# 4 Second moment method

Previously, we used $\mathbb{E}X \geq a$ to deduce $\mathbb{P}(X \geq a) > 0$. We also saw from Markov's inequality that for $X \geq 0$, if $\mathbb{E}X$ is very small, then $X$ is small with high probability.

Does $\mathbb{E}X$ being (very) large imply that $X$ is large with high probability?

No! $X$ could be almost always small but $\mathbb{E}X$ could still be large due to outliers (rare large values of $X$).

Often we want to show that some random variable is **concentrated** around its mean. This would then imply that outliers are unlikely.

We will see many methods in this course on proving concentrations of random variables. We begin with the simplest method. It is the easiest to execute, requires the least hypotheses, but only produces weak (though often useful) concentration bounds.

**Second moment method**: show that a random variable is concentrated near its mean by bounding its variance.

**Variance**: $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Notation convention: mean $\mu$, variance $\sigma^2$, standard deviation $\sigma$.

**Theorem 4.0.1** (Chebyshev's inequality)**.** Let $X$ be a random variable with mean $\mu$ and standard deviation $\sigma$. For any $\lambda > 0$

$$\mathbb{P}(|X - \mu| \geq \lambda\sigma) \leq \lambda^{-2}.$$

*Proof.* By Markov's inequality,

$$LHS = \mathbb{P}(|X - \mu|^2 \geq \lambda^2\sigma^2) \leq \frac{\mathbb{E}[(X-\mu)^2]}{\lambda^2\sigma^2} = \frac{1}{\lambda^2}. \qquad \square$$

*Remark* 4.0.2. Concentration bounds that show small probability of deviating from the mean are called **tail bounds** (also: upper tail bounds for bounding $\mathbb{P}(X \geq \mu + a)$ and lower tail bounds for bounding $\mathbb{P}(X \leq \mu - a)$). Chebyshev's inequality gives tail bounds with polynomial decay. Later on we will see tools that give much better decay (usually exponential) provided additional assumptions on the random variable (e.g., independence).

We can rewrite Chebyshev's inequality as

$$\mathbb{P}(|X - \mathbb{E}X| \geq \epsilon\mathbb{E}X) \leq \frac{\mathrm{Var}\, X}{\epsilon^2(\mathbb{E}X)^2}.$$

**Corollary 4.0.3.** If $\mathrm{Var}[X] = o(\mathbb{E}X)^2$ then $X \sim \mathbb{E}X$ whp.

*Remark* 4.0.4. We are invoking asymptotics here (so we are actually considering a sequence $X_n$ of random variables instead of a single one). The conclusion is equivalent to that for every $\epsilon > 0$, one has $|X - \mathbb{E}X| \le \epsilon \mathbb{E}X$ with probability $1 - o(1)$ as $n \to \infty$.

Variance can be calculated from pairwise covariances. Recall the **covariance**

$$\mathrm{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

So $\mathrm{Var}[X] = \mathrm{Cov}[X, X]$. Covariance is bilinear in $X$ and $Y$, i.e., for constants $a_1, \dots$ and $b_1, \dots,$ one has

$$\mathrm{Cov}\left[\sum_i a_i X_i, \sum_j b_j Y_j\right] = \sum_{i,j} a_i b_j \,\mathrm{Cov}[X_i, Y_j].$$

Thus, given $X = X_1 + \cdots + X_n$ (no assumptions on dependencies between the $X_i$'s), we have

$$\mathrm{Var}[X] = \mathrm{Cov}[X, X] = \sum_{i,j \in [n]} \mathrm{Cov}[X_i, X_j] = \sum_{i \in [n]} \mathrm{Var}[X_i] + 2 \sum_{i<j} \mathrm{Cov}[X_i, X_j]$$

We have $\mathrm{Cov}[X, Y] = 0$ if $X$ and $Y$ are independent. Thus in the sum we only need to consider dependent pairs $(i, j)$.

**Example 4.0.5** (Sum of independent Bernoulli)**.** Suppose $X = X_1 + \cdots + X_n$ with $X_i$ iid $X_i \sim \mathrm{Bernoulli}(p)$, i.e., $X = 1$ with prob $p$ and $X = 0$ with prob $1 - p$.

Then $\mu = np$ and $\sigma^2 = np(1 - p)$. If $np \gg 1$ then $\sigma \ll \mu$ and thus $X = \mu + o(\mu)$ whp.

Note that the above computation remains identical even if we only knew that the $X_i$'s are *pairwise uncorrelated* (much weaker than assuming full independence).

Here the "tail probability" (the bound hidden in "whp") decays polynomially in the deviation. Later on we will derive much sharper rates of decay (exponential) using more powerful tools such as the Chernoff bound when the r.v.'s are independent.

**Example 4.0.6** (The number of triangles in a random graph)**.** Let

$$X = \text{the number of triangles in the random graph } G(n, p).$$

For vertices $i, j, k \in [n]$, denote the edge indicator variables by $X_{ij} = 1_{ij \text{ is an edge}}$. Let the triangle indicator variables be $X_{ijk} = 1_{ijk \text{ is a triangle}} = X_{ij} X_{ik} X_{jk}$. Then

$$X = \sum_{i<j<k} X_{ijk} = \sum_{i<j<k} X_{ij} X_{ik} X_{jk}.$$

Its expectation is easy to compute, since $\mathbb{E}[X_{ij}X_{ik}X_{jk}] = \mathbb{E}[X_{ij}]\mathbb{E}[X_{ik}]\mathbb{E}[X_{jk}] = p^3$ by independence. So

$$\mathbb{E}X = \binom{n}{3}p^3$$

Now we compute $\operatorname{Var} X$. Unlike in the earlier example, the summands of $X$ are not all independent. Nonetheless, it is easy to compute the variance.

Given two triples $T_1, T_2$ of vertices

$$\begin{aligned}\operatorname{Cov}[X_{T_1}, X_{T_2}] = \mathbb{E}[X_{T_1}X_{T_2}] - \mathbb{E}[X_{T_1}]\mathbb{E}[X_{T_2}] &= p^{e(T_1\cup T_2)} - p^{e(T_1)+e(T_2)} \\ &= \begin{cases} 0 & \text{if } |T_1\cap T_2| \le 1 \\ p^5 - p^6 & \text{if } |T_1\cap T_2| = 2 \\ p^3 - p^6 & \text{if } T_1 = T_2 \end{cases}\end{aligned}$$

Thus

$$\operatorname{Var} X = \sum_{T_1,T_2} \operatorname{Cov}[X_{T_1}, X_{T_2}] = \binom{n}{3}(p^3 - p^6) + \binom{n}{2}n(n-1)(p^5 - p^6) \lesssim n^3p^3 + n^4p^5$$

When do we have $\sigma \ll \mu$? It is equivalent to satisfying both $n^{3/2}p^{3/2} \ll n^3p^3$ (which gives $p \gg 1/n$) and $n^2p^{5/2} \ll n^3p^3$ (which gives $p \gg n^{-2}$). So $\sigma \ll \mu$ if and only if $p \gg 1/n$, and as we saw earlier, in this case $X \sim \mathbb{E}X$ with high probability.

*Remark* 4.0.7. Later on we will use more powerful tools (including martingale methods/Azuma-Hoeffding inequalities, and also Janson inequalities) to prove better tail bounds on triangle (and other subgraph) counts.

*Remark* 4.0.8. Actually the number $X$ of triangles in $G(n,p)$ satisfies an asymptotic central limit theorem, i.e., $(X-\mu)/\sigma \to N(0,1)$ in distribution (Rucinski 1988), initially proved via moment of moments (by showing that higher moments of $(X-\mu)/\sigma$ match those of the normal distribution). Later a different proof was found using the "method of projections."

On the other hand, for much sparser random graphs, when $p \lesssim 1/n$, $X$ is asymptotically Poisson.

## 4.1 Threshold functions for small subgraphs in random graphs

**Question 4.1.1.** For which $p = p_n$ is $K_4 \subset G(n,p)$ true with high probability (i.e., with probability $1 - o(1)$)?

There are two statements that one wants to show:

- (0-statement) if $p = p_n$ is small, then $\mathbb{P}(K_4 \subset G(n,p)) \to 0$ as $n \to \infty$.
- (1-statement) if $p = p_n$ is large, then $\mathbb{P}(K_4 \subset G(n,p)) \to 1$ as $n \to \infty$.

Let $X$ be the number of copies of $K_4$ in $G(n,p)$.

- To show the 0-statement, it suffices to have $\mathbb{E}X \to 0$, in which case Markov's inequality implies that $\mathbb{P}(X \geq 1) \leq \mathbb{E}X \to 0$ (here we are only using the first moment method).
- To show the 1-statement, it suffices to show $\operatorname{Var} X = o((\mathbb{E}X)^2)$, by the lemma below (second moment method).

For simple applications, e.g., $K_4 \subset G(n,p)$, these two methods turn out to be sufficient. Other applications may require stronger techniques (though sometimes "only" second moment, but much more difficult applications).

**Lemma 4.1.2.** For any random variable $X$,

$$\mathbb{P}(X = 0) \leq \frac{\operatorname{Var} X}{(\mathbb{E}X)^2}$$

*Proof.* By Chebyshev inequality, writing $\mu = \mathbb{E}X$,

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mu| \geq |\mu|) \leq \frac{\operatorname{Var} X}{\mu^2}. \qquad \square$$

**Corollary 4.1.3.** If $\operatorname{Var} X = o((\mathbb{E}X)^2)$, then $X > 0$ with probability $1 - o(1)$.

*Remark* 4.1.4. Here is a slightly stronger inequality in the case of nonnegative random variables. It is a special case of the Paley–Zygmund inequality. I am showing it here because it is neat. It makes no difference for our applications whether we use the next lemma or the previous one.

**Lemma 4.1.5.** For any random variable $X \geq 0$,

$$\mathbb{P}(X > 0) \geq \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]}.$$

*Proof.* We have $\mathbb{P}(X > 0) = \mathbb{E}[1_{X>0}]$. By the Cauchy–Schwarz inequality

$$\mathbb{E}[1_{X>0}]\, \mathbb{E}[X^2] \geq (\mathbb{E}[1_{X>0} X])^2 = (\mathbb{E}X)^2. \qquad \square$$

**Definition 4.1.6** (Graph properties)**.** A **graph property** $\mathcal{P}$ is a subset of all graphs. We say that $\mathcal{P}$ is **monotone (increasing)** if whenever $G \in \mathcal{P}$, then any graph obtained by adding edges to $G$ also satisfies $\mathcal{P}$. We say that $\mathcal{P}$ is **non-trivial** if for all sufficiently large $n$, there exists an $n$-vertex graph in $\mathcal{P}$ and an $n$-vertex graph not in $\mathcal{P}$.

**Example 4.1.7.** Examples of graph properties

- Contains $K_4$; i.e., $\mathcal{P} = \{G : K_4 \subset G\}$
- Connected
- Hamiltonian
- 3-colorable (a monotone decreasing property)
- Planar (monotone decreasing)
- Contains a vertex of degree 1 (not monotone increasing or decreasing)

**Definition 4.1.8** (Threshold function)**.** We say that $r_n$ is a **threshold function** for some graph property $\mathcal{P}$ if

$$\mathbb{P}(G(n,p_n) \text{ satisfies } \mathcal{P}) \to \begin{cases} 0 & \text{if } p_n/r_n \to 0, \\ 1 & \text{if } p_n/r_n \to \infty. \end{cases}$$

*Remark* 4.1.9. The above definition is most suitable for monotone increasing properties. For other types of properties one may need to adjust the definition appropriately.

*Remark* 4.1.10. From the definition, we see that if $r_n$ and $r_n'$ are both threshold functions, then they must be within a constant factor of each other. So it is fine to say "*the* threshold" of some property, with the understanding that we do not care about constant factors. Later on we will see that every monotone property *has* a threshold function.

**Theorem 4.1.11.** A threshold function for containing a $K_3$ is $1/n$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(K_3 \subset G(n,p_n)) = \begin{cases} 0 & \text{if } p_n n \to 0 \\ 1 & \text{if } p_n n \to \infty \end{cases}$$

*Proof.* Let $X$ be the number of triangles in $G(n,p)$. Then $\mu := \mathbb{E}X = \binom{n}{3}p^3 \sim n^3p^3/6$. Let $\sigma^2 = \operatorname{Var} X$.

If $p \ll 1/n$, then $\mu = o(1)$, so $\mathbb{P}(X \geq 1) = o(1)$ by Markov, and hence $X = 0$ w.h.p.

If $p \gg 1/n$, then $\mu \to \infty$, and we saw earlier that $\sigma \ll \mu$, so whp $X \sim \mu$ and thus $X > 0$ whp. □

**Question 4.1.12.** What is the threshold for containing a fixed $H$ as a subgraph?

The next calculation is similar in spirit to what we did earlier for triangles, but we would like to be more organized as there may be more interacting terms in the variance calculation.

**General setup.** Suppose $X = X_1 + \cdots + X_m$ where $X_i$ is the indicator random variable for event $A_i$. Write $i \sim j$ if $i \neq j$ and the pair of events $(A_i, A_j)$ are not independent. (For variance calculation, we are only considering pairwise dependence. Warning: later on when we study the Lovász Local Lemma, we will need a strong notion of a dependency graph.)

If $i \neq j$ and $i \not\sim j$ then $\operatorname{Cov}[X_i, X_j] = 0$. Otherwise,

$$\operatorname{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \leq \mathbb{E}[X_i X_j] = \mathbb{P}[A_i \wedge A_j].$$

Thus

$$\operatorname{Var} X = \sum_{i,j} \operatorname{Cov}[X_i, X_j] \leq \mathbb{E}X + \Delta$$

where

$$\Delta = \sum_{(i,j):i\sim j} \mathbb{P}(A_i \wedge A_j)$$

The earlier second moment results (Corollary 4.0.3) imply that

$$\text{If } \mathbb{E}X \to \infty \text{ and } \Delta = o(\mathbb{E}X)^2 \text{ then and } X \sim \mathbb{E}X \text{ and } X > 0 \text{ whp.}$$

We have

$$\sum_{(i,j):i\sim j} \mathbb{P}(A_i \wedge A_j) = \sum_i \mathbb{P}(A_i) \sum_{j:j\sim i} \mathbb{P}\left(A_j \mid A_i\right)$$

In many symmetric situations (e.g. our examples), the following quantity does not depend on $i$:

$$\Delta^* = \sum_{j:j\sim i} \mathbb{P}\left(A_j \mid A_i\right)$$

(or take $\Delta^*$ to be the maximum such value ranging over all $i$). Then

$$\Delta = \sum_i \mathbb{P}[A_i]\Delta^* = \Delta^* \mathbb{E}X$$

Thus we have

**Lemma 4.1.13.** If $\mathbb{E}X \to \infty$ and $\Delta^* = o(\mathbb{E}X)$, then $X \sim \mathbb{E}X$ and $X > 0$ whp.

**Theorem 4.1.14.** A threshold function for containing $K_4$ is $n^{-2/3}$.

*Proof.* Let $X$ denote the number of copies of $K_4$ in $G(n,p)$. Then $\mathbb{E}X = \binom{n}{4}p^6 \sim n^4p^6/24$.

If $p \ll n^{-2/3}$ then $\mathbb{E}X = o(1)$ so $X = 0$ whp

Now suppose $p \gg n^{-2/3}$, so $\mathbb{E}X \to \infty$. For each 4-vertex subset $S$, let $A_S$ be the event that $S$ is a clique in $G(n,p)$.

For each fixed $S$, one has $A_S \sim A_{S'}$ if and only if $|S \cap S'| \geq 2$.

- The number of $S'$ that share exactly 2 vertices with $S$ is $6\binom{n}{2} = O(n^2)$, and for each such $S'$ one has $\mathbb{P}(A_{S'}|A_S) = p^5$ (as there are 5 additional edges, no in the $S$-clique, that needs to appear clique to form the $S'$-clique).
- The number of $S'$ that share exactly 3 vertices with $S$ is $4(n-4) = O(n)$, and for each such $S'$ one has $\mathbb{P}(A_{S'}|A_S) = p^3$.

Summing over all above $S'$, we find Then

$$\Delta^* = \sum_{S':|S'\cap S|\in\{2,3\}} \mathbb{P}(A_{S'}|A_S) \lesssim n^2p^5 + np^3 \ll n^4p^6 \asymp \mathbb{E}X.$$

Thus $X > 0$ whp by Lemma 4.1.13. □

For both $K_3$ and $K_4$, we saw that any choice of $p = p_n$ with $\mathbb{E}X \to \infty$ one has $X > 0$ whp. Is this generally true?

**Example 4.1.15** (First moment is not enough)**.** Let $H =$  . We have $\mathbb{E}X_H \asymp n^5p^7$. If $\mathbb{E}X = o(1)$ then $X = 0$ whp. But what if $\mathbb{E}X \to \infty$, i.e., $p \gg n^{-5/7}$?

We know that if $n^{-5/7} \ll p \ll n^{-2/3}$, then $X_{K_4} = 0$ whp, so $X_H = 0$ whp since $K_4 \subset H$.

On the other hand, if $p \gg n^{-2/3}$, then whp can find $K_4$, and pick an arbitrary edge to extend to $H$ (we'll prove this).

Thus the threshold for $H =$  is actually $n^{-2/3}$, and not $n^{-5/7}$ as one might have naively predicted from the first moment alone.

Why didn't $\mathbb{E}X_H \to \infty$ give $X_H > 0$ whp? In the calculation of $\Delta^*$, one of the terms is $\asymp np$ (from two copies of $H$ with a $K_4$-overlap), and $np \not\ll n^5p^7 \asymp \mathbb{E}X_H$ if $p \ll n^{-2/3}$.

**Definition 4.1.16.** Define the **edge-vertex ratio** of a graph $H$ by $\rho(H) = e_H/v_H$. Define the **maximum edge-vertex ratio of a subgraph** of $H$:

$$m(H) := \max_{H' \subset H} \rho(H').$$

**Example 4.1.17.** Let $H =$  . We have $\rho(H) = 7/5$ whereas $\rho(K_4) = 3/2 > 7/5$. It is not hard to check that $m(H) = \rho(K_4) = 3/2$ as $K_4$ is the subgraph of $H$ with the maximum edge-vertex ratio.

**Theorem 4.1.18** (Bollobás 1981)**.** Fix a graph $H$ with $v_H$ vertices and $e_H$ edges. Then $p = n^{-1/m(H)}$ is a threshold function for containing $H$ has a subgraph. Furthermore, if $p \gg n^{-1/m(H)}$, then the number $X_H$ of copies of $H$ in $G(n,p)$ satisfies, with probability $1 - o(1)$,

$$X_H \sim \mathbb{E}X_H = \binom{n}{v_H} \frac{v_H!}{\operatorname{aut}(H)} p^{e_H} \sim \frac{n^{v_H} p^{e_H}}{\operatorname{aut}(H)}.$$

*Proof.* Let $H'$ be a subgraph of $H$ achieving the maximum edge-vertex ratio, i.e., $\rho(H') = m(H)$.

If $p \ll n^{-1/m(H)}$, then $\mathbb{E}X_{H'} \asymp n^{v_{H'}} p^{e_{H'}} = o(1)$, so $X_{H'} = 0$ whp, hence $X_H = 0$ whp.

Now suppose $p \gg n^{-1/m(H)}$. Let us count *labeled* copies of the subgraph $H$ in $G(n,p)$. Let $J$ be a labeled copy of $H$ in $K_n$, and let $A_J$ denote the event that $J$ appears in $G(n,p)$. We have, for fixed $J$,

$$\Delta^* = \sum_{J' \sim J} \mathbb{P}\left(A_{J'} \mid A_J\right) = \sum_{J' \sim J} p^{|E(J') \setminus E(J)|}$$

For any $J' \sim J$, we have

$$n^{|V(J') \setminus V(J)|} p^{|E(J') \setminus E(J)|} \ll n^{|V(J)|} p^{|E(J)|}$$

since

$$p \gg n^{-1/m(H)} \geq n^{-1/\rho(J \cap J')} = n^{-|V(J) \cap V(J')|/|E(J) \cap E(J')|}.$$

It then follows, after consider all the possible ways that $J'$ can overlap with $J$, that $\Delta^* \ll n^{|V(J)|} p^{|E(J)|} \asymp \mathbb{E}X_H$. So Lemma 4.1.13 yields the result. □

## 4.2 Existence of thresholds

**Question 4.2.1.** Does every monotone graph property $\mathcal{P}$ have a threshold function?

E.g., could it be the case that $\mathbb{P}(G(n, n^{-1/3}) \in \mathcal{P}), \mathbb{P}(G(n, n^{-1/4}) \in \mathcal{P}) \in [0.1, 0.9]$ for all sufficiently large $n$?

First, an even simpler question, why is it that if $\mathcal{P}$ is a nontrivial monotone property, then $\mathbb{P}(G(n,p) \in \mathcal{P})$ is an increasing function of $p$? This is intuitively obvious, but how to prove it?

Let us give two (related) proofs of this basic fact. Both are quite instructive.

More abstractly, this is not really about graphs, but rather about random subsets (for random graphs, we are taking random subgraphs of edges).

Given a collection $\mathcal{F}$ of subsets of $[n]$, we say that $\mathcal{F}$ is an **upward closed set** (or **up-set**) if whenever $A \subset B$ and $A \in \mathcal{F}$ then $B \in \mathcal{F}$. We say that an up-set $\mathcal{F}$ is nontrivial if $\emptyset \notin \mathcal{F}$ and $[n] \in \mathcal{F}$.

Let $[n]_p$ denote the random subset of $[n]$ obtained by including every element independently with probability $p$.

**Theorem 4.2.2.** Let $\mathcal{F}$ a nontrivial up-set of $[n]$. Then $p \mapsto \mathbb{P}([n]_p \in \mathcal{F})$ is a strictly increasing function.

The first proof is by **coupling**. Coupling is powerful probabilistic idea. Given two random variables $X$ and $Y$ with individually prescibed distributions, we "couple" them together by considering a single probabilistic process that generates both $X$ and $Y$ in a way that clarifies their relationship. More formally, we construct a joint distribution $(X, Y)$ whose marginals agree with those of $X$ and $Y$.

*Proof 1.* (By coupling) Let $0 \le p < q \le 1$. Consider the following process to generate two random subsets of $[n]$: pick a uniform random vector $(x_1, \dots, x_n) \in [0,1]^n$. Let $A = \{i : x_i \le p\}$ and $B = \{i : x_i \le q\}$. Then $A$ has the same distribution as $[n]_p$ and $B$ has the same distribution as $[n]_q$. Furthermore, we see that $A \in \mathcal{F}$ implies $B \in \mathcal{F}$. Thus

$$\mathbb{P}([n]_p \in \mathcal{F}) = \mathbb{P}(A \in \mathcal{F}) \le \mathbb{P}(B \in \mathcal{F}) = \mathbb{P}([n]_q \in \mathcal{F}).$$

To see that the inequality strict, we simply have to observe that with positive probability, one has $A \notin \mathcal{F}$ and $B \in \mathcal{F}$ (e.g., $A = \emptyset$ and $B = [n]$). □

The second proof is also uses coupling, but viewed somewhat differently. The idea is that we can obtain $[n]_p$ as the union of several independent $[n]_{p'}$ for some smaller values of $p'$.

In other words, we are exposing the random subset in several rounds.

*Proof 2.* (By two-round exposure) Let $0 \le p < q \le 1$. Note that $B = [n]_q$ has the same distribution as the union of two independent $A = [n]_p$ and $A' = [n]_{p'}$, where $p'$ is chosen to satisfy $1 - q = (1 - p)(1 - p')$. Thus

$$\mathbb{P}(A \in \mathcal{F}) \le \mathbb{P}(A \cup A' \in \mathcal{F}) = \mathbb{P}(B \in \mathcal{F}).$$

Like earlier, to observe that the inequality is strict, one observes that with positive probability, one has $A \notin \mathcal{F}$ and $A \cup A' \in \mathcal{F}$. □

The above technique (generalized from two round exposure to multiple round exposures) gives a nice proof of the following theorem (originally proved using the Kruskal–Katona theorem).

**Theorem 4.2.3** (Bollobás and Thomason 1987)**.** Every nontrivial monotone graph property has a threshold function.

*Proof.* Note that $G(n, 1-(1-p)^k)$ has the same distribution has the union of $k$ independent copies $G^1, \dots, G^k$ of $G(n, p)$. Furthermore, by the monotonicity of the property, if $G^1 \cup \cdots \cup G^k \notin \mathcal{P}$, then $G^1, \dots, G^k \notin \mathcal{P}$. By independence,

$$\mathbb{P}(G(n, 1 - (1 - p)^k) \notin \mathcal{P}) = \mathbb{P}(G^1 \cup \cdots \cup G^k \notin \mathcal{P}) \le \mathbb{P}(G^1 \notin \mathcal{P}) \cdots \mathbb{P}(G^k \notin \mathcal{P})$$

To simplify notation, let us write

$$f_p = f_p(n) = \mathbb{P}(G(n, p) \in \mathcal{P}).$$

Since $1 - (1 - p)^k \le kp$ (check by convexity), we have that for any monotone graph property $\mathcal{P}$, any positive integer $k \le 1/p$,

$$1 - f_{kp} \le 1 - f_{1-(1-p)^k} \le (1 - f_p)^k. \tag{4.1}$$

Fix any large enough $n$ (so that set of $n$-vertex graphs satisfying the property $\mathcal{P}$ is a nontrival up-set). Since $p \mapsto f_p(n)$ is a continuous strictly increasing function from 0 to 1 as $p$ goes from 0 to 1 (in fact it is a polynomial in $p$ for each fixed $n$), there is some "critical" $p_c = p_c(n)$ with $f_{p_c}(n) = 1/2$.

We claim that $p_c$ is a threshold function. Indeed, (4.1) implies, if $p = p(n) \gg p_c(n)$, then, letting $k = k(n) = \lfloor p/p_c \rfloor \to \infty$,

$$1 - f_p \le (1 - f_{p_c})^k = 2^{-k} \to 0$$

so $f_p \to 1$. Likewise, if $p \ll p_c$, then, letting $k = \lfloor p_c/p \rfloor \to \infty$, we have

$$\frac{1}{2} = 1 - f_{p_c} \le (1 - f_p)^k,$$

and thus $f_p \to 0$ as $n \to \infty$. Thus $p_c(n)$ is a threshold function for $\mathcal{P}$. □

*Remark* 4.2.4. Note that, by definition, if $p_1(n)$ and $p_2(n)$ are both threshold functions for the same property, then $cp_1(n) \le p_2(n) \le Cp_2(n)$ for some constants $0 < c < C$.

Last section we identified the threshold for the property of containing a fixed subgraph. Let us state the result (at least in the case of triangles, but similar results are known for every subgraph) a bit more precisely, where we use the fact that for a constant $c > 0$, the number of triangles in $G(n, c/n)$ converges to a Poisson distribution with mean $c^3/6$ (this can be proved using the "method of moments" but we will not do it here). So

$$\mathbb{P}\left(G\left(n, \frac{c_n}{n}\right) \text{ contains a triangle}\right) \to \begin{cases} 0 & \text{if } c_n \to 0 \\ 1 - e^{-c^3/6} & \text{if } c_n \to c \text{ constant} \\ 1 & \text{if } c_n \to \infty \end{cases}$$

What about other graph properties? It turns out that we can sometimes identity the transition very precisely.

**Example 4.2.5.** Here are some more examples of threshold functions. The first two statements are in the original Erdős–Rényi (1959) paper on random graphs. The first is an easy (and instructive) exercise in the second moment method.

- With $p = \dfrac{\log n + c_n}{n}$

$$\mathbb{P}\left(G\left(n,p\right) \text{ has no isolated vertices}\right) \to \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$$

- With $p = \dfrac{\log n + c_n}{n}$

$$\mathbb{P}\left(G\left(n,p\right) \text{ is connected}\right) \to \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$$

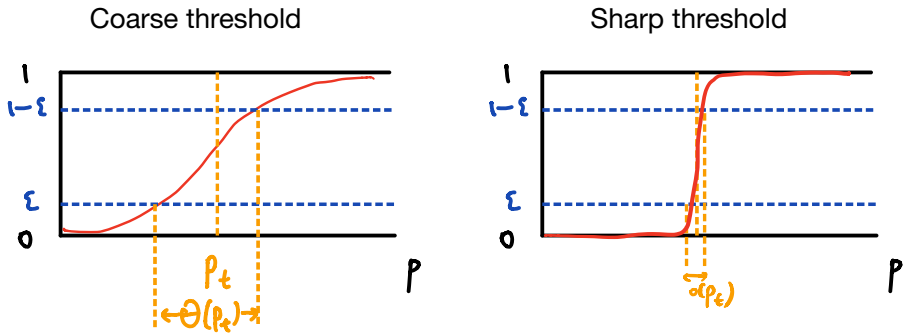


Figure 4: Examples of coarse and sharp thresholds. The vertical axis is the probability that $G(n,p)$ satisfies the property.

In fact, a much stronger statement is true, connecting the above two examples: consider a process where one adds an random edges one at a time, then with probability $1-o(1)$, the graph becomes connected as soon as there are no more isolated vertices.

- With $p = \dfrac{\log n + \log\log n + c_n}{n}$

$$\mathbb{P}(G(n,p) \text{ has a Hamiltonian cycle}) \to \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$$

Like earlier, it is true that with high probability, a random graph becomes Hamiltonian as soon as its minimum degree reaches 2.

In the above examples, the probability that $G(n,p)$ satisfies the property changes quickly and dramatically as $p$ crosses the threshold (physical analogy: similar to how the structure of water changes dramatically as the temperature drops below freezing). For example, while for connectivity, while $p = \log n/n$ is a threshold function, we see that $G(n, 0.99\log n/n)$ is whp not connected and $G(n, 1.01\log n/n)$ is whp connected, unlike the situation for containing a triangle earlier. We call this the **sharp threshold phenomenon**.

**Definition 4.2.6** (Sharp thresholds)**.** We say that $r_n$ is a **sharp threshold** for some graph property $\mathcal{P}$ if, for every $\delta > 0$,

$$\mathbb{P}(G(n,p_n) \text{ satisfies } \mathcal{P}) \to \begin{cases} 0 & \text{if } p_n \le (1-\delta)r_n, \\ 1 & \text{if } p_n \ge (1-\delta)r_n. \end{cases}$$

Equivalently, a graph property $\mathcal{P}$ exhibits a sharp threshold at $r_n$ if, for every $\epsilon > 0$,

for a given large $n$, as $p$ increases from 0 to 1, the probability $\mathbb{P}(G(n,p) \in \mathcal{P})$ increases from $\epsilon$ to $1-\epsilon$ over a short window of width $o(r_n)$ around $r_n$. On the other hand, if this transition window has width $\Omega(r_n)$ for some $\epsilon > 0$, then we say that it is a **coarse threshold**. See Figure 4.

We saw coarse thresholds for the "local" property of containing some given subgraph, whereas we saw sharp thresholds for "global" properties such as connectivity. It turns out that this is a general phenomenon.

**Friedgut's sharp threshold theorem** (1999), a deep and important result, roughly says that:

> All monotone graph properties with a coarse threshold may be approximated by a local property.

In other words, informally, if a monotone graph property $\mathcal{P}$ has a coarse threshold, then there is finite list of graph $G_1, \dots, G_m$ such that $\mathcal{P}$ is "close to" the property of containing one of $G_1, \dots, G_m$ as a subgraph.

We need "close to" since the property could be "contains a triangle and has at least $\log n$ edges", which is not exactly local but it is basically the same as "contains a triangle."

There is some subtlety here since we can allow very different properties depending on the value of $n$. E.g., $\mathcal{P}$ could be the set of all $n$-vertex graphs that contain a $K_3$ if $n$ is odd and $K_4$ if $n$ is even. Friedgut's theorem tells us that if there is a threshold, then there is a partition $\mathbb{N} = \mathbb{N}_1 \cup \cdots \cup \mathbb{N}_k$ such that on each $\mathbb{N}_i$, $\mathcal{P}$ is approximately the form described in the previous paragraph.

In the last section, we derived that the property of containing some fixed $H$ has threshold $n^{-1/m(H)}$ for some rational number $m(H)$. It follows as a corollary of Friedgut's theorem that every coarse threshold must have this form.

**Corollary 4.2.7** (of Friedgut's sharp threshold theorem)**.** Suppose $r(n)$ is a coarse threshold function of some graph property. Then there is a partition of $\mathbb{N} = \mathbb{N}_1 \cup \cdots \cup \mathbb{N}_k$ and rationals $\alpha_1, \dots, \alpha_k > 0$ such that $r(n) \asymp n^{-\alpha_j}$ for every $n \in \mathbb{N}_j$.

In particular, if $(\log n)/n$ is a threshold function of some monotone graph property (e.g., this is the case for connectivity), then we automatically know that it must be a sharp threshold, even without knowing anything else about the property. Likewise if the threshold has the form $n^{-\alpha}$ for some irrational $\alpha$.

The exact statement of Friedgut's theorem is more cumbersome. We refer those who are interested to Friedgut's original 1999 paper and his later survey for details and applications. This topic is connected more generally to an area known as the **analysis of**

**boolean functions**.

Also, it is known that the transition window of every monotone graph property is $(\log n)^{-2+o(1)}$ (Friedgut—Kalai (1996), Bourgain–Kalai (1997)).

Curiously, tools such as Friedgut's theorem sometimes allow us to prove the existence of a sharp threshold without being able to identify its exact location. For example, it is an important open problem to understand where exactly is the transition for a random graph to be $k$-colorable.

**Conjecture 4.2.8** ($k$-colorability threshold)**.** For every $k \geq 3$ there is some real constant $d_k > 0$ such that for any constant $d > 0$,

$$\mathbb{P}(G(n, d/n) \text{ is } k\text{-colorable}) \to \begin{cases} 1 & \text{if } d < d_k, \\ 0 & \text{if } d > d_k. \end{cases}$$

We do know that there *exists* a sharp threshold for $k$-colorability.

**Theorem 4.2.9** (Achlioptas and Friedgut 2000)**.** For every $k \geq 3$, there exists a function $d_k(n)$ such that for every $\epsilon > 0$, and sequence $d(n) > 0$,

$$\mathbb{P}\left(G\left(n, \tfrac{d(n)}{n}\right) \text{ is } k\text{-colorable}\right) \to \begin{cases} 1 & \text{if } d(n) < d_k(n) - \epsilon, \\ 0 & \text{if } d(n) > d_k(n) + \epsilon. \end{cases}$$

On the other hand, it is not known whether $\lim_{n\to\infty} d_k(n)$ exists, which would imply Conjecture 4.2.8. Further bounds on $d_k(n)$ are known, e.g. the landmark paper of Achlioptas and Naor (2006) showing that for each fixed $d > 0$, whp $\chi(G(n, d/n) \in \{k_d, k_d + 1\}$ where $k_d = \min\{k \in \mathbb{N} : 2k \log k > d\}$. Also see the later work of Coja-Oghlan and Vilenchik (2013).

## 4.3 Clique number of a random graph

The **clique number** $\omega(G)$ of a graph is the maximum number of vertices in a clique of $G$.

**Question 4.3.1.** What is the clique number of $G(n, 1/2)$?

Let $X$ be the number of $k$-cliques of $G(n, 1/2)$. We have

$$f(k) := \mathbb{E}X = \binom{n}{k} 2^{-\binom{k}{2}}.$$

**Theorem 4.3.2.** Let $k = k(n)$ satisfy $f(k) \to \infty$. Then $\omega(G(n, 1/2)) \geq k$ whp.

*Proof.* For each $k$-element subset $S$ of vertices, let $A_S$ be the event that $S$ is a clique. Let $X_S$ be the indicator random variable for $A_S$. Let $X = \sum_{S \in \binom{[n]}{k}} X_S$ denote the number of $k$-cliques.

For fixed $k$-set $S$, consider all $k$-set $T$ with $|S \cap T| \geq 2$:

$$\Delta^* = \sum_{\substack{T \in \binom{[n]}{k} \\ 2 \leq |S \cap T| \leq k-1}} \mathbb{P}(A_T | A_S) = \sum_{i=2}^{k-1} \binom{k}{i} \binom{n-k}{k-i} 2^{\binom{i}{2} - \binom{k}{2}} \overset{\text{omitted}}{\ll} \mathbb{E}X = \binom{n}{k} 2^{-\binom{k}{2}}.$$

It then follows from Lemma 4.1.13 that $X > 0$ (i.e., $\omega(G) \geq k$) whp. □

**Theorem 4.3.3** (Bollobás–Erdős 1976 and Matula 1976)**.** There exists a $k = k(n) \sim 2 \log_2 n$ such that $\omega(G(n, 1/2)) \in \{k, k+1\}$ whp.

*Proof.* (Sketch) For $k \sim 2 \log_2 n$,

$$\frac{f(k+1)}{f(k)} = \frac{n-k}{k+1} 2^{-k} = n^{-1+o(1)} = o(1).$$

So the value of $f(k)$ drops rapidly for $k \sim 2 \log_2 n$. Let $k_0 = k_0(n)$ be the value with $f(k_0) \geq 1 > f(k_0 + 1)$. If $n$ is such that $f(k_0) \to \infty$ while $f(k_0 + 1) \to 0$ (it turns out that this is true for most integers $n$), and thus $\omega(G) = k_0$ whp. When $f(k_0) = O(1)$, we have $f(k_0 - 1) \to \infty$ and $f(k_0 + 1) \to 0$ so one has $\omega(G(n, 1/2)) \in \{k_0 - 1, k_0\}$ whp. □

*Remark* 4.3.4. The result also implies the same about size of largest independent set in $G(n, 1/2)$ (take complement). Also extends to constant $p$: $\omega(G(n, p)) \sim 2 \log_{1/(1-p)} n$ whp.

Since the chromatic number satisfies $\chi(G) \geq n/\alpha(G)$, we have

$$\chi(G(n, 1/2)) \geq (1 + o(1)) \frac{n}{2 \log_2 n} \qquad \text{whp.}$$

Later on, using more advanced methods, we will prove $\chi(G(n, 1/2)) \sim n/(2 \log_2 n)$ whp (Bollobás 1987).

Also, later, using martingale concentration, we know show that $\chi(G(n, p))$ is tightly concentrated around its mean without a priori needing to know where the mean is located.

## 4.4 Hardy–Ramanujan theorem on the number of prime divisors

Let $\nu(n)$ denote the number of primes $p$ dividing $n$ (do not count multiplicities).

The next theorem says that "almost all" $n$ have $(1+o(1))\log\log n$ prime factors

**Theorem 4.4.1** (Hardy and Ramanujan 1917)**.** For every $\epsilon > 0$, there exists $C$ such that all but $\epsilon$-fraction of $x \in [n]$ satisfy

$$|\nu(x) - \log\log n| \le C\sqrt{\log\log n}$$

The original proof of Hardy and Ramanujan was quite involved. Here we show a "probabilistic" proof due to Turán (1934), which played a key role in the development of probabilistic methods in number theory.

*Proof.* Choose $x \in [n]$ uniformly at random. For prime $p$, let

$$X_p = \begin{cases} 1 & \text{if } p|x, \\ 0 & \text{otherwise.} \end{cases}$$

Set $M = n^{1/10}$, and (the sum is taken over primes $p$).

$$X = \sum_{p \le M} X_p$$

We have $\nu(x) - 10 \le X(x) \le \nu(x)$ since $x$ cannot have more than 10 prime factors $> n^{1/10}$. So it suffices to analyze $X$. Since exactly $\lfloor n/p \rfloor$ positive integers $\le n$ are divisible by $p$, we have

$$\mathbb{E}X_p = \frac{\lfloor n/p \rfloor}{n} = \frac{1}{p} + O\left(\frac{1}{n}\right)$$

So

$$\mathbb{E}X = \sum_{p \le M} \left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right) = \log\log n + O(1)$$

Here we are applying Merten's theorem from analytic number theory: $\sum_{p \le n} 1/p = \log\log n + O(1)$ (the $O(1)$ error term converges to the Meissel–Mertens constant).

Next we compute the variance. The intuition is that distinct primes should be have independently. Indeed, if $pq$ divides $n$, then $X_p$ and $X_q$ are independent. Then $pq$ does not divide $n$, but $n$ is large enough, then there is some small covariance contribution. (Contrast to the earlier calculations in random graphs, where there are very few nonzero

covariance terms, but each can be more significant.)

If $p \neq q$, then $X_p X_q = 1$ if and only if $pq|x$. Thus

$$\begin{aligned}|\operatorname{Cov}[X_p, X_q]| &= |\mathbb{E}[X_p X_q] - \mathbb{E}[X_p]\mathbb{E}[X_q]| \\ &= \left| \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \frac{\lfloor n/q \rfloor}{n} \right| \\ &= O\left(\frac{1}{n}\right)\end{aligned}$$

Thus

$$\sum_{p \neq q} |\operatorname{Cov}[X_p, X_q]| \lesssim \frac{M^2}{n} \lesssim n^{-4/5}$$

Also, $\operatorname{Var} X_p = \mathbb{E}[X_p] - (\mathbb{E}X_p)^2 = (1/p)(1 - 1/p) + O(1/n)$. Combining, we have

$$\begin{aligned}\operatorname{Var} X &= \sum_{p \leq M} \operatorname{Var} X_p + \sum_{p \neq q} \operatorname{Cov}[X_p, X_q] \\ &= \sum_{p \leq M} \frac{1}{p} + O(1) = \log\log n + O(1) \sim \mathbb{E}X\end{aligned}$$

Thus by Chebyshev, for every constant $\lambda > 0$

$$\mathbb{P}\left(|X - \log\log n| \geq \lambda \sqrt{\log\log n}\right) \leq \frac{(\operatorname{Var} X)^2}{\lambda^2 (\log\log n)} = \frac{1}{\lambda^2} + o(1).$$

Finally, recall that $|X - \nu| \leq 10$, so same asymptotic bound holds with $X$ replaced by $\nu$. □

**Theorem 4.4.2** (Erdős and Kac 1940)**.** With $x \in [n]$ uniformly chosen at random, $\nu(x)$ is asymptotically normal, i.e., for every $\lambda \in \mathbb{R}$,

$$\lim_{n \to \infty} \mathbb{P}_{x \in [n]} \left( \frac{\nu(x) - \log\log n}{\sqrt{\log\log n}} \geq \lambda \right) = \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-t^2/2}\, dt$$

The intuition is that the number of prime divisors $X = \sum_p X_p$ (from the previous proof) behaves like a sum of independent random variables, the central limit theorem should imply an asymptotic normal distribution.

The original proof of Erdős and Kac verifies the above intuition using some more involved results in analytic number theory. Simpler proofs have been subsequently given, and we outline one below, which is based on computing the moments of the distribution. The idea of computing moments for this problem was first used by Delange (1953), who was

apparently not aware of the Erdős–Kacs paper. Also see a more modern account by Granville and Soundararajan (2007).

The following tool from probability theory allows us to verify asymptotic normality from convergence of moments.

**Theorem 4.4.3** (Method of moments)**.** Let $X_n$ be a sequence of real valued random variables such that for every positive integer $k$, $\lim_{n\to\infty} \mathbb{E}[X_n^k]$ equals to the $k$-th moment of the standard normal distribution. Then $X_n$ converges in distribution to the standard normal, i.e., $\lim_{n\to\infty} \mathbb{P}(X_n \le a) = \mathbb{P}(Z \le a)$ for every $a \in \mathbb{R}$, where $Z$ is a standard normal.

*Remark* 4.4.4. The same conclusion holds for any probability distribution (other than normal) that is "determined by its moments," i.e., there are no other distributions sharing the same moments. Many common distributions that arise in practice, e.g., the Poisson distribution, satisfy this property. There are various sufficient conditions for guaranteeing this moments property, e.g., Carleman's condition tells us that any probability distribution whose moments do not increase too quickly is determined by its moments.

*Proof sketch of Erdős–Kacs Theorem 4.4.2.* We compare higher moments of $X = \nu(x)$ with that of an idealized $Y$ treating the prime divisors as truly random variables.

Set $M = n^{1/s(n)}$ where $s(n) \to \infty$ sufficiently slowly. As earlier, $\nu(x) - s(n) \le \nu(x) \le v(x)$.

We construct a "model random variable" mimicking $X$. Let $Y = \sum_{p \le M} Y_p$, where $Y_p \sim$ Bernoulli$(1/p)$ independently for all primes $p \le M$. We can compute:

$$\mu := \mathbb{E}Y \sim \mathbb{E}X \sim \log\log n$$

and

$$\sigma^2 := \operatorname{Var} Y \sim \operatorname{Var} X \sim \log\log n.$$

Let $\widetilde{X} = (X - \mu)/\sigma$ and $\widetilde{Y} = (Y - \mu)/\sigma$.

By the central limit theorem (e.g., the Lindeberg CLT), $\widetilde{Y} \to N(0,1)$ in distribution. In particular, $\mathbb{E}[\widetilde{Y}^k] \sim \mathbb{E}[Z^k]$ (asymptotics as $n \to \infty$) where $Z$ is a standard normal.

Let us compare $\widetilde{X}$ and $\widetilde{Y}$. It suffices to show that for every fixed $k$, $\mathbb{E}[\widetilde{X}^k] \sim \mathbb{E}[\widetilde{Y}^k]$.

For every set of distinct primes $p_1, \dots p_r \le M$,

$$\mathbb{E}[X_{p_1} \cdots X_{p_r} - Y_{p_1} \cdots Y_{p_r}] = \frac{1}{n}\left\lfloor \frac{n}{p_1 \cdots p_r} \right\rfloor - \frac{1}{p_1 \cdots p_r} = O\left(\frac{1}{n}\right)$$

Comparing expansions of $\widetilde{X}^k$ in terms of the $X_p$'s ($n^{o(1)}$ terms), we get

$$\mathbb{E}[\widetilde{X}^k - \widetilde{Y}^k] = n^{-1+o(1)} = o(1).$$

So the moments of $\widetilde{X}$ approach those of $N(0,1)$. The method of moments theorem from probability then implies that $\widetilde{X}$ is asymptotically normally distributed. □

## 4.5 Distinct sums

**Question 4.5.1.** Let $S$ be a $k$-element subset of positive integers such that all $2^k$ subset sums of $S$. What is the minimum possible $\max S$?

E.g., $S = \{1, 2, 2^2, \ldots, 2^{k-1}\}$ (the greedy choice).

We begin with an easy pigeonhole argument. On the other hand, since all $2^k$ sums are distinct and are at most $k \max S$, we have $2^k \le k \max S$, so $\max S \ge 2^k/k$.

Erdős offered \$300 for a proof or disproof that $\max S \gtrsim 2^k$. This remains an interesting open problem.

Let us use the second moment to give a modest improvement on the earlier pigeonhole argument. The main idea here is that, by second moment, most of the subset sums lie within an $O(\sigma)$-interval, so that we can improve on the pigeonhole estimate ignoring outlier subset sums.

**Theorem 4.5.2.** Let $S$ be a $k$-element subset of positive integers such that all $2^k$ subset sums of $S$. Then $\max S \gtrsim 2^k/\sqrt{k}$.

*Proof.* Let $S = \{x_1, \ldots, x_k\}$ and $n = \max S$. Set

$$X = \epsilon_1 x_1 + \cdots + \epsilon_k x_k$$

where $\epsilon_i \in \{0, 1\}$ are chosen uniformly at random independently. We have

$$\mu := \mathbb{E}X = \frac{x_1 + \cdots + x_k}{2}$$

and

$$\sigma^2 := \operatorname{Var} X = \frac{x_1^2 + \cdots + x_k^2}{4} \le \frac{n^2 k}{4}.$$

By Chebyshev,

$$\mathbb{P}(|X - \mu| < n\sqrt{k}) \ge \frac{3}{4}.$$

Since $X$ takes distinct values for every $(\epsilon_1, \dots, \epsilon_k) \in \{0,1\}^k$, we have $\mathbb{P}(X = x) \le 2^{-k}$ for all $x$, so we have the lower bound

$$\mathbb{P}(|X - \mu| < n\sqrt{k}) \le 2^{-k}(2n\sqrt{k} + 1).$$

Putting them together, we get

$$2^{-k}(2n\sqrt{k} + 1) \le \frac{3}{4}.$$

So $n \gtrsim 2^k/\sqrt{k}$. □

Recently, this July, Dubroff–Fox–Xu gave another short proof of this result (with an improved error term $O(1)$) by applying Harper's vertex-isoperimetric inequality on the cube (this is an example of "concentration of measure", which we will explore more later this course).

Here for the "$n$-dimensional boolean cube" we consider the graph on the vertex set $\{0,1\}^n$ with an edge between every pair of $n$-tuples that differ in exactly one coordinate. Given $A \subseteq \{0,1\}^n$, let $\delta A$ be the set of all vertices outside $A$ that is adjacent to some vertex of $A$.

**Theorem 4.5.3** (Harper 1966)**.** Every $A \subset \{0,1\}^k$ with $|A| = 2^{k-1}$ has $|\partial A| \ge \binom{k}{\lfloor k/2 \rfloor}$.

*Remark* 4.5.4. Harper's theorem, more generally, gives the precise value of $\min_{A \subset \{0,1\}^n : |A| = m} |\partial A|$ for every $(n, m)$. Basically, the minimum is achieved when $A$ is a a Hamming ball (or, if $m$ is not exactly the size of some Hamming ball, then take the first $m$ elements of $\{0,1\}^n$ when ordered lexicographically).

**Theorem 4.5.5** (Dubroff–Fox–Xu)**.** If $S$ is a set of $k$ positive integers with distinct subset sums, then

$$\max S \ge \binom{k}{\lfloor k/2 \rfloor} = \left(\sqrt{\frac{2}{\pi}} + o(1)\right) \frac{2^k}{\sqrt{k}}.$$

*Remark* 4.5.6. The above bound has the currently best known leading constant factor.

*Proof.* Let $S = \{x_1, \dots, x_k\}$. Let

$$A = \left\{ (\epsilon_1, \dots, \epsilon_k) \in \{0,1\}^k : \epsilon_1 x_1 + \cdots + \epsilon_1 x_k < \frac{x_1 + \cdots + x_k}{2} \right\}.$$

Note that due to the distinct sum hypothesis, one can never have $x_1 s_1 + \cdots + x_n s_n = (s_1 + \cdots + s_n)/2$. It thus follows by symmetry that $|A| = 2^{k-1}$.

Note that every element of $\partial A$ corresponds to some subset sum in the open interval

$$\left(\frac{x_1+\cdots+x_k}{2}, \frac{x_1+\cdots+x_k}{2}+\max S\right)$$

Since all subset sums are distinct, we must have $\max S \geq |\partial A| \geq \binom{k}{\lfloor k/2 \rfloor}$ by Harper's theorem (Theorem 4.5.3). □

## 4.6 Weierstrass approximation theorem

We finish off the chapter with an application to analysis.

Weierstrass approximation theorem every continuous real function on an interval can be uniformly approximated by a polynomial.

**Theorem 4.6.1** (Weierstrass approximation theorem 1885)**.** Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function. Let $\epsilon > 0$. Then there is a polynomial $p(x)$ such that $|p(x) - f(x)| \leq \epsilon$ for all $x \in [0,1]$.

*Proof.* (Bernstein 1912) The idea is to approximate $f$ by a sum of polynomials look like "bumps":

$$P_n(x) = \sum_{i=0}^{n} E_i(x) f(i/n)$$

where $E_j(x)$ chosen as some polynomials peaks at $x = i/n$ and then decaysaway from $x = i/n$. To this end, set set

$$E_i(x) = \mathbb{P}(\mathrm{Bin}(n,x) = i) = \binom{n}{i} x^i (1-x)^{n-i} \quad \text{for } 0 \leq i \leq n.$$

For each $x \in [0,1]$, the binomial distribution $\mathrm{Bin}(n,x)$ has mean $nx$ and variance $nx(1-x) \leq n$. By Chebyshev's inequality,

$$\sum_{i:|i-nx|>n^{2/3}} E_i(x) = \mathbb{P}(|\mathrm{Bin}(n,x) - nx| > n^{2/3}) \leq n^{-1/3}.$$

Since $[0,1]$ is compact, $f$ is uniformly continuous and bounded. By rescaling, assume that $|f(x)| \leq 1$ for all $x \in [0,1]$. Also there exists $\delta > 0$ such that $|f(x) - f(y)| \leq \epsilon/2$ for all $x, y \in [0,1]$ with $|x - y| \leq \delta$.

Take $n > \max\{64\epsilon^{-3}, \delta^{-3}\}$. Then for every $x \in [0,1]$ (note that $\sum_{j=0}^{n} E_j(x) = 1$),

$$\begin{aligned}
|P_n(x) - f(x)| &\le \sum_{i=0}^{n} E_i(x)|f(i/n) - f(x)| \\
&\le \sum_{i:|i/n-x|<n^{-1/3}<\delta} E_i(x)|f(i/n) - f(x)| + \sum_{i:|i-nx|>n^{2/3}} 2E_i(x) \\
&\le \frac{\epsilon}{2} + 2n^{-1/3} \le \epsilon. \square
\end{aligned}$$

# 18.226 Probabilistic Method in Combinatorics
Fall 2020