THE BIRTHDAY PARADOX

Let $k \geq 1$ be a natural number. Assume that k people are in a room. How likely is it that two of them share the same birthday? We assume that:

(1) The birthdays of each person are independent.
(2) The birthday of each person is a number from 1 to 365 (no leap years!).
(3) The birthday of each person is equally likely to be any day.

If $k > 365$, then two of them will share the same birthday for sure (in mathematical jargon, this is a consequence of the pigeonhole principle), so assume that $k \leq 365$. The probability of the event A that two of them share the same birthday is somewhat hard to compute (because, for example, it could be that also three share the same birthday). Instead, we compute the probability of the complement A^c: this is the probability that no two share the same birthday. We then have that $P(A) = 1 - P(A^c)$.

Now $P(A^c) = \frac{|A^c|}{|\Omega|}$, were Ω denotes the sample space. We think of it as the set of sequences $\langle x_1, x_2, \dots x_k \rangle$, where x_i is a number between 1 and 365, the birthday of person i. There are 365^k such sequences. A^c is the set of such sequences where $x_i \neq x_j$ for each $i < j \leq k$. There are $365 \cdot 364 \cdot \ldots \cdot (365 - k + 1)$ such sequences. Thus:

$$
\begin{aligned}
P(A^c) &= \frac{365 \cdot 364 \cdot \ldots \cdot (365 - k + 1)}{365^k} \\
&= \frac{365}{365} \cdot \frac{364}{365} \cdot \ldots \cdot \frac{365 - k + 1}{365} \\
&= \left(1 - \frac{0}{365}\right)\left(1 - \frac{1}{365}\right) \cdots \left(1 - \frac{k-1}{365}\right) \\
&= \prod_{i=0}^{k-1} \left(1 - \frac{i}{365}\right)
\end{aligned}
$$

To estimate this product, we use the fact that $1 + x \leq e^x$, for any real number x (assignment 2). This implies that

$$
\begin{aligned}
P(A^c) &\leq \prod_{i=0}^{k-1} e^{-\frac{i}{365}} \\
&= e^{-\frac{1}{365} \sum_{i=0}^{k-1} i} \\
&= e^{-\frac{k(k-1)}{365 \cdot 2}}
\end{aligned}
$$

Where the last step used that $\sum_{i=0}^{k-1} i = 0 + 1 + 2 + \ldots + (k-1) = \frac{k(k-1)}{2}$ (to see this, use induction on k, or observe that when e.g. k is even $0 + (k-1) = 1 + (k-2) = 2 + (k-3) = \ldots = (\frac{k}{2} - 1) + \frac{k}{2}$).

Note that this implies that $P(A^c)$ decreases exponentially fast. Already when $k \geq 23$, we obtain that $P(A^c) < \frac{1}{2}$. Thus when there are at least 23 people in a room, it is more likely than not that two of them will share a birthday.