1 Gambling in the lottery

In the lottery, six numbered balls are drawn randomly from a sample of 49 balls carrying the numbers 1 to 49. The balls are drawn sequentially and are not placed back in the sample, so that each number can only occur at most once. In every drawing step, each of the remaining balls is chosen with equal probability. To win the lottery, you have to guess the six random numbers, where the order of the numbers is irrelevant.

a) Calculate the number of possible sequences of 6 out of 49.

The total number of possible number sequences is given by

$$\begin{aligned} N_{tot} &= \frac{49!}{(49-6)!} = 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 && (1) \\ &= 10068347520 && (2) \end{aligned}$$

Since the order of the numbers is irrelevant, and there are $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$ orders in which the numbers could be drawn, the total number of (order-independent) sets of lottery numbers is given by the binomial coefficient

$$\begin{aligned} N &= \frac{N_{tot}}{6!} = \frac{49!}{(49-6)! \cdot 6!} && (3) \\ &= 13983816 && (4) \end{aligned}$$

b) Calculate the probability that you guess all six numbers correctly.

Each possible sequence of numbers is drawn with equal probability. Together with the normalization criterion, it follows that

$$\begin{aligned} P &= \frac{1}{N} && (5) \\ &\approx 7.15 \times 10^{-8} && (6) \end{aligned}$$

c) Calculate the probability that you guess exactly four numbers correctly.

-

2 Birthdays

Assuming that the birthdays of a population are distributed equally over the 365 days of a year, how large is the probability that in a lecture attended by $N = 150$ students at least two of the students have their birthday on the same day of the year? How many people would have to attend the lecture so that the probability for them having their birthday on the same day is $p = 0.5$?

If N is larger than the number of days in a year, than there must be two people with the same birthday, so $p(N > 365) = 1$ (ignoring leap years). If there is only a single person, then there obviously can't be two people with the same birthday, so $p(N = 1) = 0$. It is easier to calculate the inverse probability, i.e. the probability $q = 1 - p$ for the case that no two individuals share a birthday.

If there are two people, the probability of no birthday occurring twice is

$$q_2 = \frac{364}{365}$$

For three people, it is

$$q_3 = \frac{364}{365} \cdot \frac{363}{365}$$

For $N \leq 365$ people, it is

$$q_N = \prod_{i=0}^{N-1} \frac{365 - i}{365} = \prod_{i=0}^{N-1} \left(1 - \frac{i}{365}\right) \tag{7}$$

We can quickly find the "critical" N (the number of people where the probability of two people having the same birthday is larger than 0.5) using the following python code snippet:

```python
for N in range(1, 365):
    q = 1.
    for i in range(N):
        q *= 1 - i / 365.
    print(f'q_{N} = {q}')
```

Executing this prints:

...

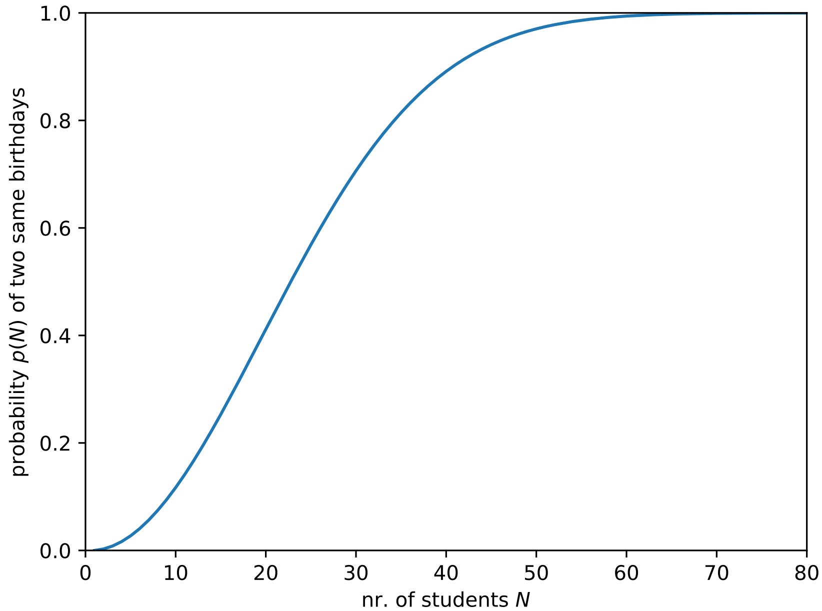
q_22 = 0.524

q_23 = 0.493

q_24 = 0.462

...

Thus, for 23 people, the probability of two people having the same birthday $p = 1 - q$ is larger than half for the first time.



3 Bayes and a-posteriori probabilities

Bayes theorem can be used to improve/revise our guesses on probabilities. Here is an example: Let us assume that - without any prior information - the probability of a terrorist attack is rather low, say $p_0 = p(A) = 5 \cdot 10^{-5}$. In the morning of 9/11, when the first plane hit one of the WTC towers (event B), it could just have been an accident, since the air traffic is rather high in Manhattan: in fact, in the 25000 days before 9/11, two such accidents occured - in 1945 the Empire State Building was hit by a plane by accident an in 1946 a building on Wall Street was hit - allowing you to estimate $p(B|\bar{A})$, the probability for a plane hitting the WTC by accident. Finally, let us assume that the terrorists are well prepared, unsuspected and hence always successful, implying $p(B|A) = 1$.

a) Use Bayes' theorem to get $p_1 = p(A|B)$ as a new estimate for the probability of a terrorist attack, using the "initial guess" $p_0 = p(A)$. This is called an "a-posteriori" probability, because it takes into account the new information (B has happened).

Bayes' theorem:

$$p(A|B) = p(B|A) \cdot \frac{p(A)}{p(B)} \tag{8}$$

$p(B|\bar{A}) \approx 2/25000 = 8 \times 10^{-5}$

$p(B) = p(B|A) \cdot p(A) + p(B|\bar{A}) \cdot p(\bar{A})$

$p(\bar{A}) = 1 - p(A)$

Plugging this into Bayes equation:

$$\Rightarrow p_1 := p(A|B) = p(B|A) \cdot \frac{p(A)}{p(B|A) \cdot p(A) + p(B|\bar{A}) \cdot (1 - p(A))} \tag{9}$$

$$\approx 38.5\ \% \tag{10}$$

b) After second plane hit the towers, one can again use Bayes to improve the "estimate" further. Use now p_1 as the new guess for $p(A)$ and calculate the new a-posteriori probability p_2. Is it still probable that the event "two planes hitting WTC" is an accident?

Here, we can use Bayes' equation almost exactly as we did before, only now in the equation we put p_1 everywhere where $p(A)$ used to be.

$$\Rightarrow p_2 = p(B|A) \cdot \frac{p_1}{p(B|A) \cdot p_1 + p(B|\bar{A}) \cdot (1 - p_1)} \tag{11}$$

$$\approx 99.9987\ \% \tag{12}$$

It is thus very unlikely that the two planes hit the towers by accident.