E0 206: Theorist's Toolkit

Naman Mishra

August 2025

Contents

Lectures

The course

Instructor: Anand Louis

Lecture 1. Wednesday August 6

Evaluation:

(35%) Roughly one homework every two weeks

(25%) Study project

(40%) Final exam

Tools which are generally useful in theoretical computer science

Chapter I

Probabilistic methods

Proposition I.1. *In any graph $G = (V, E)$, there is a cut $S \subseteq V$ such that $|E(S, V \setminus S)| \geq \frac{1}{2}|E|$.*

I will call such cuts *half-cuts*.

Proof. Pick a uniform random set $S \subseteq V$. Each vertex is in S with probability $1/2$. Thus each edge is in $E(S, V \setminus S)$ with probability $1/2$. The expected value of $E(S, V \setminus S)$ is thus $\frac{1}{2}|E|$. Therefore, there exists a subset of vertices for which this quantity is at least $\frac{1}{2}|E|$. ■

We've used the following.

Lemma I.2. *For any random variable X,*

$$\mathbf{P}(X \geq \mathbf{E}\, X) > 0.$$

Proof. Assume the contrary. Then $\mathbf{E}[X] - X > 0$ almost surely. By positivity, the equality in $0 = \mathbf{E}[\mathbf{E}[X] - X] \geq 0$ holds iff $\mathbf{E}[X] - X = 0$ almost surely. Contradiction. ■

A remark is in order regarding the nature of these notes. I will be trying to prove probabilistic results in full generality (because even after 4 probability courses I have understood absolutely nothing about the continuum). Knowing that a random variable is discrete often allows for much simpler proofs. In the above proof, I have tried to stick close to the following fact.

Fact I.3 (Expectation)**.** *There is a unique function $\mathbf{E}$ from RV_+, the space of non-negative (extended) real-valued random variables, to $[0, \infty]$ such that*

(E1) [Linearity] $\mathbf{E}[X + cY] = \mathbf{E}[X] + c\,\mathbf{E}[Y]$ for all $X, Y \in \mathsf{RV}_+$ and $c \geq 0$;

(E2) [Positivity] $\mathbf{E}[X] \geq 0$ with equality iff $X = 0$ almost surely;

(E3) [Monotone convergence] If $X_n \uparrow X$ almost surely, then $\mathbf{E}[X_n] \uparrow \mathbf{E}[X]$; and

(E4) $\mathbf{E}[\mathbf{1}_A] = \mathbf{P}(A)$ for all events A.

This is naturally extended to all real-valued random variables, and the supplied properties also hold only if the positivity is only in an almost sure sense (as with everything else in probability).

We can do slightly better than proposition I.1.

Proposition I.4. *Let S be a random cut of $G = (V, E)$ and $X = |E(S, V \setminus S)|$. Then $\mathbf{P}(X > (1-\delta)\,\mathbf{E}\,X) \geq \frac{\delta}{1+\delta}$.*

Manual proof. We know that $0 \leq X \leq |E|$. Let $A = \{X > (1-\delta)\,\mathbf{E}\,X\}$ and $p = \mathbf{P}(A)$. Then

$$\begin{aligned}\frac{|E|}{2} = \mathbf{E}\,X &= \mathbf{E}[X\mathbf{1}_A] + \mathbf{E}[X\mathbf{1}_{A^c}] \\ &\leq \mathbf{E}[|E|\mathbf{1}_A] + \mathbf{E}[(1-\delta)\,\mathbf{E}\,X\mathbf{1}_{A^c}] \\ &= p|E| + (1-p)(1-\delta)\frac{|E|}{2}.\end{aligned}$$

Rearranging,

$$\delta \leq p(1+\delta).$$ ■

In general, if $-a \leq X \leq b$ and $\mathbf{E}\,X \geq 0$, then $X > 0$ with probability at least $\frac{\mathbf{E}\,X}{b}$.

Proof by Markov's inequality. Let $Y = |E| - X \geq 0$. Now $X \leq (1-\delta)\,\mathbf{E}[X]$ is equivalent to $Y \geq |E| - \frac{1-\delta}{2}|E| = \frac{1+\delta}{2}|E|$. Since $\mathbf{E}[Y] = \frac{1}{2}|E|$, Markov's inequality gives

$$\mathbf{P}\left\{Y \geq \frac{1+\delta}{2}|E|\right\} \leq \frac{1}{1+\delta}.$$

Thus $\mathbf{P}(X > (1-\delta)\,\mathbf{E}\,X) \geq \frac{\delta}{1+\delta}$. ■

Thus, for any $\delta > 0$ and $\varepsilon > 0$, we can sample a random cut $\log_{1+\delta}\varepsilon^{-1}$ times so that the largest of these cuts has size more than $(1-\delta)\frac{|E|}{2}$ with probability at least $1-\varepsilon$. If we choose δ such that $(1-\delta)\frac{|E|}{2} \geq \frac{|E|-1}{2}$, we get a half-cut.

I.1 Derandomization

We can derandomize the algorithm discussed previously. Label the vertices 1 through n. Let X_i denote whether $i \in S$ or $i \notin S$.

$$\frac{|E|}{2} = \mathbf{E}[X] = \frac{1}{2}\mathbf{E}[X \mid 1 \in S] + \frac{1}{2}\mathbf{E}[X \mid 1 \notin S]$$

At least one of these has to be at least $\mathbf{E}\,X$. In this case, by symmetry, both are equal.

$$\begin{aligned}\mathbf{E}[X] &= \mathbf{E}[X \mid 1 \in S] \\ &= \frac{1}{2}\,\mathbf{E}[X \mid 1 \in S, 2 \in S] + \frac{1}{2}\,\mathbf{E}[X \mid 1 \in S, 2 \notin S].\end{aligned}$$

If we could compare $\mathbf{E}[X \mid X_1, \ldots, X_{k-1}, k \in S]$ and $\mathbf{E}[X \mid X_1, \ldots, X_{k-1}, k \notin S]$, we would get an algorithm, picking the "better" (the guarantee we obtain here is better, the choice may or may not be) choice at each step. These probabilities are easy to compute in $O(|E|)$ time at each step. This gives an $O(|V||E|)$ algorithm.

Here, though, we can analyze without computation. Write X as $\sum_{u\sim v} \mathbf{1}_{u,v \text{ crosses the cut}}$. Expanding out both conditional expectations shows that their difference only depends on X_u for $u \sim k$ which have already been determined. Thus it is best to simply add k to whichever of S and $V \setminus S$ has fewer of its neighbors at that time step. This gives an $O(|E|)$ algorithm.

Exercise I.5 (Local search for half-cut)**.** *Start with an arbitrary cut $S_0 \subseteq V$. If there is a vertex $v \in V$ such that moving it to the other side increases the (edge-)size of the cut, do it. This process terminates and yields a half-cut.*

Proof. Terminates because the size can only increase so much. Once it terminates, each vertex has at least half of its incident edges being a part of the cut. Thus the cut is a half-cut and the algorithm takes $O(|E|^2)$ time. ■

I will not care enough to write $|V|+|E|$ in place of $|E|$ simply because pathetic loner vertices may exist.

Lecture 2.
Monday
August 11

Theorem I.6. *Given unit vectors $v_1, \ldots, v_n \in \mathbb{R}^n$, there exist $\varepsilon_1, \ldots, \varepsilon_n \in \{\pm 1\}$ such that*

$$\left\| \sum_i \varepsilon_i v_i \right\| \le \sqrt{n}.$$

Proof. Set the ε_is to be iid with $\mathbf{P}\{\varepsilon_1 = 1\} = \mathbf{P}\{\varepsilon_1 = -1\} = \frac{1}{2}$. Let

$$\begin{aligned}X \coloneqq \left\| \sum_i \varepsilon_i v_i \right\|^2 &= \sum_i \|v_i\|^2 + \sum_{i,j} \varepsilon_i \varepsilon_j \langle v_i,\, v_j\rangle \\ &= n + \sum_{i,j} \varepsilon_i \varepsilon_j \langle v_i,\, v_j\rangle.\end{aligned}$$

Then X has expectation n, so there exists a confirguration in which it is at most n. ■

The other inequality also holds by the same proof. The norm can be made at least $\sqrt{n}$.

Lecture 2: Vector sums; Ramsey numbers; intersecting families

Exercise I.7. *Give a deterministic algorithm to yield $\varepsilon_1, \ldots, \varepsilon_n$.*

Solution. Choose $\varepsilon_1, \varepsilon_2, \ldots$ one by one, at each step choosing the option which yields the smaller current norm. The proof that this works is the same as the derandomization for half-cut shown in section I.1. ■

Prof. Anand says there is a simpler algorithm.

Definition I.8 (Ramsey numbers)**.** $R(k, l)$ is defined to be the smallest n such that any coloring of the edges of K_n using the colors Red and Blue has either a Red clique of k vertices or a Blue clique of l vertices.

Exercise I.9. *Show that $R(k, l)$ is finite. Ideally, show that $R(k, k) \leq 4^k$.*

Examples.

- $R(1, 1) = 1$.
- $R(2, 2) = 2$.
- $R(3, 3) = 6$.

  Proof. Let $v \in K_n$ be arbitrary. At least 3 of its outgoing edges are the same color. Assume that to be Red. Consider the clique formed by the three corresponding neighbors. If all of the edges are Blue, we are done. If any of them is Red, it forms a Red clique together with v.

  It is easy to color K_5 such that it avoids 3-cliques. ■

Theorem I.10. *For any k, $R(k, k) > \frac{1}{e\sqrt{2}} k 2^{k/2}$.*

Proof. Let $n \in \mathbb{N}$. To show that $R(k, k) > n$, we must obtain a "nice" coloring of K_n, that is, one that avoids monochromatic k-cliques.

Color the edges of K_n uniformly at random. Then for any k-clique, the probability that it is monochromatic is $2\left(\frac{1}{2}\right)^{\binom{k}{2}}$.

Summing this up over all possible k-cliques yields an upper bound of $2\binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}}$ on the probability that there exists any monochromatic k-clique at all. If this bound is strictly less that 1, then with some positive probability, we obtain a nice coloring of K_n. This happens when

$$\binom{n}{k} < 2^{\binom{k}{2}-1}.$$

Now $\binom{n}{k} \leq \frac{n^k}{k!} < 2^{\binom{k}{2}-1}$ when $n < \frac{1}{2^{1/k}} 2^{\frac{k-1}{2}} k!^{\frac{1}{k}}$.

Using that $k! \geq \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$, we have that

$$n < \left(\frac{\pi k}{2}\right)^{\frac{1}{2k}} \frac{1}{\sqrt{2}} 2^{\frac{k}{2}} \frac{k}{e}$$

suffices. The first term approaches 1 from above, so we will write the bound as

$$n < \frac{1}{e\sqrt{2}} k 2^{\frac{k}{2}}.$$

Thus, $R(k,k) \geq (e\sqrt{2})^{-1} k 2^{k/2}$. ■

Definition I.11 (k-intersecting family)**.** A family $F = \{S_1, \ldots, S_m\}$ of subsets from a universe U of n elements is a *k-intersecting family* if

- $|S_i| = k$ for every $i \in [m]$, and
- $S_i \cap S_j \neq \varnothing$ for every $i, j \in [m]$.

Theorem I.12. *Let $|U| = n$ and F be a k-intersecting family from U where $n \geq 2k$. Then $|F| \leq \binom{n-1}{k-1}$.*

This bound is tight. One such F would be $\{A \cup \{u\} : A \in \binom{U \setminus \{u\}}{k-1}\}$, where u is an arbitrary element of U.

Lemma I.13. *Let $|U| = n$ and F be a k-intersecting family from U where $n \geq 2k$. Fix a permutation σ of $[n]$. Let $A_{\sigma,i} = \{\sigma(i), \ldots, \sigma(i+k-1)\}$. (The addition is modulo n.) Then there are at most k values of i for which $A_{\sigma,i} \in F$.*

Proof. The permutation is irrelevant here. WLOG assume that it is the identity. Each A_i has size k.

Assume $A_i \in F$. The sets which overlap with A_i are A_{i-k+1} through A_{i+k-1}. That is, $2k-2$ other sets which could be a part of F. However, A_{i-k+1} and A_{i+1} do not overlap. Neither do A_{i-k+2} and A_{i+2}, all the way up to A_{i-1} and A_{i+k-1}. Since only one set may be present from each of these pairs, only $k-1$ of these "other sets" may be a part of F. ■

Proof of theorem I.12. Sample $\sigma \in S_n$ and $i \in [n]$ uniformly at random. Note that $A_{\sigma,i}$ is then a uniform random subset of size k. Call this A.

Thus $\mathbf{P}\{A \in F\} = |F|\binom{n}{k}^{-1}$. However, this probability is also

$$\begin{aligned} \mathbf{P}(A \in F) &= \sum_\sigma \frac{1}{n!} \mathbf{P}(A \in F \mid \sigma) \\ &\leq \sum_\sigma \frac{1}{n!} \frac{k}{n} \end{aligned}$$

by the lemma. Thus $|F| \leq \binom{n}{k}\frac{k}{n} = \binom{n-1}{k-1}$. ■

Lecture 3. Wednesday August 13

Theorem I.14. *A graph of average degree d has an independent set of size at least $n/2d$.*

Proof. Fix $p \in [0,1]$ to be optimized over later. Pick each vertex independently with probability p. Call this set S. For each edge in edge, (simultaneously) delete an arbitrary vertex in S. Call the resultant set S'.

The expected size of S is np. The expected number of edges is $|E|p^2$. Thus the expected value of $|S| - |S'|$ is at most $|E|p^2$.

$$\mathbf{E}|S'| \geq np - \frac{nd}{2}p^2.$$

The lower bound is maximised by $p = \frac{1}{d}$, yielding $\mathbf{E}|S'| \geq \frac{n}{2d}$. ■

Exercise I.15. *If G has maximum degree d, come up with a simple greedy algorithm to output an independent set of size $n/(d+1)$.*

Solution. Start with $S = \varnothing$. Pick any vertex and add it to S. Delete all its neighbours. Continue to exhaustion. ■

Theorem I.16. $R(k,k) \geq \frac{1}{e}k2^{k/2}$.

Proof. Color the edges of K_n uniformly at random. From each monochromatic clique of size k, delete one arbitrary vertex.

For each clique of size k, the probability that it is monochromatic is $2^{1-\binom{k}{2}}$. The resultant set of vertices is expected to have size at least

$$n - \binom{n}{k}2^{1-\binom{k}{2}} \geq \tilde{n} := n - \frac{n^k}{k!}2^{1-\binom{k}{2}}.$$

Thus we have the existence of a "nice" coloring of $K_{\tilde{n}}$, so that $R(k,k) > \tilde{n}$.

To optimise over n, we differentiate and select a large n so that

$$1 - k\frac{n^{k-1}}{k!}2^{1-\binom{k}{2}} > 0.$$

Using Stirling's approximation as in the proof of theorem I.10, we get that

$$n < \frac{1}{2^{\frac{1}{k-1}}}2^{k/2}\frac{k-1}{e}$$

suffices. The first term again approaches 1 from above, so we ignore it.

Substituting this in $\tilde{n}$ gives

$$\begin{aligned}\tilde{n} &= \frac{k}{e}2^{k/2} - \left(\frac{k}{e}\right)^k \frac{2^{k^2}}{k!}2^{1-\binom{k}{2}} \\ &\geq \frac{k}{e}2^{k/2} - 2^{1+\frac{k^2+k}{2}}\frac{1}{e}\end{aligned}$$

■ what?

What does a graph G with a large chromatic number $\chi(G)$? One imagines a densely connected subgraph. But such a subgraph would have short cycles.

Theorem I.17. *For any k and l, there exists a graph G with chromatic number more than k and girth more than l.*

To bound the chromatic number below, we can bound the size of the largest independent set above. Theorem I.14 gives a lower bound and is not useful here.

Lecture 4. Monday August 18

Theorem I.18 (Markov's inequality)**.** *For any non-negative random variable X,*

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}\,X}{a}.$$

Proof. $X \geq X\mathbf{1}_{X\geq a} \geq a\mathbf{1}_{X\geq a}$. Taking expectations, $\mathbf{E}\,X \geq a\,\mathbf{P}(X \geq a)$. ■

Theorem I.19 (Chebyshev's inequality)**.** *Let X be any random variable with mean μ and variance σ^2. Then*

$$\mathbf{P}(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}.$$

Proof. Apply Markov's inequality to the random variable $(X-\mu)^2$ with a set to $t^2\sigma^2$. ■

When X is integer-valued and non-negative, we can use this to our advantage. Specifically, $\mathbf{P}(X > 0) = \mathbf{P}(X \geq 1) \leq \mu$. If $\mu < 1$, then X takes the value 0 with positive probability.

In fact, if $\mu = o(1)$, then $X = 0$ with high probability. If $\mu \geq 1$, this is useless.

Chebyshev's inequality yields a bound for the *other* direction.

$$\mathbf{P}(X=0) \leq \mathbf{P}(|X-\mu| \geq \mu) \leq \frac{\sigma^2}{\mu^2}.$$

Thus, if $\sigma^2 = o(\mu^2)$, then $X > 0$ with high probability.

Lemma I.20. *If $X_1, \ldots, X_k$ are* $\mathrm{Ber}(p_1), \ldots, \mathrm{Ber}(p_k)$ *random variables, respectively, and X denotes their sum, then*

$$\mathrm{Var}(X) \leq \mathbf{E}[X] + \sum_{i\neq j} \mathbf{E}[X_i X_j].$$

Proof.

$$\begin{aligned}\operatorname{Var}(X) &= \sum_i \operatorname{Var}(X_i) + \sum_{i\neq j} \operatorname{Cov}(X_i, X_j) \\ &= \sum_i p_i(1-p_i) + \sum_{i\neq j} \mathbf{E}[X_iX_j] - \mathbf{E}[X_i]\,\mathbf{E}[X_j] \\ &\le \sum_i p_i + \sum_{i\neq j} \mathbf{E}[X_iX_j]. \end{aligned}$$ ■

If $\mu = \omega(1)$, then the first term is $o(\mu^2)$, so we only need to bound the second term to use the observation pertaining to $\sigma^2 = o(\mu^2)$ above.

I.2 Erdős–Rényi graphs

Lemma I.21. *$G(n, \frac{1}{2})$ is uniformly distributed over all all graphs with vertex set $[n]$.*

Proof. The total number of graphs is $2^{\binom{n}{2}}$. For any graph G,

$$\mathbf{P}(G(n, 1/2) = G) = \left(\frac{1}{2}\right)^{|E|}\left(\frac{1}{2}\right)^{\binom{n}{2}-|E|} = \frac{1}{2^{\binom{n}{2}}}.$$ ■

Definition I.22 (Threshold)**.** A property Π on the set of graphs on $[n]$ is said to have a *threshold* function f if for $G(n,p)$, Π holds with high probability when $p \gg f(n)$ and with low probability when $p \ll f(n)$.

We will not be properly defining $\gg$ and $\ll$. The obvious choice is $p \ll f$ iff $p = o(f)$ and $p \gg f$ iff $p = \omega(f)$, which is what we will prove shortly.

Notation. For a graph G, $\omega(G)$ will denote the size of the largest clique in G.

Theorem I.23. $\Pi(G) = \mathbf{1}_{\omega(G)\ge 4}$ *has a threshold function $n \mapsto n^{-2/3}$.*

Proof. For each $I \in \binom{[n]}{4}$, let X_I indicate whether the corresponding 4 vertices form a clique. Let $X = \sum_I X_I$. Then $\mathbf{E}[X] = \binom{n}{4}p^6 = \Theta(n^4p^6)$. Note that $p = o(n^{-2/3}) \iff \mathbf{E}[X] = o(1)$, so in that case, no 4-clique exists with high probability.

We also have that

$$\operatorname{Var}(X) \le \mathbf{E}[X] + \sum_{I\neq J} \operatorname{Cov}(X_I, X_J).$$

In the case that $|I \cap J|$ is 0 or 1, X_I and X_J are independent. Thus

$$\begin{aligned}\operatorname{Var}(X) &\le n^4p^6 + \sum_{|I\cap J|=2} \operatorname{Cov}(X_I, X_J) + \sum_{|I\cap J|=3} \operatorname{Cov}(X_I, X_J) \\ &\le n^4p^6 + C_2n^6p^{11} + C_3n^5p^9\end{aligned}$$

where C_2 and C_3 are some constants. We require n^4p^6, n^6p^{11} and n^5p^9 to be $o(n^8p^{12})$. That is, $1 = o(n^4p^6)$, $1 = o(n^2p)$ and $1 = o(n^3p^3)$. In other words, $p = \omega(n^{-2/3})$, $p = \omega(n^{-1/2})$ and $p = \omega(n^{-1})$. ■

The same proof goes through for any $k \geq 4$, with differing threshold functions.

Exercise I.24.

What goes wrong when $k = 3$?

Theorem I.25. $\omega(G(n, \frac{1}{2})) \in (2 \pm o(1)) \log n$ *with high probability.*

Proof. Fix $k \in [n]$. Let $(X_\alpha)_{\alpha \in \binom{[n]}{k}}$ be indicator random variables as before. Then $\mathbf{E}[X] = \binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}} \leq \binom{n}{k}2^{-\frac{k^2}{2}} = 2^{k \log n - k^2/2}$.

When $k = (2 + o(1)) \log n$, then $2^{k \log n - k^2/2} = o(1)$, so there are no k-cliques with high probability.

Verify the above with what comes below.

If $k = (2 + f(n)) \log n$ where $f(n) = o(1)$, then

$$\begin{aligned} k \log n - k^2/2 &= 2\log^2 n + f(n) \log n - 2\log^2 n - 2f(n) \log n - f(n)^2/2 \\ &= -f(n) \log n - f(n)^2/2 \end{aligned}$$

If f decays too rapidly, this is not limiting to $-\infty$. Since the second term goes to 0, we ignore it. We require $f = \omega(\frac{1}{\log n})$, without of course violating that $f = o(1)$.

Exercise I.26.

Prove the other direction

.

■