Universität Bielefeld
Tanja Pasurek
Tutor: Ehsan Abedi

WiSe 2024/25

Foundations of Statistics

Homework 1

(optional)

Part I. theoretical problems

Topic: Axioms of probability

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

1. Using *only* the probability axioms, show that for any events $A, B, C \in \mathcal{A}$ the following statements 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)

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

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

$$\textbf{(a)} \quad \mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) = 1 - \mathbb{P}\left(\bigcap_{n=1}^{N} A_n^{\mathrm{c}}\right);$$

$$\textbf{(b)} \quad \mathbb{P}\left(\bigcap_{n=1}^{N} A_n\right) \geq 1 - \sum_{n=1}^{N} \mathbb{P}(A_n^{\mathrm{c}}).$$

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

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

Hint: To prove (a), use the elementary inequality $p(1-p) \leq 1/4$, which is valid for any $p \in [0,1]$. To prove (b), use (a).

Topic: Conditional probability and independence

5. Suppose $A, B \in \mathcal{A}$ are events with $0 < \mathbb{P}(A) < 1$ and $0 < \mathbb{P}(B) < 1$.

(a) If A and B are disjoint, can they be independent?
(b) If A and B are independent, can they be disjoint?
(c) If $A \subset B$, can A and B be independent?
(d) If A and B are independent, can A and $A \cup B$ be independent?

6. Let $0 < \mathbb{P}(B) < 1$ and $\mathbb{P}(A|B^{\mathrm{c}}) = \mathbb{P}(A|B)$. Show that the events A and B must be independent. How do you interpret this?

Part II. practical problems

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?

8. The "one child rule" in some provincial parts of China had been changed to the following. All couples are allowed one baby. If the baby is girl, they are allowed to have exactly one more. If this rule is exactly followed (and ignoring possibilities of twins, etc.), what will be the resulting proportion of boys to girls in this community? Assume that for each child the probability of being girl is 0.5.

Hint: draw a tree diagram to determine the corresponding probabilities and then calculated the expected value for boys and girls in a typical family.

Part III. Exercises in R

9. Install the package `tidyverse` and load it:

```
install.packages("tidyverse")
library(tidyverse)
```

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?

(e) For each diamond, calculate the maximum of its three dimensions (x, y, z). Plot the maximum dimension versus price (y-axis).