Upper Tail Estimates with Combinatorial Proofs

Jan Hązła and Thomas Holenstein

ETH Zürich
Department of Computer Science, Zurich, Switzerland
{jan.hazla,thomas.holenstein}@inf.ethz.ch

Abstract

We study generalisations of a simple, combinatorial proof of a Chernoff bound similar to the one by Impagliazzo and Kabanets (RANDOM, 2010).

In particular, we prove a randomized version of the hitting property of expander random walks and use it to obtain an optimal expander random walk concentration bound settling a question asked by Impagliazzo and Kabanets.

Next, we obtain an upper tail bound for polynomials with input variables in $[0, 1]$ which are not necessarily independent, but obey a certain condition inspired by Impagliazzo and Kabanets. The resulting bound is applied by Holenstein and Sinha (FOCS, 2012) in the proof of a lower bound for the number of calls in a black-box construction of a pseudorandom generator from a one-way function.

We also show that the same technique yields the upper tail bound for the number of copies of a fixed graph in an Erdős–Rényi random graph, matching the one given by Janson, Oleszkiewicz, and Ruciński (Israel J. Math, 2002).

1998 ACM Subject Classification G.3 Probability and Statistics

Keywords and phrases concentration bounds, expander random walks, polynomial concentration

Digital Object Identifier 10.4230/LIPIcs.STACS.2015.392

1 Introduction

Motivation and previous work

Concentration bounds are inequalities that estimate the probability of a random variable assuming a value that is far from its expectation. They have a multitude of applications all across the mathematics and theoretical computer science. See, e.g., textbooks [26, 25, 4, 10] for uses in complexity theory and randomised algorithms.

A typical setting is when this variable is a function $f(x)$ of n simpler random variables $x = (x_1, \ldots, x_n)$ that possess a certain degree of independence and we try to bound said probability with a function decaying exponentially with n (or, maybe, n^ϵ for some $\epsilon > 0$).

The canonical examples are Chernoff-Hoeffding bounds [7, 13] for the sum of n independent random variables in $[0, 1]$ and Azuma's inequality [5] for martingales.

The standard technique to prove Chernoff bounds is due to Bernstein [6]. The idea is to bound $\mathrm{E}[e^{tf(x)}]$ for some appropriately chosen t, and then to apply Markov's inequality.

Recently, Impagliazzo and Kabanets [16] gave a different, combinatorial proof of Chernoff bound, arguing that its simplicity and nature provide additional insight into understanding concentration. What is more, their proof is constructive in a certain sense (see [16] for details).

The proof given by Impagliazzo and Kabanets is related to previous published results: in [28], Schmidt, Siegel and Srinivasan give a Chernoff bound which is applicable in case the random variables $x = (x_1, \ldots, x_n)$ are only m-wise independent for some large enough m. It

turns out that the expressions which appear in their computations have close counterparts in the proof in [16], but they still bound $\mathrm{E}[e^{tf(x)}]$, and it seems to us that the approach in [16] makes the concepts clearer and the calculations shorter.

Another work related to [16] is due to Janson, Oleszkiewicz and Ruciński [17], who give an upper tail bound (i.e., a one-sided concentration bound) for the number of subgraphs in an Erdős-Rényi random graph $\mathsf{G}_{n,p}$. The proof given in [17] bears much relationship to the proof given in [16]. We elaborate on that in Section 3.2.

Finally, there is a connection to an argument used by Rao to prove a concentration bound for parallel repetition of two-prover games [27]. As we will see, one of the ideas in the proof given in [16] is to consider a subset of the variables $(x_1, \ldots, x_n)$. Rao also does this, with a somewhat different purpose.

Our contributions

In this paper we modify the proof of Impagliazzo and Kabanets and introduce a more general sufficient condition for concentration which we term *growth boundedness* (Section 3). Then, we show some applications of our framework.

First, we prove a randomized version of the hitting property of expander random walks (Theorem 4.1) and use it to obtain an optimal (up to a constant factor in the exponent) expander random walk concentration bound settling a question asked in [16] (Theorem 4.2).[1] We also show that our method is quite robust: with a little more effort one can improve the constant factor to the optimal one in case of large number of steps and small deviation (Theorem 4.3).

Second, we prove an upper tail bound for polynomials with input random variables in $[0, 1]$ (Theorem 5.2). Contrary to the previous work we are aware of, we do not assume that those variables are independent, but rather that they obey a condition similar to growth boundedness.

This bound is used in a proof of a lower bound for the complexity of a black-box construction of a pseudorandom generator from a one-way function [14]. Although [14] was published earlier, the proof of the bound is not contained there, but deferred to this paper instead. We outline how the bound was used in [14] in Section 5.1.

Notation

Throughout the paper we focus on the bounds of the form $\Pr[f(x) \geq \mu(1+\epsilon)])$. We call such bounds "(multiplicative) upper tail bounds".

Typically, we consider a probability distribution P_x over some vector of random variables $x = (x_1, \ldots, x_n)$. We denote a random choice from P_x as $x \leftarrow \mathsf{P}_x$. We try to explicitly indicate randomness whenever taking probability or expectation, i.e., we write $\Pr_{x \leftarrow \mathsf{P}_x}[\ldots]$ and so on. For a finite set A, let $a \leftarrow A$ be a shorthand for a uniform random choice of an element from A.

For a natural number n, let $[n] := \{1, \ldots, n\}$. As usual, by $\binom{n}{k}$ we denote $\frac{\prod_{i=0}^{k-1}(n-i)}{k!}$ for $n \in \mathbb{R}$ and $k \in \mathbb{N}$. For $n \in \mathbb{N}$ and $0 \leq k \leq n$, we also identify $\binom{n}{k}$ with the set of subsets of $[n]$ of size k.

In particular, $(i_1, \ldots, i_m) \leftarrow [n]^m$ denotes uniform choice of m elements from $[n]$ with repetition and $M \leftarrow \binom{n}{m}$ uniform choice of a subset of $[n]$ of size m.

[1] Of course the bound itself is not new. Impagliazzo and Kabanets asked if such a concentration bound can be obtained from the hitting property, i.e., using the technique from [16].

2 A Simple Proof of a Chernoff Bound

We start by presenting a short proof of a Chernoff bound in, arguably, the most basic setting.

▶ **Theorem 2.1.** *Let $x = (x_1, \dots, x_n)$ be i.i.d. over $\{0,1\}^n$ with $\Pr[x_i = 1] = \frac{1}{2}$ and $\epsilon \in [0, \frac{1}{2}]$. Then,*

$$\Pr_{x \leftarrow \mathsf{P}_x}\left[\sum_{i=1}^{n} x_i \geq \frac{n}{2}(1+\epsilon)\right] \leq \exp\left(-\frac{\epsilon^2 n}{6}\right) .$$

Proof. Let $m := \lceil \frac{\epsilon n}{3} \rceil$. We have

$$\begin{aligned}
\mathop{\mathrm{E}}_{x \leftarrow \mathsf{P}_x}\left[\left(\sum_{i=1}^{n} x_i\right)^m\right] &= n^m \Pr_{\substack{x \leftarrow \mathsf{P}_x \\ (i_1,\dots,i_m) \leftarrow [n]^m}}\left[\forall j \in [m] : x_{i_j} = 1\right] \\
&= n^m \prod_{j=1}^{m} \Pr_{\substack{x \leftarrow \mathsf{P}_x \\ (i_1,\dots,i_m) \leftarrow [n]^m}}\left[x_{i_j} = 1 \mid \forall k < j : x_{i_k} = 1\right] \\
&\leq n^m \left(\frac{\epsilon}{3} \cdot 1 + \left(1 - \frac{\epsilon}{3}\right) \cdot \frac{1}{2}\right)^m = \left(\frac{n}{2}\right)^m \left(1 + \frac{\epsilon}{3}\right)^m .
\end{aligned}$$

Using Markov's inequality and $\frac{1+\epsilon/3}{1+\epsilon} \leq \exp\left(-\frac{\epsilon}{2}\right)$ for $\epsilon \in [0, \frac{1}{2}]$,

$$\Pr\left[\left(\sum_{i=1}^{n} x_i\right)^m \geq \left(\frac{n}{2}\right)^m (1+\epsilon)^m\right] \leq \left(\frac{1+\frac{\epsilon}{3}}{1+\epsilon}\right)^m \leq \exp\left(-\frac{\epsilon^2 n}{6}\right) .$$

◀

The above is the simplest proof of the most basic Chernoff bound we know of, and we believe that it is worthwhile to state it explicitly. It can be obtained by adapting the proof given in [16] for the given setting, although a direct adaptation yields a slightly different (and probably a bit longer) argument. Alternatively, it can be seen as an instantiation of the proof given in [17] in case one is interested in counting the number of copies of K_2 (i.e., the number of edges) in a random graph $\mathsf{G}_{n,p}$, after rather many simplifications that can be done for this very special case. Finally, it is a straightforward instantiation of our later proof given in Section 3.

3 Growth Boundedness

In this section we present the definition of growth-boundedness and prove that it implies concentration. In Section 3.1 we introduce growth boundedness without repetition: a variation of our concept that we use to prove the expander random walk bound.

▶ **Definition 3.1.** Let $\delta \geq 0$ and $m \in [n]$. A distribution P_x over $x = (x_1, \dots, x_n) \in \mathbb{R}^n_{\geq 0}$ with $\mu := \mathrm{E}_{\substack{x \leftarrow \mathsf{P}_x \\ i \leftarrow [n]}}[x_i]$ is (δ, m)-*growth bounded* if

$$\mathop{\mathrm{E}}_{x \leftarrow \mathsf{P}_x}\left[\left(\sum_{i=1}^{n} x_i\right)^m\right] \leq (\mu n)^m (1+\delta)^m .$$

Equivalently, P_x is (δ, m)-growth bounded if and only if

$$\mathop{\mathrm{E}}_{\substack{x \leftarrow \mathsf{P}_x \\ (i_1,\dots,i_m) \leftarrow [n]^m}}\left[\prod_{j=1}^{m} x_{i_j}\right] \leq \mu^m (1+\delta)^m .$$

If random variables are over $\{0,1\}$, this condition reduces to

$$\Pr_{\substack{x\leftarrow \mathsf{P}_x \\ (i_1,\ldots,i_m)\leftarrow[n]^m}}\Big[\forall j\in[m] : x_{i_j}=1\Big] \leq \mu^m(1+\delta)^m \ .$$

We now state our main theorem:

▶ **Theorem 3.2.** *Let P_x be a distribution over $\mathbb{R}^n_{\geq 0}$, $\mu := \mathrm{E}_{\substack{x\leftarrow \mathsf{P}_x \\ i\leftarrow[n]}}[x_i]$, $\mu>0$, $\epsilon\geq 0$. If P_x is (δ,m)-growth bounded, then*

$$\Pr_{x\leftarrow \mathsf{P}_x}\Big[\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)\Big] \leq \Big(\frac{1+\delta}{1+\epsilon}\Big)^m \ .$$

Proof. By Markov's inequality and growth boundedness of P_x,

$$\begin{aligned}\Pr_{x\leftarrow \mathsf{P}_x}\Big[\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)\Big] &= \Pr_{x\leftarrow \mathsf{P}_x}\Big[\big(\sum_{i=1}^n x_i\big)^m \geq (\mu n)^m(1+\epsilon)^m\Big] \\ &\leq \Big(\frac{1+\delta}{1+\epsilon}\Big)^m \ .\end{aligned}$$

◀

There is an interesting connection between this proof (inspired by [17]) and the one used in [16], for details see Section 3.2.

We obtain more convenient bounds as a corollary:

▶ **Corollary 3.3.** *Let $\epsilon\geq 0$ and P_x be an $(\frac{\epsilon}{3},m)$-growth bounded distribution over $\mathbb{R}^n_{\geq 0}$ with $\mu := \mathrm{E}_{\substack{x\leftarrow \mathsf{P}_x \\ i\leftarrow[n]}}[x_i]$, $\mu>0$.*

1. *If $\epsilon\leq\frac{1}{2}$: $\Pr_{x\leftarrow \mathsf{P}_x}\Big[\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)\Big] \leq \exp\Big(-\frac{\epsilon m}{2}\Big)$.*
2. *If $\epsilon\geq\frac{1}{2}$: $\Pr_{x\leftarrow \mathsf{P}_x}\Big[\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)\Big] \leq \Big(\frac{4}{5}\Big)^m$.*
3. *If $\epsilon\geq 3$: $\Pr_{x\leftarrow \mathsf{P}_x}\Big[\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)\Big] \leq 2^{-m}$.*

Proof. (1) follows because $\frac{1+\epsilon/3}{1+\epsilon} \leq \exp\big(-\frac{\epsilon}{2}\big)$ for $\epsilon\in[0,\frac{1}{2}]$, (2) since $\frac{1+\epsilon/3}{1+\epsilon}\leq\frac{4}{5}$ for $\epsilon\geq\frac{1}{2}$ and (3) due to $\frac{1+\epsilon/3}{1+\epsilon}\leq\frac{1}{2}$ for $\epsilon\geq 3$. ◀

For example, suppose that $x_1,\ldots,x_n$ are independent over $\{0,1\}^n$, $\Pr[x_i=1]=\mu>0$, and $\epsilon\in[0,\frac{1}{2}]$.

Using that for each M with $|M|\leq\frac{\epsilon\mu n}{3}$ we have

$$\Pr_{\substack{x\leftarrow \mathsf{P}_x \\ i\leftarrow[n]}}[x_i=1 \mid \forall j\in M : x_j=1] = \left(\frac{|M|}{n}+\left(1-\frac{|M|}{n}\right)\mu\right) \leq \frac{|M|}{n}+\mu \leq \mu\left(1+\frac{\epsilon}{3}\right) \ ,$$

we can conclude that P_x is $(\frac{\epsilon}{3},\lceil\frac{\epsilon\mu n}{3}\rceil)$-growth bounded and

$$\Pr_{x\leftarrow \mathsf{P}_x}\Big[\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)\Big] \leq \exp(-\epsilon^2\mu n/6) \ .$$

3.1 Growth boundedness without repetition

If one looks at the process in the growth boundedness definition as choosing a uniform m-tuple of indices $(i_1, \ldots, i_m)$ (with repetition), it is possible to make a similar argument for choosing a uniform set of indices of size m instead. In particular, we find it convenient in the proof of the expander random walk bound.

▶ **Definition 3.4.** Let $\delta \geq -1$ and $m \in [n]$. We say that a distribution P_x over $\{0,1\}^n$ with $\mu := \Pr_{\substack{x \leftarrow \mathsf{P}_x \\ i \leftarrow [n]}}[x_i = 1]$ is (δ, m)-*growth bounded without repetition* if

$$\Pr_{\substack{x \leftarrow \mathsf{P}_x \\ M \leftarrow \binom{n}{m}}} \Big[\forall i \in M : x_i = 1\Big] \leq \mu^m (1+\delta)^m \,.$$

▶ **Theorem 3.5.** *Let P_x be a distribution over $\{0,1\}^n$, $\mu := \Pr_{\substack{x \leftarrow \mathsf{P}_x \\ i \leftarrow [n]}}[x_i = 1]$, $\mu > 0$, $\epsilon \geq 0$, $c \in [0,1]$. If P_x is $(\delta, c\epsilon\mu n)$-growth bounded without repetition then*

$$\Pr_{x \leftarrow \mathsf{P}_x} \Big[\sum_{i=1}^{n} x_i \geq \mu n(1+\epsilon)\Big] \leq \Big(\frac{1+\delta}{1+(1-c)\epsilon}\Big)^m \,,$$

where $m := c\epsilon\mu n$.

Proof. Set $q := \Pr[\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)]$ and compute:

$$\begin{aligned}
\mu^m(1+\delta)^m &\geq \Pr_{\substack{x \leftarrow \mathsf{P}_x \\ M \leftarrow \binom{n}{m}}} [\forall i \in M : x_i = 1] \\
&\geq q \Pr_{\substack{x \leftarrow \mathsf{P}_x \\ M \leftarrow \binom{n}{m}}} [\forall i \in M : x_i = 1 \mid \textstyle\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)] \\
&\geq q \prod_{i=0}^{m-1} \frac{\mu n(1+\epsilon) - i}{n-i} \\
&\geq q\mu^m(1+(1-c)\epsilon)^m \,.
\end{aligned}$$

◀

▶ **Corollary 3.6.** *Let $\epsilon \in [0, \frac{4}{5}]$ and P_x be a distribution over $\{0,1\}^n$ that is $(\frac{\epsilon}{3}, m)$-growth bounded without repetition for some $m \leq \frac{\epsilon\mu n}{6}$ with $\mu := \Pr_{\substack{x \leftarrow \mathsf{P}_x \\ i \leftarrow [n]}}[x_i = 1]$, $\mu > 0$. Then,*

$$\Pr_{x \leftarrow \mathsf{P}_x} \Big[\sum_{i=1}^{n} x_i \geq \mu n(1+\epsilon)\Big] \leq \exp\Big(-\frac{\epsilon m}{3}\Big) \,.$$

Proof. Apply Theorem 3.5 and note that $\frac{1+\epsilon/3}{1+5\epsilon/6} \leq \exp\big(-\frac{\epsilon}{3}\big)$ for $\epsilon \in [0, \frac{4}{5}]$. ◀

3.2 Connection of [16] and [17]

Recall the proof of Theorem 3.2. In the context of [16] and [17] we find it instructive to give an alternative proof, restricted to distributions over $\{0,1\}^n$ (essentially the same as the proof of Theorem 3.5).

▶ **Theorem 3.7.** *Let P_x be a distribution over $\{0,1\}^n$, $\mu := \Pr_{\substack{x \leftarrow \mathsf{P}_x \\ i \leftarrow [n]}}[x_i = 1]$, $\mu > 0$, $\epsilon \geq 0$. If P_x is (δ, m)-growth bounded, then*

$$\Pr_{x \leftarrow \mathsf{P}_x} \Big[\sum_{i=1}^{n} x_i \geq \mu n(1+\epsilon)\Big] \leq \Big(\frac{1+\delta}{1+\epsilon}\Big)^m \,.$$

Proof. Set $q := \Pr\left[\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)\right]$, and see that[2]

$$\begin{aligned}\mu^m(1+\delta)^m &\geq \Pr_{\substack{x\leftarrow \mathsf{P}_x\\(i_1,\ldots,i_m)\leftarrow[n]^m}}[\forall j\in[m]: x_{i_j}=1]\\ &\geq q \Pr_{\substack{x\leftarrow \mathsf{P}_x\\(i_1,\ldots,i_m)\leftarrow[n]^m}}[\forall j\in[m]: x_{i_j}=1 \mid \textstyle\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)]\\ &\geq q\,\mu^m(1+\epsilon)^m\,.\end{aligned}$$

◀

The basic idea of the proof in [16] is to consider $\Pr_{x,M}[\forall i\in M: x_i=1]$, where M is a subset of $[n]$ obtained by including each element in M independently with some probability q. Then, this is compared with $\Pr_{x,M}[\forall i\in M: x_i=1 \mid \mathcal{E}]$, where $\mathcal{E}$ is the event that $\sum_{i=1}^n x_i \geq \mu n(1+\epsilon)$. In fact, we have

$$\Pr_x[\mathcal{E}] \leq \frac{\Pr_{x,M}[\forall i\in M: x_i=1]}{\Pr_{x,M}[\forall i\in M: x_i=1 \mid \mathcal{E}]}\,.$$

It is possible to show that for $m := \mathrm{E}[|M|] \ll n$ we have $\Pr_M[\forall i\in M: x_i=1 \mid \mathcal{E}] \gtrsim \mu^m(1+\epsilon)^m$. To see the intuition of this, simply note that this probability roughly equals the probability of only selecting red balls when one chooses with repetition m times out of n balls, at least $\mu n(1+\epsilon)$ of which are red.[3] Thus,

$$\Pr_x[\mathcal{E}] \lesssim \frac{\Pr_{x,M}[\forall i\in M: x_i=1]}{\mu^m(1+\epsilon)^m}\,. \tag{1}$$

Now note that this last argument only uses the probability over M, and so is independent of the distribution of x. Thus, for any distribution on which we can give a good upper bound on $\Pr_{x,M}[\forall i\in M: x_i=1]$, the technique of [16] gives a concentration result.

The argument we use is very similar, but we pick M as an m-tuple whose elements are picked independently with repetition. However, then we also have

$$n^m \Pr_{x,M}[\forall i\in M: x_i=1] = \mathop{\mathrm{E}}_{x,M}[(x_1+\ldots+x_n)^m]\,.$$

By Markov's inequality,

$$\Pr[\mathcal{E}] = \Pr\left[(x_1+\cdots+x_n)^m \geq (\mu n(1+\epsilon))^m\right] \leq \frac{\Pr_{x,M}[\forall i\in M: x_i=1]}{\mu^m(1+\epsilon)^m}\,,$$

which is almost the same as (1).

The view in (1) is the one adopted by [16]. Bounding the m-th moment and using Markov is the view adopted in [17]. The above argument shows that these views are closely related, and one can argue that the connection is given by growth boundedness.

4 Random Walks on Expanders

Overview and our results

For an introduction to expander graphs, see [15] or [30, Chapter 4]. In short, a λ-expander is a d-regular undirected graph G with the second largest (in terms of absolute value) eigenvalue of the transition matrix at most λ.

[2] Clearly $q=0$ is not a problem.

[3] The difference to the actual random experiment is that we do not keep each ball with probability m/n but instead choose exactly m times.

We consider a random walk on λ-expander starting in a uniform random vertex. It is a very useful fact in many applications that such a random walk behaves in certain respects very similarly to a random walk on the complete graph.

In particular, the so called hitting property [2, 20] states that the probability that an ℓ-step random walk on a λ-expander G stays completely inside a set $W \subseteq V := V(G)$ with $\mu := |W|/|V|$ is at most $(\mu+\lambda)^\ell$. A more general version [3] states that for each $M \subseteq [\ell]$ the probability that a random walk stays inside W in all steps from M is at most $(\mu+2\lambda)^{|M|}$.

Our first result, which may be of independent interest, can be considered as a randomized version of the hitting property. Namely, we show that, given $\epsilon > 0$, for a relatively small *random* subset $M \subseteq [\ell]$ of size m the probability that a random walk on a λ-expander stays inside W in all steps from M is at most $(\mu(1+\epsilon))^m$:

▶ **Theorem 4.1.** *Let G be a λ-expander with a distribution P_r over V^ℓ representing an $(\ell-1)$-step random walk $r = (v_1, \ldots, v_\ell)$ (with v_1 being a uniform starting vertex) and $W \subseteq V$ with $\mu := |W|/|V|$. Let $\epsilon \geq 0$ and $m \leq \min\big(1, \frac{1-\lambda}{\lambda}\frac{\epsilon\mu}{2}\big)\ell$. Then,*

$$\Pr_{\substack{r\leftarrow\mathsf{P}_r \\ M\leftarrow\binom{\ell}{m}}}\Big[\forall i \in M : v_i \in W\Big] \leq (\mu(1+\epsilon))^m \,.$$

Another important property of random walks on expander graphs is the Chernoff bound estimating the probability that the number of times a random walk visits W is far from its expectation. The first Chernoff bound for expander random walks was given by Gillman [11] and the problem was treated further in numerous works [21, 24, 1, 12, 32, 8].

Impagliazzo and Kabanets [16] apply their technique to obtain a bound for random walks on expander graphs, but in case of deviations smaller than λ they lose a factor of $\log\big(\frac{1}{\epsilon}\big)$ in the exponent. They then ask if their technique can be modified to avoid this loss.

We answer this question affirmatively: using Theorem 4.1 we immediately obtain a bound that matches the known ones and does not suffer from the additional $\log\big(\frac{1}{\epsilon}\big)$ factor while preserving the simplicity of the proof.

▶ **Theorem 4.2.** *Let the setting be as in Theorem 4.1 with $\mu > 0$. Define P_x over $\{0,1\}^\ell$ as $x_i = 1 \iff v_i \in W$ and let $\epsilon \in [0, \frac{4}{5}]$. Then,*

$$\Pr_{r\leftarrow\mathsf{P}_r}\Big[\sum_{i=1}^{\ell} x_i \geq \mu\ell(1+\epsilon)\Big] \leq 2\exp\Big(-\frac{(1-\lambda)\epsilon^2\mu\ell}{18}\Big)\,.$$

Furthermore, we demonstrate robustness of our method by improving the exponent to $\frac{1-\lambda}{1+\lambda}\frac{\mu}{1-\mu}\frac{\epsilon^2\ell}{2} + o(\epsilon^2)\ell$, which is optimal for fixed λ, μ and $\epsilon \to 0_+$ and $\ell \to \infty$:

▶ **Theorem 4.3.** *Let the setting be as in Theorem 4.1 with $\mu \in (0,1)$. Define P_x over $\{0,1\}^\ell$ as $x_i = 1 \iff v_i \in W$ and let $\epsilon \in [0,1]$. Then, there exists c_μ that depends only on μ such that*

$$\Pr_{r\leftarrow\mathsf{P}_r}\Big[\sum_{i=1}^{\ell} x_i \geq \mu\ell(1+\epsilon)\Big] \leq 2\exp\Big(-\frac{1-\lambda}{1+\lambda}\cdot\frac{\mu}{1-\mu}\cdot\frac{\epsilon^2\ell}{2} + c_\mu\epsilon^3\ln(\frac{1}{\epsilon})\ell\Big)\,.$$

For a proof of Theorem 4.3 see the full paper. In the following we prove Theorems 4.1 and 4.2.

Proofs

First, we need a coupling argument: let $m, \ell \in \mathbb{N}, m \leq \ell$ be given. We consider the distribution $\mathsf{D}_{m,\ell}$ defined by the following process:

- Pick uniformly $M \leftarrow \binom{\ell}{m}$ and let $M := \{x_1, \ldots, x_m\}$ with $x_1 < \ldots < x_m$.
- Let $d_1 := x_1$ and $d_i := x_i - x_{i-1}$ for $i > 1$.

A bijection shows that $d = (d_1, \ldots, d_m)$ is distributed uniformly on the $\binom{\ell}{m}$ m-tuples which satisfy $\sum_{i=1}^{m} d_i \leq \ell$ and $d_i > 0$. We now couple $\mathsf{D}_{m,\ell}$ with independent random variables (see full paper for the proof):

▶ **Lemma 4.4.** *Let $0 < m \leq \ell$. There exists a distribution over $(d_1, \ldots, d_m, e_1, \ldots, e_m)$ such that:*

- *$e_i \leq d_i$ for $1 \leq i \leq m$.*
- *$(d_1, \ldots, d_m)$ is distributed according to $\mathsf{D}_{m,\ell}$.*
- *$(e_1, \ldots, e_m)$ are i.i.d. with e_i in $\mathbb{N}_+$ and $\Pr[e_i = k] \leq \frac{2m}{\ell}$ for every k.*

Proof of Theorem 4.1. Pick $M \leftarrow \binom{\ell}{m}$ and let $(d_1, \ldots, d_m)$ be as in the definition of $\mathsf{D}_{m,\ell}$.

▶ **Lemma 4.5.**

$$\Pr_{\substack{r \leftarrow \mathsf{P}_r \\ M \leftarrow \binom{\ell}{m}}} \left[\forall i \in M : v_i \in W\right] \leq \mathop{\mathrm{E}}_{M \leftarrow \binom{\ell}{m}} \left[\prod_{i=1}^{m} (\mu + \lambda^{d_i})\right] .$$

Proof. Let $v := (\frac{1}{n}, \ldots, \frac{1}{n})$ be the vector of the uniform distribution on V and let P_W be a diagonal $n \times n$ matrix with $(P_W)_{uu} = 1$ if $u \in W$ and $(P_W)_{uu} = 0$ otherwise. Note that $P_W^2 = P_W$.

Let A_G be the probability transition matrix of G. Let us denote the spectral norm of a matrix with $||\cdot||$. We bound the probability of a random walk staying in W on indices of M using a standard technique. In particular, we use (for the proof see [30, Claim 4.21]):

▶ **Claim 4.6.**

$$||P_W A_G^k P_W|| \leq \mu + (1 - \mu)\lambda^k .$$

Fix M. First of all, by induction (and noting that $vA_G = v$):

$$\Pr_{r \leftarrow \mathsf{P}_r} \left[\forall i \in M : v_i \in W\right] = \left|vP_W \prod_{i=2}^{m} A_G^{d_i} P_W\right|_1 .$$

Estimate:

$$\begin{aligned}
\left|vP_W \prod_{i=2}^{m} A_G^{d_i} P_W\right|_1 &\leq \sqrt{\mu n} \cdot \left|\left|vP_W \prod_{i=2}^{m} A_G^{d_i} P_W\right|\right| && (2)\\
&\leq \sqrt{\mu n} \cdot \left||vP_W|\right| \prod_{i=2}^{m} \left||P_W A_G^{d_i} P_W|\right| && (3)\\
&= \mu \prod_{i=2}^{m} \left||P_W A_G^{d_i} P_W|\right| && (4)\\
&\leq \prod_{i=1}^{m} (\mu + \lambda^{d_i}) , && (5)
\end{aligned}$$

where (2) is due to Cauchy-Schwarz inequality (note there are at most μn non-zero coordinates in the final vector), (3) follows from $||AB|| \leq ||A|| \cdot ||B||$, (4) from $||vP_W|| = \sqrt{\frac{\mu}{n}}$ and (5) from Claim 4.6. ◀

The hope is that $(d_1, \ldots, d_m)$ behave "almost" like i.i.d. uniform random variables. This is indeed true, and by Corollary 4.4 we have $(e_1, \ldots, e_m)$ such that $e_i \leq d_i$ and e_i are i.i.d. with e_i in $\mathbb{N}_+$ and $\Pr[e_i = k] \leq \frac{2m}{\ell}$ for each k.

Putting this fact together with Lemma 4.5:

$$\begin{aligned}\Pr_{\substack{r\leftarrow \mathsf{P}_r \\ M\leftarrow\binom{\ell}{m}}}\left[\forall i \in M : v_i \in W\right] &\leq \mathrm{E}\left[\prod_{i=1}^{m}\left(\mu + \lambda^{e_i}\right)\right] \\ &= \prod_{i=1}^{m}\left(\mu + \mathrm{E}[\lambda^{e_i}]\right) \\ &\leq \left(\mu + \frac{2m}{\ell}\cdot\frac{\lambda}{1-\lambda}\right)^m \leq \mu^m(1+\epsilon)^m \,.\end{aligned}$$

◀

An immediate corollary of Theorem 4.1 is:

▶ **Corollary 4.7.** *Let the setting be as in Theorem 4.1. Define* P_x *over* $\{0,1\}^\ell$ *as* $x_i = 1 \iff v_i \in W$. *Then,* P_x *is* $(\epsilon, \min\left(\ell, \lfloor\frac{1-\lambda}{\lambda}\frac{\epsilon\mu\ell}{2}\rfloor\right))$*-growth bounded without repetition.*

Proof of Theorem 4.2. : Combine Corollary 4.7 with Corollary 3.6 (setting $m := \lfloor\frac{(1-\lambda)\epsilon\mu\ell}{6}\rfloor$). ◀

5 Polynomial Concentration

In certain applications it is desired to bound the concentration not only of the sum, but rather of a (low-degree) polynomial of some random variables.

In the case when (informally) the polynomial is such that the change in its value is bounded when the value of a single input variable is changed the Azuma's inequality can be applied to bound concentration.

If this is not so, one can use techniques that were invented by Kim and Vu [22] and developed in a body of work that followed (in particular [31, 29]). In the special case of a multilinear low-degree polynomial $p(v)$ and an independent distribution of input variables P_v their concentration bound can be expressed, very roughly speaking, as a function of $\frac{\mu_0}{\mu'}$, where μ_0 is the expectation of $p(v)$ and $\mu' = \max_{K\neq\emptyset}\mathrm{E}[\partial_K p(v)]$.

We obtain a bound in similar spirit. It is not tight in general, but can be applied to arbitrary polynomials with positive coefficients over input random variables in $[0,1]$ and is tight in the case of *elementary symmetric polynomials* $e_k(v) := \sum_{|S|=k}\prod_{i\in S} v_i$ (see the full paper for a proof).

Most importantly, as opposed to prior results, it does not require the input variables to be independent, but rather *almost independent* in a certain sense (for simplicity we limit ourselves to multilinear polynomials and inputs in {0,1}, full treatment can be found in the full paper):

▶ **Definition 5.1.** Let P_v be a distribution over $\{0,1\}^\ell$, $\delta \geq 0$ and $m \in [\ell]$. P_v is (δ, m)-almost independent if for each $M \subseteq [\ell]$ with $|M| \leq m$

$$\Pr_{v\leftarrow\mathsf{P}_v}\left[\forall i \in M : v_i = 1\right] \leq (1+\delta)^m \prod_{i\in M}\Pr_{v\leftarrow\mathsf{P}_v}\left[v_i = 1\right] \,.$$

Let us state our main theorem of this section.

Let P_v be a (δ, km)-almost independent distribution. Let $p(v)$ be a multilinear polynomial of degree k with positive coefficients. Our way to deal with dependencies in P_v is to state the bound in terms of P_v^* which is the distribution of independent variables with the same marginals as P_v (i.e., each v_i^* has the same distribution as v_i).

We express the concentration in terms of

$$\mu_i^* := \max_{\substack{K\subseteq[\ell]\\|K|=i}} \mathop{\mathrm{E}}_{v\leftarrow\mathsf{P}_v^*}[\partial_K p(v)] .$$

Note that μ_0^* is the expectation of $p(v)$ under P_v^*.

▶ **Theorem 5.2.** *Let the setting be as above and $\epsilon > 0$. Then,*

$$\Pr_{v\leftarrow\mathsf{P}_v}\left[p(v) \geq \mu_0^*(1+\epsilon)\right] \leq \left(\frac{(1+\delta)^k(1+\frac{\sum_{i=1}^k \binom{km}{i}\mu_i^*}{\mu_0^*})}{1+\epsilon}\right)^m .$$

Proof outline. Write $p(v)$ as a sum of binary random variables (corresponding to the monomials) $x_1, \ldots, x_n$. Due to Theorem 3.2 it is enough to show that $(x_1, \ldots, x_n)$ are (δ', m)-growth bounded, where $1+\delta' = (1+\delta)^k\big(1+\frac{\sum_{i=1}^k \binom{km}{i}\mu_i^*}{\mu_0^*}\big)\frac{\mu_0^*}{\mu}$.

Since P_v is (δ, km)-almost independent, this task can be further reduced to showing that if v is distributed according to P_v^* instead of P_v, then $(x_1, \ldots, x_n)$ are (δ'', m)-growth bounded, where $1+\delta'' = \big(1+\frac{\sum_{i=1}^k \binom{km}{i}\mu_i^*}{\mu_0^*}\big)$.

Fix $s < m$ and $(i_1, \ldots, i_s) \in [n]^s$ and let M be the set of all indices j such that v_j influences at least one of $x_{i_1}, \ldots, x_{i_s}$ (note that $|M| \leq km$).

We write $p(v) = \sum_{K\subseteq M:|K|\leq k} p_K(v)$, where $p_K(v)$ consists of those monomials whose variables intersected with M are exactly K. Observe that

$$\mathop{\mathrm{E}}_{v\leftarrow\mathsf{P}_v^*}\left[p_K(v) \mid \forall i \in M : v_i = 1\right] \leq \mathop{\mathrm{E}}_{v\leftarrow\mathsf{P}_v^*}\left[\partial_K p(v)\right] .$$

To get growth boundedness for $x_1, \ldots, x_n$ we proceed by induction and bound

$$\begin{aligned}\Pr_{\substack{v\leftarrow\mathsf{P}_v^*\\ i_{s+1}\leftarrow[n]}}\left[x_{i_{s+1}} = 1 \mid \forall j \in [s] : x_{i_j} = 1\right] &= \frac{1}{n}\mathop{\mathrm{E}}_{v\leftarrow\mathsf{P}_v^*}\left[p(v) \mid \forall i \in M : v_i = 1\right]\\ &\leq \frac{1}{n}\sum_{K\subseteq M:|K|\leq k}\mathop{\mathrm{E}}_{v\leftarrow\mathsf{P}_v^*}\left[\partial_K p(v)\right]\\ &\leq \frac{\mu_0^*}{n}\left(1+\frac{\sum_{i=1}^k \binom{km}{i}\mu_i^*}{\mu_0^*}\right) .\end{aligned}$$

◀

Let $\mu' := \max_{i\in[k]} \mu_i^*$. Since $\sum_{i=1}^k \binom{km}{i} \leq (km)^k$, we have:

▶ **Corollary 5.3.** *Let the setting be as in Theorem 5.2. Then,*

$$\Pr_{v\leftarrow\mathsf{P}_v}\left[p(v) \geq \mu_0^*(1+\epsilon)\right] \leq \left(\frac{(1+\delta)^k(1+\frac{(km)^k\mu'}{\mu_0^*})}{1+\epsilon}\right)^m .$$

5.1 An application in [14]

In [14] the authors prove a lower bound on the complexity of a black-box construction of a pseudorandom generator from a one-way function.

Part of their proof consists in using Theorem 5.2 to show a concentration bound for a certain polynomial. The proof of Theorem 5.2 is not included in [14], but deferred to this paper instead. Since the input variables of the polynomial are not independent, to the best of our knowledge no previous work is applicable to this case.[4]

The following random process is considered: pick a permutation $f : \{0,1\}^n \to \{0,1\}^n$ u.a.r. and consider the distribution P_g over 2^{2n} random variables $g := \{g_{x,y} : x, y \in \{0,1\}^n\}$ defined as $g_{x,y} = 1$ if $f(x) = y$ and $g_{x,y} = 0$ otherwise.

The random variables in g are not independent, but it is easy to check that they are $(1, 2^{n-1})$-almost independent. Also, the corresponding independent distribution P^*_g has expectation 2^{-n} for each $g_{x,y}$.

Fix $k \leq \frac{n}{100 \log n}$. [14] defines a certain multilinear polynomial $p(g)$ of degree at most k such that $\mu^*_0 \leq 2^{n/15}$ and $\mu' \leq 2^{n/15}$ (we omit the details).

[14] needs to show that (for n big enough):

$$\Pr_{g \leftarrow \mathsf{P}_g}\left[p(g) \geq 2^{n/10}\right] \leq 2^{-2^{n/100k}} .$$

To this end, calculate using Corollary 5.3 and setting $\delta := 1$, $\epsilon := 2^{9n/100}/\mu^*_0$ and $m := 2^{n/100k}$:

$$\begin{aligned}
\Pr_{g \leftarrow \mathsf{P}_g}\left[p(g) \geq \mu^*_0 + 2^{9n/100}\right] &\leq \left(\frac{2^k \max\left(2, \frac{2k^k 2^{n/100}\mu'}{\mu^*_0}\right)}{\frac{2^{9n/100}}{\mu^*_0}} \right)^{2^{n/100k}} \\
&\leq \left(\frac{2^{k+1} \max\left(\mu^*_0, k^k 2^{n/100}\mu'\right)}{2^{9n/100}} \right)^{2^{n/100k}} \\
&\leq 2^{-2^{n/100k}} .
\end{aligned}$$

5.2 Other applications

We note that despite the fact that the deviation for which we applied our theorem in Section 5.1 is big relative to the expectation, one can obtain meaningful bounds also for very small deviations.

This can be seen by taking a restricted version of Theorem 5.2:

▶ **Theorem 5.4.** *Let P_v be a distribution of independent variables (i.e., $\mathsf{P}_v = \mathsf{P}^*_v$) over $[0,1]^\ell$. Let $p(v)$ be as in Theorem 5.2 and $\epsilon \in [0, \frac{1}{2}]$. Then:*

$$\Pr_{v \leftarrow \mathsf{P}_v}\left[p(v) \geq \mu(1+\epsilon)\right] \leq 2\exp\left(-\frac{\epsilon}{6k}\left(\frac{\epsilon\mu}{\mu'}\right)^{1/k}\right) .$$

Proof. Note that P_v are $(0,\ell)$-almost independent. Take $m := \left\lfloor \frac{1}{k}\left(\frac{\epsilon\mu}{3\mu'}\right)^{1/k} \right\rfloor$, obtain $(\frac{\epsilon}{3}, m)$-growth boundedness as in Corollary 5.3 and apply Corollary 3.3.1. ◀

[4] It was pointed out to us that a generalisation of the result of Latała and Łochowski [23] might be applicable (together with [9]). However, moment bound in [23] is optimal only up to a constant in the exponent that depends on the degree and the degree is non-constant in our setting.

For example, in a representative setting when Azuma-like methods fail: consider the polynomial that counts the triangles in Erdős–Rényi random graph $\mathsf{G}_{n,n^{-3/4}}$, i.e., $p(v) = \sum_{\{a,b,c\}\in\binom{n}{3}} v_{ab}v_{ac}v_{bc}$. We compute $\mu = \Theta(n^{3/4})$ and $\mu' = \Theta(1)$.

For $\epsilon \in [0, \frac{3}{16}]$ Theorem 5.4 gives:

$$\Pr_{v\leftarrow\mathsf{P}_v}\left[p(v) \geq \mu(1+n^{-\epsilon})\right] \leq \exp(-\Omega(n^{1/4-4\epsilon/3}))\ .$$

This is comparable to the bound from [22] (which was the first paper to give a good bound in this setting). Better bounds are known, in particular we revisit the triangle counting in Section 6.

For some more discussion on the tightness of Theorem 5.2, see the full paper.

6 Counting Subgraphs in Random Graphs

In the proof of the polynomial concentration bound we consider values μ_i^* which are maxima of expectations of $\partial_K p(v)$ over sets K of size i. Each such value yields a contribution[5] of $\binom{km}{i}\mu_i^*$ (proportional to the number of partial derivatives of this type in the subset of input variables of size km) and the "quality" of a concentration bound depends, roughly, on the maximum such contribution.

In principle, nothing prevents us from considering a different, possibly finer, division of partial derivatives into a constant number of classes, each with its own contribution.

In particular, it is an obvious fact that the number of occurrences of a fixed subgraph H in a random Erdős–Rényi graph (for some of the work on the problem see [18, 17, 19]) can be expressed in terms of a multilinear polynomial. In this setting we may divide the partial derivatives into classes corresponding to subgraphs of H. Interestingly, this yields an upper tail bound proof that is basically isomorphic to the famous one of Janson, Oleszkiewicz and Ruciński [17].

Our result holds in the setting of almost-independent distributions, readily applicable, for example, to $\mathsf{G}_{n,m}$ random graphs (of course the proof of [17] also generalises to those settings).

For details, see the full paper.

References

1 Carlos A. León and François Perron. Optimal Hoeffding bounds for discrete reversible Markov chains. *The Annals of Applied Probability*, 14(2):958–970, 05 2004.

2 Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in LOGSPACE. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, pages 132–140, New York, NY, USA, 1987. ACM.

3 Noga Alon, Uriel Feige, Avi Wigderson, and David Zuckerman. Derandomized graph products. *Computational Complexity*, 5(1):60–75, 1995.

4 Sanjeev Arora and Boaz Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009.

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

6 Sergei N. Bernstein. On a modification of Chebyshev's inequality and of the error formula of Laplace. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math.*, 1, 1924.

[5] Think of a constant k and a family of polynomials with m going to infinity.

7 Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):pp. 493–507, 1952.

8 Kai-Min Chung, Henry Lam, Zhenming Liu, and Michael Mitzenmacher. Chernoff-Hoeffding bounds for Markov chains: Generalized and simplified. In Christoph Dürr and Thomas Wilke, editors, *STACS*, volume 14 of *LIPIcs*, pages 124–135. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2012.

9 Victor H. de la Peña and S. J. Montgomery-Smith. Bounds on the tail probability of U-statistics and quadratic forms. *Bulletin of the American Mathematical Society*, 31(2):223–227, 1994.

10 Devdatt Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

11 David Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, 1998.

12 Alexander Healy. Randomness-efficient sampling within NC^1. *Computational Complexity*, 17(1):3–37, 2008.

13 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):pp. 13–30, 1963.

14 Thomas Holenstein and Makrand Sinha. Constructing a pseudorandom generator requires an almost linear number of calls. In *FOCS*, pages 698–707. IEEE Computer Society, 2012.

15 Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the AMS*, 43(4):439–561, 2006.

16 Russell Impagliazzo and Valentine Kabanets. Constructive proofs of concentration bounds. In Maria J. Serna, Ronen Shaltiel, Klaus Jansen, and José D. P. Rolim, editors, *APPROX-RANDOM*, volume 6302 of *Lecture Notes in Computer Science*, pages 617–631. Springer, 2010.

17 Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. Upper tails for subgraph counts in random graphs. *Israel Journal of Mathematics*, 142(1):61–92, 2004.

18 Svante Janson and Andrzej Ruciński. The infamous upper tail. *Random Struct. Algorithms*, 20(3):317–342, 2002.

19 Svante Janson and Andrzej Ruciński. Upper tails for counting objects in randomly induced subhypergraphs and rooted random graphs. *Arkiv för matematik*, 49(1):79–96, 2011.

20 Nabil Kahalé. Eigenvalues and expansion of regular graphs. *J. ACM*, 42(5):1091–1106, September 1995.

21 Nabil Kahalé. Large deviation bounds for Markov chains. *Combinatorics, Probability & Computing*, 6(4):465–474, 1997.

22 Jeong Han Kim and Van H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000.

23 Rafał Latała and Rafał Łochowski. Moment and tail estimates for multidimensional chaoses generated by positive random variables with logarithmically concave tails. *Progr. Probab.*, 56:77–92, 2003.

24 Pascal Lezaud. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8(3):849–867, 1998.

25 Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, New York, NY, USA, 2005.

26 Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, New York, Melbourne, 1995. Réimpressions : 1997, 2000.

27 Anup Rao. Parallel repetition in projection games and a concentration bound. In *In Proc. 40th STOC*, pages 1–10. ACM, 2008.

28 Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff-Hoeffding bounds for applications with limited independence. *SIAM J. Discret. Math.*, 8(2):223–250, May 1995.

29 Warren Schudy and Maxim Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In Yuval Rabani, editor, *SODA*, pages 437–446. SIAM, 2012.

30 Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

31 V. H. Vu. Concentration of non-Lipschitz functions and applications. *Random Struct. Algorithms*, 20(3):262–316, May 2002.

32 Roy Wagner. Tail estimates for sums of variables sampled by a random walk. *Combinatorics, Probability and Computing*, 17:307–316, 3 2008.