Exercise Sheet 1: Data Science Methods

Technische Universität München

Winter semester 2024-2025

Instructions

You can hand in your solutions during the next tutorial on 07/11/24 or submit them earlier at the group office 1269 in the Physics Department.

1 Exercise: Probability (Kolmogorov) axioms

A *Probability Space* $(\Omega, \mathcal{F}, P)$ consists of:

- a **sample space** $\Omega \neq \varnothing$: set of possible outcomes (elementary events).
- an **event space** $\mathcal{F}$: set of events, where an event is in turn a set of possible outcomes (elementary events).
- a **probability measure** $P : \mathcal{F} \to [0, 1]$: a function that assigns a probability to every event $A \in \mathcal{F}$.

1.1 An event $A \in \mathcal{F}$ is given with known probability $P(A)$. What is the probability of its complement $P(A^c)$?
mathematical note: the complement of an event is itself an event, i.e. if $A \in \mathcal{F}$, then also $A^c \in \mathcal{F}$. Hence, $P(A^c)$ is well-defined.

Solution: The definition of the complement A^c of a set $A \subseteq \Omega$ is $A^c \equiv \{x \in \Omega | x \notin A\}$. From the definition it is clear that $A \cup A^c = \Omega$ and that the intersection is empty, $A \cap A^c = \varnothing$. Hence,

$$1 \overset{2.\ \mathrm{KA}}{=} P(\Omega) = P(A \cup A^c) \overset{3.\ \mathrm{KA}}{=} P(A) + P(A^c)$$

$$\Rightarrow P(A^c) = 1 - P(A),$$

where KA refers to the Kolmogorov axioms.

1.2 Let $A, B \in \mathcal{F}$. Show that from $A \subseteq B$ follows $P(A) \leq P(B)$.

Solution: We can always decompose a set B as the *disjoint* union $B = (B \backslash A) \cup (A \cap B)$, where the set difference is defined as $B \backslash A \equiv B \cap A^c$, i.e. the set of elements that belong to B but not to A. Furthermore, since $A \subseteq B \Rightarrow A \cap B = A$. Hence,

$$P(B) = P((B \backslash A) \cup A) \overset{3.\ \mathrm{KA}}{=} P(B \backslash A) + P(A) \overset{1.\ \mathrm{KA}}{\geq} P(A).$$

1.3 Show that for any two events $A, B \in \mathcal{F}$, the probability of the union is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Solution: Again, we make the *disjoint* decomposition $B = (B \backslash A) \cup (A \cap B)$. In terms of probability, $P(B \backslash A) \overset{3.\ \mathrm{KA}}{=} P(B) - P(A \cap B)$ and analogously for A. This yields a *disjoint* decomposition for the union $A \cup B = (A \backslash B) \cup (B \backslash A) \cup (A \cap B)$ and in terms of probability

$$P(A \cup B) = P(A \backslash B) + (B \backslash A) + P(A \cap B) = P(A) + P(B) - P(A \cap B).$$

1.4 Imagine a bag containing three different colored balls: red, blue, and green. The experiment consists of randomly pulling one ball out of the bag.

(a) Write explicitly the sample space Ω and the event space $\mathcal{F}$,
Hint: $\mathcal{F}$ should have 8 elements. Recall that if $A \in \mathcal{F}$, then also $A^c \in \mathcal{F}$.

Solution:

$$\Omega = \{\text{red}, \text{blue}, \text{green}\}.$$

$$\mathcal{F} = \{\varnothing, \{\text{red}\}, \{\text{blue}\}, \{\text{green}\}, \{\text{red}, \text{blue}\}, \{\text{red}, \text{green}\}, \{\text{blue}, \text{green}\}, \Omega\}.$$

(b) Assuming each ball is equally likely to be drawn, determine the probability $P(A)$ for all events $A \in \mathcal{F}$.

Solution:

$$P(\varnothing) = 0; P(\{\text{red}\}) = P(\{\text{blue}\}) = P(\{\text{green}\}) = 1/3.$$

$$P(\{\text{red}, \text{blue}\}) = P(\{\text{red}, \text{green}\}) = P(\{\text{blue}, \text{green}\}) = 2/3; P(\Omega) = 1.$$

2 Exercise: Probability interpretation

Consider the following two statements and explain how a frequentist and a Bayesian would interpret these probabilities, respectively.

2.1 The probability of a coin landing on heads is 50%.

Solution:
Frequentist interpretation: A frequentist would say that if the coin is flipped an infinite amount of times, 50% of the flips will land on heads. The probability is an objective measure based on repeated trials.
Bayesian interpretation: A Bayesian would view this as a subjective belief that the coin has a 50% chance of landing on heads in any given toss, based on prior knowledge or assumptions about the coin's fairness.

2.2 The probability that it rains tomorrow is 40%.

Solution:
Frequentist interpretation: Frequentists would have difficulty interpreting this single event probability because they focus on repeated experiments. However, they could argue that, in identical weather conditions repeated infinite times, rain would occur 40% of the time. So frequentists can still address this kind of questions, although a little far-fetched.
Bayesian interpretation: A Bayesian, as before, would interpret the 40% as a degree of belief, representing how confident they are that it will rain tomorrow, based on prior information like weather forecasts or historical data.

3 Exercise: Conditional probability

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A, B \in \mathcal{F}$ two events with $P(B) \neq 0$. The conditional probability of A given B is defined as

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)}. \tag{1}$$

3.1 Show that $P(A|B) + P(A^c|B) = 1$.

Solution: Since $A \cap B$ and $A^c \cap B$ are *disjoint* and $(A \cap B) \cup (A^c \cap B) = \Omega \cap B = B$, we have

$$P(A|B) + P(A^c|B) \overset{\text{eq. (1)}}{=} \frac{P(A \cap B) + P(A^c \cap B)}{P(B)} \overset{\text{3. KA}}{=} \frac{P((A \cap B) \cup (A^c \cap B))}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

3.2 Suppose that there are disjoint events $C_i \in \mathcal{F}$, i.e. $C_i \cap C_j = \varnothing$ for all $i \neq j$, such that $\bigcup_i C_i = \Omega$. Show that for any $B \in \mathcal{F}$, we can write $B = \bigcup_i (B \cap C_i)$.

Solution: $B = B \cap \Omega = B \cap \bigcup_i C_i = \bigcup_i (B \cap C_i)$. Formally, the last equation holds, since

$$x \in B \cap \bigcup_i C_i \Leftrightarrow x \in B \wedge \exists i \in \mathbb{N} : x \in C_i \Leftrightarrow \exists i \in \mathbb{N} : x \in B \wedge x \in C_i \Leftrightarrow x \in \bigcup_i (B \cap C_i).$$

3.3 Assume further that $P(C_i) \neq 0$ for all i. Show that

$$P(B) = \sum_i P(B|C_i)\, P(C_i). \tag{2}$$

Solution:

$$\sum_i P(B|C_i)\, P(C_i) \overset{\text{eq. (1)}}{=} \sum_i P(B \cap C_i) \overset{3.\text{ KA}}{=} P\left(\bigcup_i (B \cap C_i)\right) = P(B).$$

4 Exercise: Bayes' Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A, B \in \mathcal{F}$ two events with $P(B) \neq 0$. The Bayes' theorem states:

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}. \tag{3}$$

Using the result in equation (2) we can rewrite Bayes' theorem as

$$P(A|B) = \frac{P(B|A)\, P(A)}{\sum_i P(B|C_i)\, P(C_i)}, \tag{4}$$

where we assumed $C_i \in \mathcal{F}$, $C_i \cap C_j = \varnothing$ for all $i \neq j$, such that $\bigcup_i C_i = \Omega$ and $P(C_i) \neq 0$ for all i.

4.1 **Medical test accuracy:** A certain disease affects 1% of a population. A test for the disease correctly identifies 90% of those who have the disease. Furthermore, the test correctly identifies 95% of those who do not have the disease. If a person tests positive for the disease, what is the probability that he or she actually has the disease?

Solution: Given is

$$P(\text{disease}) = 0.01,\ P(+|\text{disease}) = 0.90,\ P(-|\text{no disease}) = 0.95.$$

From the latter (see Exercise 3.1) the subsequent probabilities follow

$$P(\text{no disease}) = 0.99,\ P(-|\text{disease}) = 0.10,\ P(+|\text{no disease}) = 0.05.$$

Hence, according to Bayes' theorem (4), we have

$$P(\text{disease}|+) = \frac{P(+|\text{disease})P(\text{disease})}{P(+|\text{disease})P(\text{disease}) + P(+|\text{no disease})P(\text{no disease})} \approx 0.154.$$

4.2 **Spam Email detection:** A spam filter classifies emails into "spam" or "not spam." 20% of all emails are spam. The spam filter correctly identifies 95% of spam emails and correctly identifies 90% of non-spam emails. If an email is classified as spam by the filter, what is the probability that it is actually spam?

Solution: Given is

$$P(\text{spam}) = 0.20,\ P(+|\text{spam}) = 0.95,\ P(-|\text{no spam}) = 0.90.$$

From the latter (see Exercise 3.1) the subsequent probabilities follow

$$P(\text{no spam}) = 0.80,\ P(-|\text{spam}) = 0.05,\ P(+|\text{no spam}) = 0.10.$$

Hence, according to Bayes' theorem (4), we have

$$P(\text{spam}|+) = \frac{P(+|\text{spam})P(\text{spam})}{P(+|\text{spam})P(\text{spam}) + P(+|\text{no spam})P(\text{no spam})} \approx 0.704.$$

4.3 **Unfair coins:** Consider 3 coins where two are unfair, yielding heads with probability 0.40 and 0.25, respectively, while the third coin is fair and yields heads with probability 0.50. If one randomly selects one of the coins and tosses it 3 times, yielding 3 heads, what is the probability this is the unbiased coin?

Solution: We denote the three coins as C1 (unfair), C2 (unfair), C3 (fair), according to the order in which they appear in the text. "H" stands for heads. We know that

$$P(\text{H}|\text{C1}) = 0.40,\ P(\text{H}|\text{C2}) = 0.25,\ P(\text{H}|\text{C3}) = 0.50.$$

Furthermore, because every toss is independent from the others, the probability of obtaining 3 heads out of 3 tosses, is

$$P(\text{HHH}|\text{C1}) = P(\text{H}|\text{C1})^3,\ P(\text{HHH}|\text{C2}) = \ P(\text{H}|\text{C2})^3, P(\text{HHH}|\text{C3}) = P(\text{H}|\text{C3})^3.$$

Finally, the probability of selecting one of the three coins is just

$$P(\text{C1}) = P(\text{C2}) = P(\text{C3}) = 1/3.$$

.

Hence, according to Bayes' theorem (4), we have

$$P(\text{C3}|\text{HHH}) = \frac{P(\text{HHH}|\text{C3})P(\text{C3})}{P(\text{HHH}|\text{C1})P(\text{C1}) + P(\text{HHH}|\text{C2})P(\text{C2}) + P(\text{HHH}|\text{C3})P(\text{C3})} \approx 0.611.$$

4.4 **Being on time for work:** A person uses his car 30% of the time, rides the bus 50% of the time and takes the S-Bahn 20% of the time, as he goes to work. He is late 5% of the time when he drives; he is late 10% of the time when he takes the bus; and he is late 99% of the time when he takes the S-Bahn. What is the probability he took the S-Bahn if he is on time?

Solution: We write "C" for car, "B" for bus and "S" for S-Bahn. Given is

$$P(\text{C}) = 0.30,\ P(\text{B}) = 0.50,\ P(\text{S}) = 0.20.$$

$$P(\text{late}|\text{C}) = 0.05,\ P(\text{late}|\text{B}) = 0.10,\ P(\text{late}|\text{S}) = 0.99.$$

From the latter (see Exercise 3.1) the subsequent probabilities follow

$$P(\text{no late}|\text{C}) = 0.95,\ P(\text{no late}|\text{B}) = 0.90,\ P(\text{no late}|\text{S}) = 0.01.$$

Hence, according to Bayes' theorem (4), we have

$$P(\text{S}|\text{no late}) = \frac{P(\text{no late}|\text{S})P(\text{S})}{P(\text{no late}|\text{C})P(\text{C}) + P(\text{no late}|\text{B})P(\text{B}) + P(\text{no late}|\text{S})P(\text{S})} \approx 0.0027.$$