CONSTRUCTION AND COUNTING WITH THE PROBABILISTIC METHOD - HUJI MATH STUDENT'S SEMINAR

MICHAEL SIMKIN

1. Introduction

The probabilistic method is a powerful tool introduced to combinatorics in the middle of the previous century. In this lecture we'll learn a little about it via two examples. The first is a classic construction, due to Paul Erdőos, of graphs with certain properties, for which no explicit construction is known. In fact, it is this construction that is largely responsible for the popularization of the probabilistic method. The second application is to obtain an upper-bound on the number of Latin squares, which we'll define later. This uses a more cutting-edge technique, which is still being developed.

2. Ramsey Numbers

Here is a riddle that made the rounds while I was in high school:

Riddle 2.1. *Say* 6 *people go to a party. Prove that one of the following holds:*

(a) There are 3 *people who are all mutually acquainted.*
(b) There are 3 *people none of whom know each other.*

The truth is that to this day I don't feel this is a legitimite riddle. It feels more like a motivation for graph theory. At any rate, I think I went through the following thought process:

(a) Let's model the problem as a graph. The graph will have 6 vertices, corresponding to the guests. We'll put an between two guests if they know each other. So now we have to prove that there is either a triangle or an "anti-triangle".
(b) Hmm... There are $\binom{6}{2} = 15$ potential edges, so there are 2^{15} possible graphs. So brute force is a little difficult (though we could use a computer). But actually, there are a lot of isomorphisms, and this makes it feasible to check all the possibilities. This is where I got to in high school.
(c) The truth is that there is a slightly more elegant approach: Pick a vertex. There are five other vertices, so either it has three neighbors or it has three "anti-neighbors". W.l.o.g. there are three neighbors. If any of them know each other we have a triangle; otherwise they themselves form an anti-triangle.

As mathematicians we're always looking to generalize things, and this one is very easy. We just need two basic graph-theoretic definitions:

Definition 2.1. Let $G = (V, E)$ be a graph. $U \subseteq V$ is a clique if for all $u \neq v \in U$, $uv \in E$. U is an **anti-clique** (or **independent set**) if for all $u \neq v \in U$, $uv \notin E$.

Date: 7/6/2017.

Theorem 2.2 (Ramsey 1930)**.** *For every $k \in \mathbb{N}$ there exists some $N \in \mathbb{N}$ s.t. for all $n \geq N$, every graph on n vertices contains either a clique of size k or an independent set of size k.*

I'm not going to prove this; it's a fun exercise. One thing the standard proof gives is an upper bound on $N(k)$ - the smallest N s.t. the conclusion of Theorem 2.2 holds:

$$N(k) \leq \binom{2k}{k} \leq 4^k$$

To date, this has only been improved up to subexponential multiplicative factors.

What about a lower bound? The obvious thing to do is to construct large graphs with no cliques or anti-cliques of a specified size. Unfortunately, there is no known deterministic construction of a graph with α^k vertices and no clique or anti-clique of size k, for any $\alpha > 1$. On the other hand, it's very easy to prove such graphs exist, using the probablistic method:

Theorem 2.3 (Erdőos 1947)**.** *$N(k) \geq \frac{1}{2\cdot\sqrt{2}}\sqrt{2}^k$.*

The proof essentially says: If you're trying to give an explicit construction, you're working too hard. If you just choose any old graph at random, it will probably have the desired property. In particular, such a graph exists.

Proof. Let $k \in \mathbb{N}$ and let $n < \frac{1}{2\cdot\sqrt{2}}\sqrt{2}^k$. We'll define a probability space as follows: Ω is the set of all graphs on n vertices, with the uniform distribution. There are $2^{\binom{n}{2}}$ possible graphs on n vertices, so each one has probability $2^{-\binom{n}{2}}$ of being selected. Amazingly, this distribution is the same as the following: Start with K_n, and iterate over all $\binom{n}{2}$ edges. For each one, keep it with probability $\frac{1}{2}$ and delete it with probability $\frac{1}{2}$, all choices independent. This model is known as a $G\left(n, \frac{1}{2}\right)$-random graph (the $\frac{1}{2}$ is the probability that a given edge survives, and can be replaced by any parameter appropriate for the problem at hand). In this way, for any fixed subset of m edges of K_n, the probability that all of them survive is 2^{-m}. Similarly, the probability that none survive is 2^{-m}.

So let $G \sim G\left(n, \frac{1}{2}\right)$.

Next, we define a random variable X, which is the number of k-cliques in G. We're going to show that $\mathbb{E}X \ll 1$, which we'll show implies that with high probability $X = 0$.

Let $\alpha_1, \ldots, \alpha_{\binom{n}{k}}$ be an enumeration of the k-cliques in K_n. Let X_i be the indicator random variable of the event that the ith clique survived. Then $\mathbb{E}X_i = 2^{-\binom{k}{2}}$, because a k-clique has k edges.

So $X = \sum_{i=1}^{\binom{n}{k}} X_i$. Thus, by linearity of expectation:

$$\mathbb{E}X = \sum_{i=1}^{\binom{n}{k}} \mathbb{E}X_i = \binom{n}{k} 2^{-\binom{k}{2}} \leq \left(\frac{n}{2^{\frac{k-1}{2}}}\right)^k < \left(\frac{2^{\frac{k}{2}}}{2\cdot\sqrt{2}\cdot 2^{\frac{k-1}{2}}}\right)^k \leq \frac{1}{2}$$

So the expected number of k-cliques is less than $\frac{1}{2}$. What is the probability that there are no k-cliques? Well, here we apply Markov's inequality:

Theorem 2.4 (Markov's inequality)**.** *Let Y be a non-negative random variable. Then, for every $c > 0$:*

$$\mathbb{P}\left[Y \geq c\right] \leq \frac{\mathbb{E}Y}{c}$$

X is a non-negative random variable, so we conclude:

$$\mathbb{P}\left[X \geq 1\right] < \frac{1}{2}$$

An analogous calculation shows that the probability that G contains a size k anti-clique is also less than $\frac{1}{2}$. By a union bound, the probability that G contains either a k-clique or a k-anti-clique is strictly less than 1.

Since with positive probability G doesn't satisfy the conclusion of Ramsey's theorem, there must exist such a graph on n vertices, implying $N\left(k\right) \geq \frac{1}{2\cdot\sqrt{2}}\sqrt{2}^{k}$. □

Remark 2.5. To this day, the bound on $N\left(k\right)$ has only been improved by subexponential multiplicative factors.

Remark 2.6. You might reasonably object that there was nothing inherently probabilistic in either the problem or the proof. The problem is a question about the existence of a finite graph with certain properties, and the proof, though dressed up as an analysis of random variables, is just a counting argument. We could just as easily have counted the total number of graphs on n vertices, and the the total number of graphs containing a specific k(-anti)-clique, and discovered that the former is way larger than the latter. In theory, this can be said of pretty much any argument involving the probabilistic method, but in practice it becomes very cumbersome to talk about martingales and entropy in a non-probabilistic setting.

An additional advantage of the probabilistic approach in the above problem is that it yields an algorithm for *finding* graphs with no large (anti-)cliques: Just flip a coin for each edge to determine its existence; a more careful analysis of the algorithm above tells us that the probability of containing a large clique is much smaller (for example) than the probability that a piano will fall on my head the next time I walk out of the building.

Which brings us to another point: There are *so many* graphs satisfying the no-large-clique property, that it's astounding that no non-probabilistic constructions are known. A philosopher might have something to say about the mismatch between our expectations and reality, or on the other hand, on the gap between theoretical math and reality...

3. An Upper Bound on the Number of Latin Squares via Entropy

An order n Latin square is an $n \times n$ matrix in which each row and each column is a permutation of $[n]$. For example:

Latin squares are well-known in other branches of science because they have applications in the design of experiments and in coding theory. Also, they're similar to Sudoko, so they're easy to describe to non-academics. They're also interesting mathematical objects in their own right, and were studied by Euler (and possibly before). Perhaps the most basic mathematical question is, "How many order n Latin squares are there?"

This question was answered by van Lint and Wilson in their book *A Course in Combinatorics*. Let L_n be the number of order n Latin squares.

Theorem 3.1 (van Lint, Wilson 1992)**.** $|L_n| = \left((1+o(1))\frac{n}{e^2}\right)^{n^2}$

By this I mean that if $f(n)$ satisfies the equation $|L_n| = \left((1+f(n))\frac{n}{e^2}\right)^{n^2}$ then $\lim_{n\to\infty} f(n) = 0$.

The original proof used bounds on something called the *permanent* of a matrix - if I have time at the end I can briefly describe this. The upper and lower bound are each an application of a non-trivial theorem, and are proved separately.

We'll prove the upper bound using the *entropy method*, which is an application of the probabilistic method. This is not the original proof, but it's inspired by it. The approach we take here was developed fully by Nati Linial and Zur Luria who then applied it to obtain upper bounds in many other counting problems.

We'll need a few definitions:

Definition 3.2. Let X be a random variable with finite range $R(X)$. The **entropy** of X is:

$$H(X) = -\sum_{x\in R(X)} \mathbb{P}[X=x]\log(\mathbb{P}[X=x]) =$$

By log I mean ln.

Intuitively, the entropy is the number of bits needed on average to describe the value of X. Note that entropy depends only on X's distribution, and not at all on its range. Here are a few properties of entropy:

Proposition 3.3. *$H(X) \le \log|R(X)|$ with equality iff X is distributed uniformly.*

Proof. If X is uniform, then:

$$H(X) = |R(X)|\frac{1}{|R(X)|}\log(|R(X)|) = \log(|R(X)|)$$

The "only if" part of the claim follows from the fact that log is a strongly concave function. □

Our upper bound on the number of Latin squares is going to work as follows: We'll let X be a random variable that chooses an order n LS uniformly at random. Then $H(X) = \log L_n$. It then suffices to give a good upper bound on $H(X)$. For this we'll need two more concepts and a proposition.

Definition 3.4. Let X, Y be random variables. Their **joint entropy** $H(X,Y)$is the entropy of (X,Y) considered as a single random variable.

The **conditional entropy** of X given Y is:

$$H(X|Y) = \sum_{y\in R(Y)} p_y H(X|Y=y) = \mathbb{E}_{y\sim Y}[H(X|Y=y)] = \mathbb{E}_{x,y\sim X,Y}[H(X|Y=y)]$$

Intuitively, the conditional entropy measures the average (over X and Y) number of bits needed to describe the outcome of X if the outcome of Y is known.

A useful fact about conditional entropy is the *chain rule*:

Proposition 3.5. *Let $X_1, \dots, X_n$ be random variables with finite range. Then:*

$$H\left(X_1, X_2, \dots, X_n\right) = \sum_{[i=1]}^{n} H\left(X_i | X_1, \dots, X_{i-1}\right)$$

Proof. Induction. □

We're now ready to count Latin squares:

Proposition 3.6. $L_n \leq \left(\left(1 + o\left(1\right)\right) \frac{n}{e^2}\right)^{n^2}$.

Proof. Let X be an order n Latin square chosen uniformly at random. Then:

$$H\left(X\right) = \log L_n$$

For $1 \leq i, j \leq n$, let $X_{i,j}$ be the value of X at position i, j. Then $X_{i,j}$ is also a random variable, and in fact:

$$X = \left(X_{1,1}, \dots, X_{n,n}\right)$$

Then, by the chain rule, for any ordering $\left(i_1, j_1\right), \dots, \left(i_{n^2}, j_{n^2}\right)$ of $\left[n\right]^2$ we have:

$$H\left(X\right) = \sum_{\ell=1}^{n^2} H\left(X_{i_\ell, j_\ell} | X_{i_1, j_1}, \dots X_{i_{\ell-1}, j_{\ell-1}}\right)$$

This remains true even if the ordering is itself a random variable. So, to each i, j associate a number $z_{i,j} \in [0, 1]$ chosen uniformly and independently, and order $\left[n\right]^2$ in decreasing order. Then:

$$H\left(X\right) = \mathbb{E}_z \sum_{i,j \in [n]^2} H\left(X_{i,j} | X_{k,\ell} : z_{k,\ell} > z_{i,j}\right) = \sum_{i,j \in [n]^2} \mathbb{E}_z \mathbb{E}_{x \sim X} H\left(X_{i,j} | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right)$$

By Fubini we can change the order of integration:

$$H\left(X\right) = \sum_{i,j \in [n]^2} \mathbb{E}_{x \sim X} \mathbb{E}_z H\left(X_{i,j} | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right)$$

We now come to the nice part: We want to bound the entropy. Well, we know that the entropy is maximized by the uniform distribution. We'll use this. Given $x \sim X$ and z, $\left(X_{i,j} | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right)$ is a random variable taking values on a subset of $[n]$. But we might be able to say something about this subset: if, for example, one of the previously revealed variables is in row i, then $X_{i,j}$ can't have the same value. So call $v \in [n]$ *available* if it hasn't yet appeared in row i or column j. Let $A\left(i, j | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right)$ be the number of available values, given the previously revealed variables. Then:

$$\mathbb{E}_{x \sim X} \mathbb{E}_z H\left(X_{i,j} | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right) \leq \mathbb{E}_{x \sim X} \mathbb{E}_z \log A\left(i, j | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right)$$

But log is convex, so:

$$\mathbb{E}_{x \sim X} \mathbb{E}_z H\left(X_{i,j} | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right) \leq \mathbb{E}_{x \sim X} \mathbb{E}_{z_{i,j}} \log\left(\mathbb{E}_{z_{k,\ell} : k,\ell \neq i,j} A\left(i, j | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right)\right)$$

What can we say about the inner expectation? Well, the Latin square x is fixed, so $x_{i,j}$ is always available. As for the other $n-1$ values, each of them can become unavailable in two different ways, so they are available with probability $z_{i,j}^2$. Thus:

$$\mathbb{E}_{z_{i,j}} \log \left(\mathbb{E}_{z_{k,\ell}:k,\ell \neq i,j} A\left(i,j | X_{k,\ell} = x_{k.\ell} : z_{k,\ell} > z_{i,j}\right)\right) \leq \int_0^1 \log\left(1+\left(n-1\right)z^2\right) dz$$

According to Wolfram:

$$\int_0^1 \log\left(1+\left(n-1\right)z^2\right) dz = \log n - 2 \pm o\left(1\right)$$

Thus:

$$H\left(X\right) \leq n^2 \left(\log n - 2 \pm o\left(1\right)\right)$$

Hence:

$$L_n \leq \left(\left(1+o\left(1\right)\right)\frac{n}{e^2}\right)^{n^2}$$

□

Institute of Mathematics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

E-mail address: menahem.simkin@mail.huji.ac.il