Lecture 3: Tail bounds, Probabilistic Method

Today we will see what is known as the *probabilistic method* for showing the existence of combinatorial objects. We also review basic concentration inequalities. As an example illustrating all these, we will give a randomized construction of an $\Omega(\log n)$ integrality gap instance for the Set Cover problem.

The Probabilistic Method

A simple puzzle in graph theory is the following:

In any graph on 6 vertices, show that there is either a clique or an independent set of size 3.[1]

This is a simple example of the kind of questions studied in the area of *Ramsey theory.* One of the basic results there says that for any integer k, there is an $N(k)$ such that: in any graph on $N(k)$ vertices, there is either a clique or an independent set of size k. The first proof also gave an upper bound of 2^{2k} on $N(k)$. The natural question is if one can obtain a much better bound (one that is not exponential in k, for example). Can we give an exponential in k *lower bound* on $N(k)$?

Giving a lower bound on $N(k)$ simply means constructing a graph on a large number of vertices and proving that there is neither a clique nor an independent set. This question remained open until Erdős came up with a remarkable way of proving it: he proved that a *random* graph on roughly $2^{k/2}$ vertices has neither a clique nor an independent set with non-zero probability. Thus it follows that there *exist* graphs on $2^{k/2}$ vertices without a clique or an independent set, implying $N(k) \geq 2^{k/2}$.

Note that the method is non-constructive! We cannot, in the end, come up with an explicit family of graphs for every k.

The proof is actually rather simple. Consider the following random process:

1. Start with n vertices (n will be specified below).
2. For every pair i, j, add the edge $\{ij\}$ independently with probability $1/2$.

The result of this process is a graph $\mathcal{G}$.

Theorem 1. *Let $k > 2$ be any integer, and suppose $n \leq 2^{k/2-1}$. Then with high probability, the graph $\mathcal{G}$ produced above has neither a clique nor an independent set of size k.*

Proof. The proof is rather simple. Fix a set of k vertices (call it S). What is the probability that they form either a clique or an independent set? It is precisely:

$$\frac{2}{2^{\binom{k}{2}}},$$

[1] A clique a set of vertices in which every pair of vertices has an edge, and an independent set is a set in which no two vertices have an edge.

The Probabilistic Method continued on next page...

because we have $2^{\binom{k}{2}}$ possibilities for edges between the k vertices, all equally likely, and precisely two correspond to a clique and an independent set.

Now we can take a union bound over all sets S of size k, and conclude that the probability that *at least one* of them is either a clique or an independent set is at most

$$\binom{n}{k} \cdot \frac{2}{2^{\binom{k}{2}}} < 2^{k \log_2 n + 1 - \binom{k}{2}}. \qquad \text{(we used } \binom{n}{k} < n^k\text{)}.$$

For this quantity to be < 1, we must have $k \log_2 n + 1 < \binom{k}{2} = \frac{k(k-1)}{2}$. From our choice of n, we have $\log_2 n = k/2 - 1$, for which it is easy to check that the above holds for any $k > 2$.

This completes the proof, and thus shows that $N(k) \geq 2^{k/2-1}$. □

The probabilistic method. The general technique of proving the existence of combinatorial objects with certain properties by proving that the probability of the properties holding for a *random* object is non-zero, came to be called the probabilistic method. It is used in a variety of applications, from dimension reduction and sketching, to low-congestion routings in communication networks, to Ramsey theory.

One interesting thing to note, is that computation above (with a slightly smaller n) shows that in fact, a random graph has neither a clique nor an independent set of size k, with extremely high probability. But even so, constructing an explicit graph with this property for every k is rather tricky! This is a fairly general phenomenon with probabilistic constructions, where we know that random objects possess certain properties with very high probability, but we cannot *pin down* an explicit object with the property. The phrase *finding hay in a haystack* was coined to illustrate to this seemingly strange situation.

Concentration Inequalities

A *concentration inequality*, or a *tail bound* for a random variable, is an inequality which says that the probability that the random variable deviates *too much* from its expectation is *small*.

For a well written introduction, please refer to the survey of Chung and Lu:
https://projecteuclid.org/euclid.im/1175266369

Pay particular attention to the simplest case, namely the sum of independent Bernoulli trials: let X_i be independent *coin tosses*, i.e., random variables that takes value 0 with probability 1/2, and 1 with probability 1/2. Now consider the sum $X = \sum_{i=1}^{n} X_i$.

Recall also the formal definition of independence of a collection of random variables:

0/1 random variables $X_1, \ldots, X_n$ are said to be *independent* if for any subset of the variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$, and any *assigment* $(a_1, \ldots, a_k) \in \{0,1\}^k$, we have

$$\Pr[X_{i_1} = a_1 \wedge X_{i_2} = a_2 \wedge \cdots \wedge X_{i_k} = a_k] = \prod_{j=1}^{k} \Pr[X_{i_j} = a_j].$$

The random variable $Y = X - \mathbb{E}[X]$ has a simple interpretation (scaling by a factor 2) as the position at time n, of the natural *random walk* on the line (we start at the origin, and at every time step, we toss a fair coin, move left by one unit if heads, or right by one unit if tails).

To get a sense of how large we expect the *magnitude* of Y to be, a natural measure is the *variance*, i.e., $\mathbb{E}[|Y|^2]$. Let us evaluate this quantity. To do so, first recall that $Y = Y_1 + \cdots + Y_n$, where $Y_i = X_i - \mathbb{E}[X_i] = X_i - 1/2$. We also recall the *linearity of expectation*, a fundamental notion in analyzing probabilistic processes:

For any collection of random variables $Z_1, Z_2, \ldots, Z_n$ (they need not be independent!), we have

$$E[\sum_i Z_i] = \sum_i \mathbb{E}[Z_i].$$

We can use this to conclude that

$$\mathbb{E}[Y^2] = \mathbb{E}[(\sum_i Y_i)^2] = \mathbb{E}[\sum_i Y_i^2 + \sum_{i \neq j} Y_i Y_j] = \sum_i \mathbb{E}[Y_i^2] + \sum_{i \neq j} \mathbb{E}[Y_i Y_j].$$

Now for any $i \neq j$, it is easy to see that $\mathbb{E}[Y_i Y_j] = 0$ (convince yourself if you don't see this immediately). Further, Y_i^2 is $1/4$ w.prob. 1, thus $\mathbb{E}[Y_i^2] = 1/4$. These observations imply that

$$\mathbb{E}[Y^2] = n/4.$$

So this means that *typically*, we expect the value $|Y|$ to be $\sim \sqrt{n}$. What is the probability that it is much larger than say $C\sqrt{n}$, for some large constant C? This is precisely the question answered by Chernoff bounds. It turns out to be $< e^{-C^2}$.

This also hints at the general *format* of most tail inequalities, which is the following:

> *Let X be a random variable that is an* aggregate *of several independent random variables, without depending too strongly on any of them. Then the probability that X is more than C "standard deviations" away from its mean is exponentially small in C.*

Recall that *standard deviation* is simply the square root of the variance (it is the *typical* value of $|Y|$ above). The first 10 or so pages of the Chung-Lu survey give several examples of such theorems. Apart from this, a great reference is Terry Tao's blog post:
https://terrytao.wordpress.com/2010/01/03/254a-notes-1-concentration-of-measure/

Set Cover – An Integrality Gap

Recall the set cover problem, we have topics $T_1, \ldots, T_n$, and people $P_1, \ldots, P_m$, and each person is an expert on a subset of the topics. The goal is to pick the smallest number of people, among whom there is an expert on *every* T_i.

The ILP we considered last time is the following. For every person, we have a 0/1 variable x_i that indicates if the person is picked. We then have the constraints: for each topic T_j,

$$\sum_{i: P_i \text{ expert on } T_j} x_i \geq 1.$$

(As we did last time, we will write $i \sim T_j$ to denote i being an expert on T_j.) The goal is to minimize $\sum_i x_i$.

The LP *relaxation* replaces $x_i \in \{0, 1\}$ with $0 \le x_i \le 1$. We now describe an instance in which the value of the objective ($\sum_i x_i$) is $O(1)$ for the LP, but is at least $\Omega(\log n)$ for the ILP. The instance is very simple: for each topic T_j and person i, we make i an expert on T_j with probability 1/2, independently of other pairs i, j.

What we show is the following:

Theorem 2. *Let $\mathcal{I}$ be the instance of set cover generated by the above probabilistic process. With probability at least 3/4, it satisfies the following two properties:*

1. *The optimal solution to the ILP is $\geq (1/4)\log_2 n$.*
2. *There exists a (fractional) solution to the LP with objective value ≤ 4.*

Let us now prove the theorem. We show that the probability of each of the conditions holding is $\geq 7/8$, and from there, the theorem follows (Union bound the failure events).

The first condition. Intuitively, the reason this condition holds is the following. Suppose we choose any set of k people. Now, for a topic j, the probability that *none* of the chosen people is an expert on T_j is precisely 2^{-k}. Thus the expected number of topics that are not *covered* by any of the chosen people is $n \cdot 2^{-k}$ (this is exactly like tossing n coins, each having a probability 2^{-k} of falling heads, and we want the expected number of heads). Now as long as this number is > 1, there exists an uncovered topic. Thus, we must have $k \geq \log n$ in order to have a valid cover.

The reasoning above is heuristic for two reasons. First, it was *in expectation*. Second, it was for a "fixed" set of k people. We now make the argument formal. Let us fix some k-subset of people S. The probability that S does not cover some T_j is 2^{-k}, as we have seen. These events for different j are independent, because of the way we picked the graph. Hence, the probability that they cover all the n topics is equal to

$$\left(1 - \frac{1}{2^k}\right)^n, \text{ which is } \leq e^{-n/2^k}.$$

Now, what is the probability there *exists* a k-element subsets of $[n]$ that covers all the topics? This is tricky to compute exactly, but we can upper bound the probability by $\binom{n}{k}$ times the probability that any given S covers all the elements (union bound). From the calculation above, this is at most

$$\binom{n}{k} e^{-n/2^k}.$$

To make this $< 1/8$, we must have $8\binom{n}{k} < e^{n/2^k}$. We claim that this holds whenever $k < (1/4)\log_2 n$ (for large enough n). The LHS is upper bounded by $8n^k$, which is $8e^{k \log n}$, or $O(e^{\log^2 n})$. The RHS is at least $e^{n^{3/4}}$. Since $n^{3/4} \gg \log^2 n$ (for n large), the claim follows.

Thus the probability that there exists a set of size $(1/4)\log_2 n$ that covers all the topics is $< 1/8$.

The second condition. Consider the solution $x_i = 4/n$ for all i. Now consider the constraint corresponding to some topic T_j:

$$\sum_{i \sim T_j} x_i \geq 1. \tag{1}$$

To show that this holds, we only need to prove that for every j, the size of the set $\{i : i \sim T_j\}$ is at least $n/4$. From our construction, the expected size of the set is $n/2$. What is the probability that it is $< n/4$?

We can use Chernoff bounds to show that this probability is at most $e^{-\Omega(n)}$. For n large enough, this quantity is $< 1/8n$, i.e., the probability that the constraint fails for some j is $< 1/8n$. Thus the probability that none of the constraints fail is at least $7/8$, which is what we wanted to prove.

This shows that the LP relaxation for the Set Cover problem has an integrality gap of $\Omega(\log n)$.