Statistics for Data Science Exercise 1
Wintersemester 2023/24
Please complete these problems before the exercise session on Tuesday 24 October, 2023, 8:30. Please be prepared to present your solutions to any problems that you completed successfully.

1. Let Ω be a finite non-empty set, let $\mathcal{F} = \mathcal{P}(\Omega) := \{A \mid A \subset \Omega\}$ be a σ-algebra, and let $p\colon \Omega \to \mathbb{R}$ be a function such that

   (i) $p(\omega) \geq 0$ for all $\omega \in \Omega$,

   (ii) $\sum_{\omega\in\Omega} p(\omega) = 1$.

   Show that the function

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega), \quad A \in \mathcal{F},$$

   is a probability measure.

2. An urn contains 40 balls enumerated from 1 to 40. In a lottery, 6 balls are drawn without replacement from the urn. Tickets bearing the correct sequence of numbers, up to a permutation of the numbers, win a T-shirt, while the ticket with the correct ordered sequence wins a car. Compute the probability of winning a T-shirt and the probability of winning a car.

3. A class in primary school is composed of 25 pupils, all born outside of a leap year. We are interested in the probability that two or more children in the class have the same birthday.

   (a) Model this experiment with an appropriate sample space Ω and probability measure $\mathbb{P}$.

   (b) Represent the event that two of more children in the class have the same birthday by an appropriate $A \subset \Omega$, and compute its probability. Check numerically (using Python, MATLAB, R, WolframAlpha, etc.) that this probability is greater than 1/2.

   Hint: You may set Ω to be the set of all maps from $\{1, 2, \dots, m\}$ to $\{1, \dots, N\}$ for suitably chosen numbers m and N. Recall also the fact that the number of maps $f : \{1, 2, \dots, m\} \to \{1, \dots, N\}$ such that $f(i) \neq f(j)$ for all $i \neq j$ is given by

$$\frac{N!}{(N-m)!}.$$

4. The medical test for disease D has outcomes + (positive) and − (negative). We assume that

   - the probability for an individual to have the disease is 0.01,

- the probability of a positive test, given that the individual has the disease, is 0.9,
- the probability of a negative test, given that the individual does not have the disease, is 0.9.

Compute the probability that an individual has the disease, given that the individual has tested positive. Comment on the quality of the test.

Some useful formulae

Recall some formulae which may come in handy to compute probabilities.

- If $N \geq 1$ and $0 \leq m \leq N$, the number of ways of picking m numbers $f(1), \ldots, f(m)$ from $1, \ldots, N$ is given by

  $$N^m.$$

  Mathematically, this corresponds to the number of maps f from $\{1, \ldots, m\}$ to $\{1, \ldots, N\}$.

- If $N \geq 1$ and $0 \leq m \leq N$, the number of ways of picking m *distinct* numbers $f(1), \ldots, f(m)$ from $1, \ldots, N$ is given by

  $$\frac{N!}{(N-m)!} = N(N-1)\cdots(N-m+1).$$

  Mathematically, this corresponds to the number of maps f from $\{1, \ldots, m\}$ to $\{1, \ldots, N\}$ that are *injective*, i.e., that satisfy

  $$i \neq j \quad \Rightarrow \quad f(i) \neq f(j).$$

- If $N \geq 1$ and $0 \leq m \leq N$, the number of ways of picking a subset of m *distinct* elements from $1, \ldots, N$ is given by

  $$\binom{N}{m} := \frac{N!}{(N-m)!m!}.$$

  Mathematically, this corresponds to the number of subsets of the set $\{1, \ldots, N\}$ which have cardinality m.