Balls in boxes; visual model covers many different stories.

N boxes and k balls. Put each ball independently into a random box.

We'll study the event A_k:
"first k balls all in different boxes".

$$\begin{aligned} P(A_2) &= \tfrac{N-1}{N} \\ P(A_3|A_2) &= \tfrac{N-2}{N} \\ P(A_4|A_3) &= \tfrac{N-3}{N} \\ &\cdots \\ P(A_k|A_{k-1}) &= \tfrac{N-(k-1)}{N} \end{aligned}$$

and so

$$P(A_3) = P(A_3|A_2) \times P(A_2) = \tfrac{N-2}{N} \times \tfrac{N-1}{N}$$

$$P(A_4) = P(A_4|A_3) \times P(A_3) = \tfrac{N-3}{N} \times \tfrac{N-2}{N} \times \tfrac{N-1}{N}$$

$$P(A_k) = P(A_k|A_{k-1}) \times P(A_{k-1}) =$$

$$= \tfrac{N-(k-1)}{N} \times \tfrac{N-(k-2)}{N} \times \ldots \times \tfrac{N-1}{N} \times \tfrac{N}{N}$$

Birthday problem. k people in a room. What is the chance some 2 people have the same birthday?

model: each person's birthday is equally likely to be any of the 365 days, independently.

Under this model, situation is same as in "balls in boxes" model with $N = 365$ boxes:

$$P(\text{some 2 people have the same birthday})$$

$$= 1 - P(\text{all } k \text{ people have different birthdays})$$

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

A well-known **surprising fact** is that, for this chance to be $\approx 50\%$, you need only $k = 23$ people.

The birthday problem gives a nice illustration of the use of **calculus approximations**. For small x

$$e^{-x} \approx 1 - x$$

$$\log(1 - x) \approx -x.$$

Looking at the "balls in boxes" formula for the event A_k: "first k balls all in different boxes":

$$\begin{aligned} \log P(A_k) &= \sum_{i=1}^{k-1} \log(1 - \tfrac{i}{N}) \\ &\approx \sum_{i=1}^{k-1} -\tfrac{i}{N} = -\tfrac{(k-1)k}{2N} \end{aligned}$$

and so

$$P(A_k) \approx \exp(-\tfrac{(k-1)k}{2N}).$$

Useful because we can see, for large N, how large k must be to make this chance be 1/2 say: solve

$$\tfrac{1}{2} = \exp(-\tfrac{(k-1)k}{2N})$$

to get $k \approx \sqrt{2 \log 2}\ \sqrt{N} \approx 1.18\sqrt{N}$.