Randomized Algorithms

Lecture 11: basic tools of the probabilistic method

Kousha Etessami

Tools of the probabilistic method

Recall again, the **general idea of the probabilistic method:** to show the existence of a hard-to-find object with a desired property, Q, try to construct a probability distribution over a sample space Ω of objects, and show that with positive probability a randomly chosen object in Ω has the property Q.

In this lecture we will highlight several commonly used tools and techniques for applying the probabilistic method, some of which we have seen and used already.

- The Expectation argument.
- "Sample and Modify" arguments.
- The Second Moment Method.
- The Lovasz Local Lemma.

The Expectation Argument

Some basic facts we use in the "expectation argument":

Proposition (Lemma 6.2)

For any random variable X with finite expectation, $\mathrm{E}[X]$*,*

$$\Pr[X \geq \mathrm{E}[X]] > 0 \quad \textit{and} \quad \Pr[X \leq \mathrm{E}[X]] > 0.$$

Proof.

For a discrete r.v., X, we have $\mathrm{E}[X] = \sum_x x \cdot \Pr[X = x]$, where the sum is over all x in the range of X. But if $\Pr[X \geq \mathrm{E}[X]] = 0$, then we have

$$\begin{aligned} \mathrm{E}[X] &= \sum_{x < \mathrm{E}[X]} x \cdot \Pr[X = x] \\ &< \sum_{x < \mathrm{E}[X]} \mathrm{E}[X] \cdot \Pr[X = x] = \mathrm{E}[X] \left(\sum_{x < \mathrm{E}[X]} \Pr[X = x] \right) = \mathrm{E}[X]. \end{aligned}$$

Contradiction. Likewise, assuming $\Pr[X \leq E[X]] = 0$ yields a contradiction. □

The Expectation Argument – applications

We've already seen several applications of the expectation argument:

- **MaxCut**: For a random cut of the vertices of any graph $G = (V, E)$ with $m = |E|$ edges, into two sets $(S, V \setminus S)$, the expected number of edges that cross the cut is $m/2$.
  Therefore such a cut exists.
- **MaxSat**: For a random truth assignment to the variables of any boolean k-CNF formula, φ, with m clauses and exactly k literals in each clause, the expected number of clauses that are satisfied is $m(1 - \frac{1}{2^k})$.
  Therefore such a truth assignment exists.

The Expectation Argument – another simple fact

Proposition

For any non-negative, integer *random variable X with finite expectation $\mathrm{E}[X]$, we have*

$$\Pr[X > 0] = \Pr[X \geq 1] \ \leq \ \mathrm{E}[X].$$

Proof.

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{i=0}^{\infty} i \cdot \Pr[X = i] \\
&= \sum_{i=1}^{\infty} i \cdot \Pr[X = i] \\
&\geq \sum_{i=1}^{\infty} \Pr[X = i] \\
&= \Pr[X \geq 1] = \Pr[X > 0] \qquad \text{(because } X \text{ is an integer).}
\end{aligned}$$

“Sample and modify” arguments

- Sometimes attempting to directly generate the desired object purely randomly doesn’t work.
- Instead, it sometimes pays off to do things in two stages:
  1. First, randomly generate an object. It doesn’t necessarily have the property, but it is likely to get you “close”.
  2. Then, modify the randomly generated object by hand, fixing it so that, with positive probability, it has the property.

"Sample and modify" arguments – an application

Theorem. *Any connected graph $G = (V, E)$ with n vertices and m edges has an independent set of size $\frac{n^2}{4m}$.*

Proof. Let $d = \frac{2m}{n}$ be the average degree of a vertex.

1. Randomly delete each $v \in V$ (and its edges), independently, with probability $(1 - \frac{1}{d})$.
2. Remove any remaining edge and one of its two endpoints.

What's left is an independent set. Let X be the number of vertices that survive step (1.). We have $\mathrm{E}[X] = n \cdot \frac{1}{d} = \frac{n}{d}$. Let Y be the number of edges that survive step (1.).

$$\mathrm{E}[Y] = m \cdot (\frac{1}{d})^2 = \frac{nd}{2} \cdot (\frac{1}{d})^2 = \frac{n}{2d}.$$

The second step removes all remaining edges, and at most Y vertices. So, the algorithm terminates with an independent set of size at least $X - Y$. But by linearity of expectation

$$E[X - Y] = \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d} = \frac{n^2}{4m}.$$

□

Second Moment Method

Recall: the *second moment* of a random variable X is $\mathrm{E}[X^2]$. And the *variance* is $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.

Sometimes, we can use the second moment/variance, together with Chebyshev's inequality, to bound probabilities of bad events.

Theorem (Theorem 6.7)

For any r.v., X, with finite $\mathrm{E}[X] \neq 0$ and finite $\mathrm{Var}[X]$, we have

$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{(\mathrm{E}[X])^2}.$$

Proof. Easy consequence of Chebyshev's inequality:

$$\Pr[X = 0] \leq \Pr[|X - \mathrm{E}[X]| \geq \mathrm{E}[X]] \leq \frac{\mathrm{Var}[X]}{(\mathrm{E}[X])^2}.$$

□

Threshold for 4-cliques in $G_{n,p}$

Recall the random graph model $G_{n,p}$.

We are interested in whether a randomly drawn graph $G \leftarrow G_{n,p}$ contains a 4-clique or not.

Clearly the graph is more likely to have a 4-clique if p has a higher value (since $G \leftarrow G_{n,p}$ is likely to have more edges).

Let the probability $p = p(n)$ be a function of n.
We will show that there is a precise *threshold* for $p(n)$, for the property "$G \leftarrow G_{n,p(n)}$ has a 4-clique" to hold or not hold.

Theorem (Theorem 6.8)

1. *If $p(n) = o(n^{-2/3})$, then*

$$\lim_{n\to\infty} \Pr[G \leftarrow G_{n,p(n)} \text{ has a 4-clique}] = 0.$$

2. *If $p(n) = \omega(n^{-2/3})$, then*

$$\lim_{n\to\infty} \Pr[G \leftarrow G_{n,p(n)} \text{ has a 4-clique}] = 1.$$

Threshold for 4-cliques in $G_{n,p}$ – Proof Sketch

Proof. Let $G \leftarrow G_{n,p(n)}$, and let X be the number of 4-cliques in $G = (V, E)$.

Let $C_1, C_2, \ldots, C_{\binom{n}{4}} \subseteq V$, be a listing of all 4-vertex subsets of V.

For $1 \leq i \leq \binom{n}{4}$, define r.v. X_i so that $X_i = 1$ if C_i forms a clique, and $X_i = 0$ otherwise. Clearly, $X = \sum_i X_i$. Then by linearity of expectation:

$$\mathrm{E}[X] = \sum_{i=1}^{\binom{n}{4}} \mathrm{E}[X_i] = \binom{n}{4}(p(n))^6 = \Theta(n^4 \cdot (p(n))^6)$$

Now, notice that

1. if $p(n) = o(n^{-2/3})$, then $\mathrm{E}[X] \approx n^4 \cdot o(n^{-4}) \to 0$, as $n \to \infty$.
2. if $p(n) = \omega(n^{-2/3})$, $\mathrm{E}[X] \approx n^4 \cdot \omega(n^{-4}) \to \infty$, as $n \to \infty$.

Hence, (1.): if $p(n) = o(n^{-2/3})$, then since X is a non-negative integer r.v., we know $\Pr[X \geq 1] \leq \mathrm{E}[X] \to 0$. Hence $\lim_{n\to\infty} \Pr[X \geq 1] = 0$.

Threshold for 4-cliques in $G_{n,p}$ – Proof

Proof sketch continued. (2.) Suppose $p(n) = \omega(n^{-2/3})$. In that case $\mathrm{E}[X] \approx n^4\omega(n^{-4}) \to \infty$ as $n \to \infty$. However, this does not imply a lower bound for $\Pr[X > 0]$. We need a second moment argument.

We want to calculate $\mathrm{Var}[X]$, and show that $\frac{\mathrm{Var}[X]}{(\mathrm{E}[X])^2} \to 0$ as $n \to \infty$.

Note that $(\mathrm{E}[X])^2 = \Theta((n^4 \cdot (p(n))^6)^2) = \Theta(n^8(p(n))^{12})$.

So, if we can show that $\mathrm{Var}[X] = o(n^8(p(n))^{12})$ we are done.

It turns out this can be done. First, we need the following **Lemma**: for any r.v., $Y = \sum_i Y_i$, if the Y_i's are all $0 - 1$ random variables, then:

$$\mathrm{Var}[Y] \;\leq\; \mathrm{E}[Y] + 2\sum_{i \neq j} \mathrm{Cov}[Y_i, Y_j].$$

For X the individual covariances $\mathrm{Cov}[X_i, X_j]$ can be bounded, via a detailed case distinction based on the amount of "overlap" between the respective sets C_i and C_j of vertices.
We will not provide further details of the proof here. See [MU] Section 6.5.1. Note: the book's proof uses multinomial coefficient notation, e.g., $\binom{n}{n_1, n_2, n_3} = \frac{n!}{n_1!n_2!n_3!}$.

The Lovász Local lemma

Consider a large bunch of "bad events", $E_1, E_2, \ldots, E_n$.

Suppose that in order to show existence of a desired object, using the probabilistic method, we have to avoid all these bad events. In other words, we want to show:

$$\Pr[\bigcap_{i=1}^{n} \overline{E}_i] > 0 \qquad (1)$$

Supppose $\Pr[E_i] < 1$, for all i. (Otherwise, there's no hope.)

If the events $E_1, \ldots, E_n$ are mutually independent then (1) is easy, because:

$$\Pr[\bigcap_{i=1}^{n} \overline{E}_i] = \prod_{i=1}^{n} \Pr[\overline{E}_i] = \prod_{i=1}^{n} (1 - \Pr[E_i]) > 0.$$

Note that this could be a very small probability, e.g., if n is very large, but nevertheless it is a positive probability, so existence follows.

Unfortunately, often the bad events may not be independent.

The Lovasz Local Lemma allows us to establish (1) in contexts where there is some limited dependencies between the E_i's.

The Lovász Local lemma

Let us define a particular event E to be *mutually independent of* a set of events $\{E_1, E_2, \ldots, E_k\}$ if for all subsets $I \subseteq \{1, \ldots, k\}$, we have

$$\Pr\left[E \mid \bigcap_{i \in I} E_i\right] = \Pr[E].$$

Definition (6.1) A *dependency graph* for a set of events $E_1, \ldots, E_n$ is a directed graph $G = (V, E)$ such that $V = \{1, \ldots, n\}$ and for each $i \in V$, the event E_i is mutually independent of the set of events $\{E_j \mid (i, j) \notin E\}$. The *degree* of G is the maximum out-degree of any vertex in G.

Theorem (Lovász Local Lemma (symmetric version))

Let $E_1, \ldots, E_n$ be a set of events. Suppose that for some $p \in (0, 1)$ and some $d \in \mathbb{N}$ the following conditions hold:

1. *For all i, $\Pr[E_i] \leq p$;*
2. *A dependency graph on $\{E_1, \ldots, E_n\}$ has degree $\leq d$;*
3. *$4dp \leq 1$.*

Then

$$\Pr\left[\bigcap_{i=1}^{n} \overline{E}_i\right] > 0.$$

Important application: satisfiability of k-CNF formulas

Recall the **k-SAT problem**: Given a k-CNF boolean formula, φ, where each clause has exactly k literals, decide whether φ is satisfiable. Recall, **k-SAT** is **NP-complete**, already for $k = 3$.

Theorem. *If no variable in a k-CNF formula φ appears in more than $\frac{2^k}{4k}$ clauses, then φ is satisfiable.*

Proof. Randomly and independently assign each boolean variable x_i either 0 or 1 with probability $1/2$ each. Suppose there are m clauses, $C_1, \ldots, C_m$, in φ. Let E_i, $i = 1, \ldots, m$, denote the event that C_i is not satisfied. Since each C_i has k literals, we have $\Pr[E_i] = 2^{-k}$.

But E_i is independent of all E_j for which C_i and C_j don't share any variables. Since each of the k variables in C_i appears in $\leq \frac{2^k}{4k}$ clauses, there is a dependency graph for the E_i's with degree $d \leq k \cdot \frac{2^k}{4k} = \frac{2^k}{4}$. Letting $p = 2^{-k}$, we have $4dp \leq 4 \cdot \frac{2^k}{4} \cdot 2^{-k} = 1$. So we can apply the Lovasz Local Lemma to conclude:

$$\Pr[\bigcap_{i=1}^{m} \overline{E}_i] > 0,$$

meaning φ is satisfiable. □

Outlook

- In the last lecture for this course, we will prove the Lovasz Local Lemma.
- When we do, we will first give a classic, but non-constructive proof.
- Then we will describe a more recent beautiful algorithmic proof by Moser (2009) (later generalized by Moser & Tardos (2010)), which gives us, in particular, a randomized (Las Vegas) polynomial time algorithm for computing a satisfying assignment for k-SAT instances that satisfy the conditions of the theorem we stated on the prior slide.
- Read Chapter 6, sections 6.1-6.7, and section 6.10.
- Starting in the next lecture, Raul will cover Markov chains and their uses in randomized algorithms.