Lecture 24: Counting (contd)

Combinations and the Selection Principle

Before we formally introduce combinations, here is a slight variation of a problem that we saw earlier:

> From Opening Day to September 1, every MLB team can carry up to 25 players on their roster. How many different ways are there to choose a 9-player team?

Recall that the number of ways to pick a *specific starting order* is $P(25, 9)$. However, the above problem is different from the one in the previous lecture since we don't have to specify a starting *order* – only a starting 9. So how do we estimate this new quantity?

One way to solve it is using the division rule. You can think of picking a 9-player starting order as a 2 step process: (1) pick a *set* of 9 starters; (2) line up the 9 starters in order.

Once you pick a set of 9 starters, there are *precisely* $P(9, 9) = 9!$ ways to choose a particular order with these players; this is simply an application of the Arrangement Principle with $n = 9, r = 9$. Therefore, if $k = 9!$, then one can imagine a k-to-1 onto mapping from the set of starting orders to the set of 9 starters (without order specified.)

Therefore, by the Division rule, we get that the number of starting orders equals the number of ways to pick a starting $9 \times k$, where $k = 9!$. In other words, our desired answer = $P(25, 9)/9! = 25!/((25 - 9)! \cdot 9!) = 25!/(16!\,9!)$.

In fact, for picking r items out of a set of size n, we will denote this quantity two special symbols: $C(n, r)$ or $\binom{n}{r}$. The following counting rule formalizes this intuition.

> *Selection Principle*: The number of ways to choose a subset of r elements from a set of n elements is
> $$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

The quantity $\binom{n}{r}$ is read as "n choose r", and is called a *Binomial coefficient*.

Though we have motivated the binomial coefficient in our discussions on combinatorics, know that there are lots of interesting connections between binomial coefficents and *algebraic* identities involving polynomials of a single variable. More on this later on.

Counting with repetitions

Counting rules are by-and-large simple applications of logic. But the most common pitfalls occur when we have a counting problem where some of the objects under consideration are repeated, i.e., are *indistinguishable* from the others. In such cases we need to be a bit careful when applying the counting rules.

Consider the following problem:

> Suppose you are planning a 20-mile walk, which should include 5 northward miles, 5 eastward miles, 5 southward miles, and 5 westward miles. How many different walks are possible?

The first step in solving such problems is to translate into a setting more suitable for enumeration. Strings of characters are often used for this purposes. For example, for the above problem let us represent each walk as a string of length-20 consisting of 5 occurrences of “N”, “E”, “S”, and “W” each. We need to count the number of such possible strings.

We can now solve this counting problem as follows. Imagine 20 empty slots for the characters and filling in these slots with 5 “N”s, 5 “E”s, 5 “W”s, and 5 “S”. First, we can place the locations of the “N”s in

$$\binom{20}{5} = \frac{20!}{15! \cdot 5!}$$

ways. Once we do that, there are 15 empty slots remaining. We can now place the “E”s in

$$\binom{15}{5} = \frac{15!}{10! \cdot 5!}$$

ways. There are 10 empty slots remaining. We can choose the “W”s in

$$\binom{10}{5} = \frac{10!}{5! \cdot 5!}$$

ways and the remaining necessarily have to be “S”s. Therefore, the total number is given by:

$$\begin{aligned} \binom{20}{5}\binom{15}{5}\binom{10}{5} &= \frac{20!}{15! \cdot 5!}\frac{15!}{10! \cdot 5!}\frac{10!}{5! \cdot 5!} \\ &= \frac{20!}{5! \cdot 5! \cdot 5! \cdot 5!}. \end{aligned}$$

More generally, if we want to permute n objects of which k_1 objects are of type 1, k_2 objects are of type 2, and so on – the

$$\frac{n!}{k_1! \cdot k_2! \ldots k_n!}$$

A proof of this result is identical to that in the above example with the 20-mile walk. We leave it as an **exercise**.

One more example.

> How many different ways are there to select 10 donuts if there are 4 varieties to choose from?

One can solve it using a variant of the above method, but the catch here is that before you knew exactly how many occurrences of each type were present in your selection, while here it can be anywhere between 0 and 10.

A generalization of this example would be:

> We have r varieties of donuts and need to choose a set of n donuts in total. In how many ways can we do this?

Let's focus on the 10-donut-4-variety case first. A fairly ingenious way of solving this problem is as follows. Suppose we choose 3 chocolate donuts, 4 Boston kremes, 2 lemon donuts, and 1 plain donut. We can imagine writing down our choice of donuts as a *binary string* as follows:

0001000010010

In the above string, 0's denote donuts, and 1's denote dividers between different types of donuts. Note that we have exactly 10 0's (10 types) and 3 1's (4-1=3 dividers between different types.) On the other hand, if all 10 of our selected donuts were Boston kremes, then our choice would be written as

1000000000011

Again, a length-13 string with 10 0's and 3 1's! Convince yourself there is a *one-to-one* bijection between the set of possible choices of 10 donuts and 4 varieties, and the set of binary strings of length-13 with 3 1's.

Therefore, by the Selection Principle, the answer is immediate; the number of donut selections is given by:

$$\binom{13}{3}.$$

For general problems for selecting n objects from r varieties, we will have a binary string with n 0's and $r-1$ 1's. The answer, using analogous reasoning as above, is given by:

$$\binom{n+r-1}{r-1} = \binom{n+r-1}{n}.$$

Pigeon Hole Principle

We now discuss a somewhat surprising rule in counting, one that is fairly obvious. Here is an example problem:

> A drawer in a dark room contains red socks, green socks, and blue socks. How many socks must you withdraw to be sure that you have a matching pair?

Thinking about this a bit, we realize that the answer is 4. If we pick 3 socks or less, there is a chance that there is no matching pair. However, if we draw 4 socks, then we are sure to get at least one matching pair.

This kind of reasoning is called the *Pigeon Hole Principle*, which is the following aphorism:

> If there are more pigeons than holes, then at least two pigeons must be in the same hole.

What's our colorful sock collection got to do with pigeons and holes? Imagine the socks that we pick out represent "pigeons", and the colors represent "holes". Then, if we pick 4 socks (pigeons) then by the Pigeon Hole Principle at least two of them are of the same hole, i.e., the same color.

Here is a more generic application of the Pigeon Hole principle. We claim that:

> There are 86 students enrolled in CPRE 310. Prove that some group of at least 9 students will get the same letter grade in the end.

How do we arrive at this? Well, here the “holes” are the letter grades: (A,A-,B+,B,B-,C+,C,C-,D,F), and the “pigeons” are CPRE students. So if there are 10 holes it has to be that some hole gets more than 11 pigeons; if not, the number of pigeons is at most 110 (which is not true in this case).

This idea is nothing but the *Generalized* Pigeon Hole Principle, stated in set-theoretic language as follows:

> If $f : A \to B$ and $|A| > k|B|$, then f maps some $k + 1$ elements of A to the same element in B.

One caveat of the Pigeon Hole Principle is that it is non-constructive; the Principle does not give you any indication of *which* pigeons get mapped to which holes – only that there exist some pigeons that get mapped to the same hole. This can have important practical implications. For example, consider the problem:

> Suppose we generate a list of 40 10-digit numbers at random. Prove that there exist some two distinct subsets of these numbers that add up to the same value.

This is called the *Subset-Sum* problem, and has several practical applications (including in privacy and cryptography). It is hard to explicitly *find* two different subsets that add up to the same value, but not hard to argue that two such subsets *exist*.

The solution is surprisingly simple. Let A be the set of all subsets (i.e., the power set) of these 40 numbers. The number of subsets (pigeons) of the list of numbers is given by:

$$|A| = 2^{40} > 10^{12}.$$

On the other hand, let B be the set of sums that you can generate from adding up subsets. We don’t know the precise set, but we can safely say that $|B| < 40 \times 10^{10}$ since each 10-digit number is less than 10^{10} and we can add up at most 40 of them. Therefore, $|B| < 4 \times 10^{11}$.

Since $|A| > |B|$, by the Pigeon Hole Principle, some two pigeons (subsets) get mapped to the same hole (subset sums). Done!