Lecture 3: Counting, Shared Birthdays

January 29, 2019

Basic counting principles:

- $$|E_1 \cup E_2| = |E_1| + |E_2|$$
  provided $E_1 \cap E_2 = \emptyset$.

- For sets that are not necessarily disjoint:
  $$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2| \leq |E_1| + |E_2|$$

- More generally,
  $$|\bigcup_{j=1}^{n} E_j| \leq \sum_{j=1}^{n} |E_j|$$
  with equality if the E_j are pairwise disjoint (any two have an empty intersection).

- $|E_1 \times E_2| = |E_1||E_2|$.

- A more general multiplication principle: Sometimes we view a set as the results of a sequence of r choices, with each different sequence giving rise to a different element of the set. Say there are k_1 possibilities for the first choice, for each first choice there are k_2 possibilities for the second choice, for each sequence of the first two choices, k_3 possibilities for the third choice, etc. Then there are $k_1 \cdots k_r$ ways to make the choices.

  This is what our tree diagrams are: There is a 1-1 correspondence between the sequences of choices and the leaves of a tree of depth r, where every node at depth $d < r$ has k_{d+1} children. (The root node has depth 0, and the leaves all have depth r.)

Permutations. How many sequences of k elements of $\{1, \ldots, n\}$ have all k elements distinct? For example, with $n = 5$, $k = 3$, then $(4, 5, 1)$ is such a sequence, but $(1, 5, 1)$ is not. We have n choices for the first component of the sequence, and for each such choice, $n - 1$ choices for the second, etc. (Note that we must have $k \leq n$.) So by the above principle, the number of such sequences is

$$n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}.$$

This is the number of *k-permutations* of an n-element set.

If $n = k$, then $(n-k)! = 0! = 1$, so the number of n-permutations of $\{1, \ldots, n\}$ is $n!$. In this case, we just call them *permutations* of $\{1, \ldots, n\}$. If $n < k$, then the formula does not make sense; clearly in this case there are no k-permutations of $\{1, \ldots, n\}$.

Example. As we saw earlier, the number of sequences of 2 distinct cards drawn from a deck of cards (sampling without replacement) is

$$52 \times 51 = \frac{52!}{50!}.$$

Example. Birthday problem. What is the probability that with k people in a room, all k have different birthdays? Model sample space S as set of all sequences of k elements of $\{1, 2, 3, .., 365\}$, with uniform distribution. In other words we are thinking of the experiment as choosing k people in succession and noting the sequence of birthdays. The total number of people is so large that we can view it as a question of sampling with replacement. The event E is then the set of all k-permutations of $\{1, 2, \ldots, 365\}$. With $k = 30$, for example,

$$|S| = 365^{30}, |E| = 365 \times \cdots \times 336,$$

so

$$P(E) = \frac{|E|}{|S|} = \frac{365}{365} \cdot \frac{364}{365} \cdots \frac{336}{365}.$$

The probability that there are two people with the same birthday is then the probability of the complementary event,

$$1 - P(E).$$

This is about 0.71. (In the in-class demo we had 34 people, for which the probability of a coincidental birthday should be about 0.8. As it turns out, we did not get a birthday in common, a 1-in-5 occurrence.)

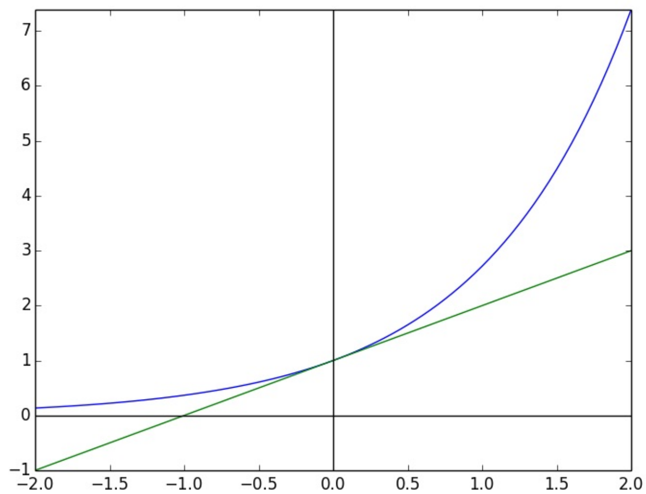

Figure 1: $y = 1 + x$ is the tangent line to $y = e^x$ at $x = 0$, making $1 + x$ a very close approximation to e^x.

Exponential Approximation.

The tangent line to the graph of $y = e^x$ at $x = 0$ is $y = 1 + x$. (See Figure 1.) This means that for small values of x, either negative or positive, $1 + x$ is a very good approximation to e^x, *and vice-versa*. In fact, for $-1 < x < 0$, the error is less than $\frac{x^2}{2,}$ so for example, if $x \approx 0.1$, then the error is less than 0.005. The tangent line lies completely below the graph of $y = e^x$, so that $1 + x < e^x$ for all $x \neq 0$.

We can apply this to get an easy-to-compute approximate solution to the birthday problem. More importantly, we can apply this to get such a solution to problems that have the same structure, like finding the probability of a collision when we hash k items into a hash table of size n, but where the numbers are too large to make an exact computation impractical.

We saw above that the probability that no two people in a group of k people share a birthday is

$$\frac{365 \cdot 364 \cdots (365 - k + 1)}{365^k}.$$

We can rewrite this, and apply the approximation, to get

$$\begin{aligned}
\frac{364}{365}\cdot\frac{363}{365}\cdots\frac{365-(k-1)}{365} &= \big(1-\frac{1}{365}\big)\cdot\big(1-\frac{2}{365}\big)\cdots\big(1-\frac{k-1}{365}\big) \\
&< e^{\frac{-1}{365}}\cdots e^{-(k-1)/365} \\
&= e^{-(1+2+\cdots+(k-1))/365} \\
&= e^{-k(k-1)/(2\cdot 365)} \\
&< e^{-(k-1)^2/730}.
\end{aligned}$$

Observe that as the number k of people gets larger, this probability gets smaller. If we want to figure out at what point it falls below, say, 0.1, we solve

$$e^{-(k-1)^2/730} = 0.1.$$

We take reciprocals:

$$e^{(k-1)^2/730} = 10,$$

then take natural logs:

$$\frac{(k-1)^2}{730} = \ln 10,$$

so

$$k = 1 + \sqrt{730 \cdot \ln 10} \approx 41.$$

Thus if 41 people are present, the probability of a coincidental birthday is *at least* 0.9. The calculation was made just using the fact that $1 + x < e^{-x}$, so it just gives a lower bound for the probability of a coincidental birthday. But in fact, this is very close to the exact probability: With 41 people present, the probability of a coincidence is 0.903.

How big is $n!$?

The factorial function grows *really fast*: It is obvious that $n! < n^n$ for all positive integers n, and it is almost obvious that $n! > a^n$ for all fixed $a > 1$, as long as n is sufficiently large—*how* large depends on a. Thus $n!$ grows more rapidly than any exponential function, although not as rapidly as n^n.

There is a rather amazing formula—called *Stirling's formula*–that in a sense captures exactly how fast $n!$ grows. It is

$$n! \sim \frac{n^n}{e^n}\cdot\sqrt{2\pi n},$$

in the following sense:

$$\lim_{n\to\infty} \frac{n!}{\frac{n^n}{e^n} \cdot \sqrt{2\pi n}} = 1.$$

(You might wonder where the e, and, especially, where the π come from!)

For example, the number of ways to arrange a deck of cards is, to five significant digits,

$$52! = 8.0658 \times 10^{67}.$$

Stirling's formula with $n = 52$ gives 8.0529×10^{67}.