CSE 312: Foundations of Computing II

Birthday "Paradox" Notes

We will then explore a common probability problem that leads to some u-intuitive outcomes. This is why why it is commonly called the birthday "paradox" (even though no logical contradictions will occur).

How many people would you expect you would need in a room to have a 50% chance of having two people with the same birthday? There are a lot of possible birthdays (365), so you might expect you would need a lot of people to make it just as likely than not to have two people with a birthday in a room. It turns out, you only need 23! If you have 23 or more people, there is a greater than 50% that two people in the room have the same birthday. Let's explore this formally and try to find out why.

Problem

Suppose we have a collection of n people in a room. What is the probability that at least 2 people share a a birthday? We will assume there are 365 possible birthdays, with uniform probability for each day.

Solution

Here we will walk through the steps introduced in class today to solve one of these probability problems when working with **uniform probability spaces**.

  (1) Define the sample space Ω and event E

  (2) Count the number of outcomes in Ω

  (3) Compute $P(E)$ by counting $|E|$

(1) Define Sample Space Ω and event E

For convenience, define the set $U = \{1, 2, 3, ..., 365\} = [365]$ be the set of possible birthdays each person can have. If we have a collection of n people in the room, we can think of their birthdays as being a list of n birthdays $[b_1, b_2, ..., b_n]$, where each birthday b_i is drawn from U. Importantly for the space of all possible birthday assignments, we should let the birthdays be drawn from U **with replacement** so that it is possible for two people to have the same birthday.

In other words, let our sample space Ω be the set of all possible birthday assignments:

$$\Omega = \{(b_1, b_2, ..., b_n) \mid \forall_i \ b_i \in U\}$$

We will then define the event of interest E as the set of birthday assignments where no two people share a birthday. If we wanted to write this out mathematically, we could write:

$$E = \{\text{birthday assignments where at least 2 people have same birthday}\} = \{(b_1, b_2, ..., b_n) \mid \forall_i \ b_i \in U, \ \exists_i \exists_{j \neq i} \ b_i = b_j\}$$

(2) Count the number of outcomes in Ω

The number of outcomes in Ω is the number of possible birthday assignments in total. We can think of counting this set as applying the product rule over all choices in an assignment:

$$|\Omega| = (\text{Number of choices for } b_1) \cdot (\text{Number of choices for } b_2) \cdot ... \cdot (\text{Number of choices for } b_n)$$

Since each birthday is chosen from U (with replacement), we have 365 options for each choice. This means:

$$|\Omega| = \underbrace{365 \cdot 365 \cdot ... \cdot 365}_{\text{n terms}} = 365^n$$

This then means that each individual outcome (possible birthday assignments) occurs with probability

$$P(\omega) = \frac{1}{365^n}$$

since we are working with a uniform probability space.

(3) Compute $P(E)$ by counting $|E|$

Unfortunately, counting the size of E is quite difficult, since we might have to account for every possible number of people that have colliding birthdays (e.g., number of ways to have only 2 people share a birthday, number of ways for only 3 people to share a birthday, etc.). Like we saw with the dice example in class, a common trick to trying to solve problems with "at least" is to flip it around and try counting the complement.

In other words, instead of trying to find

$$P(E = \text{at least 2 people have some birthday})$$

we will compute

$$1 - P(E^C = \text{no two people share a birthday})$$

The sample space for this probability is the same as above, but not we need to consider the set of birthday assignments such that no to people share a birthday.

Solving this approach is similar to counting the sample space Ω, but now we can't re-use a birthday once we have chosen it earlier. This is the same as taking 365 possible birthdays, and finding the ways to permute n of them.

$$|E^C| = \underbrace{365 \cdot 364 \cdot ... \cdot 365 - n + 1}_{\text{n terms}} = \frac{365!}{(365-n)!}$$

A very important question here is: Why do we use permutations here instead of combinations? Does the order of the birthdays matter?

Order does matter here! Our sample space Ω is defined to be all the possible assignments of birthdays, and we accounted for 365 possible choices for each person's birthday. If you consider $n = 3$, our counting of $|\Omega| = 365^3$ counts $(\text{Jan 1}, \text{Feb 2}, \text{March 3})$ as a different outcome than $(\text{March 3}, \text{Feb 2}, \text{Jan 1})$. Convince yourself that's true by writing out the sample space for a smaller number of possible birthdays (e.g., everyone can only be born on day 1, 2, 3).

So now we can compute the probability of interest, $P(E)$

$$P(E) = 1 - P(E^C) \tag{1}$$

$$= 1 - \frac{|E^C|}{|\Omega|} \tag{2}$$

$$= 1 - \frac{365!/(365-n)!}{365^n} \tag{3}$$

(1) is true from Corollary 1 (Complementation), (2) is from the definition of the probability of an event, and (3) is plugging in the values for $|E^C|$ and $|\Omega|$ from our earlier work. That's it! We have no calculated the probability as a formula in terms of the number of people n.

Exploring Solution

So while the math is all laid out, what does it mean? Well we now have this formula we can plug n into to find the probability of having at least 2 people in a room of n people sharing a room. This formula is exactly how we got the number shared in the intro paragraph. If we let $n = 23$, then plugging this formula into a calculator yield $P(E) \approx 0.507$.

What's even more surprising, is this probability grows very quickly! If you double the number of people in the room to $n = 46$, $P(E) \approx 0.948$! This is pretty shocking since 46 people is *much smaller* than the total of number of birthdays (365) but has such a high probability of having people with the same birthday. You actually only need to go up to $n = 57$ to get a probability of ≥ 0.99.

If you plug this formula into WolframAlpha, you can see how this probability grows as n increases.

(n from 0 to 120)
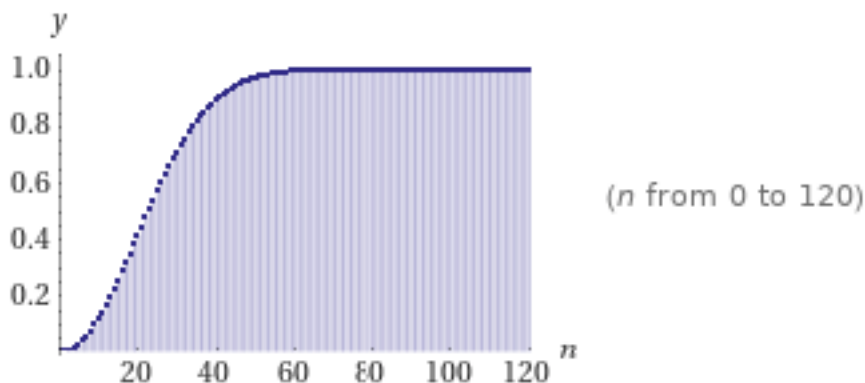
Why does the math betray our expectations? There are two main reasons for why our intuition does not match reality.

First, thinking in probabilities is not always easiest thing for our brains to do. We like certainty and so it takes lots of practice to think of things as probabilities. This is precisely why we need formal systems like probability theory to analyze random processes to find correct conclusions, even if they aren't intuitive.

For this problem in particular, one reason we underestimate the likelihood of conflicting birthdays is that we generally think a bit selfishly. You probably thought something like "I've met a lot of people in my life and only a handful of them have the same birthday as I do, therefore it's unlikely if we had a room of random people, they would share my birthday." In fact, that's somewhat true, it's not very likely that someone will share **your** birthday (check yourself: there is a $\frac{1}{365}$ probability that someone else will have your particular birthday).

However, the question had nothing to do with **your** birthday! It asked if there were **any pair of people** that have the same birthday. As we increase n, we quickly increase the number of pairs of people possible that could have conflicting birthdays. The number of pairs of people from n is exactly $\binom{n}{2}$, which grows quadratically in n. This is exactly why we see this number approach $P(E) \rightarrow 1$ so quickly!.