2 Lecture 2, October 1, 2014

2.1 Application: the probabilistic method

A *tournament* of size n is a directed complete graph. We may think of a tournament T equivalently as a skew-symmetric mapping $T : \{1, \ldots, n\} \times \{1, \ldots, n\} \to \{1, 0, -1\}$ that is 0 only on the diagonal.

A *Hamilton path* in a tournament (or a digraph more generally) is a directed simple path through all the vertices.

Lemma 1 *There exists a tournament with at least $n!2^{-n+1}$ Hamilton paths.*

This certainly isn't true for all tournaments—as an extreme case, the totally ordered tournament has only one H-path.

Proof: Consider a random tournament. (Each edge is directed independently.) Any particular permutation of the vertices has probability 2^{-n+1} of being a H-path, so the expectation of the indicator rv for this event is 2^{-n+1}. The indicator rvs are far from independent, but anyway, by linearity of expectation, the expected number of H-paths is $n!2^{-n+1}$. So some tournament has at least this many H-paths. □

Exercise: explicit construction.
Describe a specific tournament with $n!(2 + o(1))^{-n}$ Hamilton paths.

2.2 Union Bound

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \leq \Pr(A) + \Pr(B)$$

The bound applies also to countable unions:

Lemma 2 $\Pr(\bigcup_1^\infty A_i) \leq \sum_1^\infty P(A_i)$.

Proof: First note that by induction the bound applies to any finite union. Now, if the right-hand side is at least 1, the result is immediate. If not, consider any counterexample; then there is a finite k for which $\Pr(\bigcup_1^k A_i) > \sum_1^k P(A_i)$. Contradiction. □

Later in the lecture we'll use the following which, while trivial, has the whiff of assigning a value to ∞/∞:

Corollary 3 *If a countable list of events $A_1, \ldots$ all satisfy $\Pr(A_i) = 0$, then $\Pr(\bigcup A_i) = 0$. Likewise if for all i, $\Pr(A_i) = 1$, then $\Pr(\bigcap A_i) = 1$.*

Now let's revisit the birthday paradox 5. For a year of n days and a class of r students, if we calculate the *expected* number of pairs who share a birthday,

$$E(\#\text{common birthdays}) = \binom{r}{2}\frac{1}{n}$$

suggests that the probability of some joint birthday may be a constant once r is large enough that $r \sim \sqrt{n}$.

Let's be more formal. An upper bound on P(some common birthday) is

$$\binom{r}{2}\frac{1}{n}$$

by the union bound.

So this gives a one-sided estimate: for the probability of a common birthday to be bounded away from 0, we must require $r \in \Omega(\sqrt{n})$. It remains to be seen whether this condition is also sufficient.

It will turn out that it is; fundamentally this is because there is not much intersection between the $\binom{r}{2}$ different events. We'll show below how to carry out this intuition rigorously.

2.3 Using the union bound in the probabilistic method: Ramsey Theory

Theorem 4 (Ramsey [41] and see [17]) *Fix any nonnegative integers k, ℓ. There is a finite "Ramsey number" $R(k, \ell)$ such that every graph on $R(k, \ell)$ vertices contains either a clique of size k or an independent set of size ℓ.*

Numerous generalizations of Ramsey's argument have since been developed—see the book [39].

Proof: (of Theorem 4) This is outside our main line of development but we include it for completeness. First, $R(k, 1) = R(1, k) = 1$. Now consider a graph with $R(k, \ell - 1) + R(k - 1, \ell) + 1$ vertices and pick any vertex v. Let V_Y denote the vertices connected to v by an edge, and let V_N denote the remaining vertices.

If $|V_N| \geq R(k, \ell - 1)$ then the graph spanned by $V_N \cup \{v\}$ contains either a k-clique (not using v) or an independent set of size ℓ (using v).

On the other hand if $|V_Y| \geq R(k-1, \ell)$ then the graph spanned by $V_Y \cup \{v\}$ contains either a k-clique (using v) or an independent set of size ℓ (not using v).

If you work out the numbers, the proof of Ramsey's theorem gives the bound $R(k, k) \leq \binom{2k-2}{k-1} \in O(4^k/\sqrt{k})$. What we use deviation bounds for is to show a converse:

Theorem 5 (Erdös [16]) *If $\binom{n}{k} < 2^{\binom{k}{2}-1}$ then $R(k, k) > n$. Thus $R(k, k) \geq \frac{2^{k/2}k}{e\sqrt{2}}(1 - o(1))$.*

This leaves an exponential gap. Actually this gap is small by the standards of Ramsey theory. The gap has been slightly tightened since Erdös's work, as we will show later in the course, but remains exponential, and is a major open problem in combinatorics.

Proof: (of Theorem 5) Color the edges uniformly iid. Any particular subgraph of k vertices has probability only $2^{1-\binom{k}{2}}$ of being monochromatic. Take a union bound over all subgraphs. □

2.4 Bonferroni inequalities

The union bound is a special case of the Bonferroni inequalities:

Let $A_1, \ldots, A_n \subseteq \Omega$ be events, and A_i^c their complements. Let $[n]$ denote the set $\{1, \ldots, n\}$. For $S \subseteq [n]$ let $A_S = \bigcap_{i \in S} A_i$. Let $B_S = (\bigcap_{i \in S} A_i) \cap (\bigcap_{i \notin S} A_i^c)$.

For $0 \leq j \leq n$ let $\binom{[n]}{j}$ denote the subsets of $[n]$ of cardinality j.

Lemma 6 *For $j \geq 1$ let $m_j = \sum_{S \in \binom{[n]}{j}} \Pr(A_S)$ and let $M_k = \sum_{j=1}^{k} (-1)^{j+1} m_j$.*

Then:

$$M_2, M_4, \ldots \leq \Pr(\bigcup A_i) \leq M_1, M_3, \ldots$$

Moreover, $\Pr(\bigcup A_i) = M_n$; this is known as the inclusion-exclusion principle.

Comment: Often, larger values of k give better bounds, but not always.

Proof: The sample space is partitioned into 2^n measurable sets B_S. Note that $A_S = \bigcup_{S \subseteq T} B_T$, which, since the B_T's are disjoint, gives $\Pr(A_S) = \sum_{S \subseteq T} \Pr(B_T)$.

$$m_j = \sum_{S \in \binom{[n]}{j}} \Pr(A_S) = \sum_{S \in \binom{[n]}{j}} \sum_{S \subseteq T} \Pr(B_T) = \sum_T \Pr(B_T) \binom{|T|}{j}$$

$$M_k = \sum_{j=1}^{k} (-1)^{j+1} m_j = \sum_T \Pr(B_T) \sum_{j=1}^{k} (-1)^{j+1} \binom{|T|}{j}$$

$$M_k - \Pr(\bigcup A_i) = \sum_{T\neq\emptyset} \Pr(B_T) \sum_{j=0}^{k} (-1)^{j+1} \binom{|T|}{j}$$

The inclusion-exclusion principle is implied because the alternating sum of binomial coefficients is 0 for $k \geq |T|$, in particular, once $k \geq n$.

To see the Bonferroni inequalities we naturally consider even and odd k separately. In either case we use the identity $\binom{t}{j} - \binom{t}{j-1} = \binom{t-1}{j} - \binom{t-1}{j-2}$ (which holds for $t, j \geq 1$ with the interpretation $\binom{a}{b} = 0$ for $a \geq 0, b < 0$). For k even we write

$$M_k - \Pr(\bigcup A_i) = \sum_{T\neq\emptyset} \Pr(B_T) \sum_{j=0}^{k} (-1)^{j+1} \binom{|T|}{j} = \sum_{T\neq\emptyset} \Pr(B_T) \left(-1 - \sum_{\ell=1}^{k/2} \left(\binom{|T|}{2\ell} - \binom{|T|}{2\ell-1} \right)\right)$$

$$= \sum_{T\neq\emptyset} \Pr(B_T) \left(-1 - \sum_{\ell=1}^{k/2} \left(\binom{|T|-1}{2\ell} - \binom{|T|-1}{2\ell-2} \right)\right) = -\sum_{T\neq\emptyset} \Pr(B_T) \binom{|T|-1}{k} \leq 0.$$

For k odd we write

$$M_k - \Pr(\bigcup A_i) = \sum_{T\neq\emptyset} \Pr(B_T) \sum_{j=0}^{k} (-1)^{j+1} \binom{|T|}{j} = \sum_{T\neq\emptyset} \Pr(B_T) \sum_{\ell=0}^{(k+1)/2} \left(\binom{|T|}{2\ell+1} - \binom{|T|}{2\ell} \right)$$

$$= \sum_{T\neq\emptyset} \Pr(B_T) \sum_{\ell=0}^{(k+1)/2} (\binom{|T|-1}{2\ell+1} - \binom{|T|-1}{2\ell-1}) = \sum_{T\neq\emptyset} \Pr(B_T) \binom{|T|-1}{k+2} \geq 0.$$

□

Exercise
If there is a λ such that, in some limit $n \to \infty$, $m_j \to \lambda^j/j!$, then $\Pr(\bigcup A_i) \to 1 - e^{-\lambda}$.

Now let's apply Bonferroni to the birthday paradox 5. For $1 \leq a < b \leq n$ let $A_{ab} =$ [people a, b share a birthday]. Then

$$\Pr(\bigcup A_{ab}) \geq \sum_{ab} \Pr(A_{ab}) - \sum_{ab\neq cd} \Pr(A_{ab} \wedge A_{cd}) \quad (2)$$

$$= \binom{r}{2}\frac{1}{n} - \binom{r}{4}\frac{2}{n^2} - \binom{r}{3}\frac{3}{n^2} \quad (3)$$

$$\cong \frac{r^2}{2n} - \frac{r^4}{12n^2} - \frac{r^3}{2n^2} \quad (4)$$

At, say, $r = \rho\sqrt{\rho n}$, ρ any constant, this is

$$= \rho/2 - \rho^2/12 - O(1/\sqrt{n}) > 0 \quad \text{for any } 0 < \rho < 6.$$

The main point is that this establishes that for any $r \in \Omega(\sqrt{n})$, the probability of a common birthday is bounded away from 0. That is to say, we have found a converse to our earlier conclusion from the union bound.

Of course the probability of a common birthday is monotone in ρ. The breakdown at $\rho = 6$ (and even earlier, when the bound starts decreasing, at $\rho = 3$) is a feature of the proof technique, not of the problem. It can be remedied by going to deeper terms in the Bonferroni sequence.

2.5 Borel-Cantelli lemmas

Here is a very fundamental application of the union bound. Let $B = \{B_1, \ldots\}$ be a countable collection of events. Let $\limsup B$ be the event that infinitely many of the events B_i occur.

Lemma 7 (Borel Cantelli I) *Let* $\sum_{i\geq 1}\Pr(B_i)<\infty$. *Then* $\Pr(\limsup B)=0$.

$\limsup B$ is what is called a *tail event*: a function of infinitely many other events (in this case the $B_1,\ldots$) that remains undetermined by the outcomes of any finite subset of them.

Proof: It is helpful to write $\limsup B$ as

$$\limsup B=\bigcap_{i\geq 0}\bigcup_{j\geq i}B_j.$$

For every i, $\limsup B\subseteq\bigcup_{j\geq i}B_j$, so $\Pr(\limsup B)\leq\inf_i\Pr(\bigcup_{j\geq i}B_j)$. By the union bound, the latter is $\leq\inf_i\sum_{j\geq i}\Pr(B_j)=0$. □

We would be remiss not to wonder whether this lemma has a converse. If there is not, then that is saying that there exist events B_i for which $\sum_{i\geq 1}\Pr(B_i)=\infty$ but $\Pr(\limsup B)=0$. Is that possible?

Yes! Here is an example. Pick a point x uniformly from the unit interval. Let B_i be the event $x<1/i$.

You will notice that in this example the events are not independent. That is crucial, for we have:

Lemma 8 (Borel Cantelli II) *Suppose that* $B_1,\ldots$ *are independent events and that* $\sum_{i\geq 1}\Pr(B_i)=\infty$. *Then* $\Pr(\limsup B)=1$.

Proof: We'll show that $(\limsup B)^c$, the event that only finitely many B_i occur, occurs with probability 0. Write $(\limsup B)^c=\bigcup_{i\geq 0}\bigcap_{j\geq i}B_j^c$.

By Cor. 3, it is enough to show that $\Pr(\bigcap_{j\geq i}B_j^c)=0$ for all i. Of course, for any $I\geq i$, $\Pr(\bigcap_{j\geq i}B_j^c)\leq\Pr(\bigcap_{I\geq j\geq i}B_j^c)$.

By independence, $\Pr(\bigcap_{I\geq j\geq i}B_j^c)=\prod_i^I\Pr(B_j^c)$, so what we remains to show is that

$$\text{For any } i,\quad \lim_{I\to\infty}\prod_i^I\Pr(B_j^c)=0. \tag{5}$$

(Note the LHS is decreasing in I.)

There's a classic inequality we often use:

$$1+x\leq e^x \tag{6}$$

which follows because the RHS is concave and the two sides agree in value and first derivative at a point (namely at $x=0$).

Consequently if a sequence x_i satisfies $\sum x_i\geq 1$ then $\prod(1-x_i)\leq 1/e$.

Supposing 5 is false, fix i for which it fails, let q_i be the limit of the LHS, and let I be sufficient that $\prod_i^I\Pr(B_j^c)\leq 2q_i$. Let I' be sufficient that $\sum_{I+1}^{I'}\Pr(B_j)\geq 1$. Then $\prod_i^{I'}\Pr(B_j^c)\leq 2q_i/e$. Contradiction. □

2.6 Digression

A beautiful fact about tail events is the famous 0-1 law.

Theorem 9 (Kolmogorov) *If* B_i *is a sequence of independent events and* C *is a tail event of the sequence, then* $\Pr(C)\in\{0,1\}$.

We won't be using this theorem, and its proof requires some measure theory, so I'll merely offer one classic example of its application.

Bond Percolation

Fix a parameter $0\leq p\leq 1$. Start with a fixed infinite graph H, for instance the grid graph $\mathbb{Z}^2$ (nodes (i,j)

and (i', j') are connected if $|i - i'| + |j - j'| = 1$) and form the graph G by including each edge of the grid in G independently with probability p. The graph is said to "percolate" if there is an infinite connected component.

It is easy to see that $\Pr(\text{percolation})$ is monotone nondecreasing in p. Due to the 0-1 law, there exists a "critical" p_H such that $\Pr(\text{percolation}) = 0$ for $p < p_H$ and $\Pr(\text{percolation}) = 1$ for $p > p_H$. A lot of work in probability theory has gone into determining values of p_H for various graphs, and also into figuring out whether $\Pr(\text{percolation})$ is 0 or 1 at p_H.