Statistics for Data Science Exercise 2
Wintersemester 2023/24
Please complete these problems before the exercise session on Tuesday 31 October, 2023, 8:30. Please be prepared to present your solutions to any problems that you completed successfully.

1. Let X be a continuous real-valued random variable with the probability density function $f\colon \mathbb{R} \to \mathbb{R}$,
$$f(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{4x\mathrm{e}^{-x^2}}{(1+\mathrm{e}^{-x^2})^2} & \text{if } x \geq 0. \end{cases}$$
   (a) Compute the cumulative distribution function $F(x) = \mathbb{P}(X \leq x)$ of X for $x \in \mathbb{R}$.
   (b) Solve the quantile function $F^{-1}(q)$ of X for $q \in (0, 1)$.
   (c) Compute the probability $\mathbb{P}(0 < X < 1)$. Which value $a \in \mathbb{R}$ satisfies $\mathbb{P}(X \leq a) = 0.95$?

   Hint: In part (b), note that the restriction $F|_{(0,\infty)}\colon (0, \infty) \mapsto (0, 1)$ is a bijection, so the quantile function can be obtained by solving the inverse function of F from $q = F(x)$.

2. Consider a game, where we toss two fair coins. If one coin lands on Heads and the other on Tails, then the player gets one point. If both coins land on Heads, then the player gets two points. If both coins land on Tails, then the player gets zero points. Let us denote by 1 the outcome of a coin landing on Heads and by 0 the outcome of a coin landing on Tails. Let
$$\Omega = \{\omega = (\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{0, 1\}\}$$
denote the sample space and consider the discrete real-valued random variable $X\colon \Omega \to E$, $X(\omega) = \omega_1 + \omega_2$, where $E = \{0, 1, 2\}$ is the target space.
   (a) Compute the values of the probability mass function $p(x)$ of X for $x \in E$.
   (b) Solve the cumulative distribution function $F(x) = \mathbb{P}(X \leq x)$ of X for $x \in \mathbb{R}$.
   (c) Solve the quantile function $F^{-1}(q) = \inf\{x \in \mathbb{R} \mid F(x) > q\}$ for $q \in (0, 1)$.

   Hint: The quantile function in part (c) is a piecewise constant function, where the jumps occur precisely at $q \in \{F(0), F(1)\}$. What are the values of $F^{-1}(q)$ in the intervals $0 < q < F(0)$, $F(0) \leq q < F(1)$, and $F(1) \leq q < 1$?

3. Consider tossing two *four-sided dice*, which we assume to be fair. Let
$$\Omega = \{\omega = (\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, 2, 3, 4\}\}$$
denote the sample space and consider the discrete real-valued random variable $X\colon \Omega \to E$, $X(\omega) = \omega_1 + \omega_2$, where $E = \{2, 3, 4, 5, 6, 7, 8\}$ is the target space.

(a) Compute the values of the probability mass function $p(x)$ of X for $x \in E$.

(b) Solve the cumulative distribution function $F(x) = \mathbb{P}(X \leq x)$ of X for $x \in \mathbb{R}$.

(c) Solve the quantile function $F^{-1}(q) = \inf\{x \in \mathbb{R} \mid F(x) > q\}$ for $q \in (0, 1)$.

Hint: The quantile function in part (c) is a piecewise constant function, where the jumps occur precisely at $q \in \{F(2), F(3), \ldots, F(7)\}$. What are the values of $F^{-1}(q)$ in the intervals $0 < q < F(2)$, $F(2) \leq q < F(3)$, $F(3) \leq q < F(4)$, $\ldots$, $F(6) \leq q < F(7)$, and $F(7) \leq q < 1$?

4. For a real-valued random variable X, we say that $m \in \mathbb{R}$ is a *median* for X if $\mathbb{P}(X \geq m) \geq 1/2$ and $\mathbb{P}(X \leq m) \geq 1/2$.

(a) If X is a continuous random variable with CDF F, show that $F^{-1}(1/2)$ is a median.

(b) Compute a median for

(i) a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$.

(ii) an exponential random variable $X \sim \mathrm{Exp}(\lambda)$ with parameter $\lambda > 0$.