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

Foundations of Statistics

Homework 2

Topic: Conditional probability and independence

Part I. theoretical problems

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

1. 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?

2. 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 $(\Omega, \mathcal{A})$. 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

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

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

(a) Use exercise 2 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) $\mathbb{P}(A|B) + \mathbb{P}(A^c|B^c) = 1$.

4. Let $0 < \mathbb{P}(B) < 1$ and $\mathbb{P}(A|B^c) = \mathbb{P}(A|B)$.
Show that the events A and B must be independent.

Part II. practical problems

5 (demographic problem). 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.

6. (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.

In the following we consider the events

$$\text{“}+\text{”} = \{\text{positive test}\}, \quad \text{“}\mathbf{C}\text{”} = \{\text{Covid Infection}\}.$$

(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, given that 1 in 150 people in your area are currently infected with Covid (which is close to the actual 7-day incidence in Bielefeld and NRW on 21.10.2022).

(c) Now use Bayes’ theorem to calculate the probability that a person who tests positive really has Covid (under the assumptions made in **(b)**).

(d) Conversely, if a person tests negative, what is the probability of being infected anyway?

(e) Check your result from **(c)** using a probability tree in which you assume that 1 500 000 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 **BI2020**. By repeating the same calculations, find the power of test if 1 in 10 people were infected with Covid (this e.g. corresponds to the incedence rate in New York in Spring 2020). Mark the point again in the plot with **NY2020**.

(g) At what proportion of actually infected persons would a positive test be equivalent to a 50/50 situation, i.e. $\mathbb{P}(\mathbf{C}|+) = 0.5$? 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 page 26 of Ch. 1.2 in the lecture notes).