1. Let $C \subset \mathbb{R}^n$ be a closed, convex set. Prove that if $x \in \mathbb{R}^n$ and $x \notin C$, then there is a hyperplane separating $x$ from $C$. In other words, show that there exist $y \in \mathbb{R}^n$ and $r \in \mathbb{R}$ such that $y^t x > r$ while $y^t z < r$ for all $z \in C$. (Here $t$ denotes the transpose, so $y^t x$ is the inner product of $x$ and $y$.)

2. Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Prove that either there exists $x \in \mathbb{R}^n$ with $Ax = b$ and $x \geq 0$ (i.e., all the entries of $x$ are nonnegative) or there exists $y \in \mathbb{R}^m$ with $y^t A \geq c^t$ and $y \geq 0$, but not both.

Problems 3 and 4 refer to the following two dual linear programming problems, with $A \in \mathbb{R}^{m \times n}$, $x, c \in \mathbb{R}^n$ and $y, b \in \mathbb{R}^m$:

- maximize $c^t x$
- subject to $Ax \leq b$
- $x \geq 0$

- minimize $y^t b$
- subject to $y^t A \geq c^t$
- $y \geq 0$

Vectors $x$ and $y$ are called feasible points if they satisfy these constraints.

3. Prove that if there are feasible points $x$ and $y$, then $c^t x \leq y^t b$. This is weak duality (a feasible point in either linear program bounds the optimal value of the other).

4. Prove that if there exist $x$ and $y$ satisfying these constraints, then we can choose them so that $c^t x = y^t b$. This is strong duality (these bounds are optimal).

5. Imagine a financial model in which there are $n$ assets and $m$ different scenarios for the future, and let $A_{ij} \in \mathbb{R}$ be the value of one unit of asset $i$ in scenario $j$ (which may be positive or negative, since one could incur a liability). Suppose these values are known, but nobody knows which scenario will actually occur. Suppose also that there is some asset $i$ that has a constant value $A_{ij} = 1$ for all $j$ (think of this as just holding cash instead of investing).

A price vector $p \in \mathbb{R}^n$ is a vector giving the price of each asset (which could also be negative, in which case someone would have to pay you to convince you to take the asset). One natural way to arrive at a price vector is to write down a probability distribution $(\pi_1, \ldots, \pi_m)$ on the scenarios and let the price $p_i$ of asset $i$ be its expected value $\sum_j \pi_j A_{ij}$ under this distribution. However, some price vectors do not come from a probability distribution in this way.

Imagine an investor who owns $\lambda_i$ units of asset $i$ (if $\lambda_i < 0$, then the investor owes someone $|\lambda_i|$ units). The vector $\lambda \in \mathbb{R}^n$ is called the investor’s position. Under price vector $p$ it costs $\sum_i \lambda_i p_i$, and in scenario $j$ it is worth $\sum_i \lambda_i A_{ij}$.

Given asset values $A$ and a price vector $p$, an arbitrage opportunity is a position that in every scenario is worth strictly more than its original price.

Prove that there are no arbitrage opportunities if and only if there is a probability distribution $\pi$ on scenarios such that $p$ is equal to the expected value under $\pi$. (This gives another way in which no-arbitrage conditions naturally lead to probability distributions.)

6. Find a sequence of probability distributions on $\mathbb{R}$ whose characteristic functions converge pointwise to a discontinuous function (which cannot be a characteristic function, so the original sequence of distributions does not converge).

7. Let $x_1, \ldots, x_n$ be i.i.d. samples from a probability distribution on $\mathbb{R}$ with mean $\mu$ and variance $\sigma^2$. Prove that if $\bar{x} = (x_1 + \cdots + x_n)/n$, then

$$\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$
is an unbiased estimator of \( \sigma^2 \). In other words, the expected value of this quantity is exactly \( \sigma^2 \). (Note the \( n - 1 \) in the denominator instead of \( n \). The intuition here is that there are only \( n - 1 \) degrees of freedom, since \( x_i - \bar{x} \) is invariant under translating \( x_1, \ldots, x_n \) by a fixed amount.)

8. Imagine a coin with probability \( \theta \) of heads. If we fix \( k \) and then flip the coin repeatedly until obtaining \( k \) heads, what is the probability that this will take \( n \) flips? (Note that this is the opposite of what was done in class, where we chose \( n \) and then saw how many heads we got in \( n \) flips.)

What is an unbiased estimator of \( \theta \) given \( k \) and \( n \)?

If a Bayesian starts with a \( B(\alpha, \beta) \) prior on \( \theta \) and finds that it takes \( n \) flips to obtain \( k \) heads, what should the posterior distribution be?

Here \( B(\alpha, \beta) \) is the beta distribution, with density

\[
\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \theta^{\alpha - 1} (1 - \theta)^{\beta - 1} d\theta
\]

on \([0, 1]\).