Lecture 1: First moments CSE 525, Spring 2019
Instructor: James R. Lee

1 The probabilistic method

An old math puzzle goes: Suppose there are six people in a room; some of them shake hands. Prove that there are at least three people who all shook each others' hands or three people such that no pair of them shook hands.

Generalized a bit, this is the classic Ramsey problem. The *diagonal Ramsey numbers* $R(k)$ are defined as follows. $R(k)$ is the smallest integer n such that in every two-coloring of the edges of the complete graph K_n by red and blue, there is a monochromatic copy of K_k, i.e. there are k nodes such that all of the $\binom{k}{2}$ edges between them are red or all of the edges are blue. A solution to the puzzle above asserts that $R(3) \leqslant 6$ (and it is easy to check that, in fact, $R(3) = 6$).

In 1929, Ramsey proved that $R(k)$ is finite for every k. We want to show that $R(k)$ must grow pretty fast; in fact, we'll prove that for $k \geqslant 3$, we have $R(k) > \lfloor 2^{k/2} \rfloor$. This requires finding a coloring of K_n that doesn't contain any monochromatic K_k. To do this, we'll use the *probabilistic method:* We'll give a random coloring of K_n and show that it satisfies our desired property with positive probability. This proof appeared in a paper of Erdös from 1947, and this is the example that starts Alon and Spencer's famous book devoted to the probabilistic method.

Lemma 1.1. *If* $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$*, then* $R(k) > n$*. In particular,* $R(k) > \lfloor 2^{k/2} \rfloor$ *for* $k \geqslant 3$*.*

Proof. Consider a uniformly random 2-coloring of the edges of K_n. Every edge is colored red or blue independently with probability half each. For any fixed set of k vertices H, let $\mathcal{E}_H$ denote the event that the induced subgraph on H is monochromatic. An easy calculation yields

$$\mathbb{P}(\mathcal{E}_H) = 2 \cdot 2^{-\binom{k}{2}} .$$

Since there are $\binom{n}{k}$ possible choices for H, we can use the union bound:

$$\mathbb{P}(\text{exists } R \text{ such that } \mathcal{E}_H) \leqslant 2 \cdot 2^{-\binom{k}{2}} \cdot \binom{n}{k} .$$

Thus if $2^{1-\binom{k}{2}}\binom{n}{k} < 1$, then with positive probability, no event $\mathcal{E}_H$ occurs. Thus there must exist at least one coloring with no monochromatic K_k. One can check that if $k \geqslant 3$ and $n = \lfloor 2^{k/2} \rfloor$, then this is satisfied. □

We have employed the following basic tool.

Tool 1.2 (Union bound). If $A_1, A_2, \ldots, A_m$ are arbitrary events, then

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_m) \leqslant \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_m)$$

2 Linearity of expectation

Let's look at a couple more examples of the probabilistic method in action. We'll use a basic fact in probability: Linearity of expectation.

Tool 2.1 (Linearity of expectation)**.** If $X_1, X_2, \dots, X_n$ are discrete real-valued random variables, then

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]$$

The great fact about this inequality is that we don't need to know anything about the relationships between the random variables; linearity of expectation holds no matter what the dependence structure.

MAX-3SAT. Let's consider a 3-CNF formula over the variables $x_1, x_2, \dots, x_n$. Such a formula has the form $\varphi = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ where each clause is an OR of three literals involving distinct variables: $C_i = z_{i_1} \vee z_{i_2} \vee z_{i_3}$. A literal is a variable or its negation. For instance, $(x_2 \vee \bar{x}_3 \vee \bar{x}_4) \wedge (x_3 \vee \bar{x}_5 \vee \bar{x}_1) \wedge (x_1 \vee x_5 \vee x_4)$ is a 3-CNF formula.

Claim 2.2. If φ is a 3-CNF formula with m clauses, then there exists an assignment that makes at least $\frac{7}{8}m$ clauses evaluate to true.

Proof. We will prove this using the probabilistic method. For every variable independently, we choose a uniformly random truth assignment: true or false each with probability 1/2. Let A_i equal 1 if clause C_i is satisfied by our random assignment, and equal 0 otherwise. Then $\mathbb{P}(A_i = 1) = 7/8$ because there are 7 ways to satisfy a clause out of the 8 possible truth values for its literals.

Let $A = A_1 + \cdots + A_m$ denote the total number of satisfied clauses. By linearity of expectation, we have

$$\mathbb{E}[A] = \sum_{i=1}^{m} \mathbb{E}[A_i] = \frac{7}{8}m\,. \tag{2.1}$$

Since a random assignment satisfies $\frac{7}{8}m$ clauses in expectation, there must exist *at least one* assignment that satisfies this many clauses. □

MAX-CUT. Consider an undirected graph $G = (V, E)$. A *cut* is a subset $S \subseteq V$, and we use $E(S, \bar{S})$ to denote the set of edges *crossing the cut* S. This is the set of edges with one endpoint in S and one not in S.

Claim 2.3. In any graph $G = (V, E)$, there exists a cut $S \subseteq V$ that cuts at least half the edges, i.e., $|E(S, \bar{S})| \geqslant \frac{|E|}{2}$.

Proof. We construct a random set $S \subseteq V$ by including every vertex in S independently with probability 1/2. For an edge $e \in E$, let $A_e = 1$ if e crosses the cut S, and 0 otherwise. First, it should be apparent that $\mathbb{P}(A_e = 1) = 1/2$. Therefore by linearity of expectation,

$$\mathbb{E}\left[|E(S, \bar{S})|\right] = \sum_{e \in E} \mathbb{E}[A_e] = \frac{|E|}{2}\,.$$

Thus there must exist at least one cut S that has at least half the edges crossing it. □

2.1 The method of conditional expectation

Claim 2.2 asserts that there *exists* an assignment satisfying at least $\frac{7}{8}m$ clauses, but what if we wish to actually find one? One way is to randomly sample from the underlying distribution and then check the resulting assignment. Analyzing the probability of success will require our first *tail bound*; we'll get there in the next section.

Let's examine another way that actually results in a deterministic algorithm. Let $S(x_1, x_2, \ldots, x_n)$ denote the expected number of satisfied clauses given a partial truth assignment to the input variables, where we choose the unassigned variables uniformly at random. We will use T to denote true, F to denote false, and $\star$ to denote that no assignment has been chosen for that variable.

For instance, $S(\star, \star, \ldots, \star)$ denotes the expected number of satisfied clauses in a random assignment, and we have already seen (cf. (2.1)) that

$$S(\star, \star, \ldots, \star) = \frac{7}{8}m.$$

Note that a simple linear-time algorithm can estimate $S(x_1, x_2, \ldots, x_n)$ for any partial assignment $x_1, \ldots, x_n \in \{\mathsf{T}, \mathsf{F}, \star\}$ by simply going through the clauses one by one.

As an example, consider the clause $x_1 \vee \bar{x}_2 \vee \bar{x}_4$. The probability that a random assignment satisfies this is $7/8$. If we assign $x_1 = \mathsf{F}$, then the probability becomes $3/4$, and if we set $x_1 = \mathsf{T}$, then the probability becomes 1.

Observe that

$$S(\star, \star, \ldots, \star) = \frac{1}{2}S(\mathsf{F}, \star, \ldots, \star) + \frac{1}{2}S(\mathsf{T}, \star, \ldots, \star)\,.$$

Since $S(\star, \star, \ldots, \star) \geqslant \frac{7}{8}m$, it must hold that $S(\mathsf{F}, \star, \ldots, \star) \geqslant \frac{7}{8}m$ or $S(\mathsf{T}, \star, \ldots, \star) \geqslant \frac{7}{8}m$. As we have just argued, it's possible to compute both these quantities and figure out which is larger. We can then set x_1 to the corresponding value and keep assigning truth values recursively. Eventually, this process ends at a full assignment to the variables that satisfies at least $\frac{7}{8}m$ clauses. The key property we employed here is the ability to efficiently compute the conditional expectation of the underlying random variable under a partial assignment.

2.2 Markov's inequality

The probabilistic method shows the *existence* of an object, but it doesn't necessarily give us a randomized algorithm to construct it. If we just know that the probability of an event is non-zero, it could still be very tiny; we might need to do an arbitrarily large number of random experiments before we get a positive outcome. Sometimes we can say more.

Tool 2.4 (Markov's inequality)**.** Let X be a non-negative random variable. Then for any $\alpha > 0$, we have

$$\mathbb{P}[X \geqslant \alpha] \leqslant \frac{\mathbb{E}\, X}{\alpha}\,.$$

The proof of this lemma is easy; we leave it as an exercise.

Consider now our **MAX-3SAT** example above. Let X denote the number of *unsatisfied* clauses in a random truth assignment. We know from the preceding analysis that $\mathbb{E}[X] \leqslant \frac{1}{8}m$. Markov's inequality tells us that for any $\varepsilon > 0$,

$$\mathbb{P}\left[X > \left(\frac{1}{8} + \varepsilon\right) m\right] \leqslant \frac{m/8}{(1/8 + \varepsilon)m} = \frac{1}{1 + 8\varepsilon} \leqslant 1 - \varepsilon\,.$$

The last inequality is only true if we assume $\varepsilon \leqslant 7/8$, but for any value $\varepsilon > 7/8$, the probability is clearly zero.

This means that, with probability at least ε, we will get an assignment that satisfies at least $(7/8 - \varepsilon)$-fraction of clauses. So in expectation, after $1/\varepsilon$ samples, we will get an assignment that is very close to the one guaranteed to exist. The same kind of reasoning applies to our **MAX-CUT** analysis.

3 Crossing number inequalities

Let's look at one more application of the linearity of expectation. It is almost as elementary as the examples above, but has some powerful consequences in incidence geometry and sum-product estimates.

If $G = (V, E)$ is an undirected graph, we use the notation $\mathrm{cr}(G)$ to denote the *crossing number of G*. This is the minimum number of edge crossings required to draw G in the plane. A drawing of the graph means that the vertices are mapped to distinct points, and each edge is drawn as a closed, continuous curve of bounded length. The following result is due independently to Leighton and Atjai-Chvatal-Newborn-Szemeredi.

Theorem 3.1. *If G is a graph with n vertices and m edges, and $m \geqslant 4n$, then*

$$\mathrm{cr}(G) \geqslant \frac{m^3}{64n^2}\,.$$

Note that for dense graphs, i.e. those with $m = \Omega(n^2)$, we get $\Omega(n^4)$ crossings (the most possible up to a constant factor). We start with a basic fact: Euler's formula implies that, in every planar graph (a planar graph G is one for which $\mathrm{cr}(G) = 0$), we have $m \leqslant 3n - 6$.

Thus if $m > 3n$, we must have $\mathrm{cr}(G) \geqslant 1$. Since we can always remove one crossing from a drawing by removing one edge from the underlying graph, this gives us

$$\mathrm{cr}(G) \geqslant m - 3n\,. \tag{3.1}$$

This is still pretty weak. But now we will use random sampling to do seriously heavy amplification.

Proof of Theorem 3.1. Suppose we have a drawing of G in the plane. We will make some assumptions about this drawing (which are without loss of generality). We may assume that every edge crossing involves *four* distinct vertices. If an edge crosses itself, that can be fixed by short-circuiting the loops. If two edges emanating from the same vertex cross each other, they can be uncrossed without affecting the rest of the drawing (draw a picture to convince yourself). So we may assume that the only crossings are between edges $\{x, y\}$ and $\{u, v\}$ where x, y, u, v are all distinct vertices.

Now we will construct a (random) graph G_p by keeping every vertex of G independently with probability p. The value of p will be chosen soon. Let n_p and m_p denote the number of edges and vertices remaining in G_p, and let c_p denote the number of crossings remaining in our drawing (after the edges and vertices not remaining in G_p are removed).

Every vertex remains with probability p. By independence, an edge remains with probability p^2. Finally, a crossing remains with probability p^4 since we said that every crossing has to involve four distinct vertices. In order for a crossing to remain, all of those four vertices must be in G_p. Thus linearity of expectation gives us:

$$\mathbb{E}[n_p] = pn \tag{3.2}$$

$$\mathbb{E}[m_p] = p^2 m \tag{3.3}$$

$$\mathbb{E}[c_p] = p^4 \mathrm{cr}(G)\,. \tag{3.4}$$

But from (3.1), we know that $c_p \geqslant m_p - 3n_p$, and thus $\mathbb{E}[c_p] \geqslant \mathbb{E}[m_p] - 3\,\mathbb{E}[n_p]$. Plugging in our values above yields

$$p^4 \mathrm{cr}(G) \geqslant p^2 m - 3pn\,,$$

or equivalently

$$\operatorname{cr}(G) \geqslant \frac{m}{p^2} - \frac{3n}{p^3} .$$

Finally, we set $p = \frac{4n}{m}$ ($p \leqslant 1$ since we have assumed $m \geqslant 4n$). This yields

$$\operatorname{cr}(G) \geqslant \frac{m^3}{16n^2} - \frac{3m^3}{64n^2} = \frac{m^3}{64n^2} ,$$

completing our proof. □