Universität Bielefeld
Tanja Pasurek
Tutor: Ehsan Abedi WiSe 2024/25

Foundations of Statistics

Homework 2

Part I. Theoretical problems

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

Topic: probability measures and distribution functions

1. Prove the following statements known as the **continuity property** of probability measures (cf. Lecture Notes, Ch. 1.1, p. 30):

(a) If $(A_n)_{n=1}^{\infty} \subset \mathcal{A}$ is an increasing sequence of events (that is, $A_n \subseteq A_{n+1}$ 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, $A_n \supseteq A_{n+1}$ for all n), then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

2. Let $X : \Omega \to \mathbb{R}$ be an arbitrary real-valued random variable. Use exercise 1 to prove the following properties of its **commulative distribution function** (CDF)

$$F(x) := \mathbb{P}(X \leq x), \quad \text{for all } x \in \mathbb{R},$$

(that are listed in Ch. 1.3, p. 8).

(a) F is **non-decreasing**: $x < y$ implies $F(x) \leq F(y)$ (but not always a strict inequality "$<$").

(b) $0 \leq F(x) \leq 1$, with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$.

(c) F is **right-continuous**, that is, at each point $x \in \mathbb{R}$

$$F(x) = F(x^+) := \lim_{z \searrow x} F(z) \quad (= \textit{right limit}).$$

(d) F is **left-limited**: at each point $x \in \mathbb{R}$ there exists

$$F(x^-) := \lim_{z \nearrow x} F(z) \quad (= \textit{ left limit}).$$

Demonstrate by a counterexample that F is not necessarily left-continuous, i.e. it can also be $F(x) \neq F(x^-)$.

(e) Check that

$$\begin{aligned} &\text{(i)} \quad \mathbb{P}(x < X \le y) = F(y) - F(x); \\ &\text{(ii)} \quad \mathbb{P}(x \le X \le y) = F(y) - F(x^-); \\ &\text{(iii)} \quad \mathbb{P}(X = x) = F(x) - F(x^-) \quad (\neq \mathbf{0} \text{ , in general!}). \end{aligned}$$

Topic: Conditional probability and independence

3. Let $B \in \mathcal{A}$ be an event with $\mathbb{P}(B) > 0$. Prove that $\mathbb{Q}(\cdot) := \mathbb{P}(\cdot|B)$ is a probability measure on $(\Omega, \mathcal{A})$, see Ch. 1.2., p. 13.

In other words, show that

(a) $\mathbb{Q}(\Omega) = 1$.

(b) For any countable family of mutually disjoint sets $(A_n)_{n=1}^\infty$ with $A_n \in \mathcal{A}$, we have $\mathbb{Q}(\cup_{n=1}^\infty A_n) = \sum_{n=1}^\infty \mathbb{Q}(A_n)$. This means

$$\mathbb{P}\left(\bigcup_{n=1}^\infty A_n \,|B\right) = \sum_{n=1}^\infty \mathbb{P}\left(A_n \,|B\right).$$

4. Suppose $A, B \in \mathcal{A}$ are events with $\mathbb{P}(B) > 0$.

(a) Use exercise 3 to conclude that $\mathbb{P}(A|B) + \mathbb{P}(A^c|B) = 1$.

(b) Give counterexamples to show that in general the following statements are <u>false</u>:

(i) $\mathbb{P}(A|B) + \mathbb{P}(A|B^c) = 1$,

(ii) $\mathbb{P}(A|B) + \mathbb{P}(A^c|B^c) = 1$.

5.

(a) Let $A_1, A_2 \in \mathcal{A}$ be independent events. Show that their complements A_1^c, A_2^c are also independent.

(b) Let $A_1, A_2, ..., A_N \in \mathcal{A}$ be independent events. Show that the probability that none of the $A_1, A_2, ..., A_N$ occur is less than or equal to

$$\exp\left\{-\sum_{n=1}^N \mathbb{P}(A_n)\right\}.$$

Part II. Practical problems

6. (An epidemiologic problem). For the Roche Sars-CoV-2 Antigen Rapid Test, the following information is provided by the manufacturer on its accuracy:

– **Sensitivity** = 96.52,

– **Specificity** = 99.68.

Sensitivity is the conditional probability of a positive test when there is infection with Covid-19, and **specificity** is the conditional probability of a negative test, provided there is no infection with Covid.

Real data. We will consider the 7-day COVID-19 incidence in Bielefeld on two dates: the current date and two years ago, during its peak:

Oct. 2024	16	new infections per 100 000 inhabitants
Feb. 2022	1600	new infections per 100 000 inhabitants

Table 1: source: corona-in-zahlen.de

In addition, the following notation will be used for four events we study

$$\text{“}+\text{”} = \{\text{positive test}\},\ \text{“}\mathbf{C}\text{”} = \{\text{Covid Infection}\}.$$
$$\text{“}-\text{”} = \{\text{ negative test}\},\ \text{“}\mathbf{N}\text{”} = \{\text{No Covid Infection}\}.$$

(a) What is the probability of a (false) positive test in a non-infected person?

(b) Calculate the probability that a randomly selected person will test positive in Bielefeld in Oct. 2024.

(c) Now use Bayes' theorem to calculate the probability that a person who tests positive really has Covid (use Bielefeld data in Oct. 2024).

(d) Conversely, if a person tests negative, what is the probability of being infected anyway? (use Bielefeld data in Oct. 2024).

(e) Check your result from **(c)** using a probability tree in which you assume that 100 000 randomly selected persons ($\rightsquigarrow$ how many of them have Covid/no Covid $\rightsquigarrow$ how many of these will test positive/negative in turn).

(f) Plot the diagnostic power of the test (i.e. $\mathbb{P}(\mathbf{C}|+)$) as a function of the actual incidence (i.e. $\mathbb{P}(\mathbf{C})$) in `R` (both axis in %) and describe the dependence. Mark the point corresponding to the probability calculated in part (**c**) with **BI2024**. By repeating the same calculations, find the power of test for Bielefeld in Feb. 2022. Mark the point again in the plot with **BI2022**.

(g) At what proportion of actually infected persons would a positive test be equivalent to a 50/50 situation, i.e. $\mathbb{P}(\mathbf{C}|+) = 50\%$? Add a horizontal line into your plot and mark the corresponding point with a star.

(h) Confirm your numerical results in **(c)** by doing computer simulation with `R`. (see Ch. 1.2, p. 27).