Randomness & Computation: CS 271

Professor Alistair Sinclair
Notes by Albert Zhang

Spring 2020

Contents

1 First Moment Method

In many situations where we want to show the existence of an object with some desired property, it may be easier to show that

$$\mathbb{P}[X \text{ has some property}] > 0,$$

which would imply that there exists some point in the probability space which has the property. If, say, our property is something of the form $\{X \geq x\}$, then it also suffices to show

$$\mathbb{E}[X] \geq x,$$

so that at least one sample point must have value $\geq x$. We may also have a sequence of random variables $\{X_n\}$, and we wish to show that the probability of some "bad event" $\mathcal{B}_n$ occurs with probability tending to 0. If X is a nonnegative, discrete/integer-valued random variable, we may apply Markov's inequality to get

$$\mathbb{P}[\mathcal{B}_n] = \mathbb{P}[X_n > x] \leq \frac{\mathbb{E}[X_n]}{x},$$

where we want to show that $\mathbb{E}[X_n]/x \to 0$.

This is known as the *probabilistic method*, and more generally falls into the class of *first moment methods*. Note that we have shown existence without having ever constructed the object explicitly. In many applications, we may want to find an explicit construction–we deal with this in Section 1.6.

1.1 Ramsey Theory

Definition 1.1. The k-th (diagonal) Ramsey number $R_k = R_{k,k}$ is the smallest number n such that any 2-coloring of the edges of the complete graph K_n must contain a monochromatic k-clique.

It has been shown that $R_3 = 6$ and $R_4 = 18$. Surprisingly for R_5, we only know it lies in the interval $[43, 49]$. In general, for larger Ramsey numbers, we only have rather course bounds. In fact, there's a comical quote by Erdős saying that if aliens were to threaten to invade earth unless we solved R_5, we should marshal the world's resources towards computing it. However, if it were R_6, we should instead marshal the world's resources towards a preemptive military attack.

Theorem 1.2

By the probabilistic method, we may show the following lower bound:

$$R_k > 2^{k/2}.$$

Proof. It suffices to show that for $n = 2^{k/2}$, there exists a 2-coloring which does not contain a monochromatic k-clique. Consider the model $G \sim \mathcal{G}(n, p)$, where we take

$p = 1/2$. Then given any k-clique C in G, we have

$$\mathbb{P}[C \text{ is monochromatic}] = 2 \cdot \left(\frac{1}{2}\right)^{\binom{k}{2}}.$$

Therefore, since the total number of k-cliques in G is $\binom{n}{k}$, we get by a union bound

$$\begin{aligned}
\mathbb{P}[G \text{ has a monochromatic } k\text{-clique}] &\leq \binom{n}{k} 2^{1-\binom{k}{2}} \\
&\leq \frac{n^k}{k!} \cdot 2^{1-\binom{k}{2}} \\
&= \frac{2^{\frac{k^2}{2}+1-\frac{k^2-k}{2}}}{k!} \\
&= \frac{2^{1+\frac{k}{2}}}{k!} \\
&< 1,
\end{aligned}$$

for $k \geq 3$. Thus there must exist a point in the probability space which has no monochromatic k-cliques, given $n = 2^{k/2}$. □

It turns out the this lower bound is essentially the best known, in the sense that no bounds of the form $R_k \geq 2^{(1/2+\epsilon)k}$ or $R_k \leq 2^{(2-\epsilon)k}$ have been found.

1.2 Max Cut

Recall the Min-Cut problem, which can be solved as the dual to the Max Flow problem efficiently. On the other hand, we have the NP-hard Max Cut problem, in which we want to find the partition such that the number of cut edges is maximimzed.

Lemma 1.3

Given a graph $G = (V, E)$, there exists a cut containing at least $|E|/2$ edges.

Proof. Let $V_1 \cup V_2 = V$ be our partition. Assign each vertex to V_1 and V_2 with probability $1/2$. Define the random variable $X = \sum_{e \in E} X_e$ as the sum of indicators X_e determining whether the edge e is in the cut or not. Then

$$\mathbb{E}[X] = \sum_{e \in E} \mathbb{E}[X_e] = \frac{|E|}{2}.$$

Therefore, there must exist a partition such that the number of edges crossing the cut $X \geq |E|/2$. □

1.3 Independent Set

Given a graph $G = (V, E)$, a subset $U \subset V$ is said to be an *independent set* if no two vertices $u_1, u_2 \in U$ are adjacent in G. The problem of determining the size of the largest independent set is NP-hard. However, we can achieve a good lower bound.

Theorem 1.4

Given a graph $G = (V, E)$, the size of the largest independent set V' is at least

$$|V'| \geq \sum_{v \in V} \frac{1}{\deg(v) + 1}.$$

Proof. To each vertex v, assign a weight $w_v \sim \text{Unif}([0, 1])$. Call v a *local minimum* if $w_v < w_u$ for all neighbors u of v. Then clearly no two adjacent vertices can both be local minima (or at least, such an event has measure zero). Therefore the set of local minima forms an independent set. Furthermore, for each vertex v, we have

$$\mathbb{P}[v \text{ is a local minimum}] = \frac{1}{\deg(v) + 1},$$

so by linearity, we get

$$\mathbb{E}[X] = \sum_{v \in V} \mathbb{E}[X_v] = \sum_{v \in V} \frac{1}{\deg(v) + 1}.$$

Hence there must exist an independent set at least this size. □

1.4 Graph Crossing Number

Given a graph $G = (V, E)$, with $n = |V|$ and $m = |E|$, define the *crossing number* $c(G)$ as the minimum number of edge crossings in any planar embedding of G. So, a graph is planar if and only if $c(G) = 0$.

Note that by Euler's formula, if a graph is planar then

$$m \leq 3n = 6.$$

And so if a graph can be embedded in the plane without crossing edges, it must necessarily be quite sparse.

Lemma 1.5

For any graph G with n vertices and m edges, we have

$$c(G) \geq m - 3n + 6,$$

which generalizes Euler's formula.

Proof. The proof is purely deterministic. Consider the optimal embedding of G that achieves $c = c(G)$ edge crossings. Then this embedding must satisfy

1. No edge crosses itself.
2. No two edges cross more than once.
3. No two edges which share a vertex cross.

Now, construct a new graph $G' = (V', E')$ from G by inserting a vertex at each edge crossing. Note that the resulting graph is planar, so must satisfy Euler's formula. In particular, we have

$$\begin{aligned} m' &\leq 3n' - 6 \\ m + 2c &\leq 3(n + c) - 6 \\ c &\geq m - 3n + 6, \end{aligned}$$

where we have used the substitutions $m' = m + 2c$ (each edge crossing creates 2 new edges), and $n' = n + c$ (each edge crossing inserts 1 new vertex). □

The above result turns out to be reasonably tight for sparse graphs, where m is not much larger than $3n$. For denser graphs, where m is larger, we have the stronger lower bound using the probabilistic method:

Theorem 1.6

For any graph G with n vertices and m edges, where $m \geq 4n$, we have

$$c(G) \geq \frac{m^3}{64n^2}.$$

Proof. Consider an optimal planar embedding of G with $c = c(G)$ edge crossings. Now generate a random induced subgraph G_p of G by keeping each vertex with probability p, and keeping each edge both of whose endpoints are kept in G_p. The value of p will be optimized over later.

Denote by c_p, n_p, and m_p the respective quantities of G_p corresponding to those of G. Then by Lemma 1.5, we have

$$c_p \geq m_p - 3n_p + 6.$$

Taking expectations, we get

$$\mathbb{E}[c_p] \geq \mathbb{E}[m_p - 3n_p + 6] \geq \mathbb{E}[m_p] - 3\mathbb{E}[n_p].$$

Now, each crossing survives with probability p^4, each edge survives with probability p^2, and each vertex survives with probability p. Therefore the inequality becomes

$$cp^4 \geq mp^2 - 3np.$$

From this we get

$$c \geq \frac{m}{p^2} - \frac{3n}{p^3}.$$

Optimizing over p, we will set $p = 4n/m$ to yield

$$c \geq \frac{m^3}{64n^2},$$

as desired. □

1.5 Sample & Modify

In previous examples, we simply constructed a random object and computed first moments to show that some property exists in the sample space. We now provide two more sophisticated examples where we start with a randomized construction, and supplement it with a deterministic modification to ensure existence of the desired property.

1.5.1 Unbalancing Lights

In this example, we consider a square $n \times n$ array of lights, and a set of row and column switches. The n row switches each toggle the lights in one of the rows, and similarly for the column switches.

Note that naively, we can flip all the switches independently and u.a.r. Then each light will be on with probability $1/2$, and the states will all be pairwise independent. In particular, $\mathbb{E}[X] = \frac{n^2}{2}$, and $\text{Var}(X) = \frac{n^2}{4}$, so the difference $|\#\text{on} - \#\text{off}|$ would be $\Omega(n)$. Thus there exists a setting of the switches which achieves $\frac{n^2}{2} + \Omega(n)$ lights on. We will now show using the probabilistic method, along with a deterministic modification, that we can do better.

Theorem 1.7

For any initial configuration of the lights, there exists a setting of the switches such that the number of lights on X is asymptotically $\Omega\left(\frac{n^2}{2} + \sqrt{\frac{1}{2\pi}} \cdot n^{3/2}\right)$ as $n \to \infty$.

Proof. First set the column switches randomly and independently. Define the indicator X_{ij} to be 1 if light (i, j) is on and -1 if off. Define, for row i, the variable $Z_i = \sum_j X_{ij}$. Due to our random flipping of the columns, the lights in a given row are i.i.d., so that by a CLT type result, we have

$$\mathbb{E}\left[|Z_i|\right] \sim \sqrt{\frac{2n}{\pi}}.$$

Now, we deterministically flip each row so as to get the majority of lights on. By linearity, we have

$$\mathbb{E}\left[\#\text{on} - \#\text{off}\right] = \sum_{i=1}^{n} \mathbb{E}\left[|Z_i|\right] \sim \frac{2}{\pi} \cdot n^{3/2}.$$

Then there must exist some setting of switches which achieves this difference, so that as $n \to \infty$, the number of lights on will be asymptotically

$$\frac{n^2}{2} + \sqrt{\frac{1}{2\pi}} \cdot n^{3/2}.$$

□

1.5.2 Large Girth & Chromatic Number

Given a graph $G = (V, E)$, we define the *girth* of G to be the length of the shortest cycle in G, and the *chromatic number* of G to be the smallest number of colors needed to color the graph so that no two adjacent vertices are of the same color. Intuitively, it makes sense that girth and chromatic number of inversely related, and that graphs with large girth should have small chromatic number. However, the following result by Erdős says this is not the case.

Theorem 1.8

For any positive integers k and l, there exists a graph with girth $\geq l$ and chromatic number $\geq k$.

Proof. Consider the model $G \sim \mathcal{G}(n, p)$. We will pick $p = n^{\frac{1}{l}-1}$, for reasons we will see later.

Denote by X the number of cycles of length less than l in G. Then we have

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=3}^{l-1} \binom{n}{3} \cdot \frac{i!}{2i} \cdot p^i \\
&\leq \sum_{i=3}^{l-1} \frac{n^{i/l}}{2i} \\
&= o(n),
\end{aligned}$$

where the first line follows since $\binom{n}{i} \cdot \frac{i!}{2i}$ is the number of possible cycles of length i, and the second line follows from plugging in our choice of p. Therefore, by Markov, we get

$$\mathbb{P}\left[X \geq n/2\right] = o(1).$$

Now, for the chromatic number, note that

$$\text{chromatic } \# \geq \frac{|V|}{|\text{max independent set}|},$$

since the set of vertices that receive any given color is an independent set. Let Y be the size of the maximal independent set. Then by union bound,

$$\begin{aligned}
\mathbb{P}\left[Y \geq y\right] &\leq \binom{n}{y}(1-p)^{\binom{y}{2}} \\
&\leq \left(ne^{-p(y-1)/2}\right)^y,
\end{aligned}$$

which is $o(1)$ by setting $y = \frac{3}{p}\ln n$. Note that we used the inequalities $\binom{n}{y} \leq n^y$ and $1 + x \leq e^x$.

Together, these results say that we can take n large enough so that both $\mathbb{P}[X \geq n/2]$ and $\mathbb{P}[Y \geq \frac{3}{p}\ln n]$ are less than $1/2$. So by union bound, there exists a graph G with at most $n/2$ cycles of length $< l$, and containing a max independent set of size $< \frac{3}{p}\ln n$. Now, modify G by removing one vertex from each cycle of length at most l, and we get a graph G' satisfying

1. G' has girth $\geq l$,
2. G' has $\geq n/2$ vertices,
3. G' has chromatic number $\geq k$,

for n large enough. □

1.6 Construction

So far, the probabilistic method has only allowed us to prove the existence of an object, without giving us the object itself. In this section, we go over a simple example, discuss how to make the method algorithmic, and then how to derandomize it.

1.6.1 MAX3SAT

In the MAX3SAT problem we are given a boolean formula φ in 3CNF (conjunctive normal form) on variables $\{x_i\}_{1\leq i\leq n}$ and clauses $\{C_i\}_{1\leq i\leq m}$. We want to find the maximum number of clauses that can be satisfied with any assignment of T/F to the variables. This is an NP-hard optimization problem.

Theorem 1.9

For any formula φ, there exists an assignment satisfying at least $\frac{7m}{8}$ clauses.

Proof. Assign T/F to each variable with probability $1/2$ independently. Let X be the number of satisfied clauses in a random assignment. Then a simple argument by indicators (one for each clause C_i) gives

$$\mathbb{E}[X] = \sum_{i=1}^{m} \mathbb{E}[X_i] = \sum_{i=1}^{m} \frac{7}{8} = \frac{7m}{8}.$$

Since there must exist a point in the sample space achieving this, we are done. □

1.6.2 Monte Carlo Approach

Naively, we can directly apply the randomized construction by simply picking a random assignment and keep resampling until it satisfies a sufficient threshold of clauses. To analyze the behavior, we use Markov's inequality.

Lemma 1.10

Let X be the random variable from the proof of Theorem 1.9, i.e. the number of satisfied clauses in a random assignment. Then

$$\mathbb{P}\left[X \geq \frac{7}{8}m\right] \geq \frac{1}{m+1}.$$

Proof. First, we apply Markov to the random variable $m - X$ to get

$$\mathbb{P}\left[X \leq \left(1 - \frac{\alpha}{8}\right)m\right] = \mathbb{P}\left[m - X \geq \frac{\alpha}{8}m\right] \leq \frac{m - \mathbb{E}[X]}{\frac{\alpha}{8}m} = \frac{1}{\alpha}.$$

Now, let $\alpha = 1 + \frac{1}{m}$, which gives us

$$\begin{aligned}
\mathbb{P}\left[X < \frac{7}{8}m\right] &= \mathbb{P}\left[X < \left\lfloor\frac{7}{8}m\right\rfloor\right] \\
&= \mathbb{P}\left[X \leq \frac{7}{8}m - \frac{1}{8}\right] \\
&= \mathbb{P}\left[X \leq \left(1 - \frac{\alpha}{8}\right)m\right] \\
&\leq \frac{1}{\alpha} \\
&= \frac{m}{m+1}.
\end{aligned}$$

Note that we've used the crucial fact that X is integer-valued. Thus we have

$$\mathbb{P}\left[X \geq \frac{7}{8}m\right] \geq 1 - \frac{m}{m+1} = \frac{1}{m+1}.$$

□

Theorem 1.11

We can find an assignment satisfying at least $\frac{7}{8}m$ clauses in polynomial time with high probability.

Proof. By the lemma, we see that the Bernoulli random variable Z for producing an assignment satisfying at least $\frac{7}{8}m$ clauses stochastically dominates a geometric random variable with parameter $\frac{1}{m+1}$. Therefore in polynomial time we can achieve the expectation $\frac{7}{8}m$ with high probability. □

1.6.3 Method of Conditional Probabilities

Depending on the situation, we may be able to derandomize the random construction used in the probabilistic method, achieving the expected value or object with desired property deterministically.

For example, let's think of our random assignment of the variables $\{x_i\}$ of our 3CNF formula φ in a sequential fashion. First pick a T/F value for x_1, then for x_2, and so on. This process can be illustrated as a tree. We label each node of the tree with a formula Ψ, and denote by X_Ψ the number of clauses that are satisfed in the tree below Ψ given the fixed assignments of the variables above that node. So for instance, the root is just φ, with no variables fixed yet. The random variable X_0 is just X. The second level of the tree we have two nodes $\Psi_1 = \phi|_{x_1=T}$ and $\Psi_2 = \phi|_{x_1=F}$. The random variable X_1 counts the number of clauses that will be satisfied with a random assignment of variables $\{x_2, x_3, \ldots, x_n\}$, and likewise for X_2.

Note that we have

$$\mathbb{E}[X_\Psi] = \mathbb{P}[x_{i+1} = T] \cdot \mathbb{E}[X_{\Psi|_{x_{i+1}=T}}] + \mathbb{P}[x_{i+1} = F] \cdot \mathbb{E}[X_{\Psi|_{x_{i+1}=F}}] = \frac{1}{2}(\mathbb{E}[X_{\Psi_1}] + \mathbb{E}[X_{\Psi_2}])$$

where Ψ_1 and Ψ_2 are the children of Ψ here. Then at least one child must have expectation at least as large as $\frac{7}{8}m$. Since at the root, we started with $\mathbb{E}[X_\varphi] \geq \frac{7}{8}m$, there will be a leaf node with expectation at least $\frac{7}{8}m$. Furthermore, given a fixed assignment to some subset of the variables, we can explicitly compute $\mathbb{E}[X_\Psi]$, so that we can traverse down the tree in linear time to find the desired assignment.

This method will work whenever we can sequentially index our random choices and we have the ability to compute the conditional expectations when some of the random choices have already been made. Or we can compute the expectations approximately and proceed as before to obtain a final result that approximates the desired expectation.

2 Second Moment Method

In first moment method scenarios, we may be given a sequence of random variables $\{X_n\}$, and we wish to show that in the limit, the probability of some property goes to 0. It is much the same for the second moment method, although we are now given the ability to compute second moments as well. Intuitively speaking, second moments are usually harder to compute than first moments, so naively we should always attempt to use first moment bounds. But, there will be situations where these bounds are too weak, and we will necessarily have to turn to higher moments.

To see that in general, having access to higher moments gives us more power, consider the following example.

Example 2.1. Define the collection of random variables

$$X_n = \begin{cases} n^2 & \text{w.p. } 1/n \\ 0 & \text{o.w.} \end{cases}$$

Then $\mathbb{E}[X_n] = n \to \infty$, however $\mathbb{P}[X_n > 0] \to 0$. Thus any first moment techniques will fail here.

First, we list some of the basic tools of second moment methods. Recall the classical inequality:

Theorem 2.2 (Chebyshev's Inequality)
Let X be any random variable. Then

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \alpha\right] \leq \frac{\mathrm{Var}(X)}{\alpha^2}.$$

As immediate corollaries, we get

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \beta\mathbb{E}[X]\right] \leq \frac{\mathrm{Var}(X)}{\beta^2\mathbb{E}[X]^2}, \tag{1}$$

as well as

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \beta\sigma\right] \leq \frac{1}{\beta^2}, \tag{2}$$

where $\sigma = \sqrt{\mathrm{Var}(X)}$ is the standard deviation of X.

In many applications, we will set $\beta = 1$ in equation 1 to obtain:

Lemma 2.3
For a nonnegative, discrete random variable X,

$$1 - \mathbb{P}[X > 0] = \mathbb{P}[X = 0] \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2}.$$

In some situations, the vanilla inequality above might not be enough. As such, the following application of Cauchy-Schwarz provides an improved variant of the second moment method:

Theorem 2.4 (Paley-Zygmund Inequality)
Let X be a nonnegative random variable. For $0 < \theta < 1$,

$$\mathbb{P}\left[X \geq \theta \mathbb{E}[X]\right] \geq (1-\theta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}. \tag{3}$$

Proof. We have

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X\mathbf{1}_{X<\theta\mathbb{E}[X]}] + \mathbb{E}[X\mathbf{1}_{X\geq\theta\mathbb{E}[X]}] \\ &\leq \theta\mathbb{E}[X] + \sqrt{\mathbb{E}[X^2]\mathbb{P}[X \geq \theta\mathbb{E}[X]]},\end{aligned}$$

where in the second line we have used Cauchy-Schwarz to obtain the second term. Rearranging the inequality gives the result. □

As a corollary, we obtain another variant of the second moment method:

Lemma 2.5
Let X be a nonnegative random variable that is not identically 0. Then

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

Proof. Take $\theta \downarrow 0$ in the Paley-Zygmund inequality. By monotone or dominated convergence of the indicators $\mathbf{1}_{X\geq\theta\mathbb{E}[X]} \uparrow \mathbf{1}_{X>0}$, we see that

$$\mathbb{P}\left[X > 0\right] = \mathbb{E}[\mathbf{1}_{X>0}] \geq \lim_{\theta\to 0}(1-\theta)^2\frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} = \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

□

Note that since

$$\frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} = 1 - \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2 + \mathrm{Var}(X)},$$

compared to the vanilla second moment 2.3,

$$\mathbb{P}\left[X > 0\right] \geq 1 - \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} \leq 1 - \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2 + \mathrm{Var}(X)},$$

the one deduced from Paley-Zygmund in 2.5 is indeed stronger.

2.1 Thresholds in Random Graphs

Recall the $\mathcal{G}_{n,p}$ model where we sample a graph G of n vertices where each edge is included with probability p. We are concerned with questions such as

- Is $G \in \mathcal{G}_{n,p}$ connected?
- Does $G \in \mathcal{G}_{n,p}$ contain a Hamilton cycle?
- Does $G \in \mathcal{G}_{n,p}$ contain a 4-clique?

It turns out that for properties such as these, there exists a "point" where the answer to these questions transitions from yes to no (or vice versa) as we cross this point. More concretely, we call $p(n)$ a threshold for a property Q if as $n \to \infty$,

$$p \ll p(n) \Rightarrow \mathbb{P}\left[G \in \mathcal{G}_{n,p} \text{ has } Q\right] \to 0,$$
$$p \gg p(n) \Rightarrow \mathbb{P}\left[G \in \mathcal{G}_{n,p} \text{ has } Q\right] \to 1.$$

In this section, we will answer the third question. Let X denote the number of 4-cliques in G. For each subset C of 4 vertices in G, define the indicator X_C. Then we have

$$\mathbb{E}[X] = \sum_C \mathbb{E}[X_C] = \binom{n}{4} p^6 = \Theta(n^4 p^6).$$

Therefore, we see that

- If $p \ll n^{-2/3}$, then $\mathbb{E}[X] \to 0$.
- If $p \gg n^{-2/3}$, then $\mathbb{E}[X] \to \infty$.

Based on this observation, we guess that $p(n) = n^{-2/3}$ is the threshold for 4-cliques. Indeed, using second moment methods, we have the following result:

Theorem 2.6
The value $p(n) = n^{-2/3}$ is a threshold for G containing a 4-clique.

Proof. Let X and X_C be defined as above. The first direction follows easily from Markov. In particular, since X is integer-valued, we have

$$\mathbb{P}\left[X > 0\right] = \mathbb{P}\left[X \geq 1\right] \leq \mathbb{E}[X] \to 0$$

for $p \ll n^{-2/3}$.

For the other direction, note that $\mathbb{E}[X] \to \infty$ is not enough so show that

$$\mathbb{P}\left[G \in \mathcal{G}_{n,p} \text{ has a 4-clique}\right] \to 1,$$

since we could have $X = 0$ half the time, and X growing with n the other half of the time. Therefore we look to apply Lemma 2.3.

First, we compute

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(\sum_C X_C) \\ &= \sum_C \mathrm{Var}(X_C) + \sum_{C \neq D} \mathrm{Cov}(X_C, X_D).\end{aligned}$$

The first term is a sum over $\binom{n}{4}$ Bernoulli random variables, so we have

$$\sum_C \mathrm{Var}(X_C) = \binom{n}{4}(p^6 - p^{12}) = O(n^4 p^6).$$

The second term requires some casework.

- Case 1: $|C \cap D| \leq 1$. In this case X_C and X_D are independent, so $\mathrm{Cov}(X_C, X_D) = 0$.
- Case 2: $|C \cap D| = 2$. In this case we compute

$$\begin{aligned}\mathrm{Cov}(X_C, X_D) &\leq \mathbb{E}[X_C X_D] \\ &= \mathbb{P}\left[C,\ D \text{ are both cliques given } |C \cap D| = 2\right] \\ &= p^{11}.\end{aligned}$$

  Since there are $\binom{n}{6}\binom{6}{2}$ such pairs (C, D), the total contribution of this case is $O(n^6 p^{11})$.
- Case 3: $|C \cap D| = 3$. Here $\mathrm{Cov}(X_C, X_D) \leq p^9$, and there are $\binom{n}{5}\binom{5}{2}$ such pairs. Thus the total contribution of this case is $O(n^5 p^9)$.

Altogether, we get

$$\mathrm{Var}(X) = O(n^4 p^6) + O(n^6 p^{11}) + O(n^5 p^9).$$

Using the prior computation that $\mathbb{E}[X] = \Theta(n^4 p^6)$, we apply Lemma 2.3 to get

$$\mathbb{P}\left[X = 0\right] \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = O\left(\frac{1}{n^4 p^6}\right) + O\left(\frac{1}{n^2 p}\right) + O\left(\frac{1}{n^3 p^3}\right),$$

which vanishes to 0 as $n \to \infty$ assuming $p \gg n^{-2/3}$. Thus the probability of G having a 4-clique tends to 1, and this concludes the proof of the theorem. □

Remark. It's possible to generalize the above proof for containment of general k-cliques. In fact, it turns out that we can generalize it to any subgraph H that is *balanced*. We call H balanced if the average degree of H is greater than or equal to the average degree of any induced subgraph of H. In particular, if this is the case, then we would expect the threshold to be $p = n^{-v/e}$, where v and e are the number of vertices and edges of H respectively.

2.2 Clique Number of Random Graphs

Given a graph G, we are concerned with its clique number, the size of a largest clique in G. Finding the clique number is NP-hard. However, if we are given a random graph $G \in \mathcal{G}_{n,p}$, then the clique number is known asymptotically.

Theorem 2.7

For $G \in \mathcal{G}_{n,p}$ and any constant $p \in (0,1)$, the clique number of G is close to $2\log_{1/p}$ with probability tending to zero (the meaning of "close to" will be clarified in the proof).

Proof. For simplicity, restrict to the case where $p = 1/2$. Define X_k to be the number of k-cliques in a graph sampled from $\mathcal{G}_{n,p}$. Let $k_0(n)$ be the largest value of k such that $g(k) := \mathbb{E}[X_k] = \binom{n}{k}2^{-\binom{k}{2}} \geq 1$. A calculation shows that $k_0(n) \sim 2\log n$. We will show that for any integer constant c:

1. For $k_1(n) = k_0(n) + c$, $\mathbb{P}\left[X_{k_1(n)} > 0\right] \to 0$ as $n \to \infty$.
2. For $k_2(n) = k_0(n) - c$, $\mathbb{P}\left[X_{k_2(n)} > 0\right] \to 1$ as $n \to \infty$.

Now, to get the behavior of $\mathbb{E}[X_k]$ around $k_0(n) \sim 2\log n$, we observe that

$$\frac{g(k+1)}{g(k)} = \frac{n-k}{k+1} \cdot 2^{-k} \sim \frac{n}{2\log n} \cdot n^{-2} \to 0, \quad n \to \infty,$$

for $k = k_0$. A similar computation shows that the ratio $g(k-1)/g(k)$ goes to ∞. Therefore, in any c-neighborhood of $k_0(n)$, the graph of $g(k)$ decreases sharply as $n \to \infty$. We deduce the following first moment behaviors:

- $\mathbb{E}[X_{k_1(n)}] \to 0$ as $n \to \infty$.
- $\mathbb{E}[X_{k_2(n)}] \to \infty$ as $n \to \infty$.

Claim (1) follows by Markov, since

$$\mathbb{P}\left[X_{k_1(n)} > 0\right] = \mathbb{P}\left[X_{k_1(n)} \geq 1\right] \leq \mathbb{E}[X_{k_1(n)}] \to 0, \quad n \to \infty.$$

For similar reasons as in the previous section, claim (2) requires the second moment method. In particular, by Lemma 2.3,

$$\mathbb{P}\left[X_{k_2(n)} = 0\right] \leq \mathbb{P}\left[|X_{k_2(n)} - \mathbb{E}[X_{k_2(n)}]| \geq \mathbb{E}[X_{k_2(n)}]\right] \leq \frac{\mathrm{Var}(X_{k_2(n)})}{\mathbb{E}[X_{k_2(n)}]^2}.$$

So, it suffices to show that $\frac{\mathrm{Var}(X_{k_2(n)})}{\mathbb{E}[X_{k_2(n)}]^2} \to 0$ as $n \to \infty$. To ease the notation, from now on we write X for $X_{k_2(n)}$, and for every subset S of the vertex set of size $k_2(n)$, we

define the indicator X_S, so that $X = \sum_S X_S$. Also write $S \sim T$ if X_S and X_T are not independent–this happens whenever $S \neq T$ and $|S \cap T| \geq 2$. Then we have

$$\begin{aligned}\mathrm{Var}(X) &= \sum_S \mathrm{Var}(X_S) + \sum_{S\sim T} \mathrm{Cov}(X_S, X_T) \\ &\leq \sum_S \mathbb{E}[X_S^2] + \sum_{S\sim T} \mathbb{E}[X_S X_T] \\ &= \sum_S \mathbb{E}[X_S] + \sum_{S\sim T} \mathbb{E}[X_S X_T] \\ &= \mathbb{E}[X] + \sum_{S\sim T} \mathbb{E}[X_S X_T].\end{aligned}$$

Therefore, we have

$$\frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} \leq \frac{1}{\mathbb{E}[X]} + \frac{1}{\mathbb{E}[X]^2} \sum_{S\sim T} \mathbb{E}[X_S X_T],$$

so that it suffices to show

$$\sum_{S\sim T} \mathbb{E}[X_S X_T] = o(\mathbb{E}[X]^2).$$

We compute

$$\begin{aligned}\sum_{S\sim T} \mathbb{E}[X_S X_T] &= \sum_{S\sim T} \mathbb{P}[X_S = 1, X_T = 1] \\ &= \sum_{S\sim T} \mathbb{P}[X_S = 1] \cdot \mathbb{P}[X_T = 1 | X_S = 1] \\ &= \sum_S \mathbb{P}[X_S = 1] \sum_{T:T\sim S} \mathbb{P}[X_T = 1 | X_S = 1] \\ &= \left(\sum_S \mathbb{P}[X_S = 1]\right) \left(\sum_{T:T\sim S_0} \mathbb{P}[X_T = 1 | X_{S_0} = 1]\right) \\ &= \mathbb{E}[X] \sum_{T:T\sim S_0} \mathbb{P}[X_T = 1 | X_{S_0} = 1]\end{aligned}$$

where we have fixed a S_0 by symmetry. Now, after some counting arguments, we get

$$\begin{aligned}\frac{\sum_{T:T\sim S_0} \mathbb{P}[X_T = 1 | X_{S_0} = 1]}{\mathbb{E}[X]} &= \frac{\sum_{i=2}^{k_2-1} \binom{k_2}{i}\binom{n-k_2}{k_2-i} 2^{-\left[\binom{k_2}{2} - \binom{i}{2}\right]}}{\binom{n}{k_2} 2^{-\binom{k_2}{2}}} \\ &= \sum_{i=2}^{k_2-1} \frac{\binom{k_2}{i}\binom{n-k_2}{k_2-i} 2^{\binom{i}{2}}}{\binom{n}{k_2}} \\ &= \sum_{i=2}^{k_2-1} f(i) \\ &\leq k_2 \cdot \max_{2\leq i\leq k_2-1} f(i).\end{aligned}$$

It can be shown (through some nasty analysis) that $f(i)$ is maximized at $i = 2$, so that

$$k_2 f(2) \sim \frac{k_2^5}{n^2} \sim \frac{(2\log n)^5}{n^2} \to 0, \qquad \text{as } n \to \infty.$$

Thus we have shown claim (2), so this concludes the proof of the theorem. □

2.3 Pairwise Independence

Definition 2.8. A collection $\{X_i\}_{i=1}^n$ of discrete random variables over the same probability space is said to be k-wise independent if for every subset $I \subseteq \{1, \ldots, n\}$ with $|I| \leq k$, and for every set of values $\{a_i\}_{i=1}^n$, we have

$$\mathbb{P}\left[\bigcap_{i \in I} X_i = a_i\right] = \prod_{i \in I} \mathbb{P}[X_i = a_i].$$

We say the collection is mutually independent if they are n-wise independent.

There are many examples of random variables that are pairwise independent but not mutually (or k-wise, for $k > 2$) independent. It could be instructive to try and construct such an example.

Although pairwise independence is a weaker condition than mutual independence, in many applications pairwise independence is good enough for applying second moment methods. The main benefit is computational–it's possible to represent pairwise independence more compactly than mutual independence. Intuitively, this is because there is less randomness, and constructing randomness is costly.

Suppose we have a Monte Carlo algorithm $\mathcal{A}$ with one-sided error probability $\leq 1/2$ (it is always correct on 'yes', but wrong on 'no' with probability at most $1/2$). Then we can achieve an error probability of $\leq 2^{-t}$ if we use t independent trials. Assuming $\mathcal{A}$ requires m random bits, this implies we'll need $m \log r$ random bits to achieve an error probability of $1/r$. But we can do better.

Theorem 2.9

For any $r \leq 2^m$, we can achieve error probability $\leq 1/r$ using only $2m$ random bits, and runtime $O(rm)$.

Proof. Since $\mathcal{A}$ requires m random bits, we can represent the possible executions of $\mathcal{A}$ with bit strings from $\{0,1\}^m$. Then pick $r < 2^m$ pairwise independent uniform samples from $\{0,1\}^m$, and let X_i be the outcome of the algorithm on the i^{th} sample:

$$X_i = \begin{cases} 1 & \text{if } \mathcal{A} \text{ outputs yes on } i^{\text{th}} \text{ sample,} \\ 0 & \text{otherwise.} \end{cases}$$

Now, note that since X_i are pairwise independent, we have

$$\operatorname{Var}(X) = \sum_{i=1}^{r} \operatorname{Var}(X_i).$$

Furthermore, let $\mathbb{E}[X_i] = p$ be our one-sided error. Since each X_i is Bernoulli, we see that

$$\frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = \frac{rp(1-p)}{(rp)^2} = \frac{1-p}{p} \cdot \frac{1}{r} \leq \frac{1}{r},$$

since $p \geq 1/2$. It follows by Lemma 2.3 that

$$\mathbb{P}[X = 0] \leq \frac{1}{r}.$$

It remains to show how we can achieve a collection of r pairwise independent variables with $2m$ random bits. Let q be a prime such that $2^m < q < 2^{m+1}$. Then pick a, b uniformly at random from the field $\mathbb{Z}_q$. Consider the function $f : \mathbb{Z}_q \to \mathbb{Z}_q$ given by

$$f_{a,b}(x) = ax + b.$$

We will show that the collection

$$\mathcal{B} = \{f_{a,b}(x) : x \in \mathbb{Z}_q\}$$

is a pairwise independent family over $\mathbb{Z}_q$. Note that our collection is indexed by the input x, and not the random integers a and b.

First, note that for all $x, z \in \mathbb{Z}_q$, we have

$$\mathbb{P}_{a,b}[f_{a,b}(x) = z] = \mathbb{P}_{a,b}[ax + b = z] = \frac{1}{q}.$$

Now, for $x \neq y \in \mathbb{Z}_q$, in order to have $f_{a,b}(x) = z_1$ and $f_{a,b}(y) = z_2$, we must satisfy the linear system

$$\begin{bmatrix} x & 1 \\ y & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

which is invertible since the 2×2 is just a Vandermonde, which is invertible. Therefore there is a unique solution for a and b, so that

$$\mathbb{P}_{a,b}[f_{a,b}(x) = z_1, f_{a,b}(y) = z_2] = \frac{1}{q^2}.$$

It follows that the family $\mathcal{B}$ is indeed pairwise independent. Note that we only use $2m$ random bits for the values of a and b, so that this concludes the proof. □

Remark. Note that the above construction can be easily extended to k-wise independence. We just get a $k \times k$ Vandermonde, which is still invertible, and the rest of the proof is largely the same.

Theorem 2.10

Let A be a random $m \times n$ *Toeplitz matrix*, constructed by picking the entries of the first row and first column u.a.r. from $\{0, 1\}$, and then copying its values along each corresponding diagonal. For example, one such matrix might look like this:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Then the family

$$\mathcal{T} = \{h_{A,b}(x) = Ax + b : x \in \{0, 1\}^n\}$$

consists of pairwise independent uniform random variables over $\{0, 1\}^m$, using only $2m + n$ random bits.

Proof. Left as an exercise. □

Remark. We close off this section with a brief discussion about derandomization using k-wise independent random variables. In a previous section, we talked about the method of conditional probabilities for derandomization. This method is inherently sequential, and hence hard to parallelize. Instead, using k-wise independent families, which can be constructed in polynomial space, we can simply do an exhaustive search through the probability space. This yields a polynomial algorithm that can also be easily parallelized.

3 Chernoff/Hoeffding Bounds

4 FPRAS