1 Planetary Party

(a) Suppose we are at party on a planet where every year is 2849 days. If 30 people attend this party, what is the exact probability that two people will share the same birthday? You may leave your answer as an unevaluated expression.

(b) Give an approximation for the probability in (a) using a result you learned from lecture.

(c) What is the minimum number of people that need to attend this party to ensure that the probability that any two people share a birthday is at least 0.5?

(d) Now suppose that 100 people attend this party. What the is probability that none of these 100 individuals have the same birthday?

Solution:

(a) Let's compute the probability that no two partygoers have the same birthday. We know the second person at the party cannot share a birthday with the first person, the third person at the party cannot share a birthday with the first two, etc. Thus

$$\mathbb{P}[\text{no collision}] = \left(1 - \frac{1}{2849}\right)\left(1 - \frac{2}{2849}\right)\left(1 - \frac{3}{2849}\right)\cdots\left(1 - \frac{29}{2849}\right)$$

Thus $\mathbb{P}[\text{collision}] = 1 - \mathbb{P}[\text{no collision}] = 1 - \left(1 - \frac{1}{2849}\right)\left(1 - \frac{2}{2849}\right)\left(1 - \frac{3}{2849}\right)\cdots\left(1 - \frac{29}{2849}\right)$.

(b) From lecture, we know that given n bins and m balls, $\mathbb{P}[\text{no collision}] \approx \exp(-m^2/(2n))$. Therefore in this case, if we want to find the probability of collision, we must find $1 - \mathbb{P}[\text{no collision}]$.

$$\mathbb{P}[\text{no collision}] = \exp\left(-\frac{30^2}{2 \cdot 2849}\right) = 0.854$$

This means that there is a 0.146 chance that two people share the same birthday in the group of 30.

(c) What is the minimum number of people that need to attend this party to ensure that the probability that any two people share a birthday is at least 0.5?

We know that in order for $\mathbb{P}[\text{no collisions}] = 0.5$, $m \approx 1.2\sqrt{n}$. Therefore, the minimum number of people required for the probability of a collision to be 0.5 is $1.2\sqrt{2849} \approx 64.05$ From this, we can conclude that we need at least 65 people to attend the party to ensure a 50% chance of two people sharing the same birthday.

(d) Once again we need to find $\mathbb{P}[\text{no collisions}]$ given that $m = 100$.

$$\mathbb{P}[\text{no collision}] = \exp\left(-\frac{100^2}{2 \cdot 2849}\right) = 0.173$$

There is about a 17% chance that 100 people don't share the same birthday.

2 Throwing Balls into a Depth-Limited Bin

Say you want to throw n balls into n bins with depth $k-1$ (they can fit $k-1$ balls, after that the bins overflow). Suppose that n is a large number and $k = 0.1n$. You throw the balls randomly into the bins, but you would like it if they don't overflow. You feel that you might expect not too many balls to land in each bin, but you're not sure, so you decide to investigate the probability of a bin overflowing.

(a) Count the number of ways we can select k balls to put in the first bin, and then throw the remaining balls randomly. You should assume that the balls are distinguishable.

(b) Argue that your answer in (a) is an upper bound for the number of ways that the first bin can overflow.

(c) Calculate an upper bound on the probability that the first bin will overflow.

(d) Upper bound the probability that some bin will overflow. [*Hint*: Use the union bound.]

(e) How does the above probability scale as n gets really large?

Solution:

(a) We choose k of the balls to throw in the first bin and then throw the remaining $n-k$, giving us $\binom{n}{k}n^{n-k}$.

(b) Certainly any outcome of the ball-throwing that overflows the first bin is accounted for – we can simply choose the first k balls that land in the first bin and then simulate the rest of the outcome via random throwing. However, we are potentially overcounting: if $k+1$ balls go in the first bin, we have many choices for which k of them that could have been the "chosen" ones, and we count each one of these choices as distinct. However, they correspond to the same configuration, namely the one where $k+1$ balls are in the first bin. Hence we get an upper bound.

(c) We divide by the total number of ways the balls could have fallen into the bins, with order, so we get

$$\frac{\binom{n}{k}n^{n-k}}{n^n} = \frac{\binom{n}{k}}{n^k}.$$

(d) Let A_i denote the event that bin i overflows. By symmetry $\mathbb{P}(A_i) = \mathbb{P}(A_1)$ for all i. By the union bound we have

$$\mathbb{P}(\cup_i A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i) \leq n\mathbb{P}(A_i) \leq n \cdot \frac{\binom{n}{k}}{n^k}.$$

(e) We get

$$n \cdot \frac{\binom{n}{k}}{n^k} = n \cdot \frac{n \cdot (n-1) \cdots (n-k+1)}{k! n^k} \leq n \cdot \frac{n^k}{k! n^k} = \frac{n}{k!} = \frac{n}{(0.1n) \cdot (k-1)!} = \frac{10}{(0.1n-1)!}.$$

Clearly, as n gets large this probability is going to 0. Note that this same analysis would work with $k = cn$ for any constant $0 < c < 1$. Hence, using some very coarse upper bounds, we can see that as the number of balls and bins grows, we have that it is very unlikely that we get a constant fraction of the balls in any single bin.