Discrete mathematics MATH-260

Birthday paradox

or probabilistic pigeonhole principle

Question

Suppose that there are 25 students in a math class. What are the chances that there is a pair of students who share the same birthday?

Question

Suppose that there are 25 students in a math class. What are the chances that there is a pair of students who share the same birthday?

Solution

Imagine a list of students birthdays:

Alice	29.02
Bob	1.08
Charlotte	01.01

Number of possible lists is 366^{25}.

Number of lists without coincidences is $366 \cdot 365 \cdots 342$.

About probability of birthdays

Idealized assumptions

- Birthdays are uniformly distributed over 366 days.
- Birthdays or different people are independent events.
- Corollary: all birthday lists are equally likely.

Answer

The answer to our question is given by the formula:

$$P = 1 - \frac{366 \cdot 365 \cdots (366 - 24)}{366^{25}}.$$

Estimate the probability of a coincidence

Theorem

Suppose that $k \leq n$ are positive integers and each of k different people chooses 1 element from the set $[n]$. Their choices are uniformly random and independent. Then the probability $P = \frac{n!}{(n-k)!n^k}$ that they have chosen k different elements can be estimated as

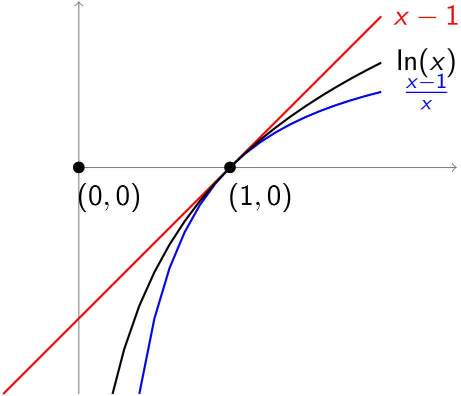
$$e^{\frac{-k(k-1)}{2(n-k+1)}} \leq P \leq e^{\frac{-k(k-1)}{2n}}.$$

Two inequalities for the logarithm

Lemma

For $x > 0$,

$$\frac{x-1}{x} \leq \ln(x) \leq x - 1.$$



Proof of the proposition

Now we estimate

$$\ln\left(\frac{n^k}{n(n-1)\cdots(n-k+1)}\right)$$
$$= \ln\left(\frac{n}{n-1}\right) + \ln\left(\frac{n}{n-2}\right) + \ldots + \ln\left(\frac{n}{n-k+1}\right)$$

Proof of the proposition

Now we estimate

$$
\begin{aligned}
&\ln\left(\frac{n^k}{n(n-1)\cdots(n-k+1)}\right)\\
&= \ln\left(\frac{n}{n-1}\right) + \ln\left(\frac{n}{n-2}\right) + \ldots + \ln\left(\frac{n}{n-k+1}\right)\\
&\geq \frac{\frac{n}{n-1}-1}{\frac{n}{n-1}} + \frac{\frac{n}{n-2}-1}{\frac{n}{n-2}} + \ldots + \frac{\frac{n}{n-k+1}-1}{\frac{n}{n-k+1}}
\end{aligned}
$$

Proof of the proposition

Now we estimate

$$
\begin{aligned}
&\ln\left(\frac{n^k}{n(n-1)\cdots(n-k+1)}\right)\\
&=\ln\left(\frac{n}{n-1}\right)+\ln\left(\frac{n}{n-2}\right)+\ldots+\ln\left(\frac{n}{n-k+1}\right)\\
&\geq \frac{\frac{n}{n-1}-1}{\frac{n}{n-1}}+\frac{\frac{n}{n-2}-1}{\frac{n}{n-2}}+\ldots+\frac{\frac{n}{n-k+1}-1}{\frac{n}{n-k+1}}\\
&=\frac{1}{n}+\frac{2}{n}+\ldots+\frac{k-1}{n}\\
&=\frac{1}{n}(1+2+\ldots+(k-1))=\frac{k(k-1)}{2n}.
\end{aligned}
$$

Proof of the proposition

Also we find

$$\begin{aligned}
&\ln\left(\frac{n^k}{n(n-1)\cdots(n-k+1)}\right)\\
&= \ln\left(\frac{n}{n-1}\right) + \ln\left(\frac{n}{n-2}\right) + \ldots + \ln\left(\frac{n}{n-k+1}\right)\\
&\leq \left(\frac{n}{n-1} - 1\right) + \left(\frac{n}{n-2} - 1\right) + \ldots + \left(\frac{n}{n-k+1} - 1\right)
\end{aligned}$$

Proof of the proposition

Also we find

$$\begin{aligned}
&\ln\left(\frac{n^k}{n(n-1)\cdots(n-k+1)}\right) \\
&= \ln\left(\frac{n}{n-1}\right) + \ln\left(\frac{n}{n-2}\right) + \ldots + \ln\left(\frac{n}{n-k+1}\right) \\
&\leq \left(\frac{n}{n-1} - 1\right) + \left(\frac{n}{n-2} - 1\right) + \ldots + \left(\frac{n}{n-k+1} - 1\right)
\end{aligned}$$

Proof of the proposition

Also we find

$$\begin{aligned}
&\ln\left(\frac{n^k}{n(n-1)\cdots(n-k+1)}\right)\\
&=\ln\left(\frac{n}{n-1}\right)+\ln\left(\frac{n}{n-2}\right)+\ldots+\ln\left(\frac{n}{n-k+1}\right)\\
&\leq\left(\frac{n}{n-1}-1\right)+\left(\frac{n}{n-2}-1\right)+\ldots+\left(\frac{n}{n-k+1}-1\right)\\
&=\frac{1}{n-1}+\frac{2}{n-2}+\ldots+\frac{k-1}{n-k+1}\\
&\leq\frac{1}{n-k+1}+\frac{2}{n-k+1}+\ldots+\frac{k-1}{n-k+1}
\end{aligned}$$

Proof of the proposition

Also we find

$$\begin{aligned}
&\ln\left(\frac{n^k}{n(n-1)\cdots(n-k+1)}\right)\\
&=\ln\left(\frac{n}{n-1}\right)+\ln\left(\frac{n}{n-2}\right)+\ldots+\ln\left(\frac{n}{n-k+1}\right)\\
&\leq\left(\frac{n}{n-1}-1\right)+\left(\frac{n}{n-2}-1\right)+\ldots+\left(\frac{n}{n-k+1}-1\right)\\
&=\frac{1}{n-1}+\frac{2}{n-2}+\ldots+\frac{k-1}{n-k+1}\\
&\leq\frac{1}{n-k+1}+\frac{2}{n-k+1}+\ldots+\frac{k-1}{n-k+1}\\
&=\frac{1}{n-k+1}(1+2+\ldots+(k-1))=\frac{k(k-1)}{2(n-k+1)}.
\end{aligned}$$

Proof of the proposition

Applying the exponential function to both sides of our estimates we get the following:

$$e^{\frac{-k(k-1)}{2(n-k+1)}} \leq \frac{n(n-1)\cdots(n-k+1)}{n^k} \leq e^{\frac{-k(k-1)}{2n}}.$$

This finishes the proof.

So the answer to the question in the beginning of this video is between 55.94% and 58.40%. More precisely, the probability is about 56.77%.