4 Second Moment

4.1 Does a typical random graph contain a triangle?

We begin with the following motivating question. Recall that the Erdős–Rényi random graph $G(n,p)$ is the n-vertex graph with edge probability p.

Question 4.1.1
For which $p = p_n$ does $G(n,p)$ contain a triangle with probability $1 - o(1)$?

(We sometimes abbreviate "with probability $1 - o(1)$ by "with high probability" or simply "whp". In some literature, this is also called "asymptotically almost surely" or "a.a.s.")

By computing $\mathbb{E}X$ (also known as the ***first moment***), we deduce the following.

Proposition 4.1.2
If $np \to 0$, then $G(n,p)$ is triangle-free with probability $1 - o(1)$.

Proof. Let X be the number of triangles in $G(n,p)$. We know from linearity of expectations that

$$\mathbb{E}X = \binom{n}{3} p^3 \asymp n^3 p^3 = o(1).$$

Thus, by Markov's inequality,

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}X = o(1).$$

In other words, $X = 0$ with probability $1 - o(1)$. □

In other words, when $p \ll 1/n$, $G(n,p)$ is triangle-free with high probaiblity (recall that $p \ll 1/n$ means $p = o(1/n)$; see asymptotic notation guide at the beginning of these notes).

What about when $p \gg 1/n$? Can we conclude that $G(n,p)$ contains a triangle with high probability? In this case $\mathbb{E}X \to \infty$, but this is not enough to conclude that

$\mathbb{P}(X \geq 1) = 1 - o(1)$, since we have not ruled out the probability that X is almost always zero but extremely large with some tiny probability.

An important technique in probabilistic combinatorics is 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. In this chapter, we begin with the simplest method. It is usually easiest to execute and it requires not much hypotheses. The downside is that it only produces relatively weak (though still useful enough) concentration bounds.

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

Definition 4.1.3 (Variance)
The ***variance*** of a random variable X is

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

The ***covariance*** of two random variables X and Y (jointly distributed) is

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

(Exercise: check the second equality in the definitions of variance and covariance above).

Remark* 4.1.4** (Notation convention)**.** It is standard to use the Greek letter μ for the mean, and σ^2 for the variance. Here $\sigma \geq 0$ is the ***standard deviation.

The following basic result provides a concentration bound based on the variance.

Theorem 4.1.5 (Chebyshev's inequality)
Let X be a random variable with mean μ and variance σ^2. 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* 4.1.6.** Concentration bounds that show small probability of deviating from the mean are called ***tail bounds (more precisely: upper tail for $X \geq \mu + a$ and lower tail

for $\mathbb{P}(X \le \mu - a)$). Chebyshev's inequality gives tail bounds that decays quadratically. Later on we will see tools that give much better decay (usually exponential) provided additional assumptions on the random variable (e.g., independence).

We are often interested in upper bounding the probability of non-existence, i.e., $\mathbb{P}(X = 0)$. Chebyshev's inequality yields the following bound.

Corollary 4.1.7 (Chebyshev bound on the probability of non-existence)
For any random variable X,

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

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

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

Corollary 4.1.8
If $\mathbb{E}X > 0$ and $\operatorname{Var} X = o(\mathbb{E}X)^2$, then $X > 0$ and $X \sim \mathbb{E}X$ with probability $1 - o(1)$.

***Remark* 4.1.9** (Asymptotic statements)**.** The above statement is really referring to not a single random variable, but a sequence of random variables X_n. It is saying that if $\mathbb{E}X_n > 0$ and $\operatorname{Var} X_n = o(\mathbb{E}X_n)^2$, then $\mathbb{P}(X_n > 0) \to 1$ as $n \to \infty$, and for any fixed $\delta > 0$, $\mathbb{P}(|X_n - EX_n| > \delta \mathbb{E}X_n) \to 0$ as $n \to \infty$.

In many situations, it is not too hard to compute the second moment. We have $\operatorname{Var}[X] = \operatorname{Cov}[X, X]$. Also, covariance is bilinear, i.e., for random variables $X_1, \ldots$ and $Y_1, \ldots$ (no assumptions needed on their independence, etc.) and constants $a_1, \ldots$ and $b_1, \ldots$, one has

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

We are often dealing with X being the cardinality of some random set. We can usually write this as a sum of indicator functions, such as $X = X_1 + \cdots + X_n$, so that

$$\operatorname{Var} X = \operatorname{Cov}[X, X] = \sum_{i,j=1}^{n} \operatorname{Cov}[X_i, X_j] = \sum_{i=1}^{n} \operatorname{Var} X_i + 2 \sum_{i<j} \operatorname{Cov}[X_i, X_j].$$

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* 4.1.10** (Sum of independent Bernoulli)**.** Suppose $X = X_1 + \cdots + X_n$ with each X_i being an independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1 - p$. Then $\mu = np$ and $\sigma^2 = np(1-p)$ (note that $\mathrm{Var}[X_i] = p - p^2$ and $\mathrm{Cov}[X_i, X_j] = 0$ if $i \neq j$). If $np \to \infty$, then $\sigma = o(\mu)$, 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.

Let us now return to the problem of determining when $G(n, p)$ contains a triangle whp.

Theorem 4.1.11
If $np \to \infty$, then $G(n, p)$ contains a triangle with probability $1 - o(1)$.

Proof. Label the vertices by $[n]$. Let X_{ij} be the indicator random variable of the edge ij, so that $X_{ij} = 1$ if the edge is present, and $X_{ij} = 0$ if the edge is not present in the random graph. Let us write

$$X_{ijk} := X_{ij} X_{ik} X_{jk}.$$

Then the number of triangles in $G(n, p)$ is given by

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

Now we compute $\mathrm{Var}\, X$. Note that the summands of X are not all independent.

If T_1 and T_2 are each 3-vertex subsets, then

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

The number of pairs (T_1, T_2) of triangles sharing exactly on edge is $O(n^4)$. Thus

$$\begin{aligned}\operatorname{Var} X = \sum_{T_1,T_2} \operatorname{Cov}[X_{T_1}, X_{T_2}] &= O(n^3)(p^3 - p^6) + O(n^4)(p^5 - p^6) \\ &\lesssim n^3p^3 + n^4p^5 = o(n^6p^6) \quad \text{as } np \to \infty.\end{aligned}$$

Thus $\operatorname{Var} X = o(\mathbb{E}X)^2$, and hence $X > 0$ whp by Corollary 4.1.8. □

Here is what we have learned so far: for $p = p_n$ and as $n \to \infty$,

$$\mathbb{P}(G(n,p) \text{ contains a triangle}) \to \begin{cases} 0 & \text{if } np \to 0, \\ 1 & \text{if } np \to \infty. \end{cases}$$

We say that $1/n$ is a ***threshold*** for containing a triangle, in the sense that if p grows asymptotically faster than this threshold, i.e., $p \gg 1/n$, then the event occurs with probability $1 - o(1)$, while if $p \ll 1/n$, then the event occurs with probability $o(1)$. Note that the definition of a threshold ignores leading constant factors (so that it is also correct to say that $2/n$ is also a threshold for containing a triangle). Determining the thresholds of various properties in random graphs (as well as other random settings) is a central topic in probabilistic combinatorics. We will discuss thresholds in more depth later in this chapter.

What else might you want to know about the probability that $G(n,p)$ contains a triangle?

Remark* 4.1.12** (Poisson limit)**.** What if $np \to c > 0$ for some constant $c > 0$? It turns out in this case that the number of triangles of $G(n,p)$ approaches a Poisson distribution with constant mean. You will show this in the homework. It will be done via the ***method of moments: if Z is some random variable with sufficiently nice properties (known as "determined by moments", which holds for many common distributions such as the Poisson distribution and the normal distribution), and X_n is a sequence of random variables such that $\mathbb{E}X_n^k \to \mathbb{E}Z^k$ for all nonnegative integers k, then X_n converges in distribution to Z.

***Remark* 4.1.13** (Asymptotic normality)**.** Suppose $np \to \infty$. From the above proof, we also deduce that $X \sim \mathbb{E}X$, i.e., the number of triangles is concentrated around its mean. In fact, we know much more. It turns out that the number X of triangles in $G(n,p)$ is asymptotically normal, meaning that it satisfies a central limit theorem: $(X - \mathbb{E}X)/\sqrt{\operatorname{Var} X}$ converges in distribution to the standard normal $N(0,1)$ in distribution. This was shown by Rucinski (1988) via the method of moments, by computing the k-th moment of $(X - \mathbb{E}X)/\sqrt{\operatorname{Var} X}$ in the limit, and showing that it agrees with the k-th moment of the standard normal.

In the homework, you will prove the asymptotic normality of X using a later-found ***method of projections***. The idea is to show that that X close to another random variable that is already known to be asymptotically normal by checking that their difference has negligible variance. For triangle counts, when $p \gg n^{-1/2}$, we can compare the number of triangles to the number of edges after a normalization. The method can be further modified for greater generality. See the §6.4 in the book *Random Graphs* by Janson, Łuczak, and Rucinski (2000).

***Remark* 4.1.14** (Better tail bounds)**.** Later on we will use more powerful tools (including martingale methods and Azuma-Hoeffding inequalities, and also Janson inequalities) to prove better tail bounds on triangle (and other subgraph) counts.

4.2 Thresholds for fixed subgraphs

In the last section, we determined the threshold for $G(n,p)$ to contain a triangle. What about other subgraphs instead of a triangle? In this section, we give a complete answer to this question for any fixed subgraph.

Question 4.2.1
What is the threshold for containing a fixed H as a subgraph?

In other words, we wish to find some sequence q_n so that:

- (0-statement) if $p_n/q_n \to 0$ (i.e., $p_n \ll q_n$), then $G(n,p_n)$ contains H with probability $o(1)$;
- (1-statement) if $p_n/q_n \to \infty$ (i.e., $p_n \gg q_n$), then $G(n,p_n)$ contains H with probability $1-o(1)$.

(It is not a priori clear why such a threshold exists in the first place. In fact, threshold always exist for monotone properties, as we will see in the next section.)

Building on our calculations for triangles from previous section, let us consider a more general setup for estimating the variance so that we can be more organized in our calculations.

Setup 4.2.2 (for variance bound with few dependencies)
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. Define

$$\Delta^* := \max_i \sum_{j:j\sim i} \mathbb{P}\left(A_j \,\middle|\, A_i\right).$$

***Remark* 4.2.3.** (a) For many applications with an underlying symmetry between the events, the sum in the definition of Δ^* does not actually depend on i.

(b) In the definition of the dependency graph ($i \sim j$) above, we are only considering pairwise dependence. Later on when we study the Lovász Local Lemma, we will need a strong notion of a dependency graph.

(c) This method is appropriate for a collection of events with few dependencies. It is not appropriate for where there are many weak dependencies (e.g., Section 4.5 on the Hardy–Ramanujan theorem on the number of distinct prime divisors).

We have the bound

$$\operatorname{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \le \mathbb{E}[X_i X_j] = \mathbb{P}[A_i A_j] = \mathbb{P}(A_i)\mathbb{P}(A_j|A_i).$$

(Here $A_i A_j$ is the shorthand for $A_i \wedge A_j$, meaning that both events occur.) Also

$$\operatorname{Cov}[X_i, X_j] = 0 \qquad \text{if } i \neq j \text{ and } i \not\sim j.$$

Thus

$$\begin{aligned}\operatorname{Var} X = \sum_{i,j=1}^{m} \operatorname{Cov}[X_i, X_j] &\le \sum_{i=1}^{m} \mathbb{P}(A_i) + \sum_{i=1}^{m} \mathbb{P}(A_i) \sum_{j: j\sim i} \mathbb{P}(A_j|A_i) \\ &\le \mathbb{E}X + (\mathbb{E}X)\Delta^*.\end{aligned}$$

Recall from Corollary 4.1.8 that $\mathbb{E}X > 0$ and $\operatorname{Var} X = o(\mathbb{E}X)^2$ imply $X > 0$ and $X \sim \mathbb{E}X$ whp. So we have the following.

Lemma 4.2.4

In the above setup, if $\mathbb{E}X \to \infty$ and $\Delta^* = o(\mathbb{E}X)$, then $X > 0$ and $X \sim \mathbb{E}X$ whp.

Let us now determine the threshold for containing K_4.

Theorem 4.2.5

The threshold for containing K_4 is $n^{-2/3}$.

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

$$\mathbb{E}X = \binom{n}{4} p^6 \asymp n^4 p^6.$$

If $p \ll n^{-2/3}$ then $\mathbb{E}X = o(1)$, and thus $X = 0$ whp by Markov's inequality.

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 $6\binom{n-2}{2} = O(n^2)$, and for each such S' one has $\mathbb{P}(A_{S'}|A_S) = p^5$ (as there are 5 additional edges not in the S-clique that need 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

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

Thus $X > 0$ whp by Lemma 4.2.4. □

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* 4.2.6** (First moment is not enough)**.** Let $H =$  . We have $\mathbb{E}X_H \asymp n^5 p^7$. 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 $K_4 \subseteq H$.

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 our proof strategy? In the calculation of Δ^*, one of the terms is $\asymp np$ (from two copies of H with a K_4-overlap), and $np \not\ll n^5 p^7 \asymp \mathbb{E}X_H$ if $p \ll n^{-2/3}$.

The above example shows that the threshold is not always necessarily determined by the expectation. For the property of containing H, the example suggests that we should look at the "densest" subgraph of H rather than containing H itself.

Definition 4.2.7

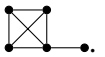
Define the ***edge-vertex ratio*** of a graph H by

$$\rho(H) := \frac{e_H}{v_H}.$$

(This is the same as half the average degree.)

Define the ***maximum edge-vertex ratio of a subgraph*** of H:

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

***Example* 4.2.8.** 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.

***Remark* 4.2.9** (Algorithm)**.** Goldberg (1984) found a polynomial time algorithm for computing $m(H)$ via network flow algorithms.

The next theorem completes determines the threshold for containing some fixed graph H. Basically, it is determined by the expected number of copies of H', where H' is the "denest" subgraph of H (i.e., with the maximum edge-vertex ratio).

Theorem 4.2.10 (Threshold for containing a fixed graph: Bollobás 1981)

Fix a graph H. Then $p = n^{-1/m(H)}$ is a threshold for containing H has a subgraph.

Proof. Let H' be a subgraph of H achieving the maximum edge-vertex ratio, i.e., $\rho(H') = m(H)$. Let X_H denote the number of copies of H in $G(n,p)$.

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,

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

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 considering 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 4.2.4 yields the result. □

***Remark* 4.2.11.** The proof also gives that if $p \gg n^{-1/m(H)}$, then the number X_H of copies of H is concentrated near its mean, i.e., with probability $1 - o(1)$,

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

4.3 Thresholds

Previously, we computed the threshold for containing a fixed H as a subgraph. In this section, we take a detour from the discussion of the second moment method and discuss thresholds in more detail.

We begin by discussing the concept more abstractly by first defining the threshold of any monotone property on subsets. Then we show that thresholds always exist.

Thresholds form a central topic in probabilistic combinatorics. For any given property, it is natural to ask the following questions:

1. Where is the threshold?
2. Is the transition sharp? (And more precisely, what is width of the transition window?)

We understand thresholds well for many basic graph properties, but for many others, it can be a difficult problem. Also, one might think that one must first understand the location of the threshold before determining the nature of the phase transition, but surprisingly this is actually not always the case. There are powerful results that can sometimes show a sharp threshold without identifying the location of the threshold.

Here is some general setup, before specializing to graphs.

Let Ω be some finite set (ground set). Let Ω_p be a random subset of Ω where each element is included with probability p independently.

An ***increasing property***, also called ***monotone property***, on subsets of Ω is some binary property so that if $A \subseteq \Omega$ satisfies the property, any superset of A automatically satisfies the property.

A property is ***trivial*** if all subsets of Ω satisfy the property, or if all subsets of Ω do not satisfy the property. From now on, we only consider non-trivial monotone properties.

A ***graph property*** is a property that only depends on isomorphism classes of graphs. Whether the random graph $G(n,p)$ satisfies a given property can be cast in our setup by viewing $G(n,p)$ as Ω_p with $\Omega = \binom{[n]}{2}$.

Here are some examples of increasing properties for subgraphs of a given set of vertices:

- Contains some given subgraph
- Connected
- Has perfect matching
- Hamiltonian
- non-3-colorable

A family $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ of subsets of Ω is called an ***up-set*** if whenever $A \in \mathcal{F}$ and $A \subseteq B$, then $B \in \mathcal{F}$. Increasing property is the same as being an element of an up-set. We will use these two terms interchangeably.

Definition 4.3.1 (Threshold)
Let $\Omega = \Omega^{(n)}$ be a finite set and $\mathcal{F} = \mathcal{F}^{(n)}$ an monotone property of subsets of Ω. We say that q_n is a ***threshold*** for $\mathcal{F}$ if,

$$\mathbb{P}(\Omega_{p_n} \in \mathcal{F}) \to \begin{cases} 0 & \text{if } p_n/q_n \to 0, \\ 1 & \text{if } p_n/q_n \to \infty. \end{cases}$$

***Remark* 4.3.2.** The above definition is only for increasing properties. We can similarly define the threshold for decreasing properties by an obvious modification. An example of a non-monotone property is sontaining some H as an induced subgraph. Some (but not all) non-monotone properties also have thresholds, though we will not discuss it here.

***Remark* 4.3.3.** From the definition, we see that if r_n and r_n' are both thresholds of the same property, then they must be within a constant factor of each other (exercise: check this). Thus it makes sense to say "the threshold" rather than "a threshold."

Existence of threshold

Question 4.3.4 (Existence of threshold)
Does every non-trivial monotone property have a threshold?

How would a monotone property not have a threshold? Perhaps one could have $\mathbb{P}(\Omega_{1/n} \in \mathcal{F})$ and $\mathbb{P}(\Omega_{(\log n)/n} \in \mathcal{F}) \in [1/10, 9/10]$ for all sufficiently large n?

Before answer this question, let us consider an even more elementary claim.

Theorem 4.3.5 (Monotonicity of satisfying probability)
Let Ω be a finite set and $\mathcal{F}$ a non-trivial monotone property of subsets of Ω. Then $p \mapsto \mathbb{P}(\Omega_p \in \mathcal{F})$ is a strictly increasing function of $p \in [0, 1]$.

Let us give two related proofs of this basic fact. Both are quite instructive. Both are based on ***coupling*** of random processes.

Proof 1. Let $0 \le p < q \le 1$. Consider the following process to generate two random subsets of Ω. For each x, generate uniform $t_x \in [0, 1]$ independently at random. Let

$$A = \{x \in \Omega : t_x \le p\} \quad \text{and} \quad B = \{x \in \Omega : t_x \le q\}.$$

Then A has the same distribution as Ω_p and B has the same distribution as Ω_q. Furthermore, since $p < q$, we always have $A \subseteq B$. Since $\mathcal{F}$ is monotone, $A \in \mathcal{F}$ implies $B \in \mathcal{F}$. Thus

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

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., if all $t_x \in (p, q]$, then $A = \varnothing$ and $B = \Omega$). □

In the second proof, the idea is to reveal a random subset of Ω in independent random stages.

Proof 2. (By two-round exposure) Let $0 \le p < q \le 1$. Note that $B = \Omega_q$ has the same distribution as the union of two independent $A = \Omega_p$ and $A' = \Omega_{p'}$, where p' is chosen to satisfy $1 - q = (1 - p)(1 - p')$ (check that the probability that each element occurs is the same in the two processes). 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).[1]

[1] (Thresholds for random subspaces of $\mathbb{F}_q^n$) The proof of the Bollobás–Thomason paper using the

Theorem 4.3.6 (Existence of thresholds: Bollobás and Thomason 1987)
Every sequence of nontrivial monotone properties has a threshold.

The theorem follows from the next non-asymptotic claim.

Lemma 4.3.7 (Multiple round exposure)
Let Ω be a finite set and $\mathcal{F}$ some non-trivial monotone property. If $p \in [0,1]$ and m is nonnegative integer. Then

$$\mathbb{P}(\Omega_p \notin \mathcal{F}) \le \mathbb{P}(\Omega_{p/m} \notin \mathcal{F})^m.$$

Proof. Consider m independent copies of $\Omega_{p/m}$, and let Y be their union. Since $\mathcal{F}$ is monotone increasing, if $Y \notin \mathcal{F}$, then none of the m copies lie in $\mathcal{F}$. Hence

$$\mathbb{P}(Y \notin \mathcal{F}) \le \mathbb{P}(\Omega_{p/m} \notin \mathcal{F})^m.$$

Note that Y has the same distribution as Ω_q for some $q \le p$. So $\mathbb{P}(\Omega_p \notin \mathcal{F}) \le \mathbb{P}(\Omega_q \notin \mathcal{F}) = \mathbb{P}(Y \notin \mathcal{F})$ by Theorem 4.3.5. Combining the two inequalities gives the result. □

Proof of Theorem 4.3.6. Since $p \mapsto \mathbb{P}(\Omega_p \in \mathcal{F})$ 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), there is some unique "critical probability" p_c so that $\mathbb{P}(\Omega_{p_c} \in \mathcal{F}) = 1/2$.

It remains to check for every $\varepsilon > 0$, there is some $m = m(\varepsilon)$ (not depending on the property) so that

$$\mathbb{P}(\Omega_{p_c/m} \notin \mathcal{F}) \ge 1 - \varepsilon \quad \text{and} \quad \mathbb{P}(\Omega_{mp_c} \notin \mathcal{F}) \le \varepsilon.$$

Kruskal–Katona theorem is still relevant. For example, there is an interesting analog of this concept for properties of subspaces of $\mathbb{F}_q^n$, i.e., random linear codes instead of random graphs. The analogue of the Bollobás–Thomason theorem was proved by Rossman (2020) via the the Kruskal–Katona approach. The multiple round exposure proof does not seem to work in the random subspace setting, as one cannot write a subspace as a union of independent copies of smaller subspaces.

As an aside, I disagree with the use of the term "sharp threshold" in Rossman's paper for describing all thresholds for subspaces—one really should be looking at the cardinality of the subspaces rather than their dimensions. In a related work by Guruswami, Mosheiff, Resch, Silas, and Wootters (2022), they determine thresholds for random linear codes for properties that seem to be analogous to the property that a random graph contains a given fixed subgraph. Here again I disagree with them calling it a "sharp threshold." It is much more like a coarse threshold once you parameterize by the cardinality of the subspace, which gives you a much better analogy to the random graph setting.

Thresholds for random linear codes seems to an interesting topic that has only recently been studied. I think there is more to be done here.

(here we write $\Omega_t = \Omega$ if $t > 1$.) Indeed, applying Lemma 4.3.7 with $p = p_c$, we have

$$\mathbb{P}(\Omega_{p_c/m} \notin \mathcal{F}) \geq \mathbb{P}(\Omega_{p_c} \notin \mathcal{F})^{1/m} = 2^{-1/m} \geq 1 - \varepsilon \quad \text{if } m \geq (\log 2)/\varepsilon.$$

Applying Lemma 4.3.7 again with $p = mp_c$, we have

$$\mathbb{P}(\Omega_{mp_c} \notin \mathcal{F}) \leq \mathbb{P}(\Omega_{p_c} \notin \mathcal{F})^m = 2^{-m} \leq \varepsilon \quad \text{if } m \geq \log_2(1/\varepsilon).$$

Thus p_c is a threshold for $\mathcal{F}$. □

Examples

We will primarily be studying monotone graph properties. In the previous notation, $\Omega = \binom{[n]}{2}$, and we are only considering properties that depend on the isomorphism class of the graph.

***Example* 4.3.8** (Containing a triangle)**.** We saw earlier in the chapter that the threshold for containing a triangle is $1/n$:

$$\mathbb{P}(G(n,p) \text{ contains a triangle}) \to \begin{cases} 0 & \text{if } np \to 0, \\ 1 - e^{-c^3/6} & \text{if } np \to c \in (0,\infty) \\ 1 & \text{if } np \to \infty. \end{cases}$$

In this case, the threshold is determined by the expected number of triangles $\Theta(n^3p^3)$, and whether this quantity goes to zero or infinity (in the latter case, we used a second moment method to show that the number of triangles is positive with high probability).

What if $p = \Theta(1/n)$? If $np \to c$ for some constant $c > 0$, then (you will show in the homework via the method of moments) that the number of triangles is asymptotically Poisson distributed, and in particular the limit probability of containing a triangle increases from 0 to 1 as a continuous function of $c \in (0,\infty)$. So, in particular, as p increases, it goes through a "window of transition" of width $\Theta(1/n)$ in order for $\mathbb{P}(G(n,p)$ contains a triangle) to increase from 0.01 to 0.99. The width of this window is on the same order as the threshold. In this case, we call it a ***coarse transition***.

***Example* 4.3.9** (Containing a subgraph)**.** Theorem 4.2.10 tells us that the threshold for containing a fixed subgraph H is $n^{-1/m(H)}$. Here the threshold is not always determined by the expected number of copies of H. Instead, we need to look at the "densest subgraph" $H' \subseteq H$ with the largest edge-vertex ratio (i.e., equivalent to largest average degree). The threshold is determined by whether the expected number of copies of H' goes to zero or infinity.

Similar to the triangle case, we have a coarse threshold.

The analysis can also be generalized to containing one of several fixed subgraphs $H_1, \dots, H_k$.

***Remark* 4.3.10** (Monotone graph properties are characterized by subgraph containment)**.** Every monotone graph property can be characterized as containing some element of $\mathcal{H}$ for some $\mathcal{H}$ that could depend on the vertex set n. For example, the property of connectivity corresponds to taking $\mathcal{H}$ to be all spanning trees. More generally, one can take $\mathcal{H}$ to be the set of all minimal graphs satisfying the property. When elements of $\mathcal{H}$ are unbounded in size, the problem of thresholds become quite interesting and sometimes difficult.

The original Erdős–Rényi (1959) paper on random graphs already studied several thresholds, such as the next two examples.

***Example* 4.3.11** (No isolated vertices)**.** With $p = \dfrac{\log n + c_n}{n}$,

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

It is a good exercise (and included in the problem set) to check the above claims. The cases $c_n \to -\infty$ and $c_n \to \infty$ can be shown using the second moment method. More precisely, when $c_n \to c$, by comparing moments one can show that the number of isolated vertices is asymptotically Poisson.

In this example, the threshold is at $(\log n)/n$. As we see above, the transition window is $\Theta(1/n)$, much narrower the magnitude of the threshold. In particular, the event probability goes from $o(1)$ to $1 - o(1)$ when p increases from $(1-\delta)(\log n)/n$ to $(1+\delta)(\log n)/n$ for any fixed $\delta > 0$. In this case, we say that the property has a ***sharp threshold*** at $(\log n)/n$ (here the leading constant factor is relevant, unlike the earlier example of a coarse threshold).

***Example* 4.3.12** (Connectivity)**.** With $p = \dfrac{\log n + c_n}{n}$

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

In fact, a much stronger statement is true, connecting the above two examples: consider a process where one adds a 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. Such stronger characterization is called a ***hitting time*** result.

A similar statement is true if we replace "is connected" by "has a perfect matching" (assuming n even).

***Example* 4.3.13** (Perfect matching in a random hypergraph: Shamir's problem)**.** Let $G^{(3)}(n,p)$ be a random hypergraph on n vertices, where each triple of vertices appears as an edge with probability p. Assume that n is divisible by 3. What is the threshold for the existence of a perfect matching (i.e., a set of $n/3$ edges covering all vertices)?

It is easy to check that the property of having no isolated vertices has a sharp threshold at $p = 2n^{-2}\log n$. Is this also a threshold for having a perfect matching? So for smaller p, one cannot have a perfect matching due to having an isolated vertex. What about larger p? This turns out to be a difficult problem known as "Shamir's problem".

A difficult result by Johansson, Kahn, and Vu (2008) (this paper won a Fulkerson Prize) showed that there is some constant $C > 0$ so that if $p \geq Cn^{-2}\log n$ then $G^{(3)}(n,p)$ contains a perfect matching with high probability. They also solved the problem much generally for H-factors in random k-uniform hypergraphs.

Recent exciting breakthroughs on the Kahn–Kalai conjecture (2007) by Frankston, Kahn, Narayanan, and Park (2021) and Park and Pham (2024) provide new and much shorter proofs of this threshold for Shamir's problem.

Recently, Kahn (2022) proved a sharp threshold result, and actually an even stronger hitting time version, of Shamir's problem, showing that with high probability, one has a perfect matching as soon as there are no isolated vertices.

Sharp transition

In some of the 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, 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***.

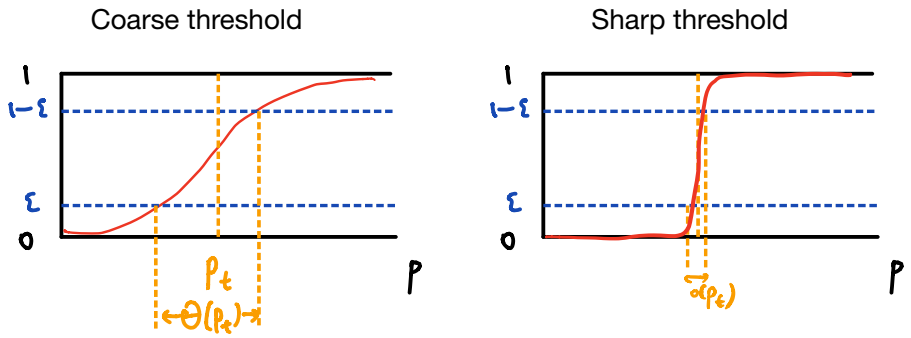


Figure 4.1: Examples of coarse and sharp thresholds. The vertical axis is the probability that $G(n,p)$ satisfies the property.

Definition 4.3.14 (Sharp thresholds)

We say that r_n is a ***sharp threshold*** for some property $\mathcal{F}$ on subsets of Ω if, for every $\delta > 0$,

$$\mathbb{P}(\Omega_{p_n} \in \mathcal{F}) \to \begin{cases} 0 & \text{if } p_n/r_n \le 1-\delta, \\ 1 & \text{if } p_n/r_n \ge 1+\delta. \end{cases}$$

On the other hand, if there is some fixed $\varepsilon > 0$ and $0 < c < C$ so that $\mathbb{P}(\Omega_{p_n} \in \mathcal{F}) \in [\varepsilon, 1-\varepsilon]$ for whenever $c \le p_n/r_n \le C$, then we say that r_n is a ***coarse threshold***.

As in Figure 4.1, the sharp/coarseness of a thresholds is about how quickly $\mathbb{P}(\Omega_p \in \mathcal{F})$ transitions from ε to $1-\varepsilon$ as p increases. How wide is the transition window for p? By the Bollobás–Thomason theorem (Theorem 4.3.6) on the existence of thresholds, this transition window always has width $O(r_n)$. If the transition window has width $\Theta(r_n)$ for some $\varepsilon > 0$, then we have a coarse threshold. On the other hand, if the transition window has width $o(r_n)$ for every $\varepsilon > 0$, then we have a sharp threshold.

From earlier examples, we saw coarse thresholds for the "local" property of containing some given subgraph, as well as 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, completely characterizes when a threshold is coarse versus sharp. We will not state Friedgut's theorem precisely here since it is rather technical (and actually not always easy to apply). Let us just give a flavor. Roughly speaking, the theorem 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 4.3.15 (of Friedgut's sharp threshold theorem)
Suppose $r(n)$ is a coarse threshold of some graph property. Then there is a partition of $\mathbb{N} = \mathbb{N}_1 \cup \cdots \cup \mathbb{N}_k$ and rationals $\alpha_1, \ldots, \alpha_k > 0$ such that $r(n) \asymp n^{-\alpha_j}$ for every $n \in \mathbb{N}_j$.

In particular, if $(\log n)/n$ is a threshold 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 α.

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 4.3.16 (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 4.3.17 (Achlioptas and Friedgut 2000)
For every $k \geq 3$, there exists a function $d_k(n)$ such that for every $\varepsilon > 0$, and sequence $d(n) > 0$,

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

On the other hand, it is not known whether $\lim_{n\to\infty} d_k(n)$ exists, which would imply Conjecture 4.3.16. 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).

4.4 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 4.4.1
What is the clique number of $G(n, 1/2)$?

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

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

Let us first do a rough estimate to see what is the critical k to get $f(n, k)$ large or small. Recall that $\left(\frac{n}{ek}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$. We have

$$\log_2 f(n, k) = k\left(\log_2 n - \log_2 k - \frac{k}{2} + O(1)\right).$$

And so the transition point is at some $k \sim 2\log_2 n$ in the sense that if $k \geq (2+\delta)\log_2 n$, then $f(n, k) \to 0$ while if $k \leq (2-\delta)\log_2 n$, then $f(n, k) \to \infty$.

The next result gives us a lower bound on the typical clique number.

Theorem 4.4.2 (Second moment bound for clique number)
Let $k = k(n)$ be some sequence of positive integers.

(a) If $f(n,k) \to 0$, then $\omega(G(n,1/2)) < k$ whp.

(b) If $f(n,k) \to \infty$, then $\omega(G(n,1/2)) \geq k$ whp.

Proof sketch. The first claim follows from Markov's inequality as $\mathbb{P}(X \geq 1) \leq \mathbb{E}X$.

For the second claim, we bound the variance using Setup 4.2.2. 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. Recall

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

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 4.2.4 that $X > 0$ (i.e., $\omega(G) \geq k$) whp. □

We can a two-point concentration result for the clique number of $G(n,1/2)$. This result is due to Bollobás–Erdős 1976 and Matula 1976.

Theorem 4.4.3 (Two-point concentration for clique number)
There exists a $k = k(n) \sim 2\log_2 n$ such that $\omega(G(n,1/2)) \in \{k, k+1\}$ whp.

Proof. For $k \sim 2\log_2 n$,

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

Let $k_0 = k_0(n) \sim 2\log_2 n$ be the value with

$$f(n,k_0) \geq n^{-1/2} > f(n,k_0+1).$$

Then $f(n,k_0-1) \to \infty$ and $f(n,k_0+1) = o(1)$. By Theorem 4.4.2, the clique number of $G(n,1/2)$ is whp in $\{k_0-1, k_0\}$. □

***Remark* 4.4.4.** By a more careful analysis, one can show that outside a very sparse

subset of integers, one has $f(n, k_0) \to \infty$, in which case one has one-point concentration $\omega(G(n, 1/2)) = k_0$ whp.

By taking the complement of the graph, we also get a two-point concentration result about the independence number of $G(n, 1/2)$. Bohman and Hofstad (2024) extended the two-point concentration result for the independence number of $G(n, p)$ to all $p \geq n^{-2/3+\varepsilon}$.

***Remark* 4.4.5** (Chromatic number)**.** Since the chromatic number satisfies $\chi(G) \geq n/\alpha(G)$, we have

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

In Theorem 8.3.2, using more advanced methods, we will prove $\chi(G(n, 1/2)) \sim n/(2 \log_2 n)$ whp, a theorem due to Bollobás (1987).

In Section 9.3, using martingale concentration, we will show that $\chi(G(n, p))$ is tightly concentrated around its mean without a priori needing to know where the mean is located.

4.5 Hardy–Ramanujan theorem on the number of prime divisors

Let $\nu(n)$ denote the number of distinct primes dividing n (not counting multiplicities).

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

Theorem 4.5.1 (Hardy and Ramanujan 1917)

For every $\varepsilon > 0$, there exists C such that for all sufficiently large n, all but ε-fraction of $x \in [n]$ satisfy

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

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

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

4 *Second Moment*

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

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

We quote Merten's theorem from analytic number theory:

$$\sum_{p \le n} 1/p = \log\log n + O(1).$$

(A more precise result says that $O(1)$ error term converges to the Meissel–Mertens constant.) So

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

Next we compute the variance. The intuition is that divisibility by distinct primes should behave somewhat independently. Indeed, if pq divides n, then X_p and X_q are independent (e.g., by the Chinese Remainder Theorem). If pq does not divide n, but n is large enough, then there is some small covariance contribution. (In contrast to the earlier variance calculations in random graphs, here we have many weak dependices.)

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

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

Thus

$$\sum_{p \ne q} \left|\mathrm{Cov}[X_p, X_q]\right| \lesssim \frac{M^2}{n} \lesssim n^{-4/5}.$$

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

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

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

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

Finally, recall that $|X - \nu| \le 10$. So the same asymptotic bound holds with X replaced by ν. □

The main idea from the above proof is that the number of prime divisors $X = \sum_p X_p$ (from the previous proof) behaves like a sum of independent random variables.

A sum of independent random variables often satisfy a central limit theorem (i.e., asymptotic normality, convergence to Gaussian), assuming some mild regularity hypotheses. In particular, we have the following corollary of the Lindenberg–Feller central limit theorem (see Durrett, Theorem 3.4.10):

Theorem 4.5.2 (Central limit theorem for sums of independent Bernoullis)
If X_n is a sum of independent Bernoulli random variables, and $\operatorname{Var} X_n \to \infty$ as $n \to \infty$, then $(X_n - \mathbb{E}X_n)/\sqrt{\operatorname{Var} X}$ converges to the normal distribution.

(Note that the divergent variance hypothesis is necessary and sufficient.)

In the setting of prime divisibility, we do not have genuine independence. Nevertheless, it is natural to expect that $\nu(x)$ still satisfies a central limit theorem. This is indeed the case, and can be proved by comparing moments against a genuine sum of independent random Bernoulli random variables.

Theorem 4.5.3 (Asymptotic normality: Erdős and Kac 1940)
With $x \in [n]$ uniformly chosen at random, $\nu(x)$ is asymptotically normal, i.e., for every $\lambda \in \mathbb{R}$,

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

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 such proof 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 4.5.4 (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* 4.5.5.** The same conclusion holds for any probability distribution 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. (See Durrett §3.3.5).

Proof of Erdős–Kacs Theorem 4.5.3. 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, $\nu(x) - s(n) \le \nu(x) \le \nu(x)$.

We construct a "model random variable" mimicking X. Let $Y = \sum_{p\le M} Y_p$, where $Y_p \sim \text{Bernoulli}(1/p)$ 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 $\widetilde{Y} = (Y - \mu)/\sigma$.

A consequence of the Lindeberg–Feller central limit theorem is that a sum of independent Bernoulli random variables with divergent variance satisfies the central limit theorem. So $\widetilde{Y} \to N(0,1)$ 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

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

It follows that $\widetilde{X}$ is asymptotically normal. □

4.6 Distinct sums

What is the largest subset of $[n]$ all of whose subsets have distinct sums? Equivalently:

Question 4.6.1
For each k, what is the smallest n so that there exists $S \subseteq [n]$ with $|S| = k$ and all 2^k subset sums of S are distinct?

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

We begin with an easy pigeonhole argument. Since all 2^k sums are distinct and are at most kn, we have $2^k \le kn$. Thus $n \ge 2^k/k$.

Erdős offered $300 for a proof or disproof of the following. It remains open.

Conjecture 4.6.2 (Erdős)
$n \gtrsim 2^k$

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 4.6.3
If there is a k-element subset of $[n]$ with distinct subset sums. Then $n \gtrsim 2^k/\sqrt{k}$.

Proof. Let $S = \{x_1, \dots, x_k\}$ be a k-element subset of $[n]$ with distinct subset sums. Set

$$X = \varepsilon_1 x_1 + \cdots + \varepsilon_k x_k$$

where $\varepsilon_i \in \{0,1\}$ 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's inequality,

$$\mathbb{P}(|X - \mu| \ge 2\sigma) \le \frac{1}{4},$$

and thus

$$\mathbb{P}(|X - \mu| < n\sqrt{k}) = \mathbb{P}(|X - \mu| < 2\sigma) \ge \frac{3}{4}.$$

Since X takes distinct values for every $(\varepsilon_1, \ldots, \varepsilon_k) \in \{0,1\}^k$, we have $\mathbb{P}(X = x) \le 2^{-k}$ for all x. Since there are $\le 2n\sqrt{k}$ elements in the interval $(\mu - n\sqrt{k}, \mu + n\sqrt{k})$, we have

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

Putting the upper and lowers bounds on $\mathbb{P}(|X - \mu| < n\sqrt{k})$ together, we get

$$2n\sqrt{k}2^{-k} \le \frac{3}{4}.$$

So $n \gtrsim 2^k/\sqrt{k}$. □

Dubroff, Fox, and Xu (2021) gave another short proof of this result 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).

Consider the graph representing the n-dimensional boolean cube, with vertex set $\{0,1\}^n$ with an edge between every pair of n-tuples that differ in exactly one coordinate. Given $A \subseteq \{0,1\}^n$, write ∂A for the set of all vertices outside A that is adjacent to some vertex of A.

Theorem 4.6.4 (Vertex-isomperimetric inequality on the hypercube: Harper 1966) Every $A \subseteq \{0,1\}^k$ with $|A| = 2^{k-1}$ has

$$|\partial A| \ge \binom{k}{\lfloor k/2 \rfloor}.$$

***Remark* 4.6.5.** A stronger form of Harper's theorem gives the precise value of

$$\min_{A\subseteq\{0,1\}^n:|A|=m} |\partial A|$$

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 A consists of the lexicographically first m elements of $\{0, 1\}^n$.

Theorem 4.6.6 (Dubroff–Fox–Xu 2021)
If there is a k-element subset of $[n]$ with distinct subset sums, then

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

***Remark* 4.6.7.** The above bound has the currently best known leading constant factor, matching an earlier result by Aliev (2009).

Proof. Let $S = \{x_1, \dots, x_k\}$ be a subset of $[n]$ with distinct sums. Let

$$A = \left\{(\varepsilon_1, \dots, \varepsilon_k) \in \{0, 1\}^k : \varepsilon_1 x_1 + \cdots + \varepsilon_1 x_k < \frac{x_1 + \cdots + x_k}{2}\right\}.$$

Note that due to the distinct sum hypothesis, one can never have $\varepsilon_1 x_1 + \cdots + \varepsilon_k x_k = (x_1 + \cdots + x_n)/2$. It thus follows by symmetry that $|A| = 2^{k-1}$.

Note that every element of ∂A corresponds to some sum of the form $z + x_i > (x_1 + \cdots + x_l)/2$ for some $z < (x_1 + \cdots + x_k)/2$, and thus $z + x_i$ lies 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 $n \geq |\partial A| \geq \binom{k}{\lfloor k/2\rfloor}$ by Harper's theorem (Theorem 4.6.4). □

4.7 Weierstrass approximation theorem

We finish off the chapter with an application to analysis.

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

Theorem 4.7.1 (Weierstrass approximation theorem 1885)
Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function. Let $\varepsilon > 0$. Then there is a polynomial $p(x)$ such that $|p(x) - f(x)| \le \varepsilon$ for all $x \in [0,1]$.

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

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

where

$$E_i(x) = \mathbb{P}(\text{Binomial}(n,x) = i) = \binom{n}{i} x^i (1-x)^{n-i} \quad \text{for } 0 \le i \le n$$

is chosen as some polynomials peaks at $x = i/n$ and then decays as x moves away from i/n.

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

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

(In the next chapter, we will see a much better tail bound.)

Since $[0,1]$ is compact, f is uniformly continuous and bounded. By multiplying by a constant, we assume that $|f(x)| \le 1$ for all $x \in [0,1]$. Also there exists $\delta > 0$ such that $|f(x) - f(y)| \le \varepsilon/2$ for all $x, y \in [0,1]$ with $|x - y| \le \delta$.

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

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

□

Exercises

1. Let X be a nonnegative real-valued random variable. Suppose $\mathbb{P}(X = 0) < 1$. Prove that
$$\mathbb{P}(X = 0) \le \frac{\operatorname{Var} X}{\mathbb{E}[X^2]}.$$

2. Let X be a random variable with mean μ and variance σ^2. Prove that for all $\lambda > 0$,
$$\mathbb{P}(X \geq \mu + \lambda) \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}.$$

3. *Threshold for k-APs.* Let $[n]_p$ denote the random subset of $\{1, \dots, n\}$ where every element is included with probability p independently. For each fixed integer $k \geq 3$, determine the threshold for $[n]_p$ to contain a k-term arithmetic progression.

4. What is the threshold function for $G(n, p)$ to contain a cycle?

5. Show that, for each fixed positive integer k, there is a sequence p_n such that
$$\mathbb{P}(G(n, p_n) \text{ has a connected component with exactly } k \text{ vertices}) \to 1 \quad \text{as } n \to \infty.$$
Hint: Make the random graph contain some specific subgraph but not some others.

6. *Poisson limit.* Let X be the number of triangles in $G(n, c/n)$ for some fixed $c > 0$.
   a) For every nonnegative integer k, determine the limit of $\mathbb{E}\binom{X}{k}$ as $n \to \infty$.
   b) Let $Y \sim \text{Binomial}(n, \lambda/n)$ for some fixed $\lambda > 0$. For every nonnegative integer k, determine the limit of $\mathbb{E}\binom{Y}{k}$ as $n \to \infty$, and show that it agrees with the limit in (a) for some $\lambda = \lambda(c)$.

      We know that Y converges to the Poisson distribution with mean λ. Also, the Poisson distribution is determined by its moments.
   c) Compute, for fixed every nonnegative integer t, the limit of $\mathbb{P}(X = t)$ as $n \to \infty$.

      (In particular, this gives the limit probability that $G(n, c/n)$ contains a triangle, i.e., $\lim_{n\to\infty} \mathbb{P}(X > 0)$. This limit increases from 0 to 1 continuously when c ranges from 0 to $+\infty$, thereby showing that the property of containing a triangle has a coarse threshold.)

7. *Central limit theorem for triangle counts.* Find a real (non-random) sequence a_n so that, letting X be the number of triangles and Y be the number of edges in the random graph $G(n, 1/2)$, one has
$$\operatorname{Var}(X - a_n Y) = o(\operatorname{Var} X).$$
Deduce that X is asymptotically normal, that is, $(X - \mathbb{E}X)/\sqrt{\operatorname{Var} X}$ converges to the normal distribution.
(You can solve for the minimizing a_n directly similar to ordinary least squares linear regression,

or first write the edge indicator variables as $X_{ij} = (1 + Y_{ij})/2$ and then expand. The latter approach likely yields a cleaner computation.)

8. *Isolated vertices.* Let $p_n = (\log n + c_n)/n$.

   a) Show that, as $n \to \infty$,

   $$\mathbb{P}(G(n, p_n) \text{ has no isolated vertices}) \to \begin{cases} 0 & \text{if } c_n \to -\infty, \\ 1 & \text{if } c_n \to \infty. \end{cases}$$

   b) Suppose $c_n \to c \in \mathbb{R}$, compute, with proof, the limit of LHS above as $n \to \infty$, by following the approach in 6.

9. ⋆ Is the threshold for the bipartiteness of a random graph coarse or sharp?
   (You are not allowed to use any theorems that we did not prove in class/notes.)

10. *Triangle packing.* Prove that, with probability approaching 1 as $n \to \infty$, $G(n, n^{-1/2})$ has at least $cn^{3/2}$ edge-disjoint triangles, where $c > 0$ is some constant.

Hint: rephrase as finding a large independent set

11. *Nearly perfect triangle factor.* Prove that, with probability approaching 1 as $n \to \infty$,

    a) $G(n, n^{-2/3})$ has at least $n/100$ vertex-disjoint triangles.

    b) *Simple nibble.* $G(n, Cn^{-2/3})$ has at least $0.33n$ vertex-disjoint triangles, for some constant C.

Hint: view a random graph as the union of several independent random graphs & iterate (a)

12. *Permuted correlation.* Recall that the *correlation* of two non-constant random variables X and Y is defined to be $\operatorname{corr}(X, Y) := \operatorname{Cov}[X, Y]/\sqrt{(\operatorname{Var} X)(\operatorname{Var} Y)}$.

    Let $f, g \in [n] \to \mathbb{R}$ be two non-constant functions. Prove that there exist permutations π and τ of $[n]$ such that, with Z being a uniform random element of $[n]$,

    $$\operatorname{corr}(f(\pi(Z)), g(Z)) - \operatorname{corr}(f(\tau(Z)), g(Z)) \geq \frac{2}{\sqrt{n-1}}.$$

    Furthermore, show that equality can be achieved for even n.

Hint: Compute the variance of the correlation for a random permutation.

13. Let $v_1 = (x_1, y_1), \ldots, v_n = (x_n, y_n) \in \mathbb{Z}^2$ with $|x_i|, |y_i| \leq 2^{n/2}/(100\sqrt{n})$ for all $i \in [n]$. Show that there are two disjoint sets $I, J \subseteq [n]$, not both empty, such that $\sum_{i \in I} v_i = \sum_{j \in J} v_j$.

14. ⋆ Prove that there is an absolute constant $C > 0$ so that the following holds. For every prime p and every $A \subseteq \mathbb{Z}/p\mathbb{Z}$ with $|A| = k$, there exists an integer x so that $\{xa : a \in A\}$ intersects every interval of length at least $Cp/\sqrt{k}$ in $\mathbb{Z}/p\mathbb{Z}$.

15. ⋆ Prove that there is a constant $c > 0$ so that every hyperplane containing the origin in $\mathbb{R}^n$ intersects at least c-fraction of the 2^n closed unit balls centered at $\{-1, 1\}^n$.