CS 70 Spring 2007 — Discrete Mathematics for CS — Luca Trevisan — Lecture 15

Counting and Probability

The topic for the next major portion of the course is Discrete Probability. Suppose you toss a fair coin a thousand times. How likely is it that you get exactly 500 heads? And what about 1000 heads? It turns out that the chances of 500 heads are roughly 5%, whereas the chances of 1000 heads are so infinitesimally small that we may as well say that it is impossible. But before you can learn to compute or estimate odds or probabilities you must learn to count! That is the subject of this lecture.

We will learn how to count the number of outcomes while tossing coins, rolling dice and dealing cards. Many of the questions we will be interested in can be cast in the following simple framework called the occupancy model:

Balls in Bins: We have a set of k balls. We wish to place them into n bins. How many different possible outcomes are there?

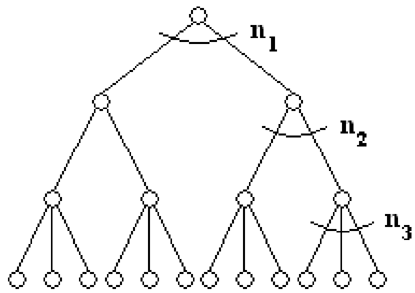
How do we represent coin tossing and card dealing in this framework? Consider the case of $n = 2$ bins labelled H and T, corresponding to the two outcomes of a coin toss. The placement of the k balls correspond to the outcomes of k successive coin tosses. To model card dealing, consider the situation with 52 bins corresponding to a deck of cards. Here the balls correspond to successive cards in a deal.

The two examples illustrate two different constraints on ball placements. In the coin tossing case, different balls can be placed in the same bin. This is called *sampling with replacement*. In the cards case, no bin can contain more than one ball (i.e., the same card cannot be dealt twice). This is called *sampling without replacement*. As an exercise, what are n and k for rolling dice? Is it sampling with or without replacement?

We are interested in counting the number of ways of placing k balls in n bins in each of these scenarios. This is easy to do by applying the first rule of counting:

First Rule of Counting: If an object can me made by a succession of choices, where there are n_1 ways of making the first choice, and *for every* way of making the first choice there are n_2 ways of making the second choice, and *for every* way of making the first and second choice there are n_3 ways of making the third choice, and so on up to the n_k-th choice, then the total number of distinct objects that can be made in this way is the product $n_1 \cdot n_2 \cdot n_3 \cdots n_k$.

Here is another way of picturing this rule: consider a tree with branching factor n_1 at the root, n_2 at every node at the second level, ..., n_k at every node at the k-th level. Then the number of leaves in the tree is the product $n_1 \cdot n_2 \cdot n_3 \cdots n_k$. For example, if $n_1 = 2$, $n_2 = 2$, and $n_3 = 3$, then there are 12 leaves (i.e., outcomes):



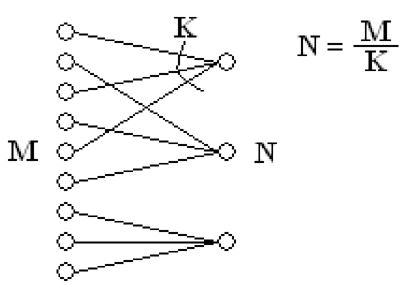
Let us apply this counting rule to figuring out the number of ways of placing k balls in n bins with replacement. This is easy; it is just n^k: n choices for the first ball, n for the second, ...

The rule is more interesting in the case of sampling without replacement. Now there are n ways of placing the first ball, and *no matter* where it is placed there are exactly $n-1$ bins in which the second ball may be placed (exactly which $n-1$ depends upon which bin the first ball was placed in), and so on. So as long as $k \leq n$, the number of placements is $n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$. By convention we assume that $0! = 1$.

Counting Unordered Sets

While dealing a hand of cards, say a poker hand, it is more natural to count the number of distinct hands (i.e. the set of 5 cards dealt in the hand), rather than the order in which they were dealt. To count this number we use the second rule of counting:

Second Rule of Counting: If an object is made by a succession of choices, and the order in which the choices is made does not matter, count the number of ordered objects (pretending that the order matters), and divide by the following number — the number of ordered objects per unordered object. Note that this rule can only be applied if the number of ordered objects is the same for every unordered object.

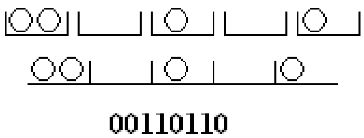


Let us continue with our example of a poker hand. We wish to calculate the number of ways of choosing 5 cards out of a deck of 52 cards. So we first count the number of ways of dealing a 5 card hand pretending that we care which order the cards are dealt in. This is exactly $\frac{52!}{47!}$ as we computed above. Now we ask for a given poker hand how many ways could it have been dealt? The 5 cards in the given hand could have been dealt in any one of 5! ways. Therefore by the second rule of counting, the number of poker hands is $\frac{52!}{47!5!}$.

This quantity $\frac{n!}{(n-k)!k!}$ is used so often that there is special notation for it: $\binom{n}{k}$, pronounced *n choose k*. This is the number of ways of placing k balls in n bins (without replacement), where the order of placement does not matter.

What about the case of sampling with replacement? How many ways are there of placing k balls in n bins with replacement when the order does not matter? Let us try to use the second rule of counting. There are

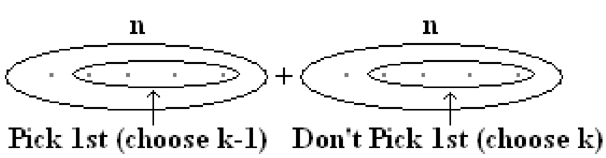
n^k ordered placements. How many ordered placements are there per unordered placement? Consider the case $k = 2$. If the two balls are in distinct bins, then there are n ways to place the first ball, and $n-1$ ways to place the second ball, giving us $n(n-1)$ ways where order matters. Now, by the second rule of counting, we divide by 2! and get $\frac{n(n-1)}{2}$ ways to place two balls in distinct bins. The number of ways for the two balls to be placed into the same bin is exactly n. Thus, there are $\frac{n(n-1)}{2} + n$ ways to place two balls into n bins where order does not matter. For larger values of k, it seems hopelessly complicated. Yet there is a remarkably slick way of calculating this number. Represent each of the balls by a 0 and the separations between boxes by 1's. So we have k 0's and $(n-1)$ 1's. Each placement of the k balls in the n boxes corresponds uniquely to a binary string with k 0's and $(n-1)$ 1's. Here is a sample placement of $k = 4$ balls into $n = 5$ bins and how it can be represented as a binary string:



But the number of such binary strings is easy to count: we have $n+k-1$ positions, and we must choose which k of them contain 0's. So the answer is $\binom{n+k-1}{k}$.

Combinatorial Proofs

Combinatorial arguments are interesting because they rely on intuitive counting arguments rather than algebraic manipulation. For example, it is true that $\binom{n}{k} = \binom{n}{n-k}$. Though you may be able to prove this fact rigorously by definition of $\binom{n}{k}$ and algebraic manipulation, some proofs are actually much more tedious and difficult. Instead, we will try to discuss what each term means, and then see why the two sides are equal. When we write $\binom{n}{k}$, we are really counting how many ways we can choose k objects from n objects. But each time we choose any k objects , we must also leave behind $n-k$ objects, which is the same as choosing $n-k$ (to leave behind). Thus, $\binom{n}{k} = \binom{n}{n-k}$. Some facts are less trivial. For example, it is true that $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$. The two terms on the right hand side are splitting up choosing k from n objects into two cases: we either choose the first element, or we do not. To count the number of ways where we choose the first element, we have $k-1$ objects left to choose, and only $n-1$ objects to choose from, and hence $\binom{n-1}{k-1}$. For the number of ways where we don't choose the first element, we have to pick k objects from $n-1$ this time.



We can also prove even more complex facts, such as $\binom{n}{k+1} = \binom{n-1}{k} + \binom{n-2}{k} + \cdots + \binom{k}{k}$. What does the right hand side really say? It is splitting up the process into cases of which object we select first. In other words:

$$\textbf{First element selected is either} \begin{cases} \text{element } 1, & \binom{n-1}{k} \\ \text{element } 2, & \binom{n-2}{k} \\ \text{element } 3, & \binom{n-3}{k} \\ \quad\vdots & \\ \text{element}(n-k), & \binom{k}{k} \end{cases}$$

The last combinatorial proof we will do is the following: $\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}=2^n$. To see this, imagine that we have a set S with n elements. In the left hand side, the i^{th} term is counting the number of ways of choosing a subset of size i, while the right hand side is counting how many ways we can either select each element or not. You may have already figured out that the number of binary strings of length n = $\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}$. Let us look at an example, where $S=\{1,2,3\}$ (so $n=3$). Now, enumerate all possible subsets of S: $\{\{\},\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$. The term $\binom{3}{0}$ is counting how many ways we can have subsets of S with 0 elements. Indeed, there is only one such subset: the empty set. There are 3 ways of choosing subsets with 1 element (i.e., $\binom{3}{1}$), 3 ways of choosing subsets with 2 elements or $\binom{3}{2}$, and 1 way of choosing subsets with 3 elements (i.e., $\binom{3}{3}$. This is the case when the entire set S is considered as a subset). Thus, summing them all up, we get $8=2^3$. The right hand side basically treats each subset as a binary string of length n. A one in the i^{th} position indicates that the i^{th} element of S is in the subset, and a zero indicates that it is not. So, in our example, the subset of S $\{1,2\}$ can be represented by the binary string 110_2.