CS265/CME309: Randomized Algorithms and Probabilistic Analysis

Lecture #11: The Probabilistic Method Continued: Second-Moment Method and Lovasz Local Lemma

Gregory Valiant*

November 4, 2019

1 Introduction

In Lecture 10, we saw several surprisingly powerful applications of the Probabilistic Method. For those applications, we leveraged one of two different techniques:

1. The "probability > 0 argument": if we can design a random variable X such that $\Pr[X = w] > 0$, then w must exist. [We used this approach to show that the kth Ramsey number is at least $2^{k/2}$.]

2. The "expectation argument": if we can design a random process such that creates some (random) object X, and function f for which $\mathbf{E}[f(X)] \geq \alpha$, then there must exist an object w for which $f(w) \geq \alpha$. [We used this approach to argue that for every graph, there is a partition that cuts at least half the edges, and for any k-SAT formula, there exists an assignment that satisfies at least a $(1 - 1/2^k)$ fraction of clauses.]

In this lecture, we will see two additional techniques that we can add to the above arsenal. The first technique is known as the *second-moment method*, which is a convenient way of bounding the probability that a random variable can equal 0 (and hence is often useful for proving things like "there will be at least one blah blah blah"). The second technique, the *Lovasz Local Lemma* (aka the "LLL") applies when there are a number of random variables which only depend "locally" on each other. This second technique takes more effort to describe, prove, and apply, though can be extremely powerful.

2 The Second-Moment Method

The second moment method is a way of bounding the probability that a random variable is non-zero, via Chebyshev's inequality.

Theorem 1. *For a real-valued random variable, X,*

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

Proof. By Chebyshev's inequality:

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

$\square$

The second-moment method is especially well suited to settings where X represents a quantity of interest, and we know that $\mathbf{E}[X]$ is fairly large. Just the fact that $\mathbf{E}[X]$ is large does not necessarily mean that $\Pr[X = 0]$ is very small. If, however, we can bound the variance of X, then we are in business. This technique has been fruitfully used to analyze phase transitions in random constraint satisfaction settings. Specifically, consider making a *random* $3 - SAT$ formula over n variables, by forming each clause by randomly selecting 3 variables from $x_1, \ldots, x_n$ and choosing to negate or not negate each variable independently with probability $1/2$. If the number of clauses is small in comparison to n, then the formula will be satisfiable with probability close to 1. If the number of clauses is very large, then the formula will, with high probability, not be satisfiable. This prompts the question of understanding if/where there is a "sharp" threshold in the number of clauses such that the probability of satisfiability goes from nearly 1 to nearly zero as one crosses this threshold. (This is analogous to the theorem you proved on problem-set 4, where you showed that there is a sharp threshold for the emergence of a large connected component in $G_{n,p}$ when the edge probability $p = 1/n$.) We won't cover this result, though feel free to check out the recent line of work on these questions (e.g. [1, 2]).

Below we sketch one simple application of the second-moment method to show a sharp threshold for the emergence of a clique of size 4 in $G_{n,p}$ when $p = n^{-2/3}$.

Theorem 2. *There are constants $c_1, c_2 > 0$ such that for sufficiently large n,*

$$\Pr[G_{n,p \leq c_1 n^{-2/3}} \textit{ has a clique of size 4}] < 0.1 \textit{ and } \Pr[G_{n,p \geq c_2 n^{-2/3}} \textit{ has a clique of size 4}] > 0.9.$$

Proof. Let $p = cn^{-2/3}$ for some constant c. For the first direction, since a 4-clique contains 6 edges, letting X denote the number of 4-cliques in $G_{n,p}$, by linearity of expectation we have $\mathbf{E}[X] = \binom{n}{4} p^6 \leq c^6$. Since X is a non-negative valued random variable that takes integer values, $\Pr[X \geq 1] \leq \mathbf{E}[X]$, and hence if $c^6 < 0.1$, the probability that there exists a 4-clique is at most 0.1.

For the second direction, we will apply the second moment method. For sufficiently large n, and $p = cn^{-2/3}$, $\mathbf{E}[X] = \binom{n}{4} p^6 \geq c^6/100$, and hence for sufficiently large c, this expectation can be made arbitrarily large. To apply the second moment method, we need to show that $\mathbf{Var}[X] \ll (\mathbf{E}[X])^2$. To do this, letting $X_1, \ldots, X_{\binom{n}{4}}$ denote indicator random variables for whether or not each potential 4-clique exists. If the X_i's were independent, this analysis would be straightforward (because the variance of the sum of independenent random variables is the sum of the variances). These random variables are not quite independent—if X_i and X_j refer to cliques that share at most 1 node, then they are independent, otherwise they are (positively) correlated. We'll skip the details of the variance calculation (you should do this on your own!), though the intuition is that since a $1 - o(1)$ fraction of pairs X_i, X_j are independent, the variance will be not be too much worse than if they were all independent, and in particular, the variance can be made arbitrarily smaller than the square of the expectation by taking sufficiently large constant c. $\square$

3 The Lovasz Local Lemma

The Lovasz Local Lemma, originally due to Erdos and Lovasz [3], is a clever way of bounding the probability that any number of "bad" events occur", in the setting that the events are not independent, but where each only depends on a small number of other events. [If Erdos's name began with "L" then this would probably be known as the "LLLL", though as it stands his name is usually left off this.]

To motivate the theorem, consider a setting where we have some probability space, and have identified a set of "bad" events, $A_1, \ldots, A_n$, such that $\Pr[A_i] \leq p < 1$. If the events were independent, then

$$\Pr[\cap_i \overline{A_i}] \geq (1-p)^n > 0,$$

and there is a nonzero probability that none of the bad events happen. [For 0/1 random variables (i.e. "events") the notation $\cap_i \overline{X_i}$ just mean "(not X_1) and (not X_2) and $\ldots$."]

If the events A_i are not independent, then we can always do a union bound, to conclude that

$$\Pr[\cap_i \overline{A_i}] \geq 1 - np,$$

though the problem is that in many cases, $np > 1$, and hence this bound is useless. The LLL provides a way around the overly-pessimistic union bound, in the setting where the events A_i are not independent, but where each only depends on a few other events.

Definition 1. *Given events B and $B_1, \ldots, B_k$ defined over some probability space, B is* mutually independent *of events $\{B_1, \ldots, B_k\}$ if the probability of B does not change if we condition on* any *subset of $B_1, \ldots, B_k$. Formally, for any subset $J \subseteq \{1, \ldots, k\}$,*

$$\Pr[B] = \Pr[B | \cap_{i \in J} B_i].$$

Theorem 3. *Let $A_1, \ldots, A_n$ denote a set of events such that, for all i, $\Pr[A_i] \leq p$, and where each A_i is mutually independent of all but d other events. Then,*

1. *(Version I) Provided $pd \leq 1/4$,* $\quad \Pr[\cap_i \overline{A_i}] \geq (1-2p)^n > 0.$
2. *(Version II) Provided $p(d+1) \leq 1/e$,* $\quad \Pr[\cap_i \overline{A_i}] \geq (1 - \frac{1}{d+1})^n > 0.$

To give an interpretation of this theorem, consider the second condition that asserts that $p(d+1) \leq 1/e$. The expression $p(d+1)$ is simply a union bound over $d+1$ dependent events (e.g. event A_i and the $\leq d$ other events that are not mutually independent of A_i). If there were only $d+1$ events in total, we would just need $p(d+1) < 1$ to ensure that $\Pr[\cap_i \overline{A_i}] > 0$. The theorem, however, says that if each union bound over the dependent neighborhoods of size $d+1$ are satisfied, with an extra constant factor room-to-spare, then it is possible to piece together all these interconnected neighborhoods!

Before proving this theorem, we will see an application of this theorem to k-SAT.

3.1 Application: k-SAT

Theorem 4. *Given a k-SAT formula over variables $x_1, \ldots, x_n$ such that 1) each clause has exactly k distinct literals [variables], and 2) each variable occurs in at most $\frac{2^{k-2}}{k}$ clauses, then the formula is satisfiable, no matter the number of variables or clauses!!*

Proof. To prove this theorem via the LLL and probabilistic method, we first need to define a probability space and set of "bad" events. Consider assigning each variable x_i to be true or false, independently with probability $1/2$. Define the events $A_1, \ldots, A_m$ so that $A_i = 1$ if the ith clause *is not satisfied* by the assignment, and is 0 otherwise. Hence, for all i,

$$\Pr[\overline{A_i}] = 1 - \frac{1}{2^k}.$$

Now we need to understand the dependency structure of the bad events. Letting $vbl(A_i)$ denote the subset of variables $\{x_j\}$ that are present in the ith clause, we claim that A_i is mutually independent of the set

$$S_i = \{A_j : vbl(A_i) \cap vbl(A_j) = \emptyset\}.$$

To see why this is the case, note that even if we condition on all the events $A_j \in S_i$ *and the values of every variable* $x_j \notin vbl(A_i)$, it is still the case that $\Pr[\overline{A_i}] = 1 - 1/2^k$, as the ith clause only depends on the values taken by the k variables in $vbl(A_i)$. By assumption, each variable in $vbl(A_i)$ occurs in at most $\frac{2^{k-2}}{k}$ other clauses, and hence A_i is mutually independent of all but at most $d = k \cdot \frac{2^{k-2}}{k}$ events. By the LLL, provided $(\max_i \Pr[A_i])\, d = \frac{1}{2^k} \cdot k \frac{2^{k-2}}{k} = 1/4 \leq 1/4$, the formula is satisfiable. □

To see some concrete implications, if $k = 4$, then the assumptions of the theorem assert that each variable occurs in at most $2^{4-2}/4 = 1$ clause, in which case the theorem is trivially satisfiable. When k is larger, the theorem starts to give interesting conclusions: for example, the theorem implies that any instance of 10-SAT in which each variable occurs in at most 26 clauses, is satisfiable.

3.2 Proof of LLL

We will just prove the first of the two statements of Theorem 3, namely that if $pd \leq 1/4$, then the statement holds. The proof of the second assertion is almost identical.

The following lemma will be the central portion of the proof:

Lemma 2. *We will prove that for any set $S \subset \{1, \ldots, n\}$, and any $i \notin S$,*

$$\Pr[A_i | \cap_{j \in S} \overline{A_j}] \leq 2p.$$

Given this lemma, the proof follows from writing out the probability we care about:

$$\Pr[\cap_i \overline{A_i}] = (1 - \Pr[A_1])(1 - \Pr[A_2 | \overline{A_1}])(1 - \Pr[A_3 | \cap_{j \leq 2} \overline{A_j}]) \cdot \ldots .$$

Lemma 2 guarantees that each of the n terms in the above expression are at least $1 - 2p$, and hence

$$\Pr[\cap_i \overline{A_i}] \geq (1 - 2p)^n > 0,$$

proving the theorem.

We now prove Lemma 2:

Proof of Lemma 2. We prove this lemma inductively on the size of set S. For the base case, when $|S| = 0$ is the empty set, $\Pr[A_i | \cap_{j \in S} \overline{A_j}] = \Pr[A_i] \leq p \leq 2p$ by the definition of p. For the inductive step, assume the lemma holds for all sets S with $|S| \leq k$. Consider some set S with $|S| = k + 1$, and an event A_i with $i \notin S$. Let set S_i denote a set of events that is mutually independent of A_i, for which $|\{j \in \{1, \dots, n\} | j \notin S_i\}| \leq d$. We now partition S into the events that intersect S_i, and those that do not: let $S^{ind} = S \cap S_i$ and $S^{dep} = S \setminus S^{ind}$. These represent the partition of S into the subset upon which A_i depends, and is independent from. If $|S^{ind}| = k+1$, then $S = S^{ind}$ and from the definition of mutual independence,

$$\Pr[A_i | \cap_{j \in S} \overline{A_j}] = \Pr[A_i] \leq p \leq 2p,$$

and we are done. Henceforth, assume $|S^{ind}| \leq k$. For any events A, B, C by the definition of conditional probability $\Pr[A|B] = \Pr[A \text{ and } B] / \Pr[B]$, and similarly, $\Pr[A|B, C] = \Pr[A \text{ and } B|C] / \Pr[B|C]$. Hence, applying this to the partition of S into S^{ind} and S^{dep}, we have:

$$\Pr[A_i | \cap_{j \in S} \overline{A_j}] = \frac{\Pr[A_i \text{ and } \cap_{j \in S^{dep}} \overline{A_j} | \cap_{j \in S^{ind}} \overline{A_j}]}{\Pr[\cap_{j \in S^{dep}} \overline{A_j} | \cap_{j \in S^{ind}} \overline{A_j}]}. \tag{1}$$

Because $|S^{ind}| \leq k$, we can apply our inductive assumption to the condition in the denominator, to bound the denominator. Using a union bound, we have the following bound on the denominator:

$$\Pr[\cap_{j \in S^{dep}} \overline{A_j} | \cap_{j \in S^{ind}} \overline{A_j}] \geq 1 - \sum_{j \in S^{dep}} \Pr[A_j | \cap_{j \in S^{ind}} \overline{A_j}] \geq 1 - |S^{dep}|(2p) \geq 1 - \frac{1}{4p}(2p) = 1/2,$$

where we used the fact that, by assumption $|S^{dep}| \leq 1/4p$ since S^{dep} is subset of the events upon which S_i depends, and hence this set is bounded in size by d, which is defined to satisfy $pd \leq 1/4$.

Given that the denominator of Equation 1 is at least $1/2$, we now turn to upper bounding the numerator. Since the probability that a set of events all occur is at most the probability of any one of them, we have

$$\Pr[A_i \text{ and } \cap_{j \in S^{dep}} \overline{A_j} | \cap_{j \in S^{ind}} \overline{A_j}] \leq \Pr[A_i | \cap_{j \in S^{ind}} \overline{A_j}] \leq p,$$

where we have used the fact that, by definition, A_i is mutually independent of the events in S^{ind}. Putting together our upper bound of p on the numerator of Equation 1 and lower bound of $1/2$ on the denominator, yields that the expression is at most $2p$, and we have completed our induction argument. □

3.3 Asymmetric LLL

The formulation of the LLL that we gave in Theorem 3 just has two parameters, p and d, to describe the probabilities of the events and sizes of the dependent neighborhoods. You might wonder if we can get a more general statement that can deal with settings where a few events might have higher probability, or where a couple of events are dependent on a large number of other events. The following theorem is precisely this statement. The proof is analogous to the proof of Theorem 3, provided one does a little more book-keeping.

Theorem 5 (Asymmetric LLL). *Let $A_1, \dots, A_n$ denote a set of events, and for each A_i, let $S_i \subset \{A_1, \dots, A_n\}$ denote a set that is mutually independent of A_i. If there exists a set of numbers $r_1, \dots, r_n \in [0, 1)$ such that*

$$\text{for all } i, \Pr[A_i] \leq r_i \prod_{j \notin S_i} (1 - r_j),$$

then

$$\Pr[\cap_i \overline{A_i}] \geq \prod_i (1 - r_i).$$

We can show that this recovers Theorem 3. To see this, observe that under the conditions of Theorem 3, $\max_i |\{j | A_j \notin S_i\}| = d$. Let $r_i = 1/(d+1)$. Under the conditions of the second version of Theorem 3, for any event A_i,

$$\begin{aligned} &\Pr[A_i](d+1) \leq 1/e, \\ \implies &\Pr[A_i] \leq \frac{1}{d+1} \cdot \frac{1}{e} \leq \frac{1}{d+1} \left(1 - \frac{1}{d+1}\right)^d \\ &\leq \frac{1}{d+1} \prod_{j \notin S_i} \left(1 - \frac{1}{d+1}\right) = r_i \prod_{j \notin S_i} (1 - r_j). \end{aligned}$$

Applying Theorem 5, we get that

$$\Pr[\cap_i \overline{A_i}] \geq \prod_i (1 - r_i) = \left(1 - \frac{1}{d+1}\right)^n > 0$$

which recovers the second condition of Theorem 3.

Applying this more general version of the LLL is a bit more involved, as we need to come up with the values $\{r_i\}$ ourselves.

References

[1] Dimitris Achlioptas and Cristopher Moore. Random k-sat: Two moments suffice to cross a sharp threshold. *SIAM Journal on Computing*, 36(3):740–762, 2006.

[2] Amin Coja-Oghlan. The asymptotic k-sat threshold. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*, STOC '14, pages 804–813, 2014.

[3] Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *COLLOQUIA MATHEMATICA SOCIETATIS JANOS BOLYAI 10. INFINITE AND FINITE SETS, KESZTHELY (HUNGARY)*, 1973.