# Probability and Statistics

## Exercise sheet 1

**Exercise 1.1** Each of 3 people toss a fair coin. What is the probability of someone being the "odd man out"? This means that 2 of the players obtain the same outcome, while the third gets a different one.

**Solution 1.1**

$$\begin{aligned} p &:= P(\text{one of the players is the "odd man out"}) \\ &= 1 - P(\text{all 3 players get the same outcome}) \\ &= 1 - P(\text{HHH or TTT}) \\ &= 1 - P(\text{HHH}) - P(\text{TTT}) \\ &= 1 - \frac{1}{2^3} \times 2 \\ &= 1 - \frac{1}{4} \\ &= \frac{3}{4}. \end{aligned}$$

Here, $P(\text{HHH}) = P(\text{TTT}) = \frac{1}{\text{card}(\Omega)}$, where $\text{card}(\Omega) = 2 \times 2 \times 2 = 2^3 = 8$, the number of all possible outcomes for the 3 tosses.

**Exercise 1.2** An urn contains 5 red, 5 black and 5 white balls. 3 balls are chosen without replacement at random. What is the probability that they are of exactly two different colors?

**Solution 1.2** There are at least two methods to solve this question.

**Method 1:** Let $p := P(A)$ be the probability of the event

$$A := \{\text{the 3 balls are of exactly 2 different colors}\}.$$

Then

$$\begin{aligned} p &= 1 - P(\text{the 3 balls have the same color or all have different colors}) \\ &= 1 - P(S) - P(D) \end{aligned}$$

where $S = \{\text{same color}\}$ and $D = \{\text{all different colors}\}$. Using the notation $\binom{n}{k}$ $\left(= C_n^k\right)$ for binomial coefficients, we compute

$$\begin{aligned}
P(S) &= P(\text{all red}) + P(\text{all black}) + P(\text{all white}) \\
&= \frac{\binom{5}{3}}{\binom{15}{3}} + \frac{\binom{5}{3}}{\binom{15}{3}} + \frac{\binom{5}{3}}{\binom{15}{3}} \\
&= \frac{3 \times \frac{5\times4\times3!}{3!\times2}}{\frac{15\times14\times13}{3\times2}} \\
&= \frac{3\times6\times10}{15\times14\times13} \\
&= \frac{180}{2730} \\
&= \frac{6}{91};
\end{aligned}$$

$$\begin{aligned}
P(D) &= P(\{\text{white, red, black}\}) \\
&= \frac{\binom{5}{1}\times\binom{5}{1}\times\binom{5}{1}}{\binom{15}{3}} \\
&= \frac{125\times6}{15\times14\times13} \\
&= \frac{750}{2730} \\
&= \frac{25}{91}.
\end{aligned}$$

In conclusion,

$$p = 1 - \frac{6+25}{91} = \frac{60}{91} \approx 0.66.$$

**Method 2:** More directly, we want to compute the probability $p$ of the event

$$\begin{aligned}
A = \{ &\{\text{red, red, white}\}, \{\text{red, red, black}\}, \{\text{white, white, red}\}, \\
&\{\text{white, white, black}\}, \{\text{black, black, white}\}, \{\text{black, black, red}\} \}.
\end{aligned}$$

Since $\text{card}(A) = 6 \times \binom{5}{2} \times \binom{5}{1}$,

$$p = \frac{6\times\binom{5}{2}\times\binom{5}{1}}{\binom{15}{3}} = \frac{6\times\frac{5\times4}{2}\times5}{\frac{15\times14\times13}{6}} = \frac{60}{91} \approx 0.66.$$

**Exercise 1.3** 3 families with three members each organize themselves randomly into a line, to take a random picture. What is the probability that members of each family appear together in the picture? (i.e., not separated by members from other families.)

**Solution 1.3** Write

$$\begin{aligned}
F_1 &= \{M_1, M_2, M_3\} \\
F_2 &= \{N_1, N_2, N_3\} \\
F_3 &= \{J_1, J_2, J_3\}
\end{aligned}$$

for the members of families 1, 2, 3 respectively.

To count the possible positionings where the members of each family are together, first note that there are 3! possibilities for the positions of the families $(F_1F_2F_3, F_1F_3F_2, ..., F_3F_2F_1)$, and that for each of these possibilities, the members within each family can be organized in $(3!)^3$ ways.

Hence, the cardinality of the event of interest is

$$3! \times (3!)^3 = (3!)^4 = 6^4.$$

Furthermore, the number of all possibilities for a random picture is 9!, and hence

$$p = \frac{6^4}{9!} \approx 0.0036.$$

**Exercise 1.4** In a building with 6 floors (plus the ground floor), an elevator starts with 4 people at the ground floor. What is the probability that these people get off at exactly 2 floors?

**Solution 1.4** If $P_i$ $(i = 1, ..., 4)$ are the 4 people, each has 6 possibilities of where to get off. Thus, $\text{card}(\Omega) = 6^4$. Write

$$A = \{P_1, P_2, P_3, P_4 \text{ get off at exactly 2 floors}\}.$$

To compute card$(A)$, there are $\binom{6}{2}$ ways of selecting those two floors. Then, given the 2 floors, there are $\binom{4}{1} + \binom{4}{2} + \binom{4}{3}$ ways of distributing people between those floors. Therefore,

$$\begin{aligned} P(A) &= \frac{\left(\binom{4}{1} + \binom{4}{2} + \binom{4}{3}\right)\binom{6}{2}}{6^4} \\ &= \frac{(4+6+4)\frac{6\times 5}{2}}{1296} \\ &= \frac{14 \times 15}{1296} \\ &= \frac{35}{216} \\ &\approx 0.162 \end{aligned}$$

**Exercise 1.5** Consider an arbitrary sample space $\Omega$ and probability measure $P$. Using induction, prove that for any events $A_1, ..., A_n \subseteq \Omega$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le n} P\left(\bigcap_{j=1}^{k} A_{i_j}\right).$$

**Solution 1.5** For $n = 1$ this is trivial, and for $n = 2$, we have

$$A_1 \cup A_2 = A_1 \cup (A_2 \setminus A_1 \cap A_2)$$

where $A_1$ and $A_2 \setminus A_1 \cap A_2$ are disjoint.

Thus $P(A_1 \cup A_2) = P(A_1) + P(A_2 \setminus A_1 \cap A_2)$. Also, $P(A_1 \cap A_2) + P(A_2 \setminus A_1 \cap A_2) = P(A_2)$. Thus,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

which clearly matches the claimed formula for $n = 2$.

For the inductive step, suppose the formula holds for $n$ and let us show that it continues to hold for $n+1$.

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P(A_1 \cup ... \cup A_{n+1}) \\
&= P(A_1 \cup ... \cup A_n) + P(A_{n+1}) - P((A_1 \cup ... \cup A_n) \cap A_{n+1}) \\
\text{(induction)} \quad &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n} P\left(\bigcap_{j=1}^{k} A_{i_j}\right) + P(A_{n+1}) - \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n} P\left(\bigcap_{j=1}^{k} A_{i_j} \cap A_{n+1}\right) \\
&= \sum_{i=1}^{n+1} P(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n} P\left(\bigcap_{j=1}^{k} A_{i_j}\right) - \sum_{k'=2}^{n+1} (-1)^{k'} \sum_{1 \le i_1 < ... < i_{k'} = n+1} P\left(\bigcap_{j=1}^{k'} A_{i_j}\right).
\end{aligned}$$

In the last step, we separated the first term from the first sum and relabelled $k' = k+1$. Since the name of the dummy variable is irrelevant, we can rewrite the last sum:

$$\begin{aligned}
-\sum_{k'=2}^{n+1} (-1)^{k'} \sum_{1 \le i_1 < ... < i_{k'} = n+1} P\left(\bigcap_{j=1}^{k'} A_{i_j}\right) &= \sum_{k'=2}^{n+1} (-1)^{k'+1} \sum_{1 \le i_1 < ... < i_{k'} = n+1} P\left(\bigcap_{j=1}^{k'} A_{i_j}\right) \\
&= \sum_{k=2}^{n+1} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k = n+1} P\left(\bigcap_{j=1}^{k} A_{i_j}\right).
\end{aligned}$$

To conclude,

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n+1} A_i\right) &= \sum_{i=1}^{n+1} P(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n} P\left(\bigcap_{j=1}^{k} A_{i_j}\right) + \sum_{k=2}^{n+1} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k = n+1} P\left(\bigcap_{j=1}^{k} A_{i_j}\right) \\
&= \sum_{i=1}^{n+1} P(A_i) + \sum_{k=2}^{n+1} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n+1} P\left(\bigcap_{j=1}^{k} A_{i_j}\right) \\
&= \sum_{k=1}^{n+1} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n+1} P\left(\bigcap_{j=1}^{k} A_{i_j}\right).
\end{aligned}$$

Thus the formula holds by induction. This is a probabilistic version of the "inclusion-exclusion" principle.