DIMACS: Probability Crash Course - Day 5

Instructor: Pat Devlin — prd41@math.rutgers.edu

Summer, 2015

To infinity and beyond

What does Pat do: Part I — Probabilistic method

Example 1 Michelangelo and Picasso are playing a game with paint.

- They first agree on a proportion p.
- Michelangelo then has to put paint on the surface of a sphere. He can paint it in any weird way that he wants to. But the proportion of the surface area that he paints *needs* to be at least p.
- Picasso was a cubist, so after Michelangelo is done painting, Picasso then inscribes a cube in the partially painted sphere. Picasso can rotate this cube however he wants to before he puts it in.
- Picasso wins if he can inscribe the cube in such a way that all eight of its corners are touching a painted part of the sphere. Otherwise, Michelangelo wins.

If $p \leq 0.25$, then Michelangelo can easily win (how?). On the other hand, if Michelangelo is required to paint *the entire sphere*, then Picasso will win without even trying!

(a) What's the *biggest* value of p you can find where Michelangelo will definitely be able to win?

(b) What's the *smallest* value of p you can find where Picasso will definitely be able to win?

The idea Pat showed for part (b) in the last question is called the ***probabilistic method***. This is a really clever way of proving something exists, and it's *really* fun to use this to prove things that have nothing to do with probability! This technique was pioneered by Paul Erdős (1913–1996). The idea can be described as follows.

You are standing in front of a haystack, and you want to prove that there's a needle somewhere inside it. Finding the needle itself is perhaps much too difficult and hard to think about. So the technique is to reach into the haystack and randomly pull something out. Then you argue that the probability that you grab a needle is not literally equal to 0. Therefore, there must be at least some chance of pulling out a needle, and therefore, there must be at least one needle somewhere in the haystack! (End of proof.) To find out more about this idea, the book The Probabilistic Method by Alon and Spencer is the absolute best. That book covers how to use this idea in discrete mathematics, number theory, and especially graph theory.

Example 2 Let G be a graph with m edges. Show that you can always find a bipartite subgraph of G having at least $m/2$ edges.

Example 3 (Erdős, Rényi; 1947) Lower bounds on Ramsey numbers.

Definition: A *sum-free set* is a set T such that the equation $x + y = z$ has no solutions over T.

Example 4 Let S be a subset of the integers modulo n. Then there is a sum-free subset $T \subseteq S$ with $|T| \geq |S|/3$. (The equations $x + y = z$ are all understood to be modulo n.)

Definition: A number is called *algebraic* if it can be described using the language of algebra (i.e., if it can be written as the solution to a polynomial equation with only integer coefficients). Otherwise, the number is called *transcendental*, which makes it a very strange type of number.

Example 5 Transcendental numbers exist. Moreover, there are numbers that exist but that could never even be defined, described, or computed in any exact way whatsoever.

Definition: A *t-design* is a collection, $\mathcal{F}$, of subsets of $\{1, 2, \ldots, n\}$ (these subsets are called *blocks*) such that there are fixed numbers k, r and λ where

- if $B \in \mathcal{F}$, then $|B| = k$ (i.e., each block has size k),
- each number $1 \leq x \leq n$ appears in exactly r blocks, and
- if $T \subseteq \{1, 2, \ldots, n\}$ has size $|T| = t$, then T appears in exactly λ blocks.

Example 6 The *Fano plane* is a 2-design on $n = 7$ points with $k = 3$, $r = 3$, and $\lambda = 1$.

Question: (Jakob Steiner, 1853) What t-designs exist?

- This is the oldest open question in discrete math
- Today, none whatsoever are known for $t > 5$
- If t-designs exist, then they have to be very special and delicate
- Finally answered by Peter Keevash in January 2014 using probability. But, we *still* don't know any actual examples!

What does Pat do: Part II — Random graphs

Studying "random graphs" is studying *typical* graphs. This goes back to Paul Erdős and Alfréd Rényi in a 1959 publication. A *random graph*, denoted $G_{n,p}$, is a graph that we make by first drawing n vertices. Then for each possible edge, we flip a biased coin. We draw that edge with probability p, and otherwise, we don't draw it. This notion is one of the most useful ideas in combinatorics, and it helps you answer *many* questions that have nothing to do with probability.

See The Probabilistic Method book for much more detail on this.

Example 7 Say $G_{n,p}$ is a random graph. How many triangles do we expect this graph to have?

Example 8 (Achlioptas, Naor; 2005) The chromatic number is *very hard* to figure out. However, 99.999% of graphs on n vertices have *the exact same* chromatic number! This number is $\lfloor n/(2\log_2(n)) \rfloor$.

The *girth* of a graph is the size of the smallest cycle, so graphs with high girth don't have *any* small cycles.

Example 9 (Erdős) For all g and l, there are graphs, G, with $girth(G) \geq g$ and $\chi(G) \geq l$.

What does Pat do: Part III — Randomness and computer science

Example 10 I flip a coin 3 times. Let X be the number of H that show up. You are allowed to ask yes-no questions to figure out X, but you pay a dollar for each question.

(a) Show that there's a strategy that *always* asks at most 2 questions.

(b) Show that there is *not* a strategy that *always* asks at most 1 question.

(c) Show that there's a strategy that asks fewer than 2 questions *on average.*

Definition: This is called *information theory*, and the basic building block of this is a tool called *entropy*[1]. This was all discovered by Claude Shannon in his ludicrously ground-breaking 1948 paper *A mathematical theory of communication* (this single paper has been cited an astounding 75,154 times) (this paper is where the word *bit* comes from). It's entirely about using probability as an extremely powerful tool to understand the theoretical framework and limitations of computers.

Claude Shannon also made practical seminal contributions to computer science, including the development of digital circuitry. He also served as a brilliant code-breaker in World War II, where he met and interacted with Alan Turing. Claude Shannon's work is an absolute pillar of computer science and cryptography. He worked just 15 miles from Rutgers at Bell Labs as did many other great discrete mathematicians including Neil Sloane.

To learn more on entropy, read Elements of Information Theory by Cover and Thomas or read the online notes called Entropy and Counting by Radhakrishnan. These same techniques give many other beautiful results.

Example 11 Suppose we have n distinct points in $\mathbb{R}^3$ with n_1 projections onto the (x, y)-plane, n_2 projections onto the (x, z)-plane, and n_3 projections onto the (y, z)-plane. Show that $n^2 \leq n_1 n_2 n_3$.

Example 12 (Brégman 1973) Suppose G is a bipartite graph with sides A and B. Let $\Phi(G)$ count the number of perfect matchings of G. Then we have

$$\Phi(G) \leq \prod_{v \in A} (d(v)!)^{1/d(v)}.$$

Example 13 (Kahn, Lawrenz; 1999) Suppose G is a d-regular bipartite graph with sides A and B where $|A| = |B| = n$. Then the number of independent sets of G is at most $\left(2^{d+1} - 1\right)^{n/d}$.

[1] Sorry, there's no connection between this mathematical idea and the physics principle.

Infinity is a prickly pear

Example 14 (Buffon's needle) [Copy from board.]

Example 15 (Bertand's paradox) [Copy from board.]

Example 16 (There are dragons) Non-measurable sets.

Example 17 (Banach-Tarski paradox) [Copy from board]

Other things to think about

These are either good questions to ponder, or cool topics to google (or both!). I tried to get many different flavors to please as many people as possible. So if you don't like some topic or problem, try another one! When I was your age, I would have *loved* to have somebody to talk to about math stuff. To that end, feel free to email me any time. I know *a lot* about everything you've been learning about (except for number theory and cryptography, which I don't know nearly as well). `prd41@math.rutgers.edu`

- Look through these notes and the notes from earlier days. Try to solve the problems and examples I gave. You can do the same with the homework problems.
- Markov chains / random walks (this is a *really* cool application of matrices)
- Binomial r.v.s, Poisson r.v.s, geometric r.v.s, et cetera
- Probability generating functions
- If we pick an integer x from 1 to N, what is the probability that x is not divisible by any perfect squares greater than 1? (Think of N as very, very big. This connects to the Riemann zeta function.)
- Quick sort algorithm
- Monte Carlo method
- Random Walks and Electric Networks (free online) by Doyle and Snell
- "The secretary problem" (aka the marriage problem, sultan's dowry problem, et cetera)
- Saint Petersburg paradox
- "The two envelopes problem" (not the same as what's right below this)
- There are two envelopes on the table. Each one has a different real number between 0 and 1 written in it. You're allowed to pick up either envelope and open it to see what number is inside. You then need to guess which envelope has the larger number in it. Show that there is a strategy that always works *strictly more* than 50% of the time ***no matter what numbers*** happen to be written in each envelope. (This is *outrageous*. Ask me for hints.)
- Variance, higher moments, moment generating functions
- "Poisson paradigm"
- the "bell curve" / Central limit theorem
- Stirling's approximation
- random processes
- queueing theory
- renewal theory
- statistics
- continuous r.v.s (requires calculus)
- measure theory (requires that you be so good at calculus that you're bored senseless)