The Probabilistic Method

Cameron Derwin and Margalit Glasgow
Mentor: Chiheon Kim

Fourth Annual MIT Primes Conference

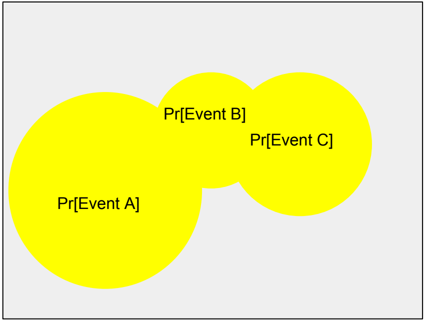
Saturday, May 17, 2014

What is the Probabilistic Method?

- A method of proving that a certain structure must exist.
- The structure exists if a randomly chosen element in the probability space has the desired structure with **positive** probability.
- We often look at the probability that a random element does **not** have the desired structure.

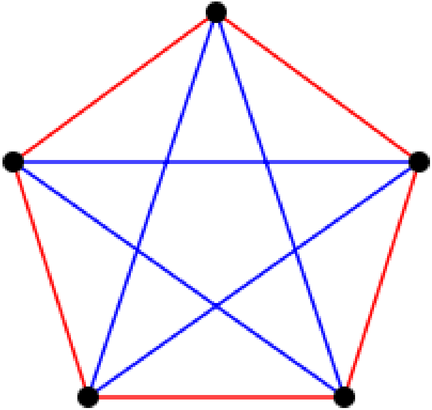
A Useful Rule

$$\sum_{e \in E} Pr(A_e) \geq Pr(\bigvee_{e \in E} A_e)$$



Ramsey Numbers

The *Ramsey Number* $R(k, l)$ is the smallest number n such that given any coloring of the complete graph K_n by red and blue, there is either a red K_k or blue K_l. For example, $R(3, 3) \neq 5$ because K_5 can be colored as



.

Ramsey Numbers

One of the first applications of the Probabilistic method was to prove the following theorem:

Theorem (Erdös)

For integers $k < n$, if $\binom{n}{k}2^{1-\binom{k}{2}} < 1$ then $R(k,k) > n$ so $R(k,k) > \lfloor 2^{k/2} \rfloor$.

- Each edge is colored red with probability $\frac{1}{2}$ and blue otherwise
- A_R is the event that the induced complete subgraph on R is monochromatic where R is a subset of V of size k

Then the probability that K_n contains a monochromatic subgraph of size k is at most

$$\sum_{\substack{R \subset V \\ |R|=k}} Pr\left[A_R\right] = \binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

.

Expected Value

- Linearity of Expectaion: $E[A+B] = E[A] + E[B]$.
- Often counted using indicator random variables.
- If the expected number of events occuring is less than k, there must be a case in which at most $k-1$ events occur.

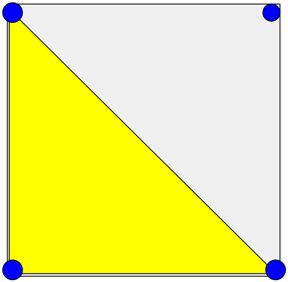
The Smallest Triangle Problem

Given a set n points in a unit square (the set S),

Let T(S) = the area of the smallest triangle defined by three of the n points in S.

What is the maximum value of T(S) over all sets S of n points?

For example, the max over 4 points is $\frac{1}{2}$.



The Smallest Triangle Problem

Theorem (Komlós, Pintz, Szemerédi)

There is a set S of n points in the unit square U such that $T(S) \geq 1/(100n^2)$.

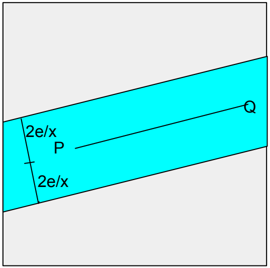
Overview

- Determine the probibility that a random 3 points make a triange of area less than $1/(100n^2)$.
- Find the expected number of "too small" triangles in a random set of $2n$ points.
- Remove points from "too small" triangles to create optimal set.

The Smallest Triangle Problem

Given 3 random points P, Q, R in U, what is the probability that area of $\Delta PQR \leq 1/(100n^2)$?

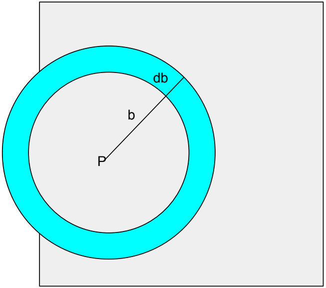
- Area $\Delta PQR = \mu$
- Condition on the distance x between P and Q.
- $\Pr[\mu \leq \epsilon] \leq 4\sqrt{2} \cdot \frac{\epsilon}{x}$



The Smallest Triangle Problem

What if the distance x is random?

$$Pr[b \le x \le b + db] \le 2\pi b db$$



$$Pr[\mu \le \epsilon] \le \int_0^{\sqrt{2}} (4\sqrt{2} \cdot \frac{\epsilon}{b})(2\pi b) db = 16\pi\epsilon \le \frac{0.6}{n^2}$$

The Smallest Triangle Problem

Making Alterations

Randomly Place $2n$ points.

Expected number of triangles with area $\leq 1/(100n^2) =$

$$\binom{2n}{3} \cdot \frac{0.6}{n^2} < \frac{8n^3}{6} \cdot \frac{0.6}{n^2} < n$$

There must be a case with less than n "bad" triangles of area less than $\frac{1}{100n^2}$.

Remove one point from each at most n "bad" triangles. We are left with a set of n points and no triangles with area less than $\frac{1}{100n^2}$.

Hardy - Ramanujan Theorem

In 1934 Turán proved the following result:

- $\nu(n)$ denotes the number of primes dividing n
- $\omega(n) \to \infty$ arbitrarily slowly
- $\pi(n)$ is the number of x in $\{1, \ldots, n\}$ such that

$$|\nu(x) - \ln \ln n| > \omega(n)\sqrt{\ln \ln n}$$

Theorem (Turán 1934)

$\pi(n) = O(n)$, *meaning that*

$$\lim_{n \to \infty} \frac{\pi(n)}{n} = 0.$$

Variance and Covariance

Variance is the measure of how spread out a specific random variable is.

- Definition: for a random variable X,

$$\mathrm{Var}\left[X\right] = E\left[\left(X - E[X]\right)^2\right]$$

- By linearity of expectation, this is equivalent to

$$\mathrm{Var}\left[X\right] = E\left[X^2\right] - E\left[X\right]^2$$

- *Covariance*, of which variance is a special case, is a measure of correlation between two random variables
- For random variables X and Y,

$$\mathrm{Cov}\left[X, Y\right] = E\left[XY\right] - E\left[X\right]E\left[Y\right]$$

Variance and Covariance

- X_i is series of indicator variables
- Random variable $X = \sum_i X_i$
- Variance is

$$\mathrm{Var}[X] = \sum_i \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j]$$

If $Pr[X_i] = p_i$,

$$\mathrm{Var}[X_i] = p_i(1 - p_i) \leq p_i = E[X_i]$$

So

$$\mathrm{Var}[X] \leq E[X] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j]$$

Chebyschev's Inequality

Theorem (Chebyshev)

- *X is a random variable*
- $\mu = E[X]$
- $\sigma = Var[X]$
- $\lambda > 0$

$$Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

This is derived from Markov's inequality, which states that

$$Pr[X \geq \lambda] \leq \frac{E[X]}{\lambda}.$$

Hardy - Ramanujan Theorem

- x is randomly chosen from $\{1, \ldots, n\}$
- X_p is 1 if $p|x$ and 0 otherwise
- $X = X_2 + X_3 + X_5 + \cdots + X_p$ where $p \leq M = n^{1/10}$
- No x may have more than 10 prime factors exceeding M

Hence,

$$\nu(x) - 10 \leq X(x) \leq \nu(x).$$

This roughly translates to asymptotic bounds on the variation of $\nu(x)$ by bounding the variation of $X(x)$:

$$X(x) \leq \nu(x) \leq X(x) + 10.$$

Hardy - Ramanujan Theorem

- Begin by finding $E[X]$
- $E[X_p]$ is

$$\frac{\#\text{ of multiples of } p \text{ less than } n}{n} = \frac{\lfloor n/p \rfloor}{n}$$

- $t - 1 \leq \lfloor t \rfloor \leq t$, so

$$E[X_p] = \frac{1}{p} + O(1/n)$$

- Then the total expectation is

$$E[X] = \sum_{p \leq M} \left(\frac{1}{p} + O(1/n) \right) = \ln \ln n + O(1)$$

Hardy - Ramanujan Theorem

- Now find the variance
- Start with

$$\text{Var}[X] = \sum_{p \leq M} \text{Var}[X_p] + \sum_{p \neq q} \text{Cov}[X_p, X_q]$$

- Since $\text{Var}[X_p] = \frac{1}{p}\left(1 - \frac{1}{p}\right) + O(1/n)$,

$$\sum_{p \leq M} \text{Var}[X_p] = \sum_{p \leq M} \left(\frac{1}{p} - \frac{1}{p^2} + O(1/n)\right) = \ln\ln n + O(1)$$

- This leaves only covariances

Hardy - Ramanujan Theorem

The covariance of X_p and X_q (p and q prime) is

$$\begin{aligned}
\mathrm{Cov}\left[X_p, X_q\right] &= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n} \\
&\leq \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{q} - \frac{1}{n}\right) \\
&\leq \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right)
\end{aligned}$$

Hardy - Ramanujan Theorem

- Then

$$\begin{aligned}\sum_{p\neq q} \operatorname{Cov}[X_p, X_q] &\leq \frac{1}{n}\sum_{p\neq q}\left(\frac{1}{p}+\frac{1}{q}\right)\\ &\leq \frac{2M}{n}\sum_{p}\frac{1}{p}\\ &= O(n^{-9/10}\ln\ln n)\\ &= o(1)\end{aligned}$$

- Likewise $\sum_{p\neq q} \operatorname{Cov}[X_p, X_q] \geq -o(1)$
- $\operatorname{Var}[X] = \ln\ln n + O(1)$.

Hardy - Ramanujan Theorem

- Apply Chebyshev's inequality with $\mu = \ln \ln n + O(1)$, $\sigma = \sqrt{\ln \ln n + O(1)}$:

$$Pr\left[|X - \ln \ln n| \geq \lambda\sqrt{\ln \ln n}\right] \leq \frac{1}{\lambda^2} \qquad (1)$$

- Letting $\lambda = \omega(n)$, the probability goes to 0 as $n \to \infty$ and the theorem is proven. □

Thank You

We would like to thank the following people for making the following reading project possible:

Our mentor, Chiheon Kim, for helping us detangle our very terse and confusing text

Our parents, for taking us out here each week even at the most inconvenient times

Most of this content we learned about was from the book "The Probabilistic Method" by Noga Alon and Joel Spencer.