Unexpected Expectations

Evan Chen evanchen@mit.edu

Andrew Critch critch@acritch.com

June, 2015

Sometimes the easiest way to solve a problem is to define some random variables whose expectations make things more intuitive to think and write about. When the problem doesn't involve any randomness at its outset, folks call this a "probabilistic method" because you introduce probability where it wasn't invited. But actually this method is cool even when it's not the only probabilistic thing going on, so we'll look at some of those applications, too.

Note from the second author: I've made very few changes to Evan's original version of these notes; all the credit for assembling this excellent collection of problems goes to him!

1 Definitions and Notation

A **random variable** is just a quantity that that depends on some world or input that we assign probabilities to. For instance, consider a six-sided die roll (with equal probability assigned to each outcome), and let D_6 be the number that ends up on the top face of the die. Here D_6 is a random variable. The subscript "6" isn't needed; it's just a common way to indicate a random variable having 6 equally likely outcomes.

We can discuss the **probability** of certain events, which we'll denote $\mathbb{P}(\bullet)$. For instance, we can write things like

$$\mathbb{P}(D_6 = 1) = \mathbb{P}(D_6 = 2) = \cdots = \mathbb{P}(D_6 = 6) = \frac{1}{6}$$

$$\mathbb{P}(D_6 = 0) = 0$$

$$\mathbb{P}(D_6 \geq 4) = \frac{1}{2}.$$

If we let D_6^* be the number that ends up on the bottom of the die, we can also say things like

$$\mathbb{P}(D_6 = D_6^*) = 0 \quad \text{and} \quad \mathbb{P}(D_6 + D_6^* = 7) = 1$$

Now suppose we rolled the same die again, and called the bottom face E_6. Even though E_6 has the same distribution over its outputs as D_6^* (namely, $1 \dots 6$ each happening with probability $1/6$), it interacts with D_6 in very different ways:

$$\mathbb{P}(D_6 = E_6) = \frac{1}{6} \quad \text{and} \quad \mathbb{P}(D_6 + D_6^* = 7) = \frac{1}{6}$$

This is why it's important to think of a random variable as a *function of a random input*, rather than just a list of values with a probability associated to each one, because in the latter sense, D_6^* and E_6 would be identical. Folks often call the outputs of a random variable its "values", which suppresses awareness that each random variable is actually a function. But to avoid confusion one needs to remember that the variables are functions, either implicitly or explicitly.

The **conditional probability** of Y given X is defined as

$$\mathbb{P}(X = x \mid Y = y) := \mathbb{P}(X = x \text{ and } Y = y)/\mathbb{P}(Y = y)$$

and is undefined when $\mathbb{P}(Y = y) = 0$. In our dice example, $\mathbb{P}[D_6 = 3 \mid D_6^* = 4] = 1$

We say X is **independent** of Y and write $X \perp\!\!\!\perp Y$ if for all x, y,

$$\mathbb{P}(X = x \text{ and } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

This is equivalent to the statement

$$\mathbb{P}(X = x \mid Y = y) = \mathbb{P}(X = x)$$

i.e., that finding out the value of Y tells you nothing about the distribution of X, and by symmetry, conversely. In our dice example, you can check that

$$D_6 \perp\!\!\!\perp E_6 \quad \text{and} \quad D_6 \not\perp\!\!\!\perp D_6^*$$

The **expected value** or **expectation** of a random variable X is the probability-weighted average of its values:

$$\mathbb{E}[X] \stackrel{\text{def}}{=} \sum_x \mathbb{P}(X = x) \cdot x.$$

For our dice roll D_6,

$$\mathbb{E}[D_6] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots + \frac{1}{6} \cdot 6 = 3.5.$$

1.1 Graph theory terminology

A number of problems here will refer to an **independent set**, which is a set of vertices (in a graph) for which no two are connected by an edge. See `http://en.wikipedia.org/wiki/Graph_theory#Definitions` if you aren't yet familiar with basic terms from graph theory, like vertices, edges, subgraphs, and so on.

2 Properties of Expected Value

2.1 A Motivating Example

It is an unspoken law that any introduction to expected value begins with the following classical example.

Example 2.1. At MOP, there are n people, each of who has a name tag. We shuffle the name tags and randomly give each person one of the name tags. Let S be the number of people who receive their own name tag. Prove that the expected value of S is 1.

This result might seem surprising, as one might intuitively expect $\mathbb{E}[S]$ to depend on the choice of n.

For simplicity, let us call a person a *fixed point* if they receive their own name tag.[1] Thus S is just the number of fixed points, and we wish to show that $\mathbb{E}[S] = 1$. If we're

[1] This is actually a term used to describe points which are unchanged by a permutation. So the usual phrasing of this question is "what is the expected number of fixed points of a random permutation?"

interested in the expected value, then according to our definition we should go through all $n!$ permutations, count up the total number of fixed points, and then divide by $n!$ to get the average. Since we want $\mathbb{E}[S] = 1$, we expect to see a total of $n!$ fixed points.

Let us begin by illustrating the case $n = 4$ first, calling the people W, X, Y, Z.

	W	X	Y	Z	Σ
1	W	X	Y	Z	4
2	W	X	Z	Y	2
3	W	Y	X	Z	2
4	W	Y	Z	X	1
5	W	Z	X	Y	1
6	W	Z	Y	X	2
7	X	W	Y	Z	2
8	X	W	Z	Y	0
9	X	Y	W	Z	1
10	X	Y	Z	W	0
11	X	Z	W	Y	0
12	X	Z	Y	W	1
13	Y	W	X	Z	1
14	Y	W	Z	X	0
15	Y	X	W	Z	2
16	Y	X	Z	W	1
17	Y	Z	W	X	0
18	Y	Z	X	W	0
19	Z	W	X	Y	0
20	Z	W	Y	X	1
21	Z	X	W	Y	1
22	Z	X	Y	W	2
23	Z	Y	W	X	0
24	Z	Y	X	W	0
Σ	6	6	6	6	24

We've listed all $4! = 24$ permutations, and indeed we see that there are a total of 24 fixed points, which are hilighted. Unfortunately, if we look at the rightmost column, there doesn't seem to be a pattern, and it seems hard to prove that this holds for other values of n.

However, suppose that *rather than trying to add by rows, we add by columns.* There's a very clear pattern if we try to add by the columns: we see a total of 6 fixed points in each column. Indeed, the six fixed W points correspond to the $3! = 6$ permutations of the remaining letters X, Y, Z. Similarly, the six fixed X points correspond to the $3! = 6$ permutations of the remaining letters W, Y, Z.

This generalizes very nicely: if we have n letters, then each letter appears as a fixed point $(n-1)!$ times. Thus the expected value is

$$\mathbb{E}[S] = \frac{1}{n!}\left(\underbrace{(n-1)! + (n-1)! + \cdots + (n-1)!}_{n \text{ times}}\right) = \frac{1}{n!} \cdot n \cdot (n-1)! = 1.$$

Cute, right? Now let's bring out the artillery.

2.2 Linearity of Expectation

The crux result of this section is the following theorem.

Theorem 2.2 (Linearity of Expectation)**.** *Given any random variables $X_1, X_2, \dots, X_n$, and constants a_i, we always have*

$$\mathbb{E}[a_1 X_1 + a_2 X_2 + \cdots + a_n X_n] = a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2] + \cdots + a_n \mathbb{E}[X_n].$$

This theorem is highly intuitive if the $X_1, X_2, \dots, X_n$ are independent of each other – if we roll 100 dice, we expect an average of 350. The wonderful thing is that this holds *even if the variables are not independent.* And the basic idea is just the double-counting we did in the earlier example: even if the variables depend on each other, if you look only at the expected value, you can still add just by columns. The proof of the theorem is just a bunch of sigma signs which say exactly the same thing, so we won't bother including it.

Anyways, we can now nuke our original problem. The trick is to define **indicator variables** as follows: for each $i = 1, 2, \dots, n$ let

$$S_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if person } i \text{ gets his own name tag} \\ 0 & \text{otherwise.} \end{cases}$$

Obviously,

$$S = S_1 + S_2 + \cdots + S_n.$$

Moreover, it is easy to see that $\mathbb{E}[S_i] = \mathbb{P}(S_i = 1) = \frac{1}{n}$ for each i: if we look at any particular person, the probability they get their own name tag is simply $\frac{1}{n}$. Therefore,

$$\mathbb{E}[S] = \mathbb{E}[S_1] + \mathbb{E}[S_2] + \cdots + \mathbb{E}[S_n] = \underbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}_{n \text{ times}} = 1.$$

Now that was a lot easier! By working in the context of expected value, we get a framework where the "double-counting" idea is basically automatic. In other words, linearity of expectation lets us only focus on small, local components when computing an expected value, without having to think about a lot of interactions between cases and quantities that would otherwise distract us.

2.3 More Examples

Example 2.3 (HMMT 2006)**.** At a nursery, 2006 babies sit in a circle. Suddenly, each baby randomly pokes either the baby to its left or to its right. What is the expected value of the number of unpoked babies?

Solution. Number the babies 1, 2, …, 2006. Define

$$X_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if baby } i \text{ is unpoked} \\ 0 & \text{otherwise.} \end{cases}$$

We seek $\mathbb{E}[X_1 + X_2 + \cdots + X_{2006}]$. Note that any particular baby has probability $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$ of being unpoked (if both its neighbors miss). Hence $\mathbb{E}[X_i] = \frac{1}{4}$ for each i, and

$$\mathbb{E}[X_1 + X_2 + \cdots + X_{2006}] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_{2006}] = 2006 \cdot \frac{1}{4} = \frac{1003}{2}. \qquad \square$$

Seriously, this should feel like cheating.

2.4 Conditional expectations

While $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ always holds, in general $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$. However, there is *something* you can say in general about the expectation of a product, for which we need a definition:

The **conditional expectation of** X **given** Y is a function of X that says what the expectation of Y is once X is known:

$$\mathbb{E}[Y \mid X] = \sum_y y \cdot \mathbb{P}(Y = y \mid X = x)$$

For example, if X is a (uniformly) randomly chosen face of a die, and Y is a randomly chosen other face of that die, then

$$\mathbb{E}[Y \mid X] = \sum_{i \neq X} i \cdot \frac{1}{6} = \frac{21 - X}{6}$$

To emphasize again, $\mathbb{E}[Y \mid X]$ denotes a *function depending on the value of* X, which in particular makes it a random variable. So now we can ask about the value of an expression like $\mathbb{E}[\mathbb{E}[Y \mid X]]$. It's an exercise in comfort-with-notation to prove that, in general,

Proposition 2.4 (Conditional expectations). *For any random variables X and Y,*

$$\mathbb{E}[f(X)Y] = \mathbb{E}[f(X) \cdot \mathbb{E}[Y \mid X]]$$

In particular, when $f(X) = X$ and $f(X) = 1$, we get

$$\mathbb{E}[XY] = \mathbb{E}[X \cdot \mathbb{E}[Y \mid X]] \quad \textit{and} \quad \mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$$

To illustrate with our dice example,

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[X \cdot \mathbb{E}[Y \mid X]] \\
&= \mathbb{E}[X \cdot \frac{21 - X}{6}] \\
&= \mathbb{E}[\frac{21X - X^2}{6}] \\
&= \frac{21}{6}\mathbb{E}[X] - \frac{1}{6}\mathbb{E}[X^2] \\
&= \frac{21}{6} \cdot \frac{21}{6} - \frac{1}{6} \cdot \frac{91}{6} \ldots = \frac{175}{3}
\end{aligned}$$

Proposition 2.5 (Independence and multiplicative expectations). *Given any random variables X and Y,*

$$X \perp\!\!\!\perp Y \quad \Rightarrow \quad \mathbb{E}[Y \mid X] = \mathbb{E}[Y] \quad \Rightarrow \quad \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

where (exercise) each of these implications is strict, i.e., the reverse does not hold.

It's reasonably common to forget that these implications are not reversible, so it's a reasonable exercise to come up with examples to illustrate it. But while you're thinking about that, *don't forget*: unlike when multiplying two random variables, *linearity* of expectation *does not* require independence!

2.5 Practice Problems

The first two problems are somewhat straightforward applications of the methods described above.

Problem 2.6 (AHSME 1989)**.** Suppose that 7 boys and 13 girls line up in a row. Let S be the number of places in the row where a boy and a girl are standing next to each other. For example, for the row $GBBGGGBGBGGGBGBGGBGG$ we have $S = 12$. Find the expected value of S.

Problem 2.7 (AIME 2006 #6)**.** Let $\mathcal{S}$ be the set of real numbers that can be represented as repeating decimals of the form $0.\overline{abc}$ where a, b, c are distinct digits. Find the sum of the elements of $\mathcal{S}$.

The next few problems are harder; in these problems linearity of expectation is not the main idea of the solution. All problems below were written by Lewis Chen.

Problem 2.8 (NIMO 4.3)**.** One day, a bishop and a knight were on squares in the same row of an infinite chessboard, when a huge meteor storm occurred, placing a meteor in each square on the chessboard independently and randomly with probability p. Neither the bishop nor the knight were hit, but their movement may have been obstructed by the meteors. For what value of p is the expected number of valid squares that the bishop can move to (in one move) equal to the expected number of squares that the knight can move to (in one move)?

Problem 2.9 (NIMO 5.6)**.** Tom has a scientific calculator. Unfortunately, all keys are broken except for one row: `1`, `2`, `3`, `+` and `-`. Tom presses a sequence of 5 random keystrokes; at each stroke, each key is equally likely to be pressed. The calculator then evaluates the entire expression, yielding a result of E. Find the expected value of E.

(Note: Negative numbers are permitted, so `13-22` gives $E = -9$. Any excess operators are parsed as signs, so `-2-+3` gives $E = -5$ and `-+-31` gives $E = 31$. Trailing operators are discarded, so `2++-+` gives $E = 2$. A string consisting only of operators, such as `-++-+`, gives $E = 0$.)

3 Direct Existence Proofs

In its simplest form, we can use expected value to show existence as follows: suppose we know that the average score of the USAMO 2014 was 12.51. Then there exists a contestant who got at least 13 points, and a contestant who got at most 12 points. This is similar in spirit to the pigeonhole principle, but the probabilistic phrasing is far more robust.

3.1 A First Example

Let's look at a very simple example, taken from the midterm of a class at the San Jose State University.[2]

Example 3.1 (SJSU M179 Midterm)**.** Prove that any subgraph of $K_{n,n}$ with at least $n^2 - n + 1$ edges has a perfect matching.

We illustrate the case $n = 4$ in the figure.

[2] For a phrasing of the problem without graph theory: given n red points and n blue points, suppose we connect at least $n^2 - n + 1$ pairs of opposite colors. Prove that we can select n segments, no two of which share an endpoint.

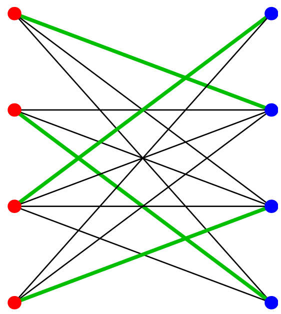

Figure 1: The case $n = 4$. There are $n^2 - n + 1 = 13$ edges, and the matching is highlighted in green.

This problem doesn't "feel" like it should be very hard. After all, there's only a total of n^2 possible edges, so having $n^2 - n + 1$ edges means we have practically all edges present.[3]

So let's be really careless and just *randomly* pair off one set of points with the other, regardless of whether there is actually an edge present. We call the *score* of such a pairing the number of pairs which are actually connected by an edge. We wish to show that some pairing has score n, as this will be the desired perfect matching.

So what's the expected value of a random pairing? Number the pairs $1, 2, \dots, n$ and define

$$X_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if the } i\text{th pair is connected by an edge} \\ 0 & \text{otherwise.} \end{cases}$$

Then the score of the configuration is $X = X_1 + X_2 + \cdots + X_n$. Given any red point and any blue point, the probability they are connected by an edge is at least $\frac{n^2-n+1}{n^2}$. This means that $\mathbb{E}[X_i] = \frac{n^2-n+1}{n^2}$, so

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] \\ &= n \cdot \mathbb{E}[X_1] \\ &= \frac{n^2 - n + 1}{n} \\ &= n - 1 + \frac{1}{n}. \end{aligned}$$

Since X takes only integer values, there must be some configuration which achieves $X = n$. Thus, we're done.

3.2 Application: Ramsey Numbers

Let's do another simple example. Before we begin, I will quickly introduce a silly algebraic lemma, taken from [5, page 30].

Lemma 3.2. *For any positive integers n and k,*

$$\binom{n}{k} < \frac{1}{e}\left(\frac{en}{k}\right)^k.$$

Here $e \approx 2.718\dots$ is Euler's constant.

[3] On the other hand, $n^2 - n + 1$ is actually the best bound possible. Can you construct a counterexample with $n^2 - n$?

Proof. Do $\binom{n}{k} < \frac{n^k}{k!}$ and then use calculus to prove that $k! \geq e(k/e)^k$. Specifically,

$$\ln 1 + \ln 2 + \cdots + \ln k \geq \int_{x=1}^{k} \ln x \, dx = k \ln k - k + 1$$

whence exponentiating works. $\square$

Algebra isn't much fun, but at least it's easy. Let's get back to the combinatorics.

Example 3.3 (Ramsey Numbers)**.** Let n and k be integers with $n \leq 2^{k/2}$ and $k \geq 3$. Then it is possible to color the edges of the complete graph on n vertices with the following property: one cannot find k vertices for which the $\binom{k}{2}$ edges among them are monochromatic.

Remark. In the language of Ramsey numbers, prove that $R(k,k) > 2^{k/2}$.

Solution. Again we just randomly color the edges and hope for the best. We use a coin flip to determine the color of each of the $\binom{n}{2}$ edges. Let's call a collection of k vertices *bad* if all $\binom{k}{2}$ edges are the same color. The probability that any collection is bad is

$$\left(\frac{1}{2}\right)^{\binom{k}{2}-1}.$$

The number of collections in $\binom{n}{k}$, so the expected number of bad collections is

$$\mathbb{E}[\text{number of bad collections}] = \frac{\binom{n}{k}}{2^{\binom{k}{2}-1}}.$$

We just want to show this is less than 1. You can check this fairly easily using Lemma 3.2; in fact, we have a lot of room to spare. $\square$

3.3 Practice Problems

The first two problems are from [2]; the last one is from [4].

Problem 3.4. Show that one can construct a (round-robin) tournament outcome with more than 1000 people such that for any set of 1000 people, some contestant outside that set beats all of them.

Problem 3.5 (BAMO 2004)**.** Consider a set of n real numbers, not all zero, with sum zero. Prove that one can label the numbers as a_1, a_2, $\ldots$, a_n such that

$$a_1a_2 + a_2a_3 + \cdots + a_na_1 < 0.$$

Problem 3.6 (Russia 1996)**.** In the Duma there are 1600 delegates, who have formed $16,000$ committees of 80 people each. Prove that one can find two committees having no fewer than four common members.

4 Heavy Machinery

Here are some really nice ideas used in modern theory. Unfortunately I couldn't find many olympiad problems that used them. If you know of any, please let me know!

4.1 Alteration

In previous arguments we often proved a result by showing $\mathbb{E}[\text{bad}] < 1$. A second method is to select some things, find the expected value of the number of "bad" situations, and subtract that off. An example will make this clear.

Example 4.1 (Weak Turán)**.** A graph G has n vertices and average degree d. Prove that it is possible to select an independent set of size at least $\frac{n}{2d}$.

Proof. Rather than selecting $\frac{n}{2d}$ vertices randomly and hoping the number of edges is 1, we'll instead select each vertex with probability p. (We will pick a good choice of p later.)

That means the expected number of vertices we will take is np. Now there are $\frac{1}{2}nd$ edges, so the expected number of "bad" situations (i.e. an edge in which both vertices are taken) is $\frac{1}{2}nd \cdot p^2$.

Now we can just get rid of all the bad situations. For each bad edge, delete one of its endpoints arbitrarily (possibly with overlap). This costs us at most $\frac{1}{2}nd \cdot p^2$ vertices, so the expected value of the number of vertices left is

$$np - \frac{1}{2}ndp^2 = np\left[1 - \frac{1}{2}dp\right].$$

It seems like a good choice of p is $\frac{1}{d}$, which now gives us an expected value of $\frac{n}{2d}$, as desired. $\square$

A stronger result is Problem 5.5.

4.2 Union Bounds and Markov's Inequality

A second way to establish existence is to establish a nonzero probability. One way to do this is using a union bound.

Proposition 4.2 (Union Bound)**.** *Consider several events $A_1, A_2, \dots, A_k$. If*

$$\mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_k) < 1$$

then there is a nonzero probability that none of the events occur.

The following assertion is sometimes useful for this purpose.

Theorem 4.3 (Markov's Inequality)**.** *Let X be a random variable taking only nonnegative values. Suppose $\mathbb{E}[X] = c$. Then*

$$\mathbb{P}(X \ge rc) \le \frac{1}{r}.$$

This is intuitively obvious: if the average score on the USAMO was 7, then at most $\frac{1}{6}$ of the contestants got a perfect score. The inequality is also sometimes called *Chebyshev's inequality* or *the first Chebyshev inequality.*

4.3 Lovász Local Lemma

The Lovász Local Lemma (abbreviated LLL) is in some sense a refinement of the union bound idea – if the events in question are "mostly" independent, then the probability no events occur is still nonzero.

We present below the "symmetric" version of the Local Lemma. An asymmetric version also exists (see Wikipedia).

Theorem 4.4 (Lovász Local Lemma). *Consider several events, each occurring with probability at most p, and such that each event is independent of all the others except at most d of them.* [4] *Then if*

$$epd \leq 1$$

the probability that no events occur is positive. (Here $e = 2.71828\ldots$ is Euler's constant.)

Note that we don't use the number of events, only the number of dependencies.

As the name implies, the local lemma is useful in situations where in a random algorithm, it appears that things do not depend much on each other. The following Russian problem is such an example.

Example 4.5 (Russia 2006). At a tourist camp, each person has at least 50 and at most 100 friends among the other persons at the camp. Show that one can hand out a T-shirt to every person such that the T-shirts have (at most) 1331 different colors, and any person has 20 friends whose T-shirts all have pairwise different colors.

The constant $C = 1331$ is extremely weak. We'll reduce it to $C = 48$ below.

Solution. Give each person a random T-shirt. For each person P, we consider the event $E(P)$ meaning "P's neighbors have at most 19 colors of shirts". We wish to use the Local Lemma to prove that there is a nonzero probability that no events occur.

If we have two people A and B, and they are neither friends nor have a mutual friend (in graph theoretic language, the distance between them is at least two), then the events $E(A)$ and $E(B)$ do not depend on each other at all. So any given $E(P)$ depends only on friends, and friends of friends. Because any P has at most 100 friends, and each of these friends has at most 99 friends other than P, $E(P)$ depends on at most $100+100\cdot 99 = 100^2$ other events. Hence in the lemma we can set $d = 100^2$.

For a given person, look at their $50 \leq k \leq 100$ neighbors. The probability that there are at most 19 colors among the neighbors is clearly at most

$$\binom{C}{19} \cdot \left(\frac{19}{C}\right)^k.$$

To estimate the binomial coefficient, we can again use our silly Lemma 3.2 to get that this is at most

$$\frac{1}{e}\left(\frac{eC}{19}\right)^{19} \cdot \left(\frac{19}{C}\right)^k = e^{18} \cdot \left(\frac{19}{C}\right)^{k-19} \leq e^{18}\left(\frac{19}{C}\right)^{31}.$$

Thus, we can put $p = e^{18}\left(\frac{19}{C}\right)^{31}$. Thus the Lemma implies we are done as long as

$$e^{19}\left(\frac{19}{C}\right)^{31} \cdot 100^2 \leq 1.$$

It turns out that $C = 48$ is the best possible outcome here. Establishing the inequality when $C = 1331$ just amounts to some rough estimation with the e's. $\square$

[4] More precisely, if we donate the events with binary variables $X_1, X_2, \ldots$, then we require that for each i, there is a set D_i of size at most $d+1$ containing X_i such that X_i is independent of its complement D_i^c. In other words, measuring the value of all the variables in D_i^c together will tell you nothing about the distribution of X_i.

5 More Practice Problems

These problems are mostly taken from [2, 4].

Problem 5.1 (IMC 2002)**.** An olympiad has six problems and 200 contestants. The contestants are very skilled, so each problem is solved by at least 120 of the contestants. Prove that there exist two contestants such that each problem is solved by at least one of them.

Problem 5.2 (Romania 2004)**.** Prove that for any complex numbers $z_1, z_2, \dots, z_n$, satisfying $|z_1|^2 + |z_2|^2 + \dots + |z_n|^2 = 1$, one can select $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \in \{-1, 1\}$ such that

$$\left| \sum_{k=1}^{n} \varepsilon_k z_k \right| \le 1.$$

Problem 5.3 (Shortlist 1999 C4)**.** Let A be a set of N distinct residues $\pmod{N^2}$. Prove that there exists a set B of N residues $\pmod{N^2}$ such that $A+B = \{a+b | a \in A, b \in B\}$ contains at least half of all the residues $\pmod{N^2}$.

Problem 5.4 (Iran TST 2008/6)**.** Suppose 799 teams participate in a round-robin tournament. Prove that one can find two disjoint groups A and B of seven teams each such that all teams in A defeated all teams in B.

Problem 5.5 (Caro-Wei Theorem)**.** Consider a graph G with vertex set V. Prove that one can find an independent set with size at least

$$\sum_{v \in V} \frac{1}{\deg v + 1}.$$

Remark. Note that, by applying Jensen's inequality, our independent set has size at least $\frac{n}{d+1}$, where d is the average degree. This result is called **Turán's Theorem** (or the complement thereof).

Problem 5.6 (USAMO 2012/6)**.** For integer $n \ge 2$, let $x_1, x_2, \dots, x_n$ be real numbers satisfying $x_1 + x_2 + \ldots + x_n = 0$ and $x_1^2 + x_2^2 + \ldots + x_n^2 = 1$. For each subset $A \subseteq \{1, 2, \dots, n\}$, define

$$S_A = \sum_{i \in A} x_i.$$

(If A is the empty set, then $S_A = 0$.) Prove that for any positive number λ, the number of sets A satisfying $S_A \ge \lambda$ is at most $2^{n-3}/\lambda^2$.

Problem 5.7 (Online Math Open, Ray Li)**.** Kevin has $2^n - 1$ cookies, each labeled with a unique nonempty subset of $\{1, 2, \dots, n\}$. Each day, he chooses one cookie uniformly at random out of the cookies not yet eaten. Then, he eats that cookie, and all remaining cookies that are labeled with a subset of that cookie. Compute the expected value of the number of days that Kevin eats a cookie before all cookies are gone.

Problem 5.8. Let n be a positive integer. Let a_k denote the number of permutations of n elements with k fixed points. Compute

$$a_1 + 4a_2 + 9a_3 + \dots + n^2 a_n.$$

Problem 5.9 (Russia 1999)**.** In a certain school, every boy likes at least one girl. Prove that we can find a set S of at least half the students in the school such that each boy in S likes an odd number of girls in S.

Problem 5.10. Let n be a positive integer. Suppose $11n$ points are arranged in a circle, colored with one of n colors, so that each color appears exactly 11 times. Prove that one can select a point of every color such that no two are adjacent.

References

[1] pythag011 at http://www.aops.com/Forum/viewtopic.php?f=133&t=481300

[2] Ravi B's collection of problems, available at:
http://www.aops.com/Forum/viewtopic.php?p=1943887#p1943887.

[3] Problem 6 talk ($c > 1$) by Po-Shen Loh, USA leader, at the IMO 2014.

[4] Also MOP lecture notes: http://math.cmu.edu/~ploh/olympiad.shtml.

[5] Lecture notes by Holden Lee from an MIT course:
http://web.mit.edu/~holden1/www/coursework/math/18997/notes.pdf

Thanks to all the sources above. Other nice reads that I went through while preparing this, but eventually did not use:

1. Alon and Spencer's *The Probabilistic Method*. The first four chapters are here: http://cs.nyu.edu/cs/faculty/spencer/nogabook/.
2. A MathCamp lecture that gets the girth-chromatic number result: http://math.ucsb.edu/~padraic/mathcamp_2010/class_graph_theory_probabilistic/lecture2_girth_chromatic.pdf

6 Unexpected Expectations: Solution Sketches

2.6 Answer: 9.1. Make an indicator variable for each adjacent pair...

2.7 Answer: 360. Pick a, b, c randomly and compute $\mathbb{E}[0.\overline{abc}]$. Then multiply by $|\mathcal{S}|$.

2.8 $8p = 4 \cdot \left(p + p^2 + p^3 + \dots\right)$.

2.9 Answer: 1866. Any expression with a + or - in it has a complementary expression with that sign switched, such that any numbers after the sign are cancelled out in expectation. Thus we need only consider numbers occurring before a sign. Also, any expression with a 3 in it has a complementary expression with a 1 instead of the 3, so in expectation every numeral is a 2. The probability of hitting a numeral in any keystroke is $p = \frac{3}{5}$, so the total expectation is

$$2(p + 10p^2 + \cdots + 10^4 p^5) = 2p \cdot \frac{(10p)^5 - 1}{10p - 1} = \frac{6}{5} \cdot \frac{7775}{5} = 6 \cdot 311 = 1866$$

3.4 Suppose there are n people, and decide each game outcome with a coin flip. Let U be the set of "unbeaten" subsets S of size 1000, i.e. such that nobody outside S beats all of S.

$$\begin{aligned} \mathbb{E}[|U|] &= \sum_{|S|=1000} \mathbb{P}(S \text{ is unbeaten}) \\ &= \sum_{|S|=1000} \mathbb{P}(\forall t \in S^c \; \exists s \in S \; : \; s \text{ beats } t) \\ &= \sum_{|S|=1000} \prod_{t \in S^c} 1 - 2^{-1000} \\ &= \binom{n}{1000} \cdot (1 - 2^{-1000})^{n-1000} \end{aligned}$$

which is less than 1 for very large n (exponentials eventually dominate polynomials). Hence for large n, sometimes $|U| = 0$, as needed.

3.5 Choose the ordering uniformly randomly. Then, with the convention $n + 1 = 1$,

$$\mathbb{E}[a_i a_{i+1}] = \mathbb{E}[a_i \mathbb{E}[a_{i+1} \mid a_i]] = \mathbb{E}[a_i \cdot (\frac{-a_i}{n-1})] = -\mathbb{E}[\frac{a_i^2}{n-1}] < 0$$

since they are not all zero, hence the expectation of the given sum is strictly negative, and so the sum itself is sometimes negative.

3.6 Let n_i be the number of committees which the ith delegate is in. Pick two committees A and B randomly, so

$$\mathbb{E}[|A \cap B|] = \sum_i \frac{n_i(n_i - 1)}{16,000 \cdot 15,999}$$

Letting $f(n) = n(n-1)$, by Jensen's inequality and the fact that the average person is on $\frac{16,000 \cdot 80}{1600} = 800$ committees,

$$\mathbb{E}[|A \cap B|] \geq \frac{1600 f(800)}{16,000 \cdot 15,999} = \frac{639,200}{10 \cdot 15,999} > 4$$

so sometimes thee two committees have at least 4 people in common.

5.1 Pick the two contestants, 1 and 2, randomly. Let X_i be the indicator that both contestants miss problem i, so each $\mathbb{E}[X_i] < (\frac{80}{200})^2 = 4/25$, and their expected number of both-missed problems is $24/25 < 1 \dots$

5.2 Select each of the ε_i randomly with a coin flip. Let LHS be the left-hand side of the desired inequality. Since $|z|^2 = z\bar{z}$ for any z, $LHS^2 = \sum_k |z_k|^2 + \sum_{i<j} \epsilon_i \epsilon_j (z_i \bar{z}_j + \bar{z}_i z_j)$ and since ϵ_i and ϵ_j are independent, $\mathbb{E}(\epsilon_i \epsilon_j) = 0$, so $\mathbb{E}(LHS^2) = \sum_k |z_k|^2 = 1$, hence $LHS^2 \le 1$ sometimes, as needed.

5.3 Select the elements of $B = \{b_1 \dots b_n\}$ uniformly randomly (we'll even allow repetitions, for simplicity). For each $r \pmod{N^2}$, and each $i, j \in 1 \dots n$,

$$\mathbb{P}(r \notin A + b_i) = \mathbb{P}(b_i \notin A - r) = \frac{N^2 - N}{N^2} = 1 - \frac{1}{N}$$

so $\mathbb{P}(r \notin A + B) < (1 - \frac{1}{N})^N < \frac{1}{e} < \frac{1}{2}$. Thus $\mathbb{E}[|A + B|] > (1 - \frac{1}{e})N^2 \dots$

5.4 Let D_k be the set of teams which defeat the kth team (here $1 \le k \le 799$), and $d_k = |D_k|$ Select $A = \{a_1, \dots, a_7\}$ randomly, so $\mathbb{P}(A \subseteq D_k) = \binom{d_k}{7}/\binom{799}{7}$, so letting N be the number of teams dominated by A, $\mathbb{E}[N] = \sum_k \binom{d_k}{7}/\binom{799}{7}$. The function $\binom{x}{7}/\binom{799}{7}$ is convex, and the average value of d_k is $798/2 = 398$, so by Jensen's inequality, $\mathbb{E}[N] \ge 799 \cdot \binom{398}{7}/\binom{799}{7} > 799 \cdot (\frac{1}{2})^7 > 6$, hence sometimes $N \ge 7$.

5.5 A fairly natural approach is to use a greedy algorithm: randomly choose a vertex, append it to W, remove it and its neighbors from G, repeat until nothing is left, and then W will be an independent set. One can prove by induction on $|G|$ that $\mathbb{E}[|W|]$ satisfies the given bound.

A simpler proof is to randomly order the vertices $\{v_1 \dots v_n\}$ of G, and take W to be the subset of those v_i which occur before all their neighbors. Then

$$\mathbb{E}[|W|] = \sum_i ind(v_i \in W) = \sum_i \frac{1}{deg(v) + 1}$$

5.6 Since $S_A = -S_{A^c}$, we have $\mathbb{P}(S_A > \lambda) = \mathbb{P}(S_{A^c} < -\lambda) = \mathbb{P}(S_A < -\lambda)$ Thus $\mathbb{P}(S_A > \lambda) = \frac{1}{2}\mathbb{P}(S_A^2 > \lambda^2)$, and S_A^2 is always positive, so we can apply the Markov inequality to it.

$$\begin{aligned} \mathbb{E}(S_A^2) &= \sum_i \mathbb{P}(i \in A) x_i^2 + \sum_{ij} \mathbb{P}(i, j \in A) 2x_i x_j \\ &= \frac{1}{2} \sum_i x_i^2 + \frac{1}{4} \sum_{i \ne j} 2x_i x_j \end{aligned}$$

Since $0 = S_{[n]}^2 = 1 + \sum_{i \ne j} 2x_i x_j$, we have

$$\mathbb{E}(S_A^2) = \frac{1}{2}(1) + \frac{1}{4}(-1) = \frac{1}{4}$$

hence by the Markov inequality, $\mathbb{P}(S_A^2 > \lambda^2) < \frac{1}{4\lambda^2}$, as needed.

5.7 The number of days equals the number of times a cookie is chosen (rather than merely eliminated). Let C be the set of cookies chosen by the process and $S = \{1 \dots n\}$, so

$$\mathbb{E}[\#days] = \mathbb{E}[|C|] = \sum_{A \subseteq S} \mathbb{E}[ind(A \in C)] = \sum_{A \subseteq S} \mathbb{P}(A \in C)$$

For any $A \subseteq S$, A is eliminated at the unique stage where a superset A' of A is chosen, so

$$\mathbb{P}(A \in C) = \mathbb{P}(A = A') = \frac{1}{\#\{\text{supersets of } A\}} = \frac{1}{2^{n-|A|}}$$

Thus

$$\sum_{A \subseteq S} \mathbb{P}(A \in C) = \sum_{A \subseteq S} 2^{|A|-n} = \sum_{1 \le k \le n} 2^k \cdot 2^{k-n} = \frac{4^{n+1}-4}{3 \cdot 2^n}$$

5.8 For a random permutation let X be the number of fixed points, so the required expression is exactly $n!\mathbb{E}[X^2]$. We already know $\mathbb{E}[X] = 1$ from Example 2.1, and by a similar argument, the expected number of pairs of fixed points in a random permutation is

$$\mathbb{E}[\binom{X}{2}] = \sum_{ij} \mathbb{P}(i, j \text{ both fixed}) = \binom{n}{2} \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{2}$$

Then $\mathbb{E}[X^2] = 2\mathbb{E}[\binom{x}{2}] + \mathbb{E}[X] = 2$, so the given expression is $2n!$.

5.9 Let L_b be the set of girls liked by a given boy b, and let B and G be the sets of chosen boys and girls. For fixed G, WLOG B is the set of all boys who like an odd number of girls in G, so the challenge is to choose G. Doing so uniformly randomly means each girl has probability 50% to be included, and for each boy b,

$$\mathbb{P}(b \in B) = \mathbb{P}(|L_b \cap G| \text{ is odd}) = 50\%$$

because a uniformly random subset of L_b is 50% likely to be odd. Hence $\mathbb{E}[|B \cup G|] > 50\%(total)$.

5.10 Label the points $s_1 \dots s_{11n} = s_0$ in order around the circle. Choose one point of each color randomly to form a set A, and consider the indicators $B_i = ind(s_i, s_{i+1} \in A)$ of the "bad" events where an adjacent pair occurs in A. For each i, $p = \mathbb{P}(B_i)$ is 11^{-2} if they're different colors, or 0 if they're the same color, so $p \le 11^{-2}$. Unfortunately, the bound $\mathbb{E}[\sum_i B_i] < \frac{11n}{11^2} = \frac{11}{n}$ is only below 1 if $n > 11$, so for small n this bound is not strong enough to show that sometimes all the $B_i = 0$.

However, each B_i is independent of B_j for all j except when s_j or s_{j+1} has the same color as s_i or s_{i+1}, so we can try to apply the Lovasz Local Lemma. There are 21 other pairs (s_i, s_{i+1}) sharing a color with s_i, and at most another 21 pairs sharing a color with s_{i+1}, so $B_i \perp\!\!\!\perp B_j$ for all but at most $d = 42$ values of j. Now,

$$epd = e \cdot \frac{42}{121} < \frac{28}{10} \cdot \frac{42}{121} < \frac{30 \cdot 40}{1210} < 1$$

so by LLL there is a positive probability that all the $B_i = 0$, as needed.