Due Date: Computerpraktikum am 16.05. und 17.05.2024

Verantwortlich für Vorlesung bzw. Computerpraktikum:
Dr. P. Goldenzweig, Dr. J. Kieseler, Dr. S. Stefkova
Tutoren: Dr. G. De Pietro, Dr. R. Quishpe, M. Mormile, J. Eppelt, A. Brusamolino, Dr. X. Zuo, S. Brommer

Computerpraktikum zur Vorlesung
Moderne Methoden der Datenanalyse - Blatt 2

This and the other exercises are available in the git repository

https://gitlab.etp.kit.edu/Lehre/dataanalysisexercises_forstudents.git.

If you already cloned the repository and want to get the newest exercises, you just need to update the `DataAnalysisExercises_ForStudents` repository. To do so, go to `Git -> Git Interface` while in the `dataanalysisexercises_forstudents` directory. Then press **Pull last changes** button on the top right of the Git Interface menu. If you still need to clone the repository, please follow the instructions given on the first exercise sheet.

Exercise 2: Priors and Monte Carlo

Bayes' Theorem

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$$

shows that the conditional probability $P(A|B)$ of observing an event A in case an event B has happened depends on the conditional probability $P(B|A)$ and the a-priori probabilities (called priors) $P(A)$ and $P(B)$ for the events A and B respectively.

The priors are of crucial importance for the interpretation of measurements. A theory usually predicts a probability $P(B|A)$ to observe a measurement B for a given assumption (theory / parameter set) A. However in general one is interested in the opposite: Given a measurement B what can be deduced about the theory? How "probable" is it that theory A is correct?

- **Exercise 2.1:** **voluntary**

  "Should I carry an umbrella or should I risk to get wet?" A possible answer to this question is to look at the weather forecast. But as we all know it is not always reliable.

  Let's assume that if it will rain, the forecast predicts this correctly in 80 % of the cases. If it will not rain, the forecast is assumed to be accurate in 90 % of the cases. In Sun-City the a-priori rain probability is only 5 %, in Equal-City it's 50 % and in Rain-City it's 95 %. Calculate (on a sheet of paper or with a short program) the four probabilities that it will (not) rain if (no) rain is predicted for the three cities.

  There are two different risks of a wrong decision:

- Carry an umbrella, but it does not rain.
- Don't carry an umbrella in case it rains.

Which are the three possible strategies and which of them is the optimal one to minimize the risk of a wrong decision in each of the three cities? Calculate and compare the risk for each of the three possible strategies. Determine the optimal strategy when the second risk is considered 10 or 100 times more serious.

- **Exercise 2.2:** **obligatory**

  Test your calculations for the previous exercise 'experimentally' by writing a Monte Carlo simulation: Simulate N weather events, generate uniformly distributed random numbers between 0 and 1 for the rain forecast and compare them to the corresponding probability given in the above exercise. Make sure that you only use the values given in the text in your code. Then count the number of events in each category (rain and no rain predicted). Finally, use these numbers to determine the fraction of wrong weather forecasts and compare this number to the probability calculated before. Repeat the simulation for different values of N.

- **Exercise 2.3:** **obligatory**

  A famous logical decision problem is the so-called "Monty Hall Dilemma", named after the host of an American television game show ("Let's make a deal"). In German it is referred to as the "Ziegenproblem". Imagine you are contestant in a game show. There are three doors with an automobile behind one of them and goats behind the other two doors. The automobile and the two goats are assigned randomly to the three doors and you don't have any prior knowledge about where the goats or the automobile are. "First you point toward a door," says the host of the show. "Then I'll open one of the other doors to reveal a goat. After I've shown you the goat, you make your final choice whether to stick with your initial choice of doors, or to switch to the remaining door. You win whatever is behind the door." You begin by pointing to door number 1. The host shows you that door number 3 has a goat. What do you think - shall you stay with your initial choice (door 1) or switch to door number 2 to win the car? What are the probabilities to win the car? Calculate the probabilities by hand and write a programme that simulates a large number of such games to validate your calculated results.

Some useful relations about probabilities:

$$\begin{aligned}
P(A \vee B) &= P(A) + P(B) - P(A \wedge B) \\
P(A \vee \neg A) &= P(A) + P(\neg A) \;=\; 1 \qquad (\neg A = \text{not} A) \\
P(A \wedge B) &= P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \quad \Rightarrow \text{Bayes' Theorem} \\
P(B|A) &= P(B) \quad \text{if } A \text{ and } B \text{ are independent} \\
P(A) &= \sum_n P(A|B_n) \cdot P(B_n) \quad \text{Law of total probability}
\end{aligned}$$