REU 2006 · Discrete Math · Lecture 4

Instructor: László Babai
Scribe: Travis Schedler
Editor: Sourav Chakraborty
NOT PROOFREAD

June 28, 2006. Last updated June 29, 2006 at 7:30 a.m.

4.1 Sollution to Exercise 3.4.12

Recall the definition $\nu(n) :=$ the number of distinct prime divisors of n ("nu"). Let us pick j at random from $\{1, \ldots, n\} = \Omega$. The exercise was to prove

$$E(\nu(j)) \sim \ln \ln n$$

Let us try to prove it. Let

$$X_p(j) := \text{ indicator of "} p \mid j \text{" } := \begin{cases} 1, & \text{if } p \mid j, \\ 0, & \text{otherwise.} \end{cases}$$

Then note that,

$$\nu(j) = \sum_{p \leq n} X_p(j) \tag{4.1.1}$$

That is, $\nu = \sum_{p \leq n} X_p$. From linearity of expectation we have

$$E(\nu) = \sum_{p \leq n} E(X_p) = \sum_{p \leq n} P_{j \in \Omega}(j \mid p) = \sum_{p \leq n} \frac{\lfloor \frac{n}{p} \rfloor}{n}. \tag{4.1.2}$$

Exercise 4.1.1. Show that the $\sum_{p \leq n} \frac{\lfloor \frac{n}{p} \rfloor}{n}$ is asymptotically equal to $\sum_{p \leq n} \frac{\frac{n}{p}}{n} = \sum_{p \leq n} \frac{1}{p} \sim \ln \ln n$.

4.2 Sollution to Exercise 3.4.6

Recall the 2000 member question: 2000 members of a club all receive cards numbered 1 through 2000; if a person gets a card that has his birth year on it, then he is lucky and get a prize. Call such members "lucky members." Assume that everybody was born between the years 1 and 2000.

The exercise was to prove that the expected number of "lucky members" is always 1, regardless of the distribution of ages.

Sollution: let M be the set of members. For every member $m \in M$, we have $P(m \text{ is lucky}) = \frac{1}{2000}$. Let T_m be the indicator of "m is lucky". Then $X = \sum_{m \in M} T_m$ is the number of lucky members. By linearity of expectation

$$E(X) = \sum_{m \in M} E(T_m) = \sum_{m \in M} P(m \text{ is lucky}) = \sum_{m \in M} \frac{1}{2000} = 1.$$

4.3 Bufforn's Needle Problem

Let us now consider **Briffon's Needle Problem** (1777). Suppose we are dropping a needle of a fixed length onto the ground, which has evenly spaced parallel lines. What is the probability that the needle will touch one of the lines?

We define unit length to be the spacing between the lines. Suppose that the needle has length 1. Then, the probability that the needle hits a line is $\frac{2}{\pi}$.

Proof. Let the indicator variable X be:

$$X := \begin{cases} 1, & \text{if the needle hits a line,} \\ 0, & \text{if the needle does not hit a line.} \end{cases} \tag{4.3.1}$$

The probability that the needle hits a line is $E(X) =: f$. Let us define X_t to be the number of lines hit by a needle of length t. Then

$$E(X_{\frac{1}{2}}) = \frac{E(X_1)}{2} = \frac{f}{2}.$$

This is because whenever we drop a needle of length 1, we can divide it in half, and if the needle of length 1 hit the line, then each half has a 50% chance of having hit the line. So the total number of hits after dropping a needle of length $\frac{1}{2}$ many times would be half the number of hits if one drops a needle of length 1. In general, one obtains by the same argument:

$$E(X_{t+s}) = E(X_t) + E(X_s). \tag{4.3.2}$$

This is even true if the needle is not straight and/or has total length greater than one! However, in this case, the expected value will take into account that sometimes the needle will hit in more than one place (e.g. a bent needle with a very small angle will very likely hit the lines either 0 or 2 times, and not usually 1.)

Now, even if the needle is not a finite number of lines put together, but is rather is a smooth (here, "smooth" = "continuously differentiable") curve. Then, the expected number of hits is still just proportional to the length. This is not too difficult to believe, so let's just accept it.

But now instead of a needle if we use a circle of diameter one, it must hit in **exactly two points** no matter where it hits. So the expected value $E(X_\pi) = 2$, since π is the circumference of this circle. Hence, we would like to say $E(X_1) = \frac{1}{\pi}E(X_\pi) = \frac{2}{\pi}$. To do this, let's look at (4.3.2) again. We may iterate this multiple times to show that, for any rational $\alpha \in \mathbb{Q}$,

$$E(X_{\alpha t}) = \alpha t. \tag{4.3.3}$$

Again, it is plausible that this is a continuous thing, so the same equation must be true for any $\alpha \in \mathbb{R}$. This gives us $E(X_1) = \frac{2}{\pi}$ as desired.

But now, since a straight needle of length $\leq$ one must hit in either one or zero places, the expected number of hits is just the probability of hitting a line. Hence, the desired probability (for a straight needle of unit length) is just $\frac{2}{\pi}$. □

Briffon posed this as an experimental means for computing π: try dropping the needle many many times and see what the probability ends up being: it should converge to $\frac{2}{\pi}$!

4.4 Multinomial Theorem

Recall the binomial theorem,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \tag{4.4.1}$$

We also know the trinomial theorem:

$$(x+y+z)^n = \sum_{i,j,k\geq 0; i+j+k=n} \frac{n!}{i!j!k!} x^i y^j z^k. \tag{4.4.2}$$

This comes from the fact that

$$\frac{n!}{i!j!k!} = \binom{n}{i} \cdot \binom{n-i}{j} \cdot \binom{n-i-j=k}{k}, \tag{4.4.3}$$

which is the number of ways of picking i things from n, then j from the remaining $n-i$, then k from the remaining $k = n-i-j$ things.

Generalizing this, we just get

$$(x_1 + \ldots + x_t)^n = \sum_{k_i \geq 0, \sum k_i = n} \binom{n}{k_1, \ldots, k_t} x_1^{k_1} \cdots x_t^{k_t}, \tag{4.4.4}$$

where

$$\binom{n}{k_1, \ldots, k_t} := \frac{n!}{k_1! \cdots k_t!}. \tag{4.4.5}$$

Question: Estimate the number of distinct values the multinomial coefficients $\binom{n}{*,*,\ldots}$ taken as a function of n alone.

This is the same question as: Take a partition $n = k_1 + \ldots + k_t$, and consider numbers of the form let $\prod_{i=1}^{t}(k_i!)$. Let r_n be the number of distinct values of this. One clearly has

$r(n) \leq p(n)$ ($p(n)$ is the total number of partitions of n), and recall that $p(n) < c^{\sqrt{n}}$ whre $e = e^{\frac{2}{3}p}$.

The goal is to find the order of magnitude of $\log(r(n))$. We have $\ln r(n) < C\sqrt{\pi}$. There are some OPEN questions:

Research Question 4.4.1. Is $\ln r(n) = \Theta(\sqrt{n})$, that is, does there exist $c, C > 0$ such that $c\sqrt{n} \leq \ln r(n) \leq C\sqrt{n}$?

Research Question 4.4.2. Is it true that $\ln r(n) = \Theta\left(\frac{\sqrt{n}}{\sqrt{\ln n}}\right)$?

Claim 4.4.3.

$$\ln r(n) = \Omega\left(\frac{\sqrt{n}}{\sqrt{\ln n}}\right). \tag{4.4.6}$$

That is, there exists $c > 0$ such that $c\frac{\sqrt{n}}{\sqrt{\log n}} < \ln r(n)$.

Definition 4.4.4. $n = k_1 + \ldots + k_n$ is a **prime partition** if each k_i is either 1 or a prime. Let $g(n)$ be the number of prime partitions of n.

Exercise 4.4.5. $g(n) \leq r(n)$.

Exercise* 4.4.6. $\log g(n) = \Theta\left(\sqrt{\frac{n}{\log n}}\right)$. *That is,* $c_1\sqrt{\frac{n}{\log n}} < \log g(n) < c_2\left(\sqrt{\frac{n}{\log n}}\right)$.

4.5 Graph Theory - Cayley's Theorem

For more on Graph Theory refer to László Babai's notes Chapter 6

A graph is a set of vertices (nodes) with an adjacency relation. For now we will only deal with simple graphs, that is, the graphs are undirected, has no self loops or parallel edges. We will use n to denote the number of vertices and m to denote the number of edges.

K_n is the complete graph on n vertices. So that $m = \binom{n}{2}$.

Now, a **walk of length** k is a collection $v_0, v_1, \ldots, v_k \in V$ such that $v_{i-1} \sim v_i$ (i.e. these two vertices are **adjacent**)

A textbfpath is a walk without repeated vertices. For paths, we consider it the same path if we go backwards (but not for walks). So to be precise, a path is an equivalence class of two walks (going forwards or backwards in the walk that has no repeated vertices).

A **closed walk of length** k is a walk such that $v_0 = v_k$.

A **cycle** is a closed walk without repeated vertices, $k \geq 3$: for example, $1-2-3-1$ is the same cycle as $2-3-1-2$ and $3-1-2-3$; also the same as $2-1-3-2\ldots$.

Now, the number of n-cycles in K_n is $\frac{(n-1)!}{2}$. The number of paths of length $n-1$ in K_n is $\frac{n!}{2}$.

Definition 4.5.1. A graph G is connected if $(\forall x, y \in V)(\exists x \ldots y \text{ path})$.

Definition 4.5.2. G is a tree if G is connected and has no cycle.

Exercise 4.5.3. *If G is connected then G has a* ***spanning tree****: that is, a subgraph that reaches all vertices and is a tree.*

Theorem 4.5.4. *(Cayley's Formula): The number of spanning trees of K_n is n^{n-2}.*

Exercise 4.5.5. Every tree has $n-1$ edges.

Exercise 4.5.6. Lemma: if $n \geq 2$ then every tree has a vertex of degree 1.

The above exercise uses:

Definition 4.5.7. deg(v) is the number of neighbors (adjacent vertices) of v.

Exercise 4.5.6 is only true for finite graphs: if we had the infinite line, then every vertex has degree 2.

Theorem 4.5.8. *"Handshake Theorem":*

$$m = \frac{1}{2}\left(\sum_{v \in V} deg(v)\right). \tag{4.5.1}$$

Proof. We think of the vertices as people and draw an edge between two people if they shook hands. The number of edges is the number of handshakes. On the other hand, the sum of the degrees of all vertices just adds the number of times each person shook someone else's hand. It's clear that the latter is twice the number of handshakes because each handshake involves exactly two people. So dividing by two the sum of the degrees, we must get the total number of handshakes. □

Exercise 4.5.9. Suppose $d_1, \ldots, d_n \geq 1$ are such that $\sum_{i=1}^{n} d_i = 2n-2$. Prove that there exists a tree with n vertices having these degrees.

Exercise 4.5.10. Prove that the number of such trees is $\dfrac{(n-2)!}{\prod_{k=1}^{n}((d_i-1)!)}$.

Exercise 4.5.11. Use this to prove Cayley's theorem in a single line.

Another proof involves an explicit bijection with a set that has size n^{n-2}. Note that n^{n-2} is size of the set of strings of length $n-2$ whose letters are from an alphabet of size n letters. There is a way to encode all trees uniquely in this way, invented by Prüfer, and called the **Prüfer Code**.

Take a spanning tree of K_n. Let's say that the vertices of K_n are numbered $\{1, 2, \ldots, n\}$, so our spanning tree also has vertices labeled in the same way.. Look at the lowest-numbered vertex of degree 1. Cut this off and write down the neighbor. Then iterate this. We'll get a sequence of length $n-1$, but the last letter will always be n (the largest letter will never get cut). So we can ignore the last letter and get a string of length $n-2$. We claim now that this string of length $n-2$ uniquely encodes the subtree of K_n.

We can determine which vertices have degree one to begin with: these are the vertices that don't appear in the code. We can pick the smallest and we know right away which vertex it's connected to: the first number listed. Then we can delete the first number and iterate this: the next smallest number, which does not appear on this truncated list; etc.

Exercise 4.5.12. Work this out precisely, and use it to write a proof of Cayley's theorem.