The Birthday Problem

Q: What is the probability that, in a group of n people, at least 2 have the same birthday?

Let us represent the days of the year by the integers 1, 2, ..., 365. Then we choose our sample space S to be {all possible combinations of n birthdays}. That is, we include all possible combinations of n days, with repetition (up to n repetitions of the same day, where all n birthdays fall on the same day).

For example, if $n = 3$, we include:

- all the single days of the year (eg. (1, 1, 1), (2, 2, 2), (3, 3, 3), ...), in the case that all 3 birthdays fall on the same day,
- all combinations of 2 different days of the year (eg. (1, 1, 2), (1, 1, 3), (1, 1, 4), ...), in the case that 2 of the birthdays fall on the same day, and
- all combinations of 3 different days of the year (eg. (1, 2, 3), (1, 2, 4), (1, 2, 5), ...), in the case that all 3 birthdays fall on different days

Suppose that all birthdays are equally likely. Then, by the classical definition of probability,

$$\boxed{P(\{\text{at least 2 people share a birthday}\}) = \frac{|\{\text{at least } 2\ldots\}|}{|S|}}$$

We know from counting principles that

$$\begin{aligned}
|S| &= \#\text{ of ways to select the first birthday} \times \#\text{ of ways to select the second birthday} \\
&\quad \times \ldots \times \#\text{ of ways to select the } n\text{th birthday} \\
&= 365 \times 365 \times \ldots \times 365 \\
&= 365^n
\end{aligned}$$

and that

$$\begin{aligned}
|\{\text{at least } 2\ldots\}| &= \#\text{ of arrangements of } n \text{ birthdays where 2 people share a birthday} \\
&\quad + \#\text{ of arrangements of } n \text{ birthdays where 3 people share a birthday} \\
&\quad + \ldots + \#\text{ of arrangements of } n \text{ birthdays where } n \text{ people share a birthday}
\end{aligned}$$

So

$$\boxed{P(\{\text{at least 2 people share a birthday}\}) = \frac{|\{\text{2 people share a birthday}\}| + \ldots + |\{n \text{ people share a birthday}\}|}{365^n}}$$

Then we consider that

$$\begin{aligned}
&\#\text{ of arrangements of } n \text{ birthdays} \\
&\quad \text{where } r \text{ people share a birthday} = \#\text{ of ways to select the first unshared birthday} \\
&\qquad \times \#\text{ of ways to select the second unshared birthday} \\
&\qquad \times \ldots \times \#\text{ of ways to select the } (n-r)\text{th unshared birthday} \\
&\qquad \times \#\text{ of ways to select the shared birthday} \\
&\qquad \times \#\text{ of ways to arrange } n-r \text{ different birthdays and } r \text{ same birthdays} \\
&\quad = 365 \times 364 \times \ldots \times (365 - (n-r-1)) \times (365 - (n-r)) \times \frac{n!}{r!1!\ldots1!}
\end{aligned}$$

For example, if $n = 3$,

$$\begin{aligned}
\#\text{ of arr. where 2 people share a birthday} &= \#\text{ of ways to select the unshared birthday}\\
&\quad \times \#\text{ of ways to select the shared birthday}\\
&\quad \times \#\text{ of ways to arrange 1 unique birthday and 2 same birthdays}\\
&= 365 \times 364 \times \frac{3!}{2!1!}\\
&= 398\,580
\end{aligned}$$

and then

$$\begin{aligned}
P(\{\text{at least 2 people share a birthday}\}) &= \frac{|\{\text{2 people share a birthday}\}| + |\{\text{3 people share a birthday}\}|}{365^3}\\
&= \frac{398\,580 + 365}{365^3}\\
&= \frac{398\,945}{365^3}
\end{aligned}$$

HOWEVER, this seems very laborious to compute, particularly if n is large.
We can instead use the complement rule to determine that

$$\boxed{\begin{aligned}
P(\{\text{at least 2 people share a birthday}\}) &= 1 - P(\overline{\{\text{at least 2 people share a birthday}\}})\\
&= 1 - P(\{\text{no shared birthdays}\})\\
&= 1 - \frac{\#\text{ of ways to select } n \text{ unshared birthdays}}{365^n}\\
&= 1 - \frac{365 \times 364 \times \ldots \times (365 - n + 1)}{365^n}
\end{aligned}}$$

For example, if $n = 3$,

$$\begin{aligned}
P(\{\text{at least 2 people share a birthday}\}) &= 1 - \frac{\#\text{ of ways to select 3 unshared birthdays}}{365^n}\\
&= 1 - \frac{365 \times 364 \times 363}{365^3}\\
&= 1 - \frac{48\,228\,180}{365^3}\\
&= \frac{398\,945}{365^3}
\end{aligned}$$

NOTICE that, if $n > 365$, the above calculation produces

$$\begin{aligned}
P(\{\text{at least 2}\ldots\}) &= 1 - \frac{365 \times (365-1) \times \ldots \times (365-364) \times (365-365) \times \ldots \times (365-n+1)}{365^n}\\
&= 1 - \frac{365 \times \ldots \times 0 \times \ldots \times (365-n+1)}{365^n}\\
&= 1 - \frac{0}{365^n}\\
&= 1 - 0 = 1
\end{aligned}$$

Why does this make sense?

By the pigeonhole principle, if we have n objects to place in fewer than n pigeonholes, at least 1 pigeonhole will contain multiple objects. In this case, if there are more than 365 birthdays to distribute over 365 days, at least 2 birthdays will fall on the same day. Thus, the probability of at least 2 people sharing a birthday is 1, or absolutely certain.

NOTE that this counting method counts *ordered* n-tuples: for example, where $n = 3$, we consider (1, 1, 2) and (1, 2, 1) to be different combinations of birthdays.
If we were to instead consider unordered n-tuples, we could not use the classical definition of probability, since not all outcomes would be equally likely. For example, where $n = 3$, the unordered combination (1, 1, 2) is more likely than the unordered combination (1, 2, 3), since there are more ways in which the former can occur.