MthStat 465, Spring 2005, Lecture Number 7
Combinatorics

To compute probablities in classical models one must be able to count the number of elements in a set. The most important tool is the **multiplication principle**: the number of ordered pairs (a, b) that can be formed if there are A choices for a and B choices for b is AB.

For example, if we have 3 choices of pants and 4 choices of shirts we have 12 choices of pants and shirts.

This principle can be applied recursively to count n-tuples. For example, the ordered triple (a, b, c) can be viewed as an ordered pair $((a, b), c)$.

If 18 people are asked the last two digits of their social security numbers, then there are 100^{18} sequences of responses. If we were to view the responses as an experiment, the sample space S could be modeled as finite sequences of length 18, which each sequence element is a number from 0 to 99. Thus the sample space has 100^{18} elements. The classical model of probability would say that the sigma algebra was the set of all subsets of of S, and that each outcome has a probability of 100^{-18}.

A natural question (with a surprising answer!) is what is the probability in this model that no two people have the same last two digits in their social security number. In our model, this is the event consisting of 18-tuples with no two elements the same. Employing the multiplication principle, there are

$$
\begin{aligned}
& 100 \times 99 \times 98 \times 97 \times 96 \times 95 \times 94 \times 93 \times 92 \times 91 \times 90 \\
& \times 89 \times 88 \times 87 \times 86 \times 85 \times 84 \times 83 \\
= & \frac{100!}{82!}
\end{aligned}
$$

outcomes!!! So the probability of no matches is $100^{-18}100!/82! \approx .1963256577$. In other words, the probability is more than 0.80 that in a group of 18 people, at least two will have the same last two digits of their social security numbers, assuming that our model is correct.

We have seen now how to count sequences in which elements can be repeated or not. However, sometimes what we want to count is subsets, not sequences. This is the case, for example, in bridge hands, poker hands, and most state lotteries that involve choosing a set of winning numbers. Subsets can be counted by considering sequences. For example: let ${}_{15}C_5$ stand for the number of sets of 5 balls that can be picked from a set of 15 balls, each ball bearing exactly one number from 1 to 15. If the balls were to picked in order, there would be two ways to do it. One way is to pick them out one at a time, recording which ball was picked when. There are $15 \times 14 \times 13 \times 12 \times 11 = 15!/10!$ ways to do this. One the other hand, we could first scoop out 5 balls, which can be done in ${}_{15}C_5$ ways, and the put each set of five balls in order, which can be done 5! ways. Using the multiplication principle, this gives us ${}_{15}C_5 \times 5!$ orderings. So we must have

$$
\begin{aligned}
\frac{15!}{10!} &= {}_{15}C_5 \times 5! \\
\frac{15!}{10!5!} &= {}_{15}C_5
\end{aligned}
$$

This same line of reasoning shows that if N and K are integers and $0 < K < N$ and ${}_N C_K$ stands for the number of subsets of size K in a set of size N, then

$$\begin{aligned}\frac{N!}{(N-K)!} &= {}_N C_K \times K! \\ \frac{N!}{(N-K)!K!} &= {}_N C_K\end{aligned}$$

By considering the 1-1 correspondence between sets and their complements, we see that ${}_N C_K = {}_N C_{N-K}$, even when $K = N$ or $K = 0$, and that ${}_N C_N = {}_N C_0 = 1$. This leads us to define $0! := 1$.

The symbol ${}_N C_K$ is read "N choose K" and is sometimes called a binomial coefficient because

$$(x+y)^N = {}_N C_0 x^0 y^N + {}_N C_1 x^1 y^{N-1} + \cdots + {}_N C_K x^K y^{N-K} + \cdots + {}_N C_N x^N y^0.$$

The best way to see this is to think about the distributive law telling you to pick either x or y out of each of N factors of the form $(x+y)$.

Frequently you will see

$$\left(\begin{array}{c} N \\ K \end{array} \right)$$

instead of ${}_N C_K$.