Q(\alpha)\setminus Q(\beta)\subseteq\Gamma_{\mathbf{p}}\}}. \tag{47}$$

Let $\mathcal{H} = \mathcal{H}_n$ contain all subgraphs isomorphic to H in K_n, and define $Q(\alpha) = E(\alpha)$ for all $\alpha \in \mathcal{H}$ (here $Q(\alpha) \neq \alpha$ is crucial to allow for isolated vertices in H). The key observation is that, by symmetry, there is a constant $w > 0$ such that we may write

$$X_H = w \sum_{\beta\in\mathcal{G}} I_\beta \sum_{\alpha\in\mathcal{H}:\beta\subseteq\alpha} I_{\alpha\setminus\beta},$$

where $\mathbb{E}\big(\sum_{\alpha\in\mathcal{H}:\beta\subseteq\alpha} I_{\alpha\setminus\beta}\big)$ is *independent* of the choice of $\beta \in \mathcal{G}$. The point is that, since $\mathbb{E}(I_\beta I_{\alpha\setminus\beta}) = \mathbb{E}I_\beta\mathbb{E}I_{\alpha\setminus\beta}$ and $X_G = \sum_{\beta\in\mathcal{G}} I_\beta$, this allows us to factorize $\mathbb{E}X_H$ in terms of $\mathbb{E}X_G$. Indeed, for any $\tilde\beta \in \mathcal{G}$ we have

$$\mathbb{E}X_H = w\mathbb{E}\big(\sum_{\alpha\in\mathcal{H}:\tilde\beta\subseteq\alpha} I_{\alpha\setminus\tilde\beta}\big)\sum_{\beta\in\mathcal{G}}\mathbb{E}I_\beta = w\mathbb{E}X_G\mathbb{E}\big(\sum_{\alpha\in\mathcal{H}:\tilde\beta\subseteq\alpha} I_{\alpha\setminus\tilde\beta}\big).$$

Intuitively, our approach exploits that correlation inequalities can be used to obtain a similar factorization of the *conditional* expected value of X_H.

With the subgraphs example in mind, the following theorem should be interpreted under the premise that the lower bound is exponentially small in $\Theta((\varepsilon\mu)^2/\Lambda)$. In other words, the multiplicative $\gamma\varepsilon$ error-term ought to be negligible as long as, say, $\gamma\varepsilon \geqslant e^{-(\varepsilon\mu)^2/\Lambda}$ holds. The crux is that this inequality is equivalent to $(\varepsilon\mu)^2/\Lambda \geqslant \log\big(1/(\gamma\varepsilon)\big)$, which matches our usual condition up to the logarithmic factor. On first reading it might be useful to consider the important special case exemplified above, where $w_{\alpha,\beta} = w > 0$, $\mathcal{X}(\beta) = \{\alpha \in \mathcal{X} : Q(\beta) \subseteq Q(\alpha)\}$ and $\kappa = 0$.

Theorem 18. *Let $Y = \sum_{\beta\in\mathcal{Y}} I_\beta$, where $\big(Q(\beta)\big)_{\beta\in\mathcal{Y}}$ is a family of subsets of Γ. Suppose that there are $w_{\alpha,\beta} \in [0,\infty)$ and families $\big(Q(\alpha)\big)_{\alpha\in\mathcal{X}(\beta)}$ of subsets of Γ such that $X = \sum_{\beta\in\mathcal{Y}} I_\beta X_\beta$, where $X_\beta = \sum_{\alpha\in\mathcal{X}(\beta)} w_{\alpha,\beta} I_{\alpha\setminus\beta}$ satisfies $\max_{\beta\in\mathcal{Y}}\mathbb{E}X_\beta \leqslant (1+\kappa)\min_{\beta\in\mathcal{Y}}\mathbb{E}X_\beta$ for $\kappa \in [0,\infty)$. For all $\varepsilon \in [0,1]$ and $\gamma \in [0,\infty)$ satisfying $\gamma\varepsilon \geqslant 2\kappa$ and $\mathbb{1}_{\{\mathbb{E}Y=0\}}\gamma\varepsilon \leqslant 2$, with $c = 1/2$,*

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant c\gamma\varepsilon\mathbb{P}(Y \leqslant (1-(1+\gamma)\varepsilon)\mathbb{E}Y). \tag{48}$$

If $\varepsilon \nearrow 1$ or $\varepsilon = 1$ holds, then, by applying Lemma 7 to Y, we often can improve (48) via

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mathbb{E}X) \geqslant \mathbb{P}(X = 0) \geqslant \mathbb{P}(Y = 0). \tag{49}$$

The proof of Theorem 18 hinges on the following simple consequence of Harris' inequality [12], which was observed by Bollobás and Riordan (see Lemma 6 in [4]).

Claim 19. *For the probability space induced by $\Gamma_{\mathbf{p}}$, suppose that $\mathcal{D}$ is a decreasing event with $\mathbb{P}(\mathcal{D}) > 0$, and that $\mathcal{I}_1$ and $\mathcal{I}_2$ are increasing events with $\mathbb{P}(\mathcal{I}_1 \cap \mathcal{I}_2) = \mathbb{P}(\mathcal{I}_1)\mathbb{P}(\mathcal{I}_2)$. Then*

$$\mathbb{P}(\mathcal{I}_1 \cap \mathcal{I}_2 \mid \mathcal{D}) \leqslant \mathbb{P}(\mathcal{I}_1)\mathbb{P}(\mathcal{I}_2 \mid \mathcal{D}). \tag{50}$$

Proof of Theorem 18. Let $y = (1-(1+\gamma)\varepsilon)\mathbb{E}Y$ and $\mu = \mathbb{E}X$. As (48) is trivial otherwise, we henceforth assume $\gamma\varepsilon > 0$ and $\mathbb{P}(Y \leqslant y) > 0$, which since $Y \geqslant 0$ implies $y \geqslant 0$. If $\mathbb{E}Y = 0$, then $\mathbb{P}(Y = 0) = \mathbb{P}(Y \leqslant y)$, and, since we then assume $1 \geqslant \gamma\varepsilon/2$, (49) establishes (48). Henceforth we thus assume $\mathbb{E}Y > 0$, so that $y \geqslant 0$ implies $1 \geqslant (1+\gamma)\varepsilon > \max\{\varepsilon, \gamma\varepsilon\}$. Note that

$$\mathbb{P}(X \leqslant (1-\varepsilon)\mu) \geqslant \mathbb{P}(Y \leqslant y)\mathbb{P}(X \leqslant (1-\varepsilon)\mu \mid Y \leqslant y). \tag{51}$$

Since $\mathbb{E}(I_\beta I_{\alpha\setminus\beta}) = \mathbb{E}I_\beta \mathbb{E}I_{\alpha\setminus\beta}$, using the definitions of X, X_β and Y we deduce

$$\mu = \mathbb{E}X = \sum_{\beta\in\mathcal{Y}} \mathbb{E}I_\beta \mathbb{E}X_\beta \geqslant \mathbb{E}Y \min_{\beta\in\mathcal{Y}} \mathbb{E}X_\beta \geqslant (1+\kappa)^{-1}\mathbb{E}Y \max_{\beta\in\mathcal{Y}} \mathbb{E}X_\beta. \tag{52}$$

We write $\mathcal{I}_\alpha$ and $\mathcal{I}_{\alpha\setminus\beta}$ for the increasing events that $I_\alpha = 1$ and $I_{\alpha\setminus\beta} = 1$, respectively. Hence $\mathbb{P}(\mathcal{I}_{\alpha\setminus\beta}\mathcal{I}_\beta) = \mathbb{P}(\mathcal{I}_{\alpha\setminus\beta})\mathbb{P}(\mathcal{I}_\beta)$. Clearly, $Y \leqslant y$ is a decreasing event. Using Claim 19 together with (52) and $(1-(1+\gamma)\varepsilon)(1+\kappa) \leqslant 1-(1+\gamma/2)\varepsilon$, it follows that

$$\begin{aligned}\mathbb{E}(X \mid Y \leqslant y) &= \sum_{\beta\in\mathcal{Y}} \sum_{\alpha\in\mathcal{X}(\beta)} w_{\alpha,\beta}\mathbb{P}(\mathcal{I}_{\alpha\setminus\beta}\mathcal{I}_\beta \mid Y \leqslant y) \leqslant \sum_{\beta\in\mathcal{Y}} \mathbb{P}(\mathcal{I}_\beta \mid Y \leqslant y) \sum_{\alpha\in\mathcal{X}(\beta)} w_{\alpha,\beta}\mathbb{P}(\mathcal{I}_{\alpha\setminus\beta}) \\ &\leqslant \mathbb{E}(Y \mid Y \leqslant y) \max_{\beta\in\mathcal{Y}} \mathbb{E}X_\beta \leqslant y \max_{\beta\in\mathcal{Y}} \mathbb{E}X_\beta \leqslant (1-(1+\gamma/2)\varepsilon)\mu.\end{aligned} \tag{53}$$

Let $\lambda = 1+\gamma/2$. If $\mu > 0$, then, using Markov's inequality, we infer from (53)

$$\mathbb{P}(X > (1-\varepsilon)\mu \mid Y \leqslant y) \leqslant \frac{1-\lambda\varepsilon}{1-\varepsilon} = 1 - \frac{(\lambda-1)\varepsilon}{1-\varepsilon} \leqslant 1-\gamma\varepsilon/2, \tag{54}$$

which together with (51) establishes (48). Finally, if $\mu = 0$, then $\mathbb{P}(X > 0) = 0$ and (48) follows trivially from the fact $1 > \gamma\varepsilon$ established above. □

It would be desirable to use Chebyshev's inequality in (54), since this presumably would improve the seemingly suboptimal $\gamma\varepsilon$ term. Here one technical obstacle is that Claim 19 can, in general, *not* be strengthened to

$$\mathbb{P}(\mathcal{I}_1 \cap \mathcal{I}_2 \mid \mathcal{D}) \leqslant \mathbb{P}(\mathcal{I}_1 \mid \mathcal{D})\mathbb{P}(\mathcal{I}_2 \mid \mathcal{D}). \tag{55}$$

Indeed, a short calculation shows that, for $\Gamma = [n] = \{1, \ldots, n\}$ and $\mathbf{p} = (p, \ldots, p)$ with $n \geqslant 3$ and $p \in (0,1)$, the events $\mathcal{I}_i = \{i \in \Gamma_{\mathbf{p}}\}$ and $\mathcal{D} = \{|\Gamma_{\mathbf{p}}| \leqslant 1 \text{ or } \Gamma_{\mathbf{p}} = \{1,2\}\}$ provide a counterexample (where, moreover, equality holds in (50)). It would be interesting to know whether there is perhaps some approximate version of (55) that suffices for our purposes.

The existence of a symmetric decomposition may not always be obvious. We hope that the following two examples from additive combinatorics serve as inspiration for future applications of Theorem 18 (or its method of proof). In both we consider $\mathbf{p} = (p, \ldots, p)$ and $Q(\alpha) = \alpha$, and the basic idea is to 'symmetrize' X using non-uniform 'weights' $w_{\alpha,\beta}$ (and $\kappa \neq 0$). In the first example, we let $\mathcal{X}$ contain all *arithmetic progressions* of length $k \geqslant 2$ in $\Gamma = [n]$, i.e., each $\alpha \in \mathcal{X}$ equals $\{b, b+d, \ldots, b+(k-1)d\} \subseteq [n]$ for some $b = b_\alpha$ and $d = d_\alpha$ with $b_\alpha, d_\alpha \geqslant 1$. For every $\beta \in \mathcal{Y} = [n]$ we define $\mathcal{X}(\beta)$ as the set of $\alpha \in \mathcal{X}$ where $\beta = b_\alpha$ or $\beta = b_\alpha + (k-1)d_\alpha$, and set $w_{\alpha,\beta} = 1/2$. Since each $\alpha \in \mathcal{X}$ contributes to exactly two X_β, we have $X = \sum_{\beta\in\mathcal{Y}} I_\beta X_\beta$. Furthermore, careful counting yields

$$\mathbb{E}X_\beta = \frac{1}{2}\Big(\Big\lfloor\frac{n-\beta}{k-1}\Big\rfloor + \Big\lfloor\frac{\beta-1}{k-1}\Big\rfloor\Big)p^{k-1} = \Big(\frac{n}{2(k-1)} + O(1)\Big)p^{k-1},$$

so $\kappa = O(1/n)$ suffices. In the second example, we let $\mathcal{X}$ contain all *Schur triples* in $\Gamma = [n]$, i.e., each $\alpha \in \mathcal{X}$ equals $\{x, y, x+y\} \subseteq [n]$ for some $x = x_\alpha$ and $y = y_\alpha$ with $1 \leqslant x_\alpha < y_\alpha$. For every $\beta \in \mathcal{Y} = [n]$ we define

$\mathcal{X}(\beta)$ as the set of all $\alpha \in \mathcal{X}$ with $\beta \in \alpha$. We set $w_{\alpha,\beta} = 1/2$ if $\beta = x_\alpha + y_\alpha$, and $w_{\alpha,\beta} = 1/4$ otherwise. By counting triples, it is not hard to see that $X = \sum_{\beta \in \mathcal{Y}} I_\beta X_\beta$ and

$$\mathbb{E}X_\beta = \Big(\frac{1}{2}\Big\lfloor\frac{\beta-1}{2}\Big\rfloor + \frac{\max\{n-2\beta,0\}+\min\{n-\beta,\beta-1\}}{4}\Big)p^2 = \Big(\frac{n}{4}+O(1)\Big)p^2,$$

so $\kappa = O(1/n)$ suffices. Finally, in both examples routine calculations (analogous to Example 3.2 in [16]) give $\mu^2/\Lambda = \Theta(\min\{\mu, np\})$. Since $\kappa = O(1/n)$ and $\mu^2/\Lambda = O(np)$, the natural condition $(\varepsilon\mu)^2 = \Omega(\Lambda)$ thus implies $\kappa/\varepsilon = O(1/n \cdot \sqrt{\mu^2/\Lambda}) = O(\sqrt{p/n}) = o(1)$. In other words, the assumption $\gamma\varepsilon \geqslant 2\kappa$ in Theorem 18 is very mild, i.e., allows for $\gamma = o(1)$.

3.3 Vertex symmetry

In many applications the set Γ has additional structure, and here our main focus is on the case where Γ contains the edges of some hypergraph. Intuitively, 'seeing' the underlying vertices introduces quite a bit of extra symmetry, and our third approach exploits this to step aside the conditioning issue we faced in the previous subsection. As an illustration, we consider, as before, the number of copies of H in $G_{n,p}$. The basic idea is to partition the vertex set into $\mathcal{U}$ and $[n] \setminus \mathcal{U}$ with $|\mathcal{U}| \approx n/2$, and then, for suitable $G \subseteq H$, to focus on the number of copies of G completely contained in $\mathcal{U}$, which we denote by Y_G. Note that $\mathbb{E}Y_G = \Theta(\mathbb{E}X_G)$. Perhaps rashly, we would like to argue that $Y_G \leqslant (1-\varepsilon)\mathbb{E}Y_G$ typically entails $X_H \leqslant (1-\varepsilon)\mathbb{E}X_H$. However, this is overly ambitious: since Y_G is somewhat 'local', we loose a bit when going to the 'global' random variable X_H, and thus we need a slightly larger deviation of Y_G. Instead of counting all copies of H, a technical reduction allows us to focus on the number of *pairs* (H', G') of copies of H and G with $G' \subseteq H'$, $V(G') \subseteq \mathcal{U}$ and $V(H') \setminus V(G') \subseteq [n] \setminus \mathcal{U}$. Now, to make variance calculations feasible (i.e., to overcome the obstacle that (55) may fail), we do not condition on Y_G, but rather on *all edges* with both endvertices in U (satisfying additional typical properties). For technical reasons, here our argument requires that all edges in the relevant graphs $H' \setminus G'$ have at least one endvertex outside of $\mathcal{U}$, which, e.g., holds if all copies of G in H are *induced* subgraphs. Luckily, it is not hard to check (see Lemma 22) that the former condition always holds for some $G \subseteq H$ that determines the exponent, i.e., satisfies $\Lambda(X_H) = \Theta((\mathbb{E}X_H)^2/\mathbb{E}X_G)$.

In the statement of the next theorem we restrict ourselves to subgraph counts in random hypergraphs. The approach works in a more general setting, but we resist the temptation of stating a very technical theorem (that would be difficult to apply). Instead, we tried to write the proof in a way that hopefully makes the basic setup and symmetry assumptions fairly transparent. In Theorem 20 the difference between Y_G and X_G is usually irrelevant in applications where constant factors in the exponent are immaterial: the point is that $G^{(k)}_{n,p}[\mathcal{U}]$ has the same distribution as $G^{(k)}_{n',p}$ with $n' = |\mathcal{U}| \approx n/2$. In comparison with Theorem 18, the key feature of Theorem 20 is that the natural condition $(\varepsilon\mathbb{E}X_H)^2 = \Omega(\Lambda(X_H))$ suffices.

Theorem 20. *Let $G \subseteq H$ be k-graphs with $e_G \geqslant 1$, where every copy of G in H is induced. Let X_H be the number of copies of H in $G^{(k)}_{n,p}$, and let Y_G be the number of copies of G in $G^{(k)}_{n,p}[\mathcal{U}]$, where $\mathcal{U} \subseteq [n]$ satisfies $\big||\mathcal{U}| - n/2\big| \leqslant \ell$. For all $n \geqslant n_0 = n_0(H,\ell)$, $p \in [0,1]$ and $\varepsilon \in (0,1]$ satisfying $(\varepsilon\mathbb{E}X_H)^2 \geqslant \Lambda(X_H)$, with $\lambda = 2^{v_H+3}$ and $c = 2^{-(4^{v_G^2}+2)}$,*

$$\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \geqslant c\mathbb{P}(Y_G \leqslant (1-\lambda\varepsilon)\mathbb{E}Y_G). \tag{56}$$

Proof. Let $\mu = \mathbb{E}X_H$, $\Lambda = \Lambda(X_H)$, $\Gamma = E(K^{(k)}_n)$ and $\mathbf{p} = (p,\dots,p)$, so that $\Gamma_{\mathbf{p}} = E(G^{(k)}_{n,p})$. Let $\mathcal{H}$ and $\mathcal{G}$ contain all subgraphs isomorphic to H and G in $K^{(k)}_n$, respectively. Define $Q(\sigma) = E(\sigma)$ for $\sigma \in \mathcal{H} \cup \mathcal{G}$. For brevity we henceforth use $I_{(\alpha_1\cup\alpha_2)\setminus(\beta_1\cup\beta_2)} = \mathbb{1}_{\{[Q(\alpha_1)\cup Q(\alpha_2)]\setminus[Q(\beta_1)\cup Q(\beta_2)]\subseteq\Gamma_{\mathbf{p}}\}}$ and $I_{\sigma_1\cup\sigma_2} = \mathbb{1}_{\{Q(\sigma_1)\cup Q(\sigma_2)\subseteq\Gamma_{\mathbf{p}}\}}$ analogous to (47). Set $Z = \sum_{(\alpha,\beta)\in\mathcal{H}\times\mathcal{G}} \mathbb{1}_{\{\beta\subseteq\alpha\}} I_\alpha$. By symmetry, we have $\sum_{\beta\in\mathcal{G}} \mathbb{1}_{\{\beta\subseteq\alpha\}} = \tau = \tau(H,G) \geqslant 1$ for all $\alpha \in \mathcal{H}$. Hence $Z = \tau X$, $\mathbb{E}Z = \tau\mathbb{E}X_H$, $\operatorname{Var} Z = \tau^2 \operatorname{Var} X_H$ and

$$\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) = \mathbb{P}(Z \leqslant (1-\varepsilon)\mathbb{E}Z). \tag{57}$$

With foresight, we set $Z_S = \sum_{(\alpha,\beta)\in\mathcal{H}\times\mathcal{G}} \mathbb{1}_{\{\alpha\in\mathcal{H}(S,\beta) \text{ and } \beta\in\mathcal{G}(S)\}} I_\alpha$ for all $S \subseteq [n]$, where

$$\begin{aligned}\mathcal{H}(S,\beta) &= \{\alpha \in \mathcal{H} : \beta \subseteq \alpha \text{ and } V(\alpha)\setminus V(\beta) \subseteq [n]\setminus S\},\\ \mathcal{G}(S) &= \{\beta \in \mathcal{G} : V(\beta) \subseteq S\}.\end{aligned}$$

Define $R_{\mathcal{U}} = Z - Z_{\mathcal{U}}$, $z = (1-\varepsilon\lambda/2)\mathbb{E}Z_{\mathcal{U}}$ and $r = (1-\varepsilon)\mathbb{E}Z - z$. Using $Z = R_{\mathcal{U}} + Z_{\mathcal{U}}$ and Harris' inequality, it follows that

$$\mathbb{P}(Z \leqslant (1-\varepsilon)\mathbb{E}Z) \geqslant \mathbb{P}(R_{\mathcal{U}} \leqslant r \text{ and } Z_{\mathcal{U}} \leqslant z) \geqslant \mathbb{P}(R_{\mathcal{U}} \leqslant r)\mathbb{P}(Z_{\mathcal{U}} \leqslant z). \tag{58}$$

The remainder of the proof is devoted to the following two inequalities, which together with (57), (58) and $(\varepsilon\mu)^2 \geqslant \Lambda$ imply (56):

$$\mathbb{P}(R_{\mathcal{U}} \leqslant r) \geqslant 1 - \mathbb{1}_{\{\mu>0\}}\Lambda/(\Lambda + (\varepsilon\mu)^2), \tag{59}$$

$$\mathbb{P}(Z_{\mathcal{U}} \leqslant z) \geqslant \big(1 - \mathbb{1}_{\{\mu>0\}}\Lambda/(\Lambda + 2(\varepsilon\mu)^2)\big)4c\mathbb{P}(Y_G \leqslant (1-\lambda\varepsilon)\mathbb{E}Y_G). \tag{60}$$

We note first that in the trivial case $\mu = 0$, almost surely $X = 0$ and thus $Z = 0$ which implies $R_{\mathcal{U}} = Z_{\mathcal{U}} = 0$; hence also $z = 0$ and $r = 0$ so that (59)–(60) follow trivially. We may thus assume $\mu > 0$.

We next estimate $\mathbb{E}Z_{\mathcal{U}}$. Let $\mathfrak{X} \subseteq [n]$ with $|\mathfrak{X}| = |\mathcal{U}|$ be chosen uniformly at random, and independent of $\Gamma_{\mathbf{p}}$. With the definitions of $\mathcal{H}(\cdot,\beta)$ and $\mathcal{G}(\cdot)$ in mind, using linearity of expectation we deduce

$$\mathbb{E}(Z_{\mathfrak{X}} \mid \Gamma_{\mathbf{p}}) = \sum_{(\alpha,\beta)\in\mathcal{H}\times\mathcal{G}} \mathbb{1}_{\{\beta\subseteq\alpha\}}\mathbb{P}(V(\beta) \subseteq \mathfrak{X} \text{ and } V(\alpha)\setminus V(\beta) \subseteq [n]\setminus\mathfrak{X})I_\alpha, \tag{61}$$

where the measure $\mathbb{P}$ is with respect to the (random) choice of $\mathfrak{X}$. Note that, whenever $\beta \subseteq \alpha$, we have

$$\sigma_{\alpha,\beta} = \mathbb{P}(V(\beta) \subseteq \mathfrak{X} \text{ and } V(\alpha)\setminus V(\beta) \subseteq [n]\setminus\mathfrak{X}) = \frac{\binom{n-v_H}{|\mathcal{U}|-v_G}}{\binom{n}{|\mathcal{U}|}}.$$

Recall that $\big||\mathcal{U}| - n/2\big| \leqslant \ell$. For fixed ℓ, v_G and v_H a short calculation shows that $\sigma_{\alpha,\beta} \to 2^{-v_H}$ as $n \to \infty$, so that $\sigma_{\alpha,\beta} \geqslant 2^{-(v_H+1)} = 4\lambda^{-1}$ for $n \geqslant n_0(H,\ell)$. Using (61) and the definition of Z we infer $\mathbb{E}(Z_{\mathfrak{X}} \mid \Gamma_{\mathbf{p}}) \geqslant 4\lambda^{-1}Z$, so that $\mathbb{E}(Z_{\mathfrak{X}}) \geqslant 4\lambda^{-1}\mathbb{E}Z$. By definition, we have $\mathbb{E}(Z_{\mathfrak{X}} \mid \mathfrak{X} = S) = \mathbb{E}Z_S$ for all $S \subseteq [n]$ with $|S| = |\mathcal{U}|$. Since $\mathbb{E}Z_S = \mathbb{E}Z_{\mathcal{U}}$ by symmetry, we infer $\mathbb{E}Z_{\mathfrak{X}} = \mathbb{E}Z_{\mathcal{U}}$, so that

$$\mathbb{E}Z_{\mathcal{U}} \geqslant 4\lambda^{-1}\mathbb{E}Z. \tag{62}$$

Turning to (59), note that $R_{\mathcal{U}}$ is a restriction of Z to a subset of all pairs $(\alpha,\beta) \in \mathcal{H}\times\mathcal{G}$. As Harris' inequality implies $\mathbb{E}(I_{\alpha_1}I_{\alpha_2}) \geqslant \mathbb{E}I_{\alpha_1}\mathbb{E}I_{\alpha_2}$, it follows that $\operatorname{Var} R_{\mathcal{U}} \leqslant \operatorname{Var} Z = \tau^2 \operatorname{Var} X_H \leqslant \tau^2\Lambda$. Recalling $\mathbb{E}R_{\mathcal{U}} = \mathbb{E}Z - \mathbb{E}Z_{\mathcal{U}}$ and the definitions of r and z, using (62) we have $r - \mathbb{E}R_{\mathcal{U}} = (\varepsilon\lambda/2)\mathbb{E}Z_{\mathcal{U}} - \varepsilon\mathbb{E}Z \geqslant \varepsilon\mathbb{E}Z = \tau\varepsilon\mu$. So, if $\mu > 0$, then the one-sided Chebyshev's inequality (Claim 16) yields

$$\mathbb{P}(R_{\mathcal{U}} > r) \leqslant \mathbb{P}(R_{\mathcal{U}} \geqslant \mathbb{E}R_{\mathcal{U}} + \tau\varepsilon\mu) \leqslant \tau^2\Lambda/(\tau^2\Lambda + (\tau\varepsilon\mu)^2) = \Lambda/(\Lambda + (\varepsilon\mu)^2).$$

In the remainder we focus on (60). Observing that $Y_G = \sum_{\beta\in\mathcal{G}(\mathcal{U})} I_\beta$, we denote by $\mathcal{E}$ the event that $Y_G \leqslant (1-\lambda\varepsilon)\mathbb{E}Y_G$ holds. With foresight, we define $X_\beta = \sum_{\alpha\in\mathcal{H}(\mathcal{U},\beta)} I_{\alpha\setminus\beta}$ and $X_{\beta_1,\beta_2} = \sum_{(\alpha_1,\alpha_2)\in\mathcal{H}(\beta_1,\beta_2)} I_{(\alpha_1\cup\alpha_2)\setminus(\beta_1\cup\beta_2)}$, where

$$\mathcal{H}(\beta_1,\beta_2) = \big\{(\alpha_1,\alpha_2) \in \mathcal{H}(\mathcal{U},\beta_1)\times\mathcal{H}(\mathcal{U},\beta_2) : \big[Q(\alpha_1)\cap Q(\alpha_2)\big] \setminus \big[Q(\beta_1)\cup Q(\beta_2)\big] \neq \emptyset\big\}.$$

Let $\mathcal{F}$ be the family of all pairwise non-isomorphic graphs that are unions of two (not necessarily distinct) copies of G. The point is that $\mathcal{F}$ naturally defines a partition $(\mathcal{P}_F)_{F\in\mathcal{F}}$ of the set of all pairs of graphs $(\beta_1,\beta_2) \in \mathcal{G}(\mathcal{U})\times\mathcal{G}(\mathcal{U})$ with $\mathcal{H}(\beta_1,\beta_2) \neq \emptyset$ (as each $\beta_1\cup\beta_2$ is isomorphic to some $F \in \mathcal{F}$). Furthermore, since every $F \in \mathcal{F}$ satisfies $v_G \leqslant v_F \leqslant 2v_G$, we have, say, $|\mathcal{F}| \leqslant 2^{\binom{2v_G}{2}} \cdot 2^{v_G} \leqslant 4^{v_G^2}$. Let $\Psi_F = \sum_{(\beta_1,\beta_2)\in\mathcal{P}_F} I_{\beta_1\cup\beta_2}$, and define $\mathcal{D}$ as the event that $\Psi_F \leqslant 2\mathbb{E}\Psi_F$ for all $F \in \mathcal{F}$. Using Harris' inequality and Markov's inequality, we deduce

$$\mathbb{P}(\mathcal{E}\cap\mathcal{D}) \geqslant \mathbb{P}(\mathcal{E})\prod_{F\in\mathcal{F}}\mathbb{P}(\Psi_F \leqslant 2\mathbb{E}\Psi_F) \geqslant 2^{-|\mathcal{F}|}\mathbb{P}(\mathcal{E}) \geqslant 4c\mathbb{P}(\mathcal{E}). \tag{63}$$

For brevity, we write $\mathbb{P}^*$ for the conditional measure with respect to the status of all edges in $G^{(k)}_{n,p}[\mathcal{U}]$. We use $\mathbb{E}^*$ and Var^* analogously. Since $\mathcal{E}\cap\mathcal{D}$ is determined by $E(G^{(k)}_{n,p}[\mathcal{U}])$, we have

$$\mathbb{P}(Z_{\mathcal{U}} \leqslant z) \geqslant \mathbb{P}(\{Z_{\mathcal{U}} \leqslant z\}\cap\mathcal{E}\cap\mathcal{D}) = \mathbb{E}\big(\mathbb{P}^*(Z_{\mathcal{U}} \leqslant z)\mathbb{1}_{\{\mathcal{E}\cap\mathcal{D}\}}\big). \tag{64}$$

In the following we estimate $\mathbb{P}^*(Z_{\mathcal{U}} \leqslant z)$ whenever $\mathcal{E}\cap\mathcal{D}$ holds. Recall that for all $\beta \in \mathcal{G}(\mathcal{U})$ and $\alpha \in \mathcal{H}(\mathcal{U},\beta)$ we have $\beta \subseteq \alpha$, $V(\beta) \subseteq \mathcal{U}$ and $V(\alpha) \setminus V(\beta) \subseteq [n] \setminus \mathcal{U}$. Since every copy of G in H is induced, for all $f \in Q(\alpha) \setminus Q(\beta)$ we infer $f \notin E(K_n^{(k)}[\mathcal{U}])$. Using $Q(\beta) \subseteq Q(\alpha)$ it follows that $\mathbb{E}^* I_\alpha = I_\beta \mathbb{E}^* I_{\alpha\setminus\beta} = I_\beta \mathbb{E} I_{\alpha\setminus\beta}$. By symmetry, $\mathbb{E}X_\beta$ is independent of the choice of $\beta \in \mathcal{G}(U)$, and so $\mathbb{E}^* Z_{\mathcal{U}} = \sum_{\beta\in\mathcal{G}(\mathcal{U})} I_\beta \mathbb{E}X_\beta = Y_G \mathbb{E}X_{\tilde\beta}$ for any $\tilde\beta \in \mathcal{G}(\mathcal{U})$. Taking expectations, we deduce $\mathbb{E}Z_{\mathcal{U}} = \mathbb{E}Y_G \mathbb{E}X_{\tilde\beta}$. Consequently $\mathbb{E}^* Z_{\mathcal{U}} \leqslant (1-\lambda\varepsilon)\mathbb{E}Z_{\mathcal{U}}$ whenever $\mathcal{E}$ holds, in which case, using the definition of z and (62), we have

$$z - \mathbb{E}^* Z_{\mathcal{U}} \geqslant (\varepsilon\lambda/2)\mathbb{E}Z_{\mathcal{U}} \geqslant 2\varepsilon\mathbb{E}Z = 2\tau\varepsilon\mu. \tag{65}$$

Turning to the conditional variance of $Z_{\mathcal{U}}$, note that, by symmetry (analogous as for Z), we have

$$\begin{aligned}\tau^2\Lambda &= \sum_{\alpha\in\mathcal{H}} \sum_{\substack{(\beta_1,\beta_2)\in\mathcal{G}\times\mathcal{G}:\\ \beta_1\subseteq\alpha,\beta_2\subseteq\alpha}} \mathbb{E}I_\alpha + \sum_{(\alpha_1,\alpha_2)\in\mathcal{H}\times\mathcal{H}:\alpha_1\sim\alpha_2} \sum_{\substack{(\beta_1,\beta_2)\in\mathcal{G}\times\mathcal{G}:\\ \beta_1\subseteq\alpha_1,\beta_2\subseteq\alpha_2}} \mathbb{E}I_{\alpha_1\cup\alpha_2} \\ &= \sum_{(\beta_1,\beta_2)\in\mathcal{G}\times\mathcal{G}} \mathbb{E}I_{\beta_1\cup\beta_2} \sum_{\substack{(\alpha_1,\alpha_2)\in\mathcal{H}\times\mathcal{H}:\beta_1\subseteq\alpha_1,\beta_2\subseteq\alpha_2,\\ Q(\alpha_1)\cap Q(\alpha_2)\neq\emptyset}} \mathbb{E}I_{(\alpha_1\cup\alpha_2)\setminus(\beta_1\cup\beta_2)}.\end{aligned} \tag{66}$$

As before, $\mathbb{E}^* I_{\alpha_1\cup\alpha_2} = I_{\beta_1\cup\beta_2}\mathbb{E}^* I_{(\alpha_1\cup\alpha_2)\setminus(\beta_1\cup\beta_2)} = I_{\beta_1\cup\beta_2}\mathbb{E} I_{(\alpha_1\cup\alpha_2)\setminus(\beta_1\cup\beta_2)}$ for all $(\beta_1,\beta_2) \in \mathcal{G}(U)\times\mathcal{G}(U)$ and $(\alpha_1,\alpha_2) \in \mathcal{H}(U,\beta_1)\times\mathcal{H}(U,\beta_2)$. It follows that

$$\operatorname{Var}^* Z_{\mathcal{U}} \leqslant \sum_{(\beta_1,\beta_2)\in\mathcal{G}(U)\times\mathcal{G}(U)} I_{\beta_1\cup\beta_2} \sum_{\substack{(\alpha_1,\alpha_2)\in\mathcal{H}(U,\beta_1)\times\mathcal{H}(U,\beta_2):\\ [Q(\alpha_1)\cap Q(\alpha_2)]\setminus[Q(\beta_1)\cup Q(\beta_2)]\neq\emptyset}} \mathbb{E}I_{(\alpha_1\cup\alpha_2)\setminus(\beta_1\cup\beta_2)}.$$

Now, recalling the definitions of $\mathcal{H}(\beta_1,\beta_2)$, X_{β_1,β_2}, $\mathcal{F}$ and Ψ_F, we infer

$$\operatorname{Var}^* Z_{\mathcal{U}} \leqslant \sum_{F\in\mathcal{F}} \sum_{(\beta_1,\beta_2)\in\mathcal{P}_F} I_{\beta_1\cup\beta_2}\mathbb{E}X_{\beta_1,\beta_2} \leqslant \sum_{F\in\mathcal{F}} \Psi_F \max_{(\beta_1,\beta_2)\in\mathcal{P}_F} \mathbb{E}X_{\beta_1,\beta_2}.$$

By symmetry, we have $\max_{(\beta_1,\beta_2)\in\mathcal{P}_F} \mathbb{E}X_{\beta_1,\beta_2} = \min_{(\beta_1,\beta_2)\in\mathcal{P}_F} \mathbb{E}X_{\beta_1,\beta_2}$ for all $F \in \mathcal{F}$. So, with analogous considerations as above, whenever $\mathcal{D}$ holds we have

$$\begin{aligned}\operatorname{Var}^* Z_{\mathcal{U}} &\leqslant 2\sum_{F\in\mathcal{F}} \mathbb{E}\Psi_F \min_{(\beta_1,\beta_2)\in\mathcal{P}_F} \mathbb{E}X_{\beta_1,\beta_2} = 2\sum_{F\in\mathcal{F}} \sum_{(\beta_1,\beta_2)\in\mathcal{P}_F} \mathbb{E}I_{\beta_1\cup\beta_2}\mathbb{E}X_{\beta_1,\beta_2} \\ &= 2\sum_{(\beta_1,\beta_2)\in\mathcal{G}(U)\times\mathcal{G}(U)} \mathbb{E}I_{\beta_1\cup\beta_2} \sum_{(\alpha_1,\alpha_2)\in\mathcal{H}(\beta_1,\beta_2)} \mathbb{E}I_{(\alpha_1\cup\alpha_1)\setminus(\beta_1\cup\beta_2)} \leqslant 2\tau^2\Lambda,\end{aligned} \tag{67}$$

where the last inequality follows by comparison with (66). If $\mu > 0$, then, using (65), the one-sided Chebyshev's inequality (Claim 16) and (67), whenever $\mathcal{E}\cap\mathcal{D}$ holds we have

$$\mathbb{P}^*(Z_{\mathcal{U}} > z) \leqslant \mathbb{P}^*(Z_{\mathcal{U}} \geqslant \mathbb{E}^* Z_{\mathcal{U}} + 2\tau\varepsilon\mu) \leqslant 2\tau^2\Lambda/(2\tau^2\Lambda + (2\tau\varepsilon\mu)^2) = \Lambda/(\Lambda + 2(\varepsilon\mu)^2). \tag{68}$$

Inserting (68) into (64), we infer (for $\mu > 0$)

$$\mathbb{P}(Z_{\mathcal{U}} \leqslant z) \geqslant \big(1 - \Lambda/(\Lambda + 2(\varepsilon\mu)^2)\big)\mathbb{P}(\mathcal{E}\cap\mathcal{D}),$$

which together with (63) implies (60) by definition of $\mathcal{E}$. □

A variant of the proof applies to *rooted* copies of H, see, e.g., Section 3 in [19] for a precise definition. The basic idea is to map the vertex set of the root R to $[r]$, and the remaining vertices of G and H to $\mathcal{U} \subseteq [n]\setminus[r]$ and $[n]\setminus(\mathcal{U}\cup[r])$, respectively; we leave the details to the interested reader.

4 Applications

In this section we illustrate the bootstrapping approaches of Section 3 via pivotal examples from additive and probabilistic combinatorics. In Section 4.1 we consider the lower tail of the number of arithmetic progressions (and Schur triples) in random subsets of the integers. In Section 4.2 we then turn to our main example: the lower tail of subgraph counts in random hypergraphs.

4.1 Random subsets of the integers

Let $X_k = X_k(n,p)$ be the number of arithmetic progressions of length $k \geqslant 2$ in the binomial random subset $\Gamma_{\mathbf{p}}$ of the integers $\Gamma = [n] = \{1,\dots,n\}$, where $\mathbf{p} = (p,\dots,p)$. Note that $\mathbb{E}X_k = \Theta(n^2p^k)$; see also Section 3.2. The following theorem gives fair exponential bounds for the lower tail of X_k, and its proof closely follows the strategy outlined in Section 3.

Theorem 21. *Given $k \geqslant 2$, let $\Psi_k = \Psi_k(n,p) = \min\{n^2p^k, np\}$. There are positive constants c, C, D and n_0, all depending only on k, such that for all $n \geqslant n_0$, $p \in [0,1)$ and $\varepsilon \in (0,1]$ satisfying $\varepsilon^2\Psi_k \geqslant \mathbb{1}_{\{\varepsilon<1\}}D$ we have*

$$\exp\bigl\{-(1-p)^{-5}C\varepsilon^2\Psi_k\bigr\} \leqslant \mathbb{P}(X_k \leqslant (1-\varepsilon)\mathbb{E}X_k) \leqslant \exp\bigl\{-c\varepsilon^2\Psi_k\bigr\}. \tag{69}$$

Proof. Let $\mu = \mathbb{E}X_k$, $\Lambda = \Lambda(X_k)$ and $\delta = \delta(X_k)$. Routine calculations, analogous to Example 3.2 in [16], reveal that

$$\delta = \Theta(np^{k-1}+p) \quad \text{and} \quad \mu^2/\Lambda = \mu/(1+\delta) = \Theta(\Psi_k), \tag{70}$$

where the implicit constants depend only on k. Hence the upper bound of (69) is an immediate consequence of (2). For the lower bound we pick, with foresight, $D = D(k) \geqslant 1$ such that $\mathbb{E}X_k \geqslant \Psi_k/D$ and $\mu^2/\Lambda \geqslant \Psi_k/D$ for $n \geqslant n_0(k)$.

If $\Psi_k = n^2p^k$, then Theorem 2 (with $X = X_k$) yields

$$\mathbb{P}(X_k \leqslant (1-\varepsilon)\mathbb{E}X_k) \geqslant \exp\bigl\{-\Theta((1-p)^{-5}\varepsilon^2\Psi_k)\bigr\}$$

since $\varepsilon^2\mathbb{E}X_k \geqslant \varepsilon^2\Psi_k/D \geqslant \mathbb{1}_{\{\varepsilon<1\}}$, $\Pi(X_k) = p^k \leqslant p$, $\delta = O(1)$ and $\mathbb{E}X_k = \Theta(\Psi_k)$.

If $\Psi_k = np$, then Theorem 15 (with $X = X_k$) and Theorem 2 (with $X = |\Gamma_{\mathbf{p}}|$) yield, with $d = 1/2 + \mathbb{1}_{\{\varepsilon=1\}}1/2$,

$$\mathbb{P}(X_k \leqslant (1-\varepsilon)\mathbb{E}X_k) \geqslant d\mathbb{P}(|\Gamma_{\mathbf{p}}| \leqslant (1-\varepsilon)\mathbb{E}|\Gamma_{\mathbf{p}}|) \geqslant \exp\bigl\{-\mathbb{1}_{\{\varepsilon<1\}}\log 2 - \Theta((1-p)^{-5}\varepsilon^2\Psi_k)\bigr\}$$

since $(\varepsilon\mu)^2 \geqslant \Lambda\varepsilon^2\Psi_k/D \geqslant \mathbb{1}_{\{\varepsilon<1\}}\Lambda$, $\varepsilon^2\mathbb{E}|\Gamma_{\mathbf{p}}| = \varepsilon^2\Psi_k \geqslant \mathbb{1}_{\{\varepsilon<1\}}$ and $\mathbb{E}|\Gamma_{\mathbf{p}}| = \Psi_k$. This completes the proof of (69) since $\mathbb{1}_{\{\varepsilon<1\}}\log 2 \leqslant \mathbb{1}_{\{\varepsilon<1\}}D \leqslant (1-p)^{-5}\varepsilon^2\Psi_k$. $\square$

For Schur triples, which are defined in Section 3.2, the same calculations carry over (with $k = 3$; the point is that (70) holds), yielding an analogous lower tail estimate. Related results for the upper tail of arithmetic progressions and Schur triples have been established by Warnke [30].

4.2 Random hypergraphs

Finally, we consider the lower tail of the number $X_H = X_H(n,p)$ of copies of a given k-graph H in $G^{(k)}_{n,p}$, and prove Theorems 3–5. Here the following precise analysis of $\Lambda(X_H)$ is at the heart of our approach. In fact, Lemma 22 is essentially given in [15] (for $k = 2$), but the restriction to subgraphs from $\mathcal{I}_H$ is new and crucial for our purposes: the key point is that *every* copy of $G \in \mathcal{I}_H$ in H is induced. Recall that $m_k(H)$ is defined by (8).

Lemma 22. *Let H be a k-graph with $e_H \geqslant 1$. Define $\mathcal{I}_H$ as the collection of all non-isomorphic subgraphs $J \subseteq H$ which satisfy $e_J \geqslant \max\{e_K, 1\}$ for all $K \subseteq H$ with $v_K = v_J$. For all $p = p(n) \in (0,1]$ we have*

$$\Lambda(X_H) = (1+o(1))\sum_{J\in\mathcal{I}_H} C^2_{J,H}\frac{(\mathbb{E}X_H)^2}{\mathbb{E}X_J} = \Theta\Bigl(\frac{(\mathbb{E}X_H)^2}{\min_{J\in\mathcal{I}_H}\mathbb{E}X_J}\Bigr), \tag{71}$$

$$\min_{J\in\mathcal{I}_H}\mathbb{E}X_J = o\bigl(\min_{J\subseteq H, e_J\geqslant 1, J\not\in^*\mathcal{I}_H}\mathbb{E}X_J\bigr), \tag{72}$$

where $C_{J,H}$ denotes the number of copies of J in H, and $J\not\in^*\mathcal{I}_H$ means that there is no $J' \in \mathcal{I}_H$ which is isomorphic to J. In addition, $p = \omega(n^{-1/m_k(H)})$ implies $\min_{J\in\mathcal{I}_H}\mathbb{E}X_J = \binom{n}{k}p$ and $\Lambda(X_H) = (1+o(1))e_H^2(\mathbb{E}X_H)^2/[\binom{n}{k}p]$.

The fairly standard proof of Lemma 22 is deferred to Appendix A. In the following proofs of Theorems 3–5 we shall not explicitly discuss the upper bounds: once the form of $(\mathbb{E}X_H)^2/\Lambda(X_H)$ has been established, these are immediate consequences of (2).

Proof of Theorem 3. Let $d = 2^{-(4^{v_H^2}+2)}$, $\lambda = 2^{v_H+3}$ and $\varepsilon_0 = (2\lambda)^{-1}$. Since the claim is trivial otherwise, we henceforth assume $p > 0$. Furthermore, we use the convention that all implicit constants depend only on H, and tacitly assume $n \geqslant n_0(H)$ whenever necessary. Suppose that $\Phi_H = \mathbb{E}X_G$ for $G \subseteq H$ with $e_G \geqslant 1$. Using (71) and (72) we infer $G \in \mathcal{I}_H$, $(\mathbb{E}X_H)^2/\Lambda(X_H) = \Theta(\Phi_H)$ and $\delta(X_G) = O(1)$. With foresight, we pick $D = D(H) \geqslant \log(1/d)$ such that $(\mathbb{E}X_H)^2/\Lambda(X_H) \geqslant \Phi_H/D$ holds.

If $\varepsilon \in [\varepsilon_0, 1]$, then $\Pi(X_G) = p^{e_G} \leqslant p$, $\mathbb{E}X_G = \Phi_H$, $1 \leqslant \varepsilon_0^{-2}\varepsilon^2$ and (3) yield

$$\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \geqslant \mathbb{P}(X_H = 0) \geqslant \mathbb{P}(X_G = 0) \geqslant \exp\bigl\{-(1-p)^{-1}\varepsilon_0^{-2}\varepsilon^2\Phi_H\bigr\}. \tag{73}$$

It remains to establish (6) when $\varepsilon < \varepsilon_0$. We shall eventually apply Theorem 20 with $U = [\lfloor n/2 \rfloor]$, where Y_G counts the total number of copies of G whose vertex sets are completely contained in U. Since $G^{(k)}_{n,p}[\mathcal{U}]$ has the same distribution as $G^{(k)}_{n',p}$ with $n' = |\mathcal{U}| \approx n/2$, we readily deduce $3^{-v_G}\mathbb{E}X_G \leqslant \mathbb{E}Y_G \leqslant \mathbb{E}X_G$ and $\delta(Y_G) = \Theta(\delta(X_G))$. Furthermore, $G \in \mathcal{I}_H$ implies that every copy of G in H is induced. So, using $\lambda\varepsilon \leqslant 1/2$, $\Pi(Y_G) \leqslant p$, $\delta(Y_G) = O(1)$ and $\mathbb{E}Y_G = \Theta(\Phi_H)$, a combination of Theorem 20 and Theorem 2 yields

$$\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \geqslant d\mathbb{P}(Y_G \leqslant (1-\lambda\varepsilon)\mathbb{E}Y_G) \geqslant \exp\bigl\{-\log(1/d) - \Theta((1-p)^{-5}\lambda^2\varepsilon^2\Phi_H)\bigr\}$$

since $\varepsilon^2(\mathbb{E}X_H)^2 \geqslant \varepsilon^2\Phi_H\Lambda(X_H)/D \geqslant \Lambda(X_H)$ and $(\lambda\varepsilon)^2\mathbb{E}Y_G \geqslant \lambda^2 3^{-v_G}\varepsilon^2\mathbb{E}X_G \geqslant \varepsilon^2\Phi_H \geqslant D \geqslant 1$. This completes the proof of (6) since $\log(1/d) \leqslant D \leqslant (1-p)^{-5}\varepsilon^2\Phi_H$. □

Proof of Theorem 4. Since the claim is trivial otherwise, we henceforth assume $p > 0$. Furthermore, since $p = o(1)$ we have $\Pi = o(1)$. Recalling the properties of G, using (71) and (72) we infer $G \in \mathcal{I}_H$, $(\mathbb{E}X_H)^2/\Lambda(X_H) = (1+o(1))\mathbb{E}X_G$ and $\delta(X_G) = o(1)$.

In the special case $e_G = 1$, note that uniqueness of G in H implies $e_H = 1$, and that minimality of $\mathbb{E}X_G$ implies $v_G = k$. Thus $X_H = X_G\binom{n-k}{v_H-k}$ and $\delta(X_G) = 0$. Using $\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) = \mathbb{P}(X_G \leqslant (1-\varepsilon)\mathbb{E}X_G)$, the lower bound of (9) now follows from Theorem 1 (applied to X_G), where $\xi = o(1)$ by our assumptions.

Henceforth we thus assume $e_G \geqslant 2$. Now, in case of $H = G$ the lower bound of (9) follows directly from Theorem 1. In the main case, where $G \subsetneq H$ and $e_G \geqslant 2$, there exists, by assumption, $\omega = \omega(n) \to \infty$ such that $\varepsilon^2\mathbb{E}X_G \geqslant \mathbb{1}_{\{\varepsilon<1\}}\omega\log(e/\varepsilon)$. Setting $\gamma = 2\exp\{-\omega^{1/2}\} = o(1)$ we have (when $\omega \geqslant 1$) $\varepsilon^2\mathbb{E}X_G \geqslant \mathbb{1}_{\{\varepsilon<1\}}\omega^{1/2}\log(2/(\gamma\varepsilon))$, which together with Lemma 10 yields $2^{-1}\gamma\varepsilon \geqslant \mathbb{1}_{\{\varepsilon<1\}}\exp\{-2\omega^{-1/2}\varphi(-\varepsilon)\mathbb{E}X_G\}$. So, if $(1+\gamma)\varepsilon < 1$ and $3\sqrt{\gamma} < 1-\varepsilon$, then a combination of Theorem 18 (with $X = X_H$, $Y = X_G$ and $\kappa = 0$), Theorem 1 (for X_G) and Lemma 12 (with $A = 1+\gamma$) establishes (9). Otherwise $\varepsilon \geqslant 1-\max\{\gamma/(1+\gamma), 3\sqrt{\gamma}\} = 1-o(1)$ holds, and then a combination of (49) (with $X = X_H$ and $Y = X_G$) and Lemma 7 (for X_G) completes the proof. □

Proof of Theorem 5. We start with the main case $\varepsilon = o(1)$. Note that Lemma 22 implies $\min_{J\in\mathcal{I}_H}\mathbb{E}X_J = \binom{n}{k}p = \mathbb{E}|\Gamma_{\mathbf{p}}|$ and $(\mathbb{E}X_H)^2/\Lambda(X_H) = (1+o(1))\mathbb{E}|\Gamma_{\mathbf{p}}|/e_H^2$. By assumption, there is $\omega = \omega(n) \to \infty$ such that $\varepsilon \leqslant 1/\omega$ and $\varepsilon^2\binom{n}{k}p \geqslant \omega$. Let $\tau = 6e_H\omega^{-1/2} = o(1)$ and $A = (1+\tau)/e_H$, so that $\varphi(-A\varepsilon) \leqslant (1+o(1))\varphi(-\varepsilon)/e_H^2$ by Lemma 12. Since $p = o(1)$, a combination of Theorem 17 (with $X = X_H$ and $k = e_H$) and Theorem 1 (with $X = |\Gamma_{\mathbf{p}}|$) establishes (10), where the factor $c = 1/2$ is negligible due to $\varphi(-\varepsilon)\binom{n}{k}p \to \infty$.

The remaining $\varepsilon = 1-o(1)$ estimate of (10) follows from Lemma 23 below and Lemma 11 since $1-p = e^{-(1+o(1))p}$ and $\varphi(-\varepsilon) = 1+o(1)$ for $p = o(1)$ and $\varepsilon = 1-o(1)$, respectively. □

The proof above used the following lemma, which follows from results of Saxton and Thomason [25].

Lemma 23. *Let H be a k-graph with $e_H \geqslant 1$. If $p = p(n) \in [0,1]$ and $\varepsilon = \varepsilon(n) \in (0,1]$ satisfy $p = \omega(n^{-1/m_k(H)})$ and $\varepsilon = 1-o(1)$, then we have*

$$\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) = (1-p)^{(1+o(1))(1-\pi_H)\binom{n}{k}}. \tag{74}$$

Proof. For the lower bound, let $\mathcal{T}_{n,H}$ be any hypergraph which achieves equality in the definition of $\mathrm{ex}(n,H)$. As every subgraph of $\mathcal{T}_{n,H}$ is H-free, it follows that

$$\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \geqslant \mathbb{P}(X_H = 0) \geqslant \mathbb{P}(G^{(k)}_{n,p} \subseteq \mathcal{T}_{n,H}) = (1-p)^{\binom{n}{k}-e(\mathcal{T}_{n,H})}.$$

This establishes the lower bound of (74) since $e(\mathcal{T}_{n,H}) = (\pi_H + o(1))\binom{n}{k}$ and $1-\pi_H \in (0,1]$.

Turning to the corresponding upper bound, we first consider the case $e_H \geqslant 2$. Let $0 < \delta \leqslant (1-\pi_H)/3$. Theorem 9.2 in [25] implies that there is $c = c(H,\delta) > 0$ such that for $n \geqslant c$ the following holds for all $q \in [n^{-1/m_k(H)}, 1/c]$: there exists $s \leqslant c$ and a mapping $T \mapsto C(T)$ of sequences $T = (T_1, \ldots, T_s)$ with $T_i \subseteq E(K_n^{(k)})$ to sets $C(T) \subseteq E(K_n^{(k)})$ such that for every k-graph G on n vertices with less than $n^{v_H} q^{e_H}$ copies of H there exists $T = (T_1, \ldots, T_s)$ such that $E(G) \subseteq C(T)$, $|C(T)| \leqslant (\pi_H + \delta)\binom{n}{k} = F$ and further $\sum_{1 \leqslant i \leqslant s} |T_i| \leqslant cqn^k = U$ and $\bigcup_{1 \leqslant i \leqslant s} T_i \subseteq E(G)$. (Recall that $E(K_n^{(k)})$ is the set of all edges in the complete k-graph $K_n^{(k)}$. The mapping $T \mapsto C(T)$ is quite complicated; the point of it is that we can bound the number of 'containers' $C(T)$ by the number of sequences T.)

By assumption we have $1 - \varepsilon \leqslant 1/\omega$ and $p \geqslant \omega n^{-1/m_k(H)}$, where $\omega = \omega(n) \to \infty$. Let $q = \omega^{-1/e_H} p$, so that $(1-\varepsilon)\mathbb{E}X_H < \omega^{-1} n^{v_H} p^{e_H} = n^{v_H} q^{e_H}$ and $n^{-1/m_k(H)} \leqslant q \leqslant \omega^{-1/e_H} \leqslant 1/c$ for $n \geqslant n_0(c)$. Note that we can construct a superset of all possible $T = (T_1, \ldots, T_s)$ as follows: we first decide on $|\bigcup_{1 \leqslant i \leqslant s} T_i| = u$, then select u edges of $K_n^{(k)}$ and decide on all the T_i in which they appear. So, taking the union bound over all choices of T that are possible for $G = G_{n,p}^{(k)}$, using $\bigcup_{1 \leqslant i \leqslant s} T_i \subseteq E(G_{n,p}^{(k)})$ and $E(G_{n,p}^{(k)}) \setminus C(T) = \emptyset$ it follows that

$$\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \leqslant \sum_{0 \leqslant u \leqslant U} \binom{\binom{n}{k}}{u} (2^s)^u p^u (1-p)^{\binom{n}{k} - F}. \tag{75}$$

Hence, recalling the definitions of F and U, for any $\theta \in (0,1]$ we obtain

$$\begin{aligned} \mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) &\leqslant (1-p)^{\binom{n}{k} - F} \theta^{-U} \sum_{0 \leqslant u \leqslant U} \binom{\binom{n}{k}}{u} (2^s)^u p^u \theta^u \\ &\leqslant (1-p)^{\binom{n}{k} - F} \theta^{-U} \bigl(1 + 2^s p\theta\bigr)^{\binom{n}{k}} \leqslant (1-p)^{(1-\pi_H-\delta)\binom{n}{k}} e^{cqn^k \log(1/\theta) + 2^s p\theta \binom{n}{k}}. \end{aligned} \tag{76}$$

Choose $\theta = q/p = o(1)$. Then $q \log(1/\theta) = p\theta \log(1/\theta) = o(p)$, $e^p \leqslant (1-p)^{-1}$ and (76) yield, for $n \geqslant n_0(c,s,\delta)$,

$$\mathbb{P}(X_H \leqslant (1-\varepsilon)\mathbb{E}X_H) \leqslant (1-p)^{(1-\pi_H-\delta)\binom{n}{k}} e^{o\left(p\binom{n}{k}\right)} \leqslant (1-p)^{(1-\pi_H-2\delta)\binom{n}{k}}. \tag{77}$$

It follows as usual that there is some $\delta(n) \to 0$ such that (77) holds with $\delta = \delta(n)$ for $n \geqslant n_0$, which together with $1 - \pi_H \in (0,1]$ establishes the upper bound of (74) when $e_H \geqslant 2$.

Finally, in the remaining case $e_H = 1$ (where Theorem 9.2 in [25] does not apply) we have $X_H = e(G_{n,p}^{(k)})\binom{n-k}{v_H-k}$. Hence $X_H \leqslant (1-\varepsilon)\mathbb{E}X_H$ is equivalent to $e(G_{n,p}^{(k)}) \leqslant (1-\varepsilon)\binom{n}{k}p$. Since $e(G_{n,p}^{(k)}) \sim \mathrm{Bin}\bigl(\binom{n}{k}, p\bigr)$ and $\pi_H = 0$, (74) follows by standard calculations. (For example, (75) holds with $s = 0$ and $U = F = (1-\varepsilon)p\binom{n}{k}$, and the reasoning of (76)–(77) carries over since $F = o(p\binom{n}{k})$ and $U \leqslant \omega^{-1} pn^k = qn^k$.) $\square$

Acknowledgement. We would like to thank Andrew Thomason for giving us a draft of [25] together with helpful comments on it.

References

[1] N. Alon and J. Spencer. *The probabilistic method.* Third edition. *Wiley-Interscience Series in Discrete Mathematics and Optimization.* John Wiley & Sons Inc., Hoboken, NJ (2008).

[2] A.D. Barbour. Poisson convergence and random graphs. *Math. Proc. Cambridge Philos. Soc.* **92** (1982), 349–359.

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

[4] B. Bollobás and O. Riordan. Colorings generated by monotone properties. *Random Struct. Alg.* **12** (1998), 1–25.

[5] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Struct. Alg.* **40** (2012), 437–451.

[6] S. Chatterjee and A. Dembo. Nonlinear large deviations. Preprint, 2014. `arXiv:1401.3495`.

[7] S. Chatterjee and S.R.S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.* **32** (2011), 1000–1017.

[8] B. DeMarco and J. Kahn. Tight upper tail bounds for cliques. *Random Struct. Alg.* **41** (2012), 469–487.

[9] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition. Applications of Mathematics (New York)* **31**. Springer-Verlag, New York (1996).

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

[11] C.M. Fortuin, P.W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22** (1971), 89–103.

[12] T.E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.* **56** (1960), 13–20.

[13] S. Janson. Poisson approximation for large deviations. *Random Struct. Alg.* **1** (1990), 221–229.

[14] S. Janson. New versions of Suen's correlation inequality. *Random Struct. Alg.* **13** (1998), 467–483.

[15] S. Janson, T. Łuczak, and A. Ruciński. An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In *Random graphs '87 (Poznań, 1987)*, pp. 73–87, Wiley, Chichester (1990).

[16] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley-Interscience, New York (2000).

[17] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.* **142** (2004), 61–92.

[18] S. Janson and A. Ruciński. The deletion method for upper tail estimates. *Combinatorica* **24** (2004), 615–640.

[19] S. Janson and A. Ruciński. Upper tails for counting objects in randomly induced subhypergraphs and rooted random graphs. *Ark. Mat.* **49** (2011), 79–96.

[20] P. Keevash. Hypergraph Turán problems. In *Surveys in combinatorics (Exeter 2011)*, pp. 83–139, Cambridge Univ. Press, Cambridge (2011).

[21] J.H. Kim and V.H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica* **20** (2000), 417–434.

[22] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. Preprint (2014). `arXiv:1402.6011`.

[23] O. Riordan and L. Warnke. The Janson inequalities for general up-sets. *Random Struct. Alg.*, to appear. `arXiv:1203.1024`.

[24] A. Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** (1988), 1–10.

[25] D. Saxton and A. Thomason. Hypergraph containers. (Revised version of arXiv:1204.6595v2.) Preprint (2014).

[26] M. Šileikis. On the upper tail of counts of strictly balanced subgraphs. *Electron. J. Combin.* **19** (2012), Paper 4.

[27] J. Spencer. Counting extensions. *J. Combin. Theory Ser. A* **55** (1990). 247–255.

[28] W.-C. Suen. A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Struct. Alg.* **1** (1990), 231–242.

[29] V.H. Vu. A large deviation result on the number of small subgraphs of a random graph. *Combin. Probab. Comput.* **10** (2001), 79–94.

[30] L. Warnke. Upper tails for arithmetic progressions in random subsets. Preprint (2013).

A Appendix

In this appendix we prove Lemmas 10–12 and 22.

Proof of Lemma 10. By our conventions, (16) is trivial for $\varepsilon = 1$, and so we henceforth assume $\varepsilon \in [0,1)$. First, let $f(x) = 2\varphi(-x) - (1-x)\log^2(1-x)$. Since $f'(x) = \log^2(1-x) \geqslant 0$ for $x \in [0,1)$, we infer $f(\varepsilon) \geqslant f(0) = 0$. Second, let $g(x) = 2\varphi(-x) - x^2$. Since $1 - x \leqslant e^{-x}$ implies $g'(x) = -2\log(1-x) - 2x \geqslant 0$ for $x \in [0,1)$, we infer $g(\varepsilon) \geqslant g(0) = 0$. Next, let $h(x) = \log^2(1-x) - 2\varphi(-x)$. Since $h'(x) = -2x(1-x)^{-1}\log(1-x) \geqslant 0$ for $x \in [0,1)$, we infer $h(\varepsilon) \geqslant h(0) = 0$. Finally, $1 - \varepsilon \leqslant e^{-\varepsilon}$ implies $\varphi(-\varepsilon) = (1-\varepsilon)\log(1-\varepsilon) + \varepsilon \leqslant \varepsilon^2$. $\square$

Proof of Lemma 11. As (17) is trivial otherwise, we henceforth assume $\varepsilon < 1$. Since $\varphi'(x) = \log(1+x) \leqslant 0$ for $x \in [-1,0]$, we infer $\varphi(-\varepsilon) \leqslant \varphi(-1) = 1$, which establishes the first inequality of (17).

Next, define $y = 1 - \varepsilon$, and note that $y \in (0, e^{-1}]$. Let $g(x) = \phi(x-1) = 1 - x\log(e/x)$. Since $g'(x) = \log x \leqslant 0$ for $x \in (0,1]$, we infer $g(y) \geqslant g(e^{-1}) = (e-2)/e > 0$. Let $h(x) = \sqrt{x}\log(e/x)$, and note that $h(y) > 0$. Since $h'(x) = -\log(ex)/(2\sqrt{x}) \geqslant 0$ for $x \in (0, e^{-1}]$, we infer $h(y) \leqslant h(e^{-1}) = 2/\sqrt{e}$. It follows that

$$\frac{1}{\varphi(-\varepsilon)} - 1 = \frac{1-g(y)}{g(y)} = \frac{\sqrt{y}h(y)}{g(y)} \leqslant \frac{2\sqrt{ey}}{e-2} \leqslant 5\sqrt{1-\varepsilon},$$

which establishes the second inequality of (17). $\square$

Proof of Lemma 12. We first consider the case $y = A\varepsilon \leqslant 1$, so that $y \in [0,1]$. Since $\log(1-x) = -\sum_{j\geqslant 1} x^j/j \leqslant -x - x^2/2$ for $x \in [0,1)$, we see that $\varphi(-y) = (1-y)\log(1-y) + y \leqslant (1+y)y^2/2$, where the inequality is trivial for $y = 1$ due to $\varphi(-1) = 1$. By Lemma 10 we have $\varepsilon^2/2 \leqslant \varphi(-\varepsilon)$, so that

$$\varphi(-A\varepsilon) \leqslant (1 + A\varepsilon)(A\varepsilon)^2/2 \leqslant (1 + A\varepsilon)A^2\varphi(-\varepsilon).$$

Turning to the second inequality of (18) we henceforth assume $\gamma > 0$ and $\varepsilon \in [0,1)$, as the claim is trivial otherwise. Let $\rho(x) = \varphi(-x)$, and note that $\rho'(x) = -\log(1-x)$ and $\rho''(x) = 1/(1-x)$. Since $\log(1-x) \geqslant -x/(1-x)$ for $x \in [0,1)$, c.f. (14), we see that $\rho'(\varepsilon) \leqslant \varepsilon/(1-\varepsilon)$. Note that $\gamma > 0$ and $3\sqrt{\gamma} \leqslant 1-\varepsilon$ imply $0 < 3\gamma^{3/2} \leqslant \gamma - \gamma\varepsilon \leqslant 1-(1+\gamma)\varepsilon$. So, recalling $\varepsilon^2/2 \leqslant \varphi(-\varepsilon)$ and $A = 1+\gamma$, using Taylor's theorem with remainder it follows that $0 \leqslant A\varepsilon < 1$ and

$$\begin{aligned}\varphi(-A\varepsilon) &\leqslant \varphi(-\varepsilon) + \gamma\varepsilon^2/(1-\varepsilon) + (\gamma\varepsilon)^2/[2(1-(1+\gamma)\varepsilon)]\\ &\leqslant \big(1 + 2\gamma/(1-\varepsilon) + \gamma^2/(1-(1+\gamma)\varepsilon)\big)\,\varphi(-\varepsilon) \leqslant (1+\sqrt{\gamma})\varphi(-\varepsilon),\end{aligned}$$

completing the proof of (18). □

Proof of Lemma 22. Define $\mathcal{S}_H$ as the collection of all non-isomorphic subgraphs $J \subseteq H$ with $e_J \geqslant 1$. Let $N(n,H)$ denote the number of copies of H in $K_n^{(k)}$. Note that $N(n,H) = \Theta(n^{v_H})$. By double counting pairs (J', H') of copies of J and H with $J' \subseteq H' \subseteq K_n^{(k)}$, using symmetry we infer that, in $K_n^{(k)}$, there are exactly

$$\lambda_{J,H}(n) = \frac{N(n,H)C_{J,H}}{N(n,J)} = \Theta(n^{v_H - v_J}) \tag{78}$$

copies of H containing any given copy of J. Since $\mathbb{E}X_J = N(n,J)p^{e_J}$ and $C_{H,H} = 1$, by distinguishing all possible intersections of H-copies it follows that

$$\Lambda(X_H) \leqslant \mathbb{E}X_H + \sum_{J\in\mathcal{S}_H : J\neq H} N(n,J)\lambda_{J,H}^2(n)p^{2e_H - e_J} = \sum_{J\in\mathcal{S}_H} C_{J,H}^2 \frac{(\mathbb{E}X_H)^2}{\mathbb{E}X_J}. \tag{79}$$

Recall that $\mathbb{E}X_J = \Theta(n^{v_J}p^{e_J})$. By definition, for every $K \in \mathcal{S}_H \setminus \mathcal{I}_H$ there is $J \in \mathcal{I}_H$ with $v_J = v_K$ and $e_J \geqslant e_K + 1$. Using $\mathbb{E}X_K = \Omega(p^{-1}\mathbb{E}X_J)$ we infer

$$\Lambda(X_H) \leqslant \sum_{J\in\mathcal{I}_H} \big(1 + \mathbb{1}_{\{e_J\geqslant 2\}}O(p)\big)C_{J,H}^2 \frac{(\mathbb{E}X_H)^2}{\mathbb{E}X_J}. \tag{80}$$

Suppose that $\omega = \omega(n) \to \infty$ satisfies $1 \leqslant \omega \leqslant n^{1/(2m_k(H)+1)}$. Using $m_k(H) \geqslant (e_K - 1)/(v_K - k)$ when $e_K \geqslant 2$, note that for $p \geqslant \omega n^{-1/m_k(H)}$ we have

$$\min_{K\in\mathcal{S}_H : v_K > k} n^{v_K - k}p^{e_K - 1} \geqslant \min\big\{n, \min_{K\in\mathcal{S}_H : e_K\geqslant 2} \omega^{e_K - 1}\big\} \geqslant \omega. \tag{81}$$

Thus the 'edge-term' with $e_J = 1$ and $v_J = k$ dominates (80) for $p \geqslant \omega n^{-1/m_k(H)}$: indeed, $K \neq J$ implies $\mathbb{E}X_K = \Omega(\omega\mathbb{E}X_J)$. As $\omega n^{-1/m_k(H)} \leqslant \omega^{-1}$, the $1 + \mathbb{1}_{\{e_J\geqslant 2\}}O(p)$ factor in (80) can thus be replaced by $1 + O(\omega^{-1})$, establishing the upper bound of (71). Furthermore, by combining $\mathbb{E}X_K = \Omega(p^{-1}\mathbb{E}X_J)$ and $\mathbb{E}X_K = \Omega(\omega\mathbb{E}X_J)$ in an analogous way, it is not difficult to see that (72) holds. For the lower bound of (71) we argue similar as for (79), but restrict our attention to intersections in subgraphs $J \in \mathcal{I}_H$ only. Moreover, to avoid overcounting (due to additional intersections outside of J), in the case $J \neq H$ we replace $\lambda_{J,H}^2(n)$ by

$$\lambda_{J,H}(n)\Big(\lambda_{J,H}(n) - O\big(\sum_{J'\subsetneq G\subseteq H : J'\cong J} \lambda_{G,H}(n)\big)\Big) = \big(1 - O(n^{-1})\big)\lambda_{J,H}^2(n),$$

where we used (78) and that every copy of $J \in \mathcal{I}_H$ in H is induced (which implies $v_G \geqslant v_J + 1$). With these modifications, the lower bound of (71) follows. □