PHYS 226N/231N: The Birthday Paradox

Prof. Satogata, ODU, Fall 2012
Tue Aug 28, 2012

I'd asked you to try to find out whether you share a birthday with anyone else in the class during a Ponderable on Monday. You probably observed that this felt impossible to answer since you only had a few minutes and there are a lot of people in the class. Why it's challenging is also why it's very likely that many pairs of people in our class share the same birthday.

We had about 80 people in the class on Monday. This is much less than 365, so the "intuitive" guess is that the chance of two people sharing the same birthday is about 25% or one in four. The truth is that it's well over a 99.99% chance.

Why is that? Pick any two people. The first one has a birthday on a given day. The next has a 1/366 chance (including leap years) of having a birthday on the same day as the first, or a 365/366 chance of not sharing a birthday. Those aren't very good odds of them having the same birthday.

Now ask another person. They have a 2/366 chance of sharing a birthday with either of the first two, or a 364/366 chance of not. So the chances of finding a birthday match in this group of three people are

$$\frac{1}{366} + \left(\frac{365}{366}\right)\left(\frac{2}{366}\right) = 0.00818 \tag{1}$$

or about 1/122. We've nearly tripled our chances by adding one person, because they could share a birthday with *either* of the two previous people. We're making three comparisons of birthdays rather than one.

As we add more and more people, the number of comparisons skyrockets. (In fact, it goes like "n choose 2" or $\left(\begin{array}{c} n \\ 2 \end{array}\right) = \frac{n!}{2(n-2)!} = n(n-1)/2$ in combinatorial terms.) Every comparison is a separate observation, a separate chance for two people to share the same birthday. So a rough estimate of when we would have 50-50 odds is when this number of comparisons is the same as the number of days in the year:

$$\frac{n(n-1)}{2} = 366 \;\;\Rightarrow\;\; n \approx 27$$

In reality, we can keep going with the logic in Equation (1) above to find the odds for each n, and draw a plot. This is shown in Figure 1 on the next page, and indicates that the true 50% point is about 23 people. This is known as the **birthday problem** and has application ranging from winning bar bets to breaking cryptographic hash functions.

See http://en.wikipedia.org/wiki/Birthday_problem for more info.

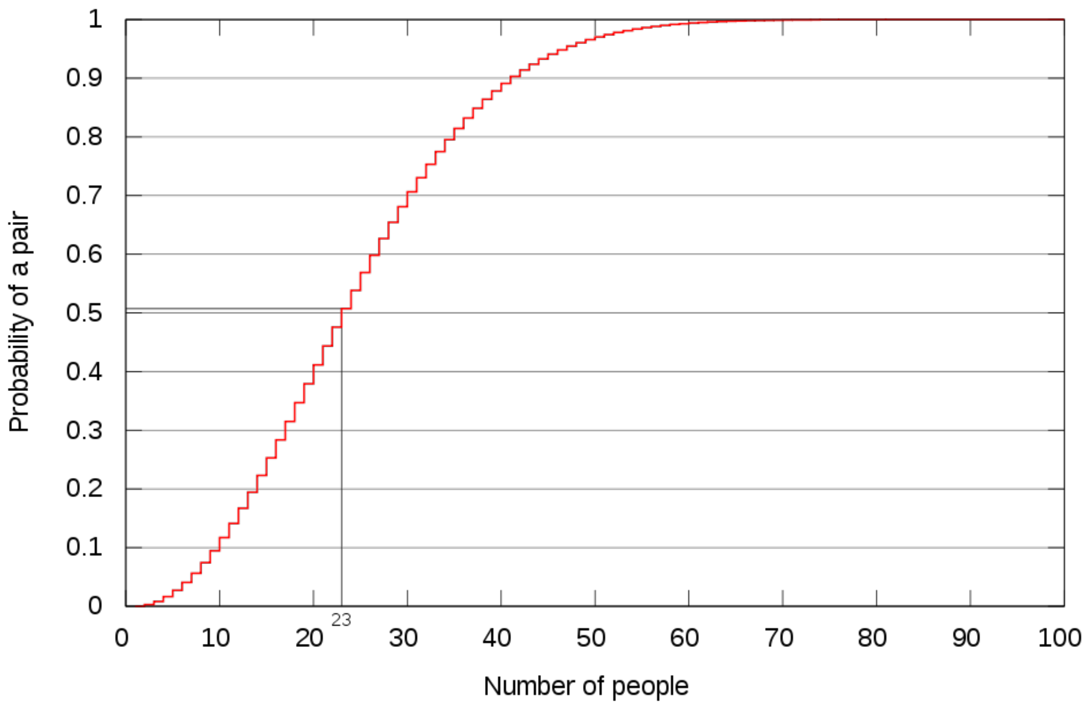

Figure 1: The approximate probability of at least two people sharing the same birthday in a randomly selected group.