142.351, 260014: Statistische Methoden der Datenanalyse

W. Waltenberger[1] (Vorlesung)
L. Einfalt[1,2] (Präsenzübung TU Wien und Universität Wien)
F. Reindl[1,2] (Onlineübung TU Wien und Universität Wien)

[1] Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften, A-1050 Wien, Nikolsdorfer Gasse 18
[2] Technische Universität Wien, Atominstitut, A-1020 Wien, Stadionallee 2

Wintersemester 2023/2024

Exercise class 1

Hand-in by: Friday, 27. October 2023 - 11:00
Class: Friday, 27. October 2023 - 14:15

Example 1.1

Let there be two events A and B with probabilities $P(A) = 1/3, P(B) = 5/7$.

a) What is the minimum value the expression $P(A \cap B)$ can attain?

b) What is $P(A \cap B)$ under the assumption that A and B are independent?

For the case of independence between A and B calculate the probability of the following events:

c) None of the events occurs

d) Exactly one of the two events occurs

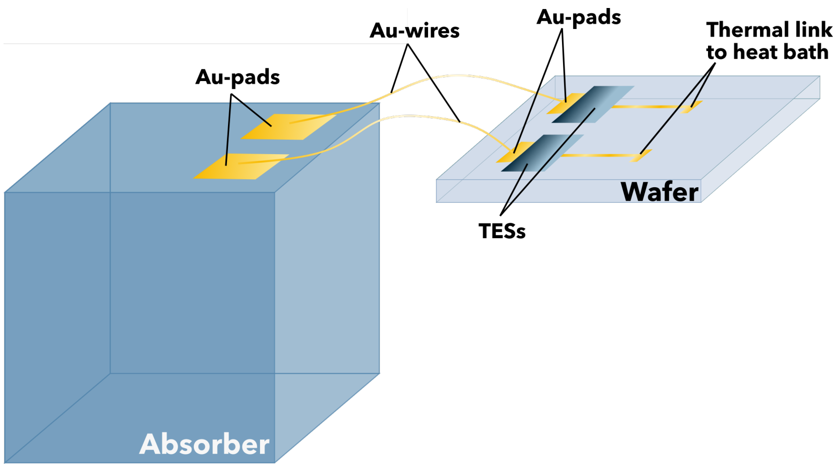
e) Both events occur

f) At least one of the two events occurs

g) At most one of the two events occurs

Example 1.2

In a novel detector concept for cryogenic calorimeters, the Transition Edge Sensor (TES - a very sensitive thermometer"based on a superconductor) is connected to the detector material by a combination of gold pads and gold bonding wires. This construction is called *remoTES* and is only operable if all three components work. The probability that a pad will not work is estimated with 10%, the probability a bonding wire will not hold with 5%.
The figure below shows a setup with two such remoTES detectors. Assuming that the rest of the experimental setup is not error-prone and that the probabilities of failure for the other components are independent of each other, calculate the probability that a signal can be measured (i.e. that at least one of the TES - wire combinations is operational).



Example 1.3

In 2022, the graduation rate for a bachelor's degree at the University of Vienna was 37.2% for women and 31.1% for men (the graduation rate indicates the number of degrees compared to all ended studies, including dropouts). Currently, 62.5% of all students at the University of Vienna are female (`https://www.univie.ac.at/fileadmin/user_upload/startseite/Dokumente/2023_UW_in_Zahlen_DT__1_.pdf`).

a) What is the probability that students will complete their bachelor's studies with a degree (independent of gender)?

b) Calculate the probability that a randomly selected dropout is a man.

Example 1.4

Compulsory for TU (142.351), optional for University (260014)!
You flip a symmetrical coin N times.

a) What is the probability of tossing **exactly** x times "heads"?

b) How does the probability of tossing **exactly** $x = N/2$ times heads behave for large N?

Example 1.5 (Prog)

Load the dataset „ClimateData.csv" from TUWEL(TU)/Moodle(Uni) and open it in the Jupyter notebook ClimateExcercis-skeleton.ipynb". The data contains three time series with one data point per decade (from 1880 to 2000) of the following variables: Global average temperature in degrees Celsius, an estimate of the number of pirates on the world's oceans, the CO2 concentration in the atmosphere in ppm (parts per million).

a) Make plots of all three time series, over time (decade) as the exogenous variable and the time series data as the endogenous variable.

b) Calculate the empirical correlation coefficient between global average temperature and the number of pirates on the world's oceans, and between global average temperature and average CO2 concentration.

c) How could these correlations be represented graphically? Interpret the results!

Example 1.6 (Prog)

You draw N random numbers from a uniform distribution in the interval [a,b].

- Calculate the expression for the mean and standard deviation of the underlying distribution.

Create a simple Jupyter notebook to run the experiment:

- Repeatedly draw N uniformly distributed random numbers and plot their mean and standard deviation as a function of N. What values do you choose for N and how many times do you need to repeat the experiment to verify your calculated results?