Theoretician's Toolkit **Spring 2019**

Lecture 2: The Probabilistic Method

Scribe: Nate Armstrong 2/11/19

The probabilistic method is a powerful tool generally used in non-constructive proofs of existence. A **non-constructive** proof of existence is a proof of the existence for some mathematical object which does not actually provide a construction of the object. This method words as follows. To show that an object with a specific property exists, we randomly sample from an appropriate collection of objects and show either that sample has the desired property with positive probability, or that the expectation of the sample has that property. In both cases, this implies the existence of at least one such object with the property. In addition to proving existence, the probabilistic method can lead to efficient randomized algorithms, which can sometimes be derandomized.

In this lecture, we discuss applications of the probabilistic method, and the proof techniques used. There is a review section on the union bound, expectation of a random variable, Markov's inequality, and Little-O notation.

2.1 Relevant Probability and Background

2.1.1 The Union Bound

The union bound, also known as Boole's inequality, is a useful (and loose) inequality about the probability of the union of many events. It can be derived from the inclusion-exclusion principle:

Definition 2.1 (Inclusion-Exclusion Principle)**.** *The* ***inclusion-exclusion principle*** *for 2 sets A, B is defined as*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

From this definition, we see that $|A \cup B| \leq |A| + |B|$. By extending this idea through induction, we arrive at the union bound.

Theorem 2.2 (Union Bound)**.** *For sets $A_1, A_2, \ldots, A_n$,*

$$\left| \bigcup_{i=1}^{n} A_i \right| \leq \sum_{i=1}^{n} |A_i| .$$

The union bound also applies to probabilities of events in the natural way, and states that

$$\Pr\left(\bigcup_{i=1}^{n} A_i \right) \leq \sum_{i=1}^{n} \Pr\left(A_i\right).$$

We use the union bound in our discussion on Ramsey and chromatic numbers of graphs.

2.1.2 Expectation of a Random Variable

The expectation of a random variable X differs in definition based on if X is a continuous or discrete random variable (or a single generalized definition in measure theory), but in all cases can be interpreted as the probability of X taking on a certain value multiplied by that value. Basically, it is the 'average' value of the random variable. In the discrete case, the expectation is written as

$$\mathbb{E}(X) = \sum_{x \in \Omega} x \cdot \Pr(X = x)$$

or in the continuous case where the probability density function of X is given by f, as

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x)\, dx.$$

One of the most important properties of the expectation of a random variable is that it is linear. In other words, it means that the expectation of a random variable has the following properties:

- $\mathbb{E}[cX] = c\mathbb{E}[X]$ for all constants c and random variables X.
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for all random variables X and Y.

In this lecture, linearity of expectation is used in max cut, independent set, and the girth/chromatic number of graphs.

Indicator Variables

A particular type of random variable that comes up frequently when working with expectation in the context of the probabilistic method is an indicator variable. It will generally occur when we are interested in the number of times that some event P occurs in a set A. For an object a in a set A, we define X_a to be 1 if P occurred for object a, and 0 otherwise. Note that $\mathbb{E}[X_a] = \Pr(X_a = 1) \cdot 1 + \Pr(X_a = 0) \cdot 0 = \Pr(X_a) = 1$. It is common to sum all the indicator variables into a larger variable $X = \sum_{a \in A} X_a$, and then use linearity of expectation to sum the probabilities of the individual variables being 1.

2.1.3 Markov's Inequality

Markov's inequality is a bound for non-negative random variables. Given a non-negative random variable X and $a > 0$,

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Markov's inequality is used in the graph girth example.

2.1.4 Little O Notation

Given two sequences (a_n) and (b_n), we write that $a_n = o(b_n)$ if

$$\lim_{n \to \infty} \frac{a_n}{b_n} = 0.$$

It is important to note that $a_n = o(1)$ means that $\lim_{n\to\infty} a_n = 0$. This is relevant in the section on graph girth and chromatic number.

2.2 Ramsey Theory

Ramsey theory is a branch of combinatorics that studies where order can appear in a random object. There is a nice result on Ramsey theory that can be proved easily using the probabilistic method.

Definition 2.3 (Ramsey (Diagonal) Number)**.** *The k-th Ramsey number, R_k, is the smallest number n such that any 2-coloring of the edges of the complete graph on n vertices, K_n, must contain a monochromatic k-clique.*

For example, $R_3 = 6$, as K_6 cannot be 2-colored without creating a monochromatic triangle.

It is known that Ramsey numbers exist for all k and that the upper bound for R_k is 2^{2k-3}. We now prove a lower bound for R_k using the probabilistic method.

Theorem 2.4. $R_k > 2^{k/2}$.

Proof. Let $n = 2^{k/2}$. We show that there exists a 2-coloring of the complete graph K_n, such that there is no monochromatic k-clique. Randomly color each edge of K_n red or blue with probability $1/2$, independently. Let C be any k-clique in the graph K_n. Then,

$$\begin{aligned}\Pr[C \text{ is monochromatic}] &= (\text{Number of colors})\Pr[\text{C is that color}]\\ &= 2\cdot 2^{-\binom{k}{2}}\\ &= 2^{1-\binom{k}{2}}.\end{aligned}$$

Since the total number of k-cliques in K_n is $\binom{n}{k}$, by the union bound, we get

$$\begin{aligned}\Pr[K_n \text{ has a monochromatic } k\text{-clique}] &\le \binom{n}{k}\cdot \Pr[\text{a given } k\text{-clique is monochromatic}]\\ &= \binom{n}{k}\cdot 2^{1-\binom{k}{2}}\\ &\le \frac{n^k}{k!}\cdot 2^{1-\binom{k}{2}}.\end{aligned}$$

For $n = 2^{k/2}$, we have

$$\begin{aligned}\Pr[K_n \text{ has a monochromatic } k\text{-clique}] &\le \frac{\left(2^{k/2}\right)^k}{k!}\cdot 2^{1-\frac{k^2-k}{2}}\\ &= \frac{1}{k!}\cdot 2^{\frac{k+2}{2}}\\ &< 1 \quad \text{for } k \ge 3.\end{aligned}$$

So, there is a nonzero probability a randomly-colored K_n does not have a monochromatic clique, so there exists some coloring without one. □

Ramsey numbers have been the topic of extensive study in combinatorics. For this reason, it may come as a surprise that this bound provided by a random selection is almost as strong as bounds produced by more sophisticated deterministic methods.

2.3 Max Cut

We consider the max cut problem, which is the problem of partitioning the vertices of a graph G into two parts P_1, P_2 such that the number of edges crossing the partition is maximized. We denote the edges of G by E, and the vertices by V. This is a well-known NP-complete problem. Even so, using the probabilistic method, we can give a lower bound on the size of the maximum cut.

Theorem 2.5. *For any graph $G = (E, V)$, there exists a cut containing at least $\frac{|E|}{2}$ edges.*

Proof. We pick a random partition, assigning each vertex to P_1 or P_2 independently with probability $\frac{1}{2}$. For each edge $e \in E$, define an indicator variable X_e as 1 if e is cut, and 0 otherwise. Let's begin by considering the expectation of a single indicator variable X_e.

$$\mathbb{E}[X_e] = \Pr[e \text{ is cut}] = \Pr[\text{the endpoints of } e \text{ are in different partitions}] = \frac{1}{2}.$$

Now, let X be a variable denoting the number of cut edges of a certain partition, and note that $X = \sum_{e \in E} X_e$. By linearity of expectation,

$$\mathbb{E}[X] = \sum_{e \in E} \mathbb{E}[X_e] = \frac{|E|}{2}.$$

Here, we can use properties of the probabilistic method. Because the expectation of X is $|E|/2$, there must exist at least one partition such that $X \geq |E|/2$. □

2.4 Independent Set

In a graph G, a subset of vertices $S \subset V$ is said to be an *independent set* if no two vertices $u, v \in S$ are adjacent in G. The problem of determining the size of the largest independent set in G is NP-hard. However, we can once again apply the probabilistic method to establish a good lower bound on the size of the maximum independent set for any graph.

Theorem 2.6. *Any graph $G = (V, E)$ contains an independent set $S \subset V$ such that $|S| \geq \sum_{v \in V} \frac{1}{\deg(v)+1}$ where $\deg(v)$ is the number of vertices adjacent to v in G.*

Proof. Assign a random weight w_v to each vertex $v \in V$, choosing the weights independently and uniform from the interval $[0, 1]$. Call $v \in V$ a local minimum if $w_v < w_u$ for each vertex u adjacent to v. Since no two adjacent vertices can be local minima, the set of local minima form an independent set. Additionally, any vertex among v and its neighbors are equally likely to have minimum weight. Therefore, for each v, the probability that v is a local minimum is $\frac{1}{\deg(v)+1}$.

Now, we again use indicator variables for each vertex. Let X be the number of local minima, and X_v be an indicator random variable for the event that v is a local minimum. By linearity of expectation,

$$\mathbb{E}[X] = \sum_{v \in V} \mathbb{E}[X_v] = \sum_{v \in V} \frac{1}{\deg(v) + 1}.$$

Therefore, there must be at least one independent set of that size. □

2.5 Graphs with large girth and chromatic number

The following is a result from Erdös in 1959. Let G be a graph with vertices V and edges E.

Definition 2.7. *The **girth** of G is the length of the shortest cycle in G.*

For example, a connected 2-regular graph of size n has girth n for $n \geq 2$. A complete graph of any size ≥ 3 has girth 3. A graph representing a grid where the intersections are vertices, with the lines edges, has girth 4.

Definition 2.8. *The **chromatic number** of G is the minimum number of colors needed to color the vertices of G such that no two adjacent vertices have the same color.*

It has been proved that the chromatic number of a planar graph is at most 4. The chromatic number of the complete graph on n vertices is n.

It should seem natural that a graph with a large girth should have a small chromatic number. It should be surprising, then, that for any girth and chromatic number, there exist graphs with at least that large of a girth and chromatic number. The proof is, once again, the probabilistic method.

Theorem 2.9. *For any positive integers k, l, there exists a graph with girth $\geq l$ and chromatic number $\geq k$.*

Proof. Pick a random graph G from $\mathcal{G}(n, p)$, where n is the number of vertices and each of the $\binom{n}{2}$ possible edges is included independently with probability p. We will choose $p = n^{1/l-1}$. This results in relatively sparse (low ratio of edges to vertices) random graphs.

Let X be the number of cycles of length $< l$ in G. Note that if $X = 0$, then the girth of G is at least l, which is what is desired. The number of possible cycles of length i is

$$\binom{n}{i} \cdot \frac{i!}{2i}$$

since for every subset of size i of the n vertices, there are $i!$ possible cycles. However, as direction and starting position do not matter, we divide by $2i$ to get the number of possible cycles. A cycle is only present if all

edges are present, which happens with probability p^i. Using this, we can now look at $\mathbb{E}[X]$. We use the indicator variable argument seen previously to find

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=3}^{l-1} \binom{n}{i} \cdot \frac{i!}{2i} \cdot p^i \\ &\leq \sum_{i=3}^{l-1} \frac{n^{i/l}}{2i} \\ &= o(n),\end{aligned}$$

The second expression follows from the fact that $\binom{n}{i} \cdot i! = \frac{n!}{(n-i)!} \leq n^i$ and $n^i p^i = n^i p^i = n^i n^{i/l-i} = n^{i/l}$. The last expression means that the expectation of X, thought of as a sequence in terms of n, is of an order of growth smaller than n. We can check this by seeing that

$$\sum_{i=3}^{l-1} \frac{n^{i/l}}{2i} \leq n^{\frac{l-1}{l}} \sum_{i=3}^{l-1} \frac{1}{2i}$$

which goes to 0 when divided by n as n grows large. Thus, it is in little-oh of n. By Markov's inequality,

$$\Pr(X \geq n/2) \leq \frac{o(n)}{n/2} \in o(1).$$

Therefore, the probability that G contains $n/2$ or more cycles of length l goes to 0 as n gets large.

We now consider the chromatic number. We can lower bound the number of colors k as

$$k \geq \frac{|V|}{\text{max. independent set size}}$$

as the set of vertices that receive any given color must form an independent set.

Let Y be the size of the maximal independent set in G. Then, by the union bound,

$$\begin{aligned}\Pr(Y \geq y) &\leq \binom{n}{y} \cdot (1-p)^{\binom{y}{2}} \\ &\leq (n \cdot \exp(-p(y-1)/2))^y \\ &\in o(1) \text{ if we set } y = \frac{3}{p} \ln n\end{aligned}$$

The second expression follows from the fact that $\binom{n}{y} \leq n^y$ and the inequality $1 + x \leq e^x$.

Putting the above calculations together, by taking n large enough we can ensure that the probabilities $\Pr(X \geq n/2)$ and $\Pr(Y \geq \frac{3}{p} \ln n)$ are both less than $1/2$, so by a union bound $\Pr(X \geq n/2 \cup Y \geq \frac{3 \ln n}{p}) < 1$. Taking a graph that follows these properties, which we have now shown exists, we can remove one vertex from every cycle of length $< l$, resulting in a graph G' such that

(i) G' has girth $\geq l$

(ii) G' has $\geq \frac{n}{2}$ vertices

(iii) G' has maximum independent set size $< \frac{3}{p} \ln n$.

Therefore, G' has chromatic number $\geq \frac{n/2}{\frac{3 \ln n}{p}} = \frac{n^{1/l}}{6 \ln n} \to \infty$ as $n \to \infty$. So by again making n larger, we can make this at least k, as desired. □

Note that, unlike in previous proofs, the graph we have shown to exist is not the graph with the desired property. Instead, we have had to take this graph that we can show exists, and then directly modify in order to complete our proof.

2.6 (Extra) Lovász Local Lemma

When working with the probabilistic method, we often want to show that we can avoid certain conditions with probability greater than 0. Let the conditions be denoted by $\mathcal{E}_1, \ldots, \mathcal{E}_n \subset \Omega$. Notationally, we want that

$$\Pr(\bigcap_{i=1}^{n} \overline{\mathcal{E}_i}) > 0.$$

If $\sum_i \Pr(\mathcal{E}_i) < 1$, then we can use the union bound to prove this. However, this is a very strong condition, as the sum can be much larger than 1 even if there is a chance of avoiding all conditions. The best case scenario is when the events are all independent; in this case, no matter what the sum of probabilities is, if all events have $p < 1$ then there is some probability that they can all not occur. The Lovasz local lemma takes this notion further: it tries to find a similar notion for events that are "mostly independent."

We first define a notion of mutual independence.

Definition 2.10 (Mutual Independence)**.** *For all integers $n > 0$, define $[n] := \{1, \ldots, n\}$. Given events $\mathcal{E}_1, \ldots, \mathcal{E}_n \subset \Omega$ and a subset $J \subset [n]$, the event $\mathcal{E}_i$ is said to be* ***mutually independent*** *of $\{\mathcal{E}_j : j \in J\}$ if for all choices of disjoint subsets $J_1, J_2 \subset J$,*

$$\Pr\left(\mathcal{E}_i \cap \bigcap_{j_1 \in J_1} \mathcal{E}_{j_1} \cap \bigcap_{j_2 \in J_2} \overline{\mathcal{E}}_{j_2}\right) = \Pr(\mathcal{E}_i) \cdot \Pr\left(\bigcap_{j_1 \in J_1} \mathcal{E}_{j_1} \cap \bigcap_{j_2 \in J_2} \overline{\mathcal{E}}_{j_2}\right)$$

Intuitively, this means that $\mathcal{E}_i$ is independent from every possible combination of events occurring. We can use this notion to define the (symmetric) Lovasz local lemma.

Theorem 2.11 (Lovasz Local Lemma)**.** *Suppose $p \in (0, 1)$, $d \geq 1$, and $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are events such that $\Pr(\mathcal{E}_i) \leq p$ for all i. If each $\mathcal{E}_i$ is mutually independent of some set of all but d other events $\mathcal{E}_j$, and $ep(d + 1) \leq 1$, where e is Euler's number, then $\Pr(\cap_{i=1}^{n} \overline{\mathcal{E}}_i) > 0$.*

In the interest of brevity, we do not prove the Lovasz local lemma. However, we introduce it as it is an extremely powerful tool for probabilistic method results. Especially of note, and the reason for the term 'local', is the fact that the lemma depends only on the degree of interconnectedness of the events, d, and not on the number of events n.

2.6.1 Example

Suppose $11n$ points are placed around a circle and colored with n different colors such that each color is applied to exactly 11 points.

Lemma 2.12. *In any such coloring, there must be a set of n points containing one point of each color but not containing any pairs of adjacent points.*

Proof. We construct a set by picking a point of each color randomly, with all points equally likely (having probability $1/11$) to be chosen. The $11n$ events we want to avoid correspond to the $11n$ pairs of adjacent points on the circle. For each pair, the odds of picking both points in that pair is at most $1/121$ (exactly $1/121$ if the two points are of different colors, and 0 otherwise). Thus, we take $p = \frac{1}{121}$.

Whether a given pair of points is chosen depends only on the colors of those points, and not on the collection of points chosen in the other $n-2$ colors. Thus, the event "a and b are both chosen" is dependent only on those pairs of adjacent points which share a color with either a or b.

There are 21 pairs other than (a, b) which include the same color as a, and the same number for b. In the worst case scenario, they are disjoint. In this case, there are at most 42 pairs of points whose selection can affect the probability of (a, b) being selected. We thus use $d = 42$. This gives $ep(d+1) = e(1/121)(42+1) \approx .97 < 1$. By the local lemma, there is thus a positive probability that none of these events occur, so a set satisfying our conditions must exist. □

References

A large portion of these notes were near directly copied from scribe notes on a lecture by Alistair Sinclair in CS271 from spring 2018. The notes can be accessed here.

The section on the Lovasz local lemma was adapted from these notes from Stanford's Math 233. The example is taken from the wikipedia page for the Lovasz local lemma.

An additional reference for the large girth and chromatic number graphs can be found here in this paper by Cheuk To Tsui.