Universität Bielefeld
Tanja Pasurek
Tutor: Ehsan Abedi WiSe 2023/24

Foundations of Statistics

Homework 1

(optional)

Part I. Axioms of probability

Throughout this text, we let $(\Omega,\, \mathcal{A},\, \mathbb{P})$ be a probability space.

1. By means of the probability axioms show that for any probability events $A, B, C \in \mathcal{A}$ the items below hold. Show also the Venn diagrams.

(a) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;

(b) $\mathbb{P}(A \triangle B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$, where $A \triangle B := (A \backslash B) \cup (B \backslash A)$ is the so-called *symmetric difference.*

(c) (union of 3 sets)

$$\begin{gathered}\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ -\mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(A \cap C) + \mathbb{P}(A \cap B \cap C).\end{gathered}$$

2. Using De Morgan's laws prove that for any $A_1, ..., A_N \in \mathcal{A}$

$$\begin{aligned}\textbf{(a)}\ \ \mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) &= 1 - \mathbb{P}\left(\bigcap_{n=1}^{N} A_n^{\mathrm{c}}\right); \\ \textbf{(b)}\ \ \mathbb{P}\left(\bigcap_{n=1}^{N} A_n\right) &\geq 1 - \sum_{n=1}^{N} \mathbb{P}(A_n^{\mathrm{c}}).\end{aligned}$$

3. Suppose we know that $A_1 \cap A_2 \cap ... \cap A_N \subset A$. Show that

$$\mathbb{P}\,(A) \geq \sum_{n=1}^{N} \mathbb{P}(A_n) - (N-1).$$

4. Prove that for any events A and B the following estimate holds:

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A) \cdot \mathbb{P}(B)| \leq \frac{1}{4},$$

which means

$$\begin{aligned} \textbf{(a)} \quad \mathbb{P}(A \cap B) - \mathbb{P}(A) \cdot \mathbb{P}(B) &\leq \frac{1}{4}; \\ \textbf{(b)} \quad \mathbb{P}(A \cap B) - \mathbb{P}(A) \cdot \mathbb{P}(B) &\geq -\frac{1}{4}. \end{aligned}$$

Hint: Use the elementary inequality $p(1-p) \leq 1/4$ valid for any $p \in [0, 1]$.

5. Prove the following statements

(a) If $(A_n)_{n=1}^{\infty} \subset \mathcal{A}$ is an increasing sequence of events (that is, $A_n \subset A_{n+1}$ for all n), then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right)$.

(b) If $(A_n)_{n=1}^{\infty} \subset \mathcal{A}$ is a decreasing sequence of events (that is, $A_n \supset A_{n+1}$ for all n), then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$.

Part II. Probability theory in practice!

6 . In a city, 65% of people drink coffee, 50% drink tea, and 25% both. What is the probability that a person chosen at random will drink at least one of coffee or tea? Will drink neither?

7. In a small town, 40% of households have at least one dog and 60% of households have at least one cat, while 20% of households have neither dogs nor cats. If a household is chosen at random, what is the probability that there is at least one cat *and* at least one dog?

Part III. Exercises in R

8. Calculate the sum of all odd numbers between 0 and 500.

9. Install the package `tidyverse` and load it. In this package, there is a dataset called diamonds that we would like to analyze. Use the command below to save the date:

```
D <- diamonds
```

(a) Try the following commands:

```
D; head(D); head(D,20); dim(D);
```

How many diamonds are there?

For each diamond, how many specifications are stored?

(b) Print the price of 10 first and 10 last diamonds in the dataset. What percentage of diamonds have prices greater than or equal to 10000?

(c) How many categories are there for variable `cut`? Can you compare them?

What percentage of diamonds have "Ideal" cuts? (Hint: you can also use the command `table`)

(d) What percentage of diamonds have "Ideal" cuts and have prices greater than or equal to 10000?