Probabilistic Method

Sam Heil

29 March 2019

Main idea: For some existence problems in graph theory, explicitly constructing an object is difficult, so it is best to select things randomly and show that it must exist in some case.

There are essentially two ideas that make most of these proofs work:

1. Linearity of Expectation: $\mathbb{E}[\sum_{n=1}^{N} X_n] = \sum_{n=1}^{N} \mathbb{E}[X_n]$.
2. "Existence from Expectation": If $\mathbb{E}_x[f(x)] = C$, then there are x_1, x_2 with $f(x_1) \leq C, f(x_2) \geq C$. Moreover, if f is integer valued, then there are x_1, x_2 with $f(x_1) \leq \lfloor C \rfloor$ and $f(x_2) \geq \lceil C \rceil$.

To apply linearity of expectation, it's often useful to take indicator variables X_A, where $X_A = 1$ if event A holds and $X_A = 0$ otherwise.

These methods are very problem-specific, so it's best to see them by starting with some problems.

Problem 1 (Vertex Cover)**.** *Given a graph G on n vertices with vertex set V, what is the smallest k such that there is a vertex set $U \subset V$ with $|U| = k$ such that every vertex $v \in V \setminus U$ is adjacent to some vertex $u \in U$?*

This problem is NP-complete in general. However, if we include a lower bound on the degrees of the vertices, we can get an upper bound on k using probability!

Theorem 1. *Let G be a graph on n vertices, such that each vertex has degree at least d. Let V be the set of vertices of G. Then, there is some subset $U \subset V$ such that every vertex $v \in V \setminus U$ is adjacent to some $u \in U$, for which*

$$|U| \leq \frac{n(1 + \log(d+1))}{d+1}.$$

Proof. It's really hard to construct U explicitly; in fact, we just saw that a similar problem is NP-complete.

Because of this, we'll just choose U randomly! Let p be some probability (to be named later), and for each $v \in V$, say that $v \in U_0$ with probability p. (Note: U_0 is approximately going to be U, but we'll have to add a few elements separately.)

It's certainly possible that U_0 won't have the desired property, since it's just something we put together randomly. If we add all the counterexamples (vertices in $V \setminus U_0$ with no neighbor in U_0) to the set U_0 though, then this must have the desired property. Let C be this set of counterexamples:

$$C = \{v \in V \setminus U_0 : (u, v) \in E(G) \implies u \notin U_0\}.$$

We'll take $U = U_0 \cup C$ and find the size of U.

Now, let v_i be any vertex, chosen a priori. The probability that we get a counterexample from vertex v_i for a specific U_0 generated according to the random process above is then equal to $(1-p)^{1+\deg(v_i)} \leq (1-p)^{d+1}$, since we need that $v_i \notin U$ and also that every neighbor of v_i is not in U_0.

Let C be the set of counterexamples with the set U_0 (as defined above). We have that the probability that each v_i is a counterexample is at most $(1-p)^{d+1}$, so the expected number of counterexamples is at most $n(1-p)^{d+1}$ by linearity of expectation. Since the expected size of U_0 itself is np, we have that

$$\mathbb{E}[|U|] = \mathbb{E}[|U_0|] + \mathbb{E}[|C|] = np + n(1-p)^{d+1}.$$

Therefore, there must exist some set U with the desired property for which $|U| \leq n(p + (1-p)^{d+1})$. We can choose $p = \text{argmin}_{0<t<1} t + (1-t)^{d+1}$ to get $|U| \leq \frac{n(1+\log(d+1))}{d+1}$. ∎

Theorem 2. *Let A be a set of n residues mod n^2. Then, there is a set B of n residues mod n^2 such that all but $\lfloor n^2/e \rfloor$ residues r mod n^2 can be written as $r = a + b$ for $a \in A, b \in B$.*

Proof. Let $A = \{a_1, a_2, \ldots, a_n\}$.

Since we have n different parameters that can be arbitrary here, it seems like it will be difficult to directly construct a set B in terms of A, so we will consider random choices of B to show that one exists.

Let X be a random variable counting the number of residues achieved for a random choice of B; we will compute $\mathbb{E}[X]$. Let $B = \{b_1, b_2, \ldots, b_n\}$, where $b_1, ..., b_n$ are selected independently from the uniform distribution on the set of residues $\{0, ..., n^2 - 1\}$ mod n^2. (If there are duplicates, we can always add more elements to get a strictly larger number of residues.)

Let X_r be the probability that some pair achieves a residue r. Then, by linearity of expectation $\mathbb{E}[X] = \sum_{r=0}^{n^2-1} E[X_r]$.

To find $\mathbb{E}[X_r]$, we start by finding the probability that there is some pair (a, b) with $a \in A, b \in B$ and $a + b = r$ using b_1 as our b. There are n possible values of b_1 that work ($n^2 - a_i$ for each i), so we've got a $n/n^2 = 1/n$ chance of getting each of them.

The probability that at least one of the b_j's takes one of these values $a_1, ..., a_n$ is $1-$(probability that none of them do), so our total probability of X_r being 1 is $1 - (1 - 1/n)^n$.

Thus, we get that $\mathbb{E}[X] = n^2\left(1 - \left(1 - \frac{1}{n}\right)^n\right)$. Therefore, there must exist a set X such that all but $n^2(1 - \frac{1}{n})^n$ of the residues are elements of X. Now, since $\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = 1/e$ and the function $f(x) = \left(1 - \frac{1}{x}\right)^x$ is strictly increasing, we have that $\left(1 - \frac{1}{n}\right)^n \leq \frac{1}{e}$. Since X is an integer, we get that at most $\lfloor n^2/e \rfloor$ elements are excluded. □

The following claim will be useful for the next problem. This is analogous to using the fact that for a nonnegative integer-valued random variable X, if $\mathbb{E}[X] > 0$ then $\Pr[X > 0] > 0$, but provides a stronger bound on the probability than it simply being positive.

Lemma 3. *Let $X_1, X_2, X_3, \ldots$ be a sequence of random variables. Then, if $Var(X_n) = o(\mathbb{E}[X_n]^2)$, then $\lim_{n\to\infty} Pr(X_n > 0) = 1$.*

Proof. Let $\mu_n = \mathbb{E}[X_n], \sigma_n^2 = \mathrm{Var}(X_n)$. Then, by Chebyshev's Inequality, we have that

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

Since $\mathrm{Var}(X_n) = o(\mathbb{E}[X_n]^2)$, for any $\epsilon > 0$ we have that $\Pr(X_n = 0) < \epsilon$ for sufficiently large n, giving the desired result. □

Theorem 4. *Let $G(n, p)$ be the random graph on n vertices where an edge is included with probability p.*

Then, the probability that $G(n, p)$ contains K_4 as a subgraph approaches 0 as $n \to \infty$ for $p = o(n^{-2/3})$ and approaches 1 as $n \to \infty$ for $p = \omega(n^{-2/3})$.

Proof. For each subset $S \subset V(G)$ with $|S| = 4$, let X_S be the indicator variable taking value 1 if the subgraph with vertices in S is a complete graph, and let $X = \sum_{S\subset V(G),|S|=4} X_S$.

Then, using linearity of expectation, we have that

$$\mathbb{E}[X] = \binom{n}{4}\mathbb{E}[X_S] = \binom{n}{4} \cdot p^{\binom{4}{2}} = \frac{n(n-1)(n-2)(n-3)}{24}p^6 = \Theta(n^4p^6).$$

If $p = o(n^{-2/3})$, then we have that $\mathbb{E}[X] = o(1)$, so for any $\epsilon > 0$ there exists n such that if $|G| = n$, then $\mathbb{E}[X] < \epsilon$. Since X is integer-valued, this implies that $\mathbb{P}(X = 0) > 1 - \epsilon$, so the probability approaches 1, as desired.

The other case is more complicated, because we have to be careful about the case where there are multiple cliques.

We'll use Lemma 3: it suffices to show that $\mathrm{Var}(X_n) = o(\mathbb{E}[X_n])^2$, where X_n is the random variable X when we choose G to be a random graph on n vertices. By abuse of notation we'll use X for X_n.

We have that

$$\mathrm{Var}(X) = \mathrm{Var}(\sum_S X_S) = \sum_S \mathrm{Var}(X_S) + \sum_{S\neq T} \mathrm{Cov}(S, T).$$

Since X_S is 1 with probability p^6 and zero otherwise, we have that $\operatorname{Var}(X_S) = p^6(1-p^6)$, so $\sum_S \operatorname{Var}(X_S) = \binom{n}{4}p^6(1-p^6) \sim n^4p^6$.

Now, we'll compute $\operatorname{Cov}(S,T) = \mathbb{E}[X_S X_T] - \mathbb{E}[X_S]\mathbb{E}[X_T]$. If $S \cap T \leq 1$, there are no overlapping edges, so X_S and X_T are independent, and this covariance is zero. Otherwise we have that $\operatorname{Cov}(S,T) \leq \mathbb{E}[X_S X_T] = \Pr(X_S = X_T = 1)$. This is equal to p^k, where k is the number of edges either between two vertices in S or two vertices in T.

We can compute this when $k = 2, 3$ (as shown), and adding things up we end up with $\operatorname{Var}(X) = o(\mathbb{E}[X]^2)$, so we conclude that $\Pr(X_n > 0) \to 1$ as $n \to \infty$, as desired. $\square$

1 Problems

1. Prove that if there is some $p \in [0,1]$ with

$$\binom{n}{k}p^{\binom{k}{2}} + \binom{n}{t}(1-p)^{\binom{t}{2}} < 1,$$

   then the Ramsey number $R(k,t)$ satisfies $R(k,t) > n$. ($R(k,t)$ is the minimal N such that any complete two-colored graph on N vertices contains either a K_k of the first color or a K_t of the second.)

2. Consider a parliament with 1600 members who have formed $16,000$ committees of 80 members each. Show that there exist two committees that each contain at least 4 common members.

3. Suppose $n \geq 4$ and let H be a hypergraph with all edges subsets of size n such that there are at most $4^{n-1}/3^n$ edges Prove there is a 4-coloring of the vertices of H so that all four colors appear in every edge.

4. Prove that if G is a graph with n vertices and m edges, then G must contain a bipartite subgraph with at least $\frac{m}{2}$ edges.

References: (Problems/theorems are my favorite examples taken from these references.)

1. N. Alon and J.H. Spencer. *The Probabilistic Method*, 3rd edition.

2. Po-Shen Loh, 2010 MOP handout: `http://www.math.cmu.edu/~lohp/docs/math/mop2010/prob-comb-soln.pdf`

3. Evan Chen. `http://web.evanchen.cc/handouts/ProbabilisticMethod/ProbabilisticMethod.pdf`