The Probabilistic Method

Po-Shen Loh

June 2010

1 Warm-up

1. (Russia 1996/4) In the Duma there are 1600 delegates, who have formed 16000 committees of 80 persons each. Prove that one can find two committees having at least four common members.

   Solution: Simply select a random 2-set of committees. Calculate expected number of people in both committees, and use convexity of $\sum x_i^2 \geq n(\text{avg})^2$.

2. (Iran Team Selection Test 2008/6) Suppose 799 teams participate in a tournament in which every pair of teams plays against each other exactly once. Prove that there exist two disjoint groups A and B of 7 teams each such that every team from A defeated every team from B.

   Solution: Sample A as a random 7-set. Let X be the number of guys that are totally dominated by A. Letting d_v^- denote the in-degree of v, we have $\mathbb{E}[X] = \sum_v \binom{d_v^-}{7} / \binom{799}{7}$. But $\sum_v d_v^- = \binom{799}{2}$, which means that the average in-degree is exactly 399. By convexity, $\mathbb{E}[X] \geq 799 \cdot \binom{399}{7} / \binom{799}{7} \approx 800 \cdot (1/2)^7 \approx 6.25$, which is enough since X is an integer. Pick 7 teams B from the dominated group.

2 Linearity of expectation

Definition. *Let X be a random variable which takes values in some finite set S. Then the **expected value** of X, denoted $\mathbb{E}[X]$, is:*

$$\mathbb{E}[X] = \sum_{x \in S} x \cdot \mathbb{P}[X = x]$$

Use the following exercises to get used to the concept of expected value.

1. What is the expected number of heads in 101 tosses of a fair coin? Prove this formally from the definition.

   Solution:

   $$\sum_{i=0}^{101} i \binom{101}{i} \frac{1}{2^{101}} = 2^{-101} \sum_{i=1}^{101} i \cdot \frac{101}{i} \binom{100}{i-1} = 2^{-101} \sum_{j=0}^{100} 101 \binom{100}{j} = 2^{-101} \cdot 101 \cdot 2^{100} = 101/2.$$

2. What is the expected number of heads in 1 toss of a coin which lands heads with probability 1/10?

3. Can you calculate the expected number of heads in 101 tosses of a coin which lands heads with probability 1/10?

   One of the most useful facts about the expected value is that if X and Y are random variables, ***not necessarily*** independent, then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. This is known as *linearity of expectation*.

Now use that fact to solve the following problems.

1. Calculate the expected number of heads in 101 tosses of a coin which lands heads with probability 1/10.

   Solution: This is the sum of 101 random variables, each of which has expectation 1/10.

2.1 That's nice, but how does this solve *combinatorics* problems?

The key lemma is the following apparently trivial result.

Lemma. *Let X be a random variable. Then there is some point in the probability space where $X \geq \mathbb{E}[X]$, and also some point in the probability space where $X \leq \mathbb{E}[X]$.*

This takes the place of the "prove-by-contradiction" approach common in many counting-in-two-ways arguments. But where can probability arise in an apparently deterministic combinatorial problem? Let us illustrate this by reworking one of the Warm-up problems.

Solution. Sample a pair of committees uniformly at random (i.e., randomly pick one of the $\binom{16000}{2}$ possible pairs). Let X be the number of people who are in both chosen committees. Note that $X = X_1 + \cdots + X_{1600}$, where each X_i is the $\{0,1\}$-random variable telling whether the i-th person was in both chosen committees. By linearity of expectation,
$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_{1600}].$$
The magic is that each $\mathbb{E}[X_i]$ is easy to calculate! Let n_i be the number of committees that the i-th person belongs to. Then, each $\mathbb{E}[X_i] = \mathbb{P}[i\text{-th person is in both picked committees}] = \binom{n_i}{2}/\binom{16000}{2}$. The only piece of information we know about the $\{n_i\}$ is that their sum $\sum_i n_i = 16000 \cdot 80$, so this suggests that we use convexity to bound $\mathbb{E}[X]$ in terms of the average of $\{n_i\}$, which we denote by $\overline{n} = (16000 \cdot 80)/1600 = 800$:

$$\mathbb{E}[X] \geq 1600 \cdot \binom{\overline{n}}{2}/\binom{16000}{2} = 1600 \cdot \binom{800}{2}/\binom{16000}{2} = 1600 \cdot \frac{800 \cdot 799}{16000 \cdot 15999} = 3.995.$$

(One could see that since $799 \approx 800$ and $15999 \approx 16000$, the last fraction is roughly $1/400$.) But by the Lemma, we know that **some outcome of the probabilistic sampling produces an $X \geq 3.995$.** Since X is always an integer, that outcome must in fact have $X \geq 4$. In particular, we conclude that some pair of committees has ≥ 4 common members. $\square$

Remarks. Observe that once we decided to select the pair of committees at random, we could essentially "follow our nose" to finish the rest of the argument. In particular, we were led to apply convexity because that was the only way to proceed, given our information. Even more conveniently, the sharpening from 3.995 to 4 was automatic!

3 Olympiad problems

1. In a 100×100 array, each of the numbers $1, 2, \ldots, 100$ appears exactly 100 times. Show that there is a row or a column in the array with at least 10 distinct numbers.

   Solution: Let $n = 100$. Choose a random row or column ($2n$ choices). Let X be the number of distinct entries in it. Now $X = \sum I_i$, where each I_i is the indicator variable of i appearing (possibly more than once) in our random row or column. Clearly, each $\mathbb{E}[I_i] = \mathbb{P}[I_i \geq 1]$. To lower-bound this, observe that the worst-case is if all n appearances of i are in some $\sqrt{n} \times \sqrt{n}$ submatrix, which gives $\mathbb{P}[I_i \geq 1] \geq 2\sqrt{n}/(2n) = 1/\sqrt{n}$. Hence by linearity, $\mathbb{E}[X] \geq \sqrt{n}$.

2. (Russia, 1999.) In a class, each boy is friends with at least one girl. Show that there exists a group of at least half of the students, such that each boy in the group is friends with an odd number of the girls in the group.

Solution: Choose girls independently with probability $1/2$, and then let the set of boys be all of those who have an odd number of friends in the girl group. Let X be the number of boys and girls selected, and break this into the sum of indicators. For each girl, obviously the indicator adds $1/2$ to the sum. For each boy, the probability that he joins is precisely the probability that $\mathrm{Bin}\,[k, 1/2]$ is odd, where k was the number of girls he knew. To see that this probability is $1/2$, note that it is the parity of the sum of k independent coin flips. In particular, the final flip independently flips or retains the final parity, hence odd with probability $1/2$.

3. (IMO Shortlist 1999/C4) Let A be any set of n residues mod n^2. Show that there is a set B of n residues mod n^2 such that at least half of the residues mod n^2 can be written as $a + b$ with $a \in A$ and $b \in B$.

   Solution: Make n independent uniformly random choices from the n^2 residues, and collect them into a set B. Note that since we use independence, this final set may have size $< n$. But if we still have $A + B$ occupying at least half of the residues, then this is okay (we could arbitrarily augment B to have the full size n).

   Let X be the number of residues achievable as $a + b$. For each potential residue i, there are exactly n ways to choose some b for which $A + b \ni i$, since $|A| = n$. Therefore, the probability that a given residue i appears in $A + B$ is precisely $1 - \left(1 - \frac{n}{n^2}\right)^n$. Then $\mathbb{E}\,[X]$ is exactly n^2 times that, because there are n^2 total residues. Hence it suffices to show that $1 - \left(1 - \frac{n}{n^2}\right)^n \geq 1/2$. But this follows from the bound $1 - \frac{1}{n} \leq e^{-1/n}$, using $e \approx 2.718$.

4. (MOP 2010.) Let G be a d-regular graph. Prove that for every $k \leq d$, there is a K_{k+1}-free induced subgraph on at least $\frac{kn}{d+1}$ vertices.

   Solution: Color with $d + 1$ colors. Take k biggest color classes, which by averaging is at least k times the average.

5. For every integer k, can you construct a graph which requires k colors to properly color, but has average degree equal to 1?

   Solution: Take K_k, giving sum of degrees $k(k - 1) = k^2 - k$ on k vertices. Then add $k^2 - 2k$ more isolated vertices to get average degree 1.

6. (MOP 2010, what should have been asked.) Let G be a graph with average degree d. Prove that for every $k \leq d$, there is a K_{k+1}-free induced subgraph on at least $\frac{kn}{d+1}$ vertices.

   Solution: Randomly permute the vertices. Use the greedy algorithm, taking each vertex if it can be added without making any K_{k+1}. Observe that we will actually take every vertex v with the property that the permutation induced on $\{v\} \cup N(v)$ has v in position $1 \ldots k$. (We might also take more.) This is because v would only have degree at most $k - 1$ back to the previously selected guys, making at most a K_k. Now the expected size of the selected set is at least

$$\sum_v \frac{k}{d_v + 1} \geq n \cdot \frac{k}{d+1}.$$

4 Famous results

These are interesting results from research mathematics (as opposed to Olympiad mathematics) that have very elegant probabilistic proofs. They come from the excellent book titled *The Probabilistic Method*, by Noga Alon and Joel Spencer.

1. (Erdős-Ko-Rado Theorem) Let $n \geq 2k$ be positive integers, and let $\mathcal{C}$ be a collection of pairwise-intersecting k-element subsets of $\{1, \ldots, n\}$, i.e., every $A, B \in \mathcal{C}$ has $A \cap B \neq \emptyset$. Prove that $|\mathcal{C}| \leq \binom{n-1}{k-1}$.
   Remark: this corresponds to the construction which takes all subsets that contain the element 1, say.

Solution: Pick a random k-set A from $2^{[n]}$ by first selecting a random permutation $\sigma \in S_n$, and then picking a random index $i \in [n]$. Then define $A = \{\sigma(i), \ldots, \sigma(i + k - 1)\}$, with indices after n wrapping around, of course. It suffices to show that $\mathbb{P}[A \in \mathcal{C}] \le k/n$.

Let us show that conditioned on any fixed σ, $\mathbb{P}[A \in \mathcal{C}|\sigma] \le k/n$, which will finish our problem. But this is equivalent to the statement that $\mathcal{C}$ can only contain $\le k$ intervals (wrapping after n) of the form $\{i, \ldots, i + k - 1\}$, which is easy to show.

2. (Sperner's Lemma) Prove that the maximum antichain in $2^{[n]}$ has size $\binom{n}{\lfloor n/2 \rfloor}$. That is, show that if $\mathcal{F}$ is a collection of subsets of $\{1, \ldots, n\}$ such that no two distinct sets $A, B \in \mathcal{F}$ satisfy $A \subset B$, then $|\mathcal{F}| \le \binom{n}{\lfloor n/2 \rfloor}$.

Solution: The idea is given as a hint in Alon-Spencer. Take a random permutation σ and let the random variable $X = \#\{i : \sigma(1), \ldots, \sigma(i) \in \mathcal{F}\}$. Consider $\mathbb{E}[X]$.

By definition of $\mathcal{F}$, X is bounded by 1, and the events $\{\sigma(1), \ldots, \sigma(K)\} \in \mathcal{F}$ are disjoint for distinct K. Let N_k be the number of subsets of size k in $\mathcal{F}$.

$$\mathbb{E}[X] = \sum_{k=1}^{n} \mathbb{P}[\{\sigma(1), \ldots, \sigma(k)\} \in \mathcal{F}] = \sum_{k=1}^{n} \frac{N_k}{\binom{n}{k}} \ge \sum_{k=1}^{n} \frac{N_k}{\binom{n}{\lfloor n/2 \rfloor}}.$$

Since $\mathbb{E}[X] \le 1$, we conclude that $\sum N_k$ is bounded by $\binom{n}{\lfloor n/2 \rfloor}$, and we are done.

3. (Crossing Lemma) No matter how you draw a graph with V vertices and E edges in the plane, there will be $\ge \frac{E^3}{64V^2}$ pairs of crossing edges, as long as $E \ge 4V$.

Solution: Since planar graphs have $E \le 3V - 6$, we automatically find that the crossing number is always $\ge E - (3V - 6) > E - 3V$. Now take a drawing with, say, t crossings, and sample vertices randomly with probability p. V goes down to pV, E goes down to $p^2 E$, and cr goes down to $p^4 t$. But this new drawing needs to satisfy the above, so

$$p^4 t > p^2 E - 3pV.$$

Substituting $p = 4V/E$, we get the desired result.

4. (Bollobás, 1965) Let $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ be distinct subsets of $\mathbb{N}$ such that:

   - every $|A_i| = r$ and every $|B_i| = s$,
   - for every i, $A_i \cap B_i = \emptyset$, and
   - for every $i \ne j$, $A_i \cap B_j \ne \emptyset$.

Prove that $n \le \binom{r+s}{r}$.

Solution: (Proof from Alon-Spencer.) Let the universe X be the union of all A_i and B_j. Take random permutation of X. Define event X_i to be that all elements of A_i precede all elements of B_i in permutation. Easy to check that all n of the events X_i are pairwise disjoint, and each $\mathbb{P}[X_i] = \binom{r+s}{r}^{-1}$. But sum of probabilities is ≤ 1, so done.

5. (Lovász, 1970) Let $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ be distinct subsets of $\mathbb{N}$ such that:

   - every $|A_i| = r$ and every $|B_i| = s$,
   - for every i, $A_i \cap B_i = \emptyset$, and
   - for every $i < j$, $A_i \cap B_j \ne \emptyset$.

Prove that $n \le \binom{r+s}{r}$. **Remark:** this is much more difficult, and the proof uses linear algebra.

6. (Erdős, 1965) A set S is called *sum-free* if there is no triple of (not necessarily distinct) elements $x, y, z \in S$ satisfying $x + y = z$. Prove that every set A of nonzero integers contains a subset $S \subset A$ of size $|S| > |A|/3$ which is sum-free.

   Solution: Choose a prime p of the form $3k + 2$ such that p is greater than twice the maximum absolute value of any element in A. Let $C = \{k + 1, \ldots, 2k + 1\}$, which is sum-free modulo p. Then pick a uniformly random $x \in \{1, \ldots, p\}$ and let B be the set obtained by multiplying each element of A by x, modulo p. For each element, probability of mapping into C is $|C|/(p - 1) > 1/3$. So expected number of elements mapping into C is $> |A|/3$, and we can take S to be those that do.

7. (Alon-Spencer, Exercise 2.9.) Suppose that every vertex of a n-vertex bipartite graph is given a personalized list of $> \log_2 n$ possible colors. Prove that it is possible to give each vertex a color from its list such that no two adjacent vertices receive the same color. (This is a statement about the *list-chromatic number* of a bipartite graph.)

   Open: If the maximum degree of the bipartite graph is Δ, is it always possible to color from lists of size $\log_2 \Delta$?

   Solution: Let the bipartition of the vertex set be $V_1 \cup V_2$. Let X be the total set of all colors in all lists. For each color, independently flip a fair coin. The idea is that we would be done if for every vertex in V_1, we can choose a color that corresponds to a Head, and for every vertex in V_2, we can choose a color that was a Tail. So, let N count the number of vertices for which this fails. Clearly, for each vertex v, the probability of failure is $2^{-|S(v)|} < 1/n$. Thus, by linearity of expectation, $\mathbb{E}[N] < 1$, and we are guaranteed a circumstance in which $N = 0$.