The Birthday Problem

If there are P people in a classroom, what is the probability that two of them have the same birthday?

We'll make the simplifying assumption that all birthdays are equally likely, and we'll forget about leap years. These shouldn't change the answer much, and it will make the problem much easier to model.

Suppose that we make a list of all P birthdays (say by alphabetical order of last names of the people who have them). Then the number of possibilities for the list $(B_1, \ldots, B_P)$ is 365^P. Since all of them are equally likely, we can calculate the probability that two of them share a birthday by counting the number of possible birthday lists with a repeat, then dividing by 365^P.

It turns out to be easier to count the number of birthday lists that <u>don't</u> have a repeat. To make such a list, there are 365 possibilities for B_1. Once B_1 is known, there are 364 possibilities for B_2. Once B_2 is known, there are 363 possibilities for B_3. In general, there are $365 - (K+1)$ possibilities for B_K. So the number of lists is

$$365 \cdot 364 \cdot 363 \cdots (365 - (K+1)) = \frac{365!}{(365-P)!}.$$

Therefore, the probability that all birthdays are different is

$$\frac{365!}{(365-P)!365^P}$$

and the probability that some birthday is shared by at least two people is

$$1 - \frac{365!}{(365-P)!365^P}.$$

Interestingly enough, setting $P = 23$ makes this probability close to one-half (intuition would probably tell you that the break-even point is a much higher P).

More generally, if P people independently choose a random thing from a set of size N, then the probability that some thing gets chosen twice is

$$1 - \frac{N!}{(N-P)!N^P}.$$

What about the probability that <u>three</u> people share a birthday? This is a harder calculation, but we can do it. Again, what I'll count is the number of ways this <u>doesn't</u> happen — i.e., the number of birthday lists in which no birthday appears more than twice.

I'm going to break the problem into cases depending on how many pairs of people share a birthday. Call this number K; it can be as small as 0 and as great as $\lfloor P/2 \rfloor$. (This notation means the integer you get by rounding down; for example, $\lfloor 7/2 \rfloor = 3$. Then the number of people who do not share a birthday with anyone else will be $P - 2K$.

First, pick a set of birthdays that will each belong to exactly two people. We can do this in $\binom{N}{K}$ ways. Write the birthdays in chronological order as $B_1, \ldots, B_K$.

Second, pick a set of birthdays that will each belong to exactly one person. We need $P - 2K$ more birthdays and have used up K of them, so we can do all this in $\binom{N-K}{P-2K}$ ways. Write these birthdays in chronological order as $C_1, \ldots, C_{P-2K}$.

Now:

- Choose two people to assign to B_1. Number of possibilities: $\binom{P}{2}$.
- Choose two people to assign to B_2. Number of possibilities: $\binom{P-2}{2}$.
- ...
- Choose two people to assign to B_K. Number of possibilities: $\binom{P-2K+2}{2}$.
- Choose one person to assign to C_1. Number of possibilities: $P - 2K$.
- Choose two people to assign to C_2. Number of possibilities: $P - 2K - 1$.

- ...
- Choose two people to assign to C_{P-2K}. Number of possibilities: 1.

Putting this all together, the number of possibilities is

$$\binom{N}{K}\binom{N-K}{P-2K}\left[\binom{P}{2}\binom{P-2}{2}\cdots\binom{P-2K+2}{2}(P-2K)(P-2K-1)\cdots 1\right].$$

This expression can be simplified. We need the general fact that $\binom{m}{2} = m(m-1)/2$ (write it as $m!/2!/(m-2)!$ and observe that most of it cancels). That means that the previous expression is

$$\binom{N}{K}\binom{N-K}{P-2K}\left[\frac{P(P-1)}{2}\frac{(P-2)(P-3)}{2}\cdots\frac{(P-2K+2)(P-2K+1)}{2}(P-2K)(P-2K-1)\cdots 1\right]$$

$$= \binom{N}{K}\binom{N-K}{P-2K}\frac{P(P-1)\cdots(P-2K+1)(P-2K)(P-2K-1)\cdots 1}{2^K}$$

$$= \binom{N}{K}\binom{N-K}{P-2K}\frac{P!}{2^K}.$$

But remember that K can be anything from 0 to $\lfloor P/2 \rfloor$. We have to sum up all these possibilities. Doing so and dividing by N^P (the total number of birthday lists) gives the answer:

Probability that no birthday occurs as many as three times

$$= \sum_{K=0}^{\lfloor P/2 \rfloor}\binom{N}{K}\binom{N-K}{P-2K}\frac{P!}{2^K \cdot N^P}.$$

For $N = 365$ and $P = 28$, this comes out to about 97.68%. So the fact that it occurred in Math 410 is quite unusual! The chance of a tripleton is very close to even (50.05%) when $P = 87$.