Probabilistic Methods in Combinatorics

Po-Shen Loh

June 2009

1 Warm-up

60. В M-элементном множестве выделили S подмножеств, содержащих a_1, a_2, ..., a_s элементов соответственно. Известно, что среди этих подмножеств ни одно не содержится в другом. Докажите, что

$$\frac{1}{\binom{M}{a_1}} + \frac{1}{\binom{M}{a_2}} + \ldots + \frac{1}{\binom{M}{a_s}} \le 1 \,,$$

где $\binom{M}{a_k} = M!/a_k!(M - a_k)!$.

2 Olympiad problems that can probably be solved

1. (Leningrad Math Olympiad 1987, Grade 10 elimination round) Let $A_1, \ldots, A_s$ be subsets of $\{1, \ldots, M\}$, and suppose that none of the A_i are subsets of each other. For each index i, let $a_i = |A_i|$. Prove that

$$\sum_{i=1}^{s} \frac{1}{\binom{M}{a_i}} \le 1.$$

   Remark. This is also known as the Lubell-Yamamoto-Meshalkin inequality.

   Solution: Generate a random permutation $\sigma = (x_1, \ldots, x_M)$ of $[M]$. Let each E_i be the event that the set A_i appears as the initial segment of σ, i.e., that $\{x_1, \ldots, x_{a_i}\} = A_i$. Since none of the A_i contain each other, the E_i are mutually exclusive. But each $\mathbb{P}[E_i] = \binom{M}{a_i}^{-1}$, so the LHS is $\mathbb{P}[\cup E_i]$, and probabilities are always ≤ 1.

2. (Very slight simplification of a problem in the MathLinks.ro 2008 contest) Let $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ be distinct finite subsets of $\mathbb{N}$ such that:

   - for every i, $A_i \cap B_i = \emptyset$, and
   - for every $i \ne j$, $(A_i \cap B_j) \cup (A_j \cap B_i) \ne \emptyset$.

   Prove that for every real number $0 \le p \le 1$,

$$\sum_i p^{|A_i|}(1 - p)^{|B_i|} \le 1.$$

   Remark. This result is due to Tuza (1985).

   Solution: Sample a random subset of the universe by taking each element with probability p. Let E_i be the event that everything in A_i got picked while everything in B_i was missed. These are disjoint events, so the LHS is the probability that some E_i occurred, hence ≤ 1.

3. (IMO Shortlist 2006/C3) Let S be a finite set of points in the plane such that no three of them are on a line. For each convex polygon P whose vertices are in S, let $a(P)$ be the number of vertices of P, and let $b(P)$ be the number of points of S which are outside P. Prove that for every real number x:

$$\sum_P x^{a(P)}(1-x)^{b(P)} = 1,$$

where the sum is taken over all convex polygons with vertices in S. **Important:** a line segment, a point, and the empty set are considered to be convex polygons of 2, 1, and 0 vertices, respectively.

Solution: Randomly color the points in black and white, with a point receiving black with probability x. For each convex polygon P, let E_P be the event that all vertices on the perimeter of P are black, and all vertices in P's exterior are white. These events are mutually exclusive, so the LHS is the probability that some E_p holds. But there is always some E_P that always holds: consider the convex hull of all of the black points.

3 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 called *linearity of expectation*.

Proposition. *For any two **not necessarily** independent random variables X and Y, we have $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.*

Proof. This is essentially a consequence of the commutativity of addition. Following the definition:

$$
\begin{aligned}
\mathbb{E}[X+Y] &= \sum_{x,y} (x+y) \cdot \mathbb{P}[X=x, Y=y] \\
&= \sum_{x,y} x \cdot \mathbb{P}[X=x, Y=y] + \sum_{x,y} y \cdot \mathbb{P}[X=x, Y=y] \\
&= \sum_x \left(x \sum_y \mathbb{P}[X=x, Y=y] \right) + \sum_y \left(y \sum_x \mathbb{P}[X=x, Y=y] \right) \\
&= \sum_x (x \mathbb{P}[X=x]) + \sum_y (y \mathbb{P}[Y=y]) \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}
$$

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.

4   Olympiad problems one can expect to solve

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 $X \leq \mathbb{E}[X]$.*

Try these problems.

1. (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. (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 $\sum x_i^2 \geq n(\text{avg})^2$.

   More carefully: 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. $\qquad\square$

3. (MOP Test 2008/7/2) Suppose that a, b, c are positive real numbers such that for every integer n,

   $$\lfloor an \rfloor + \lfloor bn \rfloor = \lfloor cn \rfloor.$$

   Prove that at least one of a, b, c is an integer.

Remark. You may use the following well-known number-theoretic fact: if x is *irrational*, then the fractional parts of the multiples of x are *equidistributed* over the interval $[0, 1]$. In particular, if we choose n uniformly among $\{1, \ldots, N\}$, then $\mathbb{E}\left[\{xn\}\right] \to 1/2$ as $N \to \infty$.

Solution: Suppose none of a, b, c are integers. Divide both sides by n and take the limit. This gives $a + b = c$, so we also know that the sum of fractional parts:

$$\{an\} + \{bn\} = \{cn\}. \tag{1}$$

If x is irrational, then $\{xn\}$ is equidistributed over the interval $[0, 1]$. In particular, if we choose n uniformly among $\{1, \ldots, N\}$, then $\mathbb{E}\left[\{xn\}\right] \to 1/2$ as $N \to \infty$. On the other hand, if x is rational with reduced form p/q, then $\{xn\}$ has expectation tending to $\frac{q-1}{2q} = \frac{1}{2} - \frac{1}{2q}$. So, it is in the interval $[1/4, 1/2)$. **Conclusion:** for any noninteger x, $\mathbb{E}\left[\{xn\}\right] \to t$, with $t \in [1/4, 1/2]$.

Taking the expectation of equation (**??**), and taking limits, we immediately see that the only way to have the equality is if $\mathbb{E}\left[\{an\}\right]$ and $\mathbb{E}\left[\{bn\}\right]$ both tend to $1/4$, and $\mathbb{E}\left[\{cn\}\right] \to 1/2$. But the only way to get expecation $1/4$ is when a, b are rational, and the only way to get the full expectation $1/2$ is when c is irrational. Yet our first deduction was that $a + b = c$, so we cannot have two rationals summing to one irrational. Contradiction.

4. (MOP Test 2007/7/1) In an $n \times n$ array, each of the numbers $1, 2, \ldots, n$ appears exactly n times. Show that there is a row or a column in the array with at least $\sqrt{n}$ distinct numbers.

   Solution: 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 $\mathbb{E}\left[I_i\right] = \mathbb{P}\left[I_i \geq 1\right]$. 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 $\mathbb{P}\left[I_i \geq 1\right] \geq 2\sqrt{n}/(2n) = 1/\sqrt{n}$. Hence by linearity, $\mathbb{E}\left[X\right] \geq \sqrt{n}$.

5. (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 $\mathbb{E}\left[X\right]$ 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 $e \approx 2.718$.

6. (Austrian-Polish Math Competition 1997/8) Let n be a natural number and M a set with n elements. Find the biggest integer k such that there exists a k-element family of 3-element subsets of M, no two of which are disjoint.

   Solution: We need to be careful about case when n is small: for $n \leq 5$, we can simply take everything. Otherwise, it is Erdős-Ko-Rado.

5 Real problems one can expect to prove

These are interesting results from research mathematics (as opposed to Olympiad mathematics) that have very elegant probabilistic proofs.

1. (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.

2. (Turán's Theorem, numerical bound) Let G be a graph with n vertices and average degree d. Prove that one can find a set S of pairwise non-adjacent vertices of size at least $\frac{n}{d+1}$.

   Remark. This is called an *independent set*, and the result is *tight* because a disjoint union of copies of K_{d+1} has no independent set larger than $\frac{n}{d+1}$.

   Solution: Start with a random ordering of the vertices. If any vertex precedes all of its neighbors according to the ordering, then take it for S. Clearly this is an independent set. Let I_v be the event that vertex v is chosen. $\mathbb{E}\left[I_v\right] = \frac{1}{d_v+1}$, so by linearity of expectation, $\mathbb{E}\left[|S|\right] = \sum_v \frac{1}{d_v+1}$. By convexity, this is at least $n \cdot \frac{1}{d+1}$.

3. (Component of research problem of Tao Jiang's masters' student) Let G be a graph whose vertices have been properly colored. Nothing is assumed about the number of colors used, except for the local fact that for every vertex v, the number of different colors which appear on v's neighbors is at most M. Prove that there is a set S of pairwise non-adjacent vertices, of size at least $\frac{n}{d+1}$.

   Solution: (Proof due to Benny Sudakov.) Start with a random ordering of the color classes. If any vertex's color class precedes all of its neighbors' color classes according to the ordering, then take the vertex for S. Again this is clearly an independent set. A proper coloring has the property that any v is differently colored from all of its neighbors, so at most $M + 1$ total colors appear on v and its neighbors. Thus the probability that v's color is earliest in the permutation is at least $\frac{1}{M+1}$. Finish by linearity of expectation.

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

   Solution: Since planar graphs have $E \leq 3V - 6$, we automatically find that the crossing number is always $\geq E - (3V - 6) > E - 3V$. 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.

6   Real problems one does not expect to prove, but are probably true

Many of these come from the excellent book titled *The Probabilistic Method*, by Noga Alon and Joel Spencer.

1. (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 $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

   Solution: Note that this is a special case of one of the first problems in this handout.

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 $\mathcal{F}$.

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

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

2. (Corollary: Littlewood-Offord Lemma) Let $x_1, \ldots, x_n$ be nonzero real numbers, not necessarily distinct. Suppose that $c_1, \ldots, c_n$ are independent random variables, each of which is ± 1 with equal probability. Prove that $\mathbb{P}[c_1 x_1 + \cdots + c_n x_n = 0] \leq O(\frac{1}{\sqrt{n}})$.

   Solution: First, we may assume all $x_i > 0$ by flipping the sign of any negative guy. (Since the c_i are ± 1 with equal probability, it won't make a difference.)

   Consider any combination of (c_i) which makes $\sum c_i x_i = 0$. Identify this with a subset of $[n]$, based on which indices have $c_i = +1$. Let the $\mathcal{F}$ be the collection of all such subsets of $[n]$. This is an antichain because if we had some S strictly contained in a larger T, then the sum $\sum c_i x_i$ corresponding to T would be strictly larger since all $x_i > 0$.

   By Sperner's Lemma, we conclude that the number of ways to have (c_i) making zero-sum is at most $\binom{n}{\lfloor n/2 \rfloor}$, and so the desired probability bound is this divided by 2^n, which is of order $\frac{1}{\sqrt{n}}$.

3. (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 $\mathbb{P}[A \in \mathcal{C}] \leq k/n$.

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

4. (ChipLiar game, Alon-Spencer Theorem 14.2.1) Paul and Carole play a game on a board with positions labeled $\{0, 1, \ldots, k\}$. Initially, n stones are at position k. Paul and Carole play for r rounds, where each round has the following structure: Paul names a subset S of the stones on the board, and then, Carole either moves all stones in S one position to the left, or moves all stones in S^c one position to the left. Any stone that is moved leftwards from 0 is discarded. If the number of stones on the board becomes 1 or 0, Carole loses. Prove that if $n \cdot \sum_{i=0}^{k} \binom{r}{i} 2^{-r} > 1$, then Carole has a winning strategy.

   Solution: Perfect information game, so one of them has a winning strategy. Suppose it is Paul, and let him play it. Carole will respond by playing randomly. Note that if Paul indeed has winning strategy, then it can even always beat a random strategy. Calculate expected number of stones left on the board after r rounds. By linearity, this is n times the probability that a chip survives for r rounds. But with random play, each chip moves left $\mathrm{Bin}(r, 1/2)$ many times. So, expected number of surviving chips is:

   $$n\mathbb{P}[\mathrm{Bin}(r, 1/2) \leq k],$$

   which we are given to be greater than 1. Hence some play sequence will leave Carole with ≥ 2 stones, contradicting existence of Paul's perfect strategy.

5. (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 $i \neq j$, $A_i \cap B_j \neq \emptyset$.

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

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.

6. (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 $i < j$, $A_i \cap B_j \neq \emptyset$.

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

7. (Alon-Spencer, Exercise 1.7) 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 $i \neq j$, $(A_i \cap B_j) \cup (A_j \cap B_i) \neq \emptyset$.

Prove that $n \leq \frac{(r+s)^{r+s}}{r^r s^s}$.

Solution: Plug into Tuza's result above with $p = \frac{r}{r+s}$ or $\frac{s}{r+s}$.

Alternatively, we can do this directly as follows. Let $X := \bigcup_{i=1}^n A_i \cup B_i$ be the base set. Define $p := r/(r+s)$, and consider a "coin" which has one side saying "A" and one side saying "B", with the A-side appearing with probability p. For each element in X, independently flip the coin. This defines a mapping $f : X \to \{A, B\}$. Define the family of events $\{E_i\}_1^n$ by having E_i occur when all elements $x \in A_i$ have $f(x) = A$, and all elements $y \in B_i$ have $f(y) = B$.

By definition of the family of sets, it is impossible for E_i and E_j to occur simultaneously if $i \neq j$, because in particular, there would exist some element in either $A_i \cap B_j$ or $A_j \cap B_i$, and it could neither be A nor B. So, just as in the previous problem, consider the probability of any E_i occurring. Trivially, $P(E_i) = p^r(1-p)^s$ for any i. Since the events are disjoint, the total probability is $np^r(1-p)^s$. Yet all probabilities are bounded by 1, so just as in the previous problem, $n \leq p^{-r}(1-p)^{-s}$, which turns out to be the desired bound. (In fact, this choice of p is optimal.)

7 Really harder problems

1. Prove that the Riemann Hypothesis has probability > 0 of being true.

References

[1] N. Alon and J. Spencer, **The Probabilistic Method**, 2nd ed., Wiley, New York, 2000.