Tutorial for Lecture Course on Modelling and System Identification (MSI)
Albert-Ludwigs-Universität Freiburg – Winter Term 2022-2023

Tutorial 3: Emergency Guide to Statistics

Prof. Dr. Moritz Diehl, Robin Verschueren, Tobias Schöls, Rachel Leuthold, and Mara Vaihinger

In this tutorial we will recall basic concepts from statistics, which are very important for modelling and identifying a system. To do so, let's begin with a fundamental understanding of the term probability and how to calculate probabilities. Later we will have a look at the analysis of experiments with random outcomes.

1 Probability

"Probability" is a very useful concept to describe an uncertain situation. It can be interpreted in various ways, for example as an expression of subjective belief, e.g. when we are talking of how sure we are that something is true. Another way of interpreting probability is to interpreted it as the frequency of occurence of an event, i.e. if you roll a fair dice many times each number occurs in average in 1/6 of all rolls.

Before digging into the theory now, say hello to Max again! Last week he made his way to the casino successfully. So let's join him and help him analyse his chances of winning in the casino. He likes to start with a simple game called "'Macao"'. It is a dice game in which the player tries to get as close as possible but not above a specific number (let's say 9). The game is similar to the card game Black Jack. Before Max places his bet, he would like to know what his chances are to get exactly the value 9 with only two rolls of a fair dice (**event** A). Namely, how possible it is, that event A occurs.

Task: How can we compute this? First, think about all possible outcomes that can occur if you roll a fair dice twice:

$$\text{Possible outcomes:} \qquad \Omega = \{\{1,1\},\{1,2\},\dots,\{6,6\}\} \\ = \{\{i,j\} \mid \forall\, i,j \in \{1,2,3,4,5,6\}\}$$

Let's call the set containing all possible outcomes of the **experiment** "'rolling a dice twice"' the **sample space** Ω. Within these outcomes, how many fulfill the condition that the values add up to 9?

Possible outcomes contained in event A: $\dots$

So there are $\dots$ elements of all $\dots$ possible outcomes, that are contained in A. The chance that event A occurs is expressed with a numerical value, the probability $P(A)$. This value can be computed as follows, taking into account that in this special case, all outcomes of the experiment are equally likely:

$$P(A) = \frac{\text{amount of elements } s_i \text{ in A}}{\text{amount of elements in } \Omega} = \qquad . \tag{1}$$

This means, that Max has a chance of $\dots$ to get exactly the value 9 with only two rolls of a fair dice.

The function $P(A)$ is called **probability law**. It assigns to a set A of possible outcomes ($A \subseteq \Omega$) a nonnegative number $P(A)$ (probability of A), which describes how probable it is that event A occurs. A probability law satisfies the following axioms if A and B are disjoint sets (i.e. do not share any element).

1. Nonnegativity: $P(A) \geq 0$ for every event A.
2. Additivity: If A and B are disjoint (mutually exclusive), then $P(A \cup B) = P(A) + P(B)$.
3. Normalization: $P(\Omega) = 1$.

From these axioms we can easily derive, that $P(\emptyset) = 0$. Note that in the following we will sometimes denote $P(A \cap B)$ as $P(A, B)$.

Task: What do these axioms mean in terms of the example above?

In general it holds that, if an experiment has a finite number of possible outcomes s_i ($S_i \in \Omega$) we speak of a **discrete probability law**, which can be computed as follows: The probability of any event $A = s_1, s_2, \dots, s_n$ is the sum of the probabilities of its elements

$$P(s_1, s_2, \dots, s_n) = P(s_1) + P(s_2) + \dots + P(s_n), \qquad \text{only if } \{s_i \cap s_j\} = \emptyset, \quad \forall i, j : i \neq j. \tag{2}$$

In the special case we've seen in the example, when all single elements s_i have the same probability $P(s_i)$, the probability of the event A is given by

$$P(A) = \frac{\text{amount of elements } s_i \text{ in A}}{\text{amount of elements in } \Omega}. \tag{3}$$

1.1 Conditional Probability

The conditional probability describes the probability that the outcome of an experiment, that belongs to the event B also belongs to some other given event A. For example what the chance is that a person has a specific disease (event A) given that a medical test was negative (event B). The conditional probability of A, given B with $P(B) > 0$, is defined as

$$P(A|B) = \frac{P(A, B)}{P(B)}. \tag{4}$$

Explained in words, out of the total probability of the elements of B, the probability $P(A|B)$ is the fraction that is assigned to possible outcomes that also belong to A. If the sample space Ω is finite and all outcomes are equally likely then

$$P(A|B) = \frac{\text{amount of elements of A} \cap \text{B}}{\text{amount of elements of B}}. \tag{5}$$

Now back to Max: The game started, Max rolled his dice once and got a 3 (event A). Compute the chance that he does not exceed the value 9 after rolling the fair dice again (event B).

Task: Determine the conditional probability $P(B|A)$ with $A = \{\textit{first roll is a } 3\}$, $B = \{\textit{sum of two rolls is } \leq 9\}$. Repeat this task for the cases that he first gets a 1, 2, 4, 5 or 6 i.e. compute the conditional probability $P(B|A_i)$, with i = 1, 2, 4, 5, 6.

Max is interested to know what the total probability $P(B)$ is, that he gets a value less than 9 with two rolls of the fair dice. He already knows the conditional probabilities $P(B|A_i)$ describing the chance of getting less than 9, when the first roll was a 1, 2, 3, 4, 5 or 6. What else does he need to compute $P(B)$ using the conditional probabilities $P(B|A_i)$? Have a look at the following theorem:

Total Probability Theorem Let $A_1, \dots, A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in one and only one of the events $A_1, \dots, A_n$), i.e. $A_i \cap A_j = \emptyset \;\; \forall i, j : i \neq j$ and $\bigcup_{i=1}^{n} A_i = \Omega$ and assume that $P(A_i) > 0$, for all $i = 1, \dots, n$. Then, for any event B, we have

$$\begin{aligned} P(B) &= P(A_1, B) + \dots + P(A_n, B) & (6)\\ &= \sum_{i=1}^{n} P(A_i, B) & (7)\\ &= P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n) & (8)\\ &= \sum_{i=1}^{n} P(A_i)P(B|A_i). & (9) \end{aligned}$$

Task: Compute the missing probabilities $P(A_i)$. Then compute the total probability $P(B)$ that he gets a value less than 9 with two rolls of the fair dice.

Max is getting more and more involved in the probabiliy stuff and asks himself: "'Given that I get less than 9 within two rolls of the fair dice (event B). What is the probability that in my first roll I got a 5?"' *Task: Can you help him? Use the following theorem to answer his question.*

Bayes Theorem Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$, for all i. Then, for any event B such that $P(B) > 0$, we have

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} \tag{10}$$

$$= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)}. \tag{11}$$

Independent events In the special case that the occurrence of an event A does not influence the probability that event B occured, those events are independent and the following relation holds

$$P(A, B) = P(A)P(B). \tag{12}$$

If two events A and B do not satisfy this relation, they are dependent. Mutually exclusive events cannot be independent.

Task: Given this definition, decide whether the events $A = \{$first roll is a 3$\}$ and $B = \{$sum of the two rolls is 9$\}$ are independent or not.

2 Discrete Random Variables

To further analyze the properties of an experiment, it is useful to introduce a **random variable** that describes the outcome of the experiment. Consider a single experiment which has a set of possible outcomes, i.e. Max playing a single round (rolling the fair dice twice). Then a random variable X is for example the sum of two rolls. The random variable asserts a particular number to each possible outcome of this experiment.

Task: For the outcome $\{1, 3\}$, what is the value of the random variable? $x = \ldots$

We refer to this number as the **numerical value** or the **experimental value** of the random variable and denote it in lower case letters, while a random variable is denoted in capital letters. Mathematically, a random variable is a real-valued function of the experimental outcome.

To get a better understanding of this concept let's have a look at another example: Given an experiment consisting of a sequence of 3 tosses of a coin, with possible oucomes {HHH}, {HHT}, {HTT}, {TTT}, {THH}, {TTH}, {THT} and {HTH}. Let the random variable X be the number of tails in the sequence. For the outcome $\{HHH\}$ the value of the random variable is $x = 0$. The sequence of heads and tails denoted in letters, can't be a random variable as it is no real-valued number.

Task: What is the value of the random variable for the outcome {THH}? $x = \ldots$

Task: Which of the following examples can be described by a random variable?

1. The sequence of heads and tails denoted in letters, when tossing coins.
2. The number of false positives of a medical test
3. The names of the people in this room.
4. The percentage of students registered to MSI, which actually come to the class.

Task: Think of another example that can be described by a random variable.

2.1 Discrete Random Variables

A random variable is discrete if its range is finite, e.g. integers between 1 and 6. It can be described by an associated **probability mass function (PMF)**:

$$P_X(x) := P(X = x), \tag{13}$$

which gives the probability of each numerical value x that the random variable X can take. The probability mass of x is the probability of the event $(X = x)$. Follow the steps below to compute a PMF:

1. Determine all possible outcomes.
2. Follow the two steps for each numerical value x:
   (a) Determine all possible outcomes which lead to the event $(X = x)$.

(b) Add the probabilities of all these outcomes to obtain $P_X(x) := P(X = x)$.

The events $(X = x)$ are disjoint and form a partition of the sample space. From the addition and normalization axioms it follows

$$\sum_x P_X(x) = 1 \tag{14}$$

$$P(X \in S) = \sum_{x \in S} P_X(x). \tag{15}$$

Example 1 Consider tossing a coin 3 times, the possible outcomes (HHH), (HHT), (HTT), (TTT), (THH), (TTH), (THT) and (HTH) are equally likely. Let the random variable X be the amount of heads in the sequence, then one possible outcome leads to $X = 0$, three possible outcomes lead to $X = 1$, three possible outcomes lead to $X = 2$ and one possible outcome leads to $X = 3$. So the PMF is

$$P_X(x) = \begin{cases} 1/8 & \text{if } x = 0 \\ 3/8 & \text{if } x = 1 \\ 3/8 & \text{if } x = 2 \\ 1/8 & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Task: Compute the PMF of the random variable X describing the sum of two rolls of a fair dice.

Functions of random variables Consider a probability model of the size of 11-year old children in Germany, define the random variable X as the size in meters. The transformation $Y = 0.3048X$, gives the size of the children in inches. The transformation is affine and can be written in a general form as $Y = g(X) = aX + b$ with parameters a and b.

If $g(X)$ is a function of a random variable X, then $Y = g(X)$ is also a random variable, since it provides a numerical value for each possible outcome. This is because every outcome in the sample space defines a numerical value x for X and hence also the numerical value $y = g(x)$ for Y. The function g may also be a nonlinear function. If X is discrete with PMF p_X, then Y is also discrete, and its PMF p_Y can be calculated using the PMF of X.

$$P_Y(y) = \sum_{\{x | g(x) = y\}} P_X(x) \tag{16}$$

Expected Value, Variance and Standard deviation Max learned to describe the outcome of rolling a dice twice by a random variable. With a PMF, he can describe the characteristics of the random variable. It is convenient to summarize these characteristics with some quantities associated to the random variable. For example the expected value and the variance of the random variable. The **expected value** of X, also **expectation** of X often denoted by μ_X, is a weighted average of the possible values of X. The weights are the associated probabilities $p_X(x)$:

$$\mathbb{E}\{X\} = \sum_x x P_X(x). \tag{17}$$

Let $g(x)$ be a real valued function and X be a random variable, then the expected value rule is defined as

$$\mathbb{E}\{g(X)\} = \sum_x g(x) P_X(x). \tag{18}$$

Useful properties of the expectation operator are

- $\mathbb{E}\{c\} = c$
- $\mathbb{E}\{X + c\} = \mathbb{E}\{X\} + c$
- $\mathbb{E}\{X + Y\} = \mathbb{E}\{X\} + \mathbb{E}\{Y\}$

- $\mathbb{E}\{cX\} = c \cdot \mathbb{E}\{X\}$

with c being a constant, and X and Y being random variables.

Task: Compute the expected value $\mathbb{E}\{X\}$ of the random variable X describing the result of rolling a fair dice.

Another quantity to characterize a random variable is the **variance**, denoted as $\mathrm{var}(X)$ written as σ_X^2. It is a measure of dispersion of X around the mean $\mathbb{E}\{X\}$, which is always nonnegative. Here $g(X) = (X - \mathbb{E}\{X\})^2$.

$$\sigma_X^2 = \mathbb{E}\{(X - \mathbb{E}\{X\})^2\} = \sum_x (x - \mathbb{E}\{X\})^2 \, P_X(x). \tag{19}$$

Its square root is called the standard deviation σ_X, which is easier to interpret because it has the same unit as X. The expected values can also be used to compute the variance: $\sigma_X^2 = \mathbb{E}\{X^2\} - (\mathbb{E}\{X\})^2$.

Task: Compute the variance σ_X^2 and the standard deviation σ_X of the random variable X describing the result of rolling a fair dice.

Task: Given the random variables $X, Y \in \mathbb{R}$, with $Y = aX + b$ (affine function) and the scalar parameters a and b, compute the expected value and the variance of Y.

2.2 Multiple Discrete Random Variables

In many situations a single experiment has more than one random variable of interest. For example in a medical diagnosis context, the results of several tests (random variables) may be significant. Then, all random variables are associated with the same sample space (here: set of possible diagnoses), the same experiment (here: medical diagnosis process)and the same probability law, namely the **joint PMF** of X and Y. It is defined as

$$P_{X,Y}(x, y) := P(X = x, Y = y), \tag{20}$$

for all pairs of numerical values (x, y) of X and Y. The joint PMF provides the probability of any event that can be expressed in terms of the random variables X and Y. If the event A is the set of all pairs (x, y) with a certain property, then $P((X, Y) \in A) = \sum_{(x,y)\in A} P_{X,Y}(x, y)$.

Task: Consider rolling the fair dice twice again. Let the random variables X and Y describe the outcome of the first and second roll respectively.

1. *Compute the joint PMF $P_{X,Y}(x, y)$ of the random variables X and Y.*
2. *What is the chance that a 6 occured once?*

The **marginal PMF** can be derived from the joint PMF as follows

$$P_X(x) = \sum_y P_{X,Y}(x, y) \tag{21}$$

and accordingly for $P_Y(y)$.

Task: Compute the marginal PMFs $P_X(x)$ and $P_Y(y)$ from the joint PMF $P_{X,Y}(x, y)$ you just computed.

For functions of multiple random variables the same concept holds as for single random variables. Let $g(X,Y)$ be a function of the random variables X and Y, which defines some other random variable $Z = g(X,Y)$. Then, the PMF of Z follows as

$$P_Z(z) = \sum_{(x,y)|g(x,y)=z} P_{X,Y}(x,y) \tag{22}$$

and its expected value is

$$\mathbb{E}\{Z\} = \sum_{x,y} g(x,y) P_{X,Y}(x,y). \tag{23}$$

2.3 Conditionals and Conditional PMF

The concept of conditional probability introduced in section 1.1 also holds for random variables. Consider a single experiment with the random variables X and Y associated to it. The **conditional PMF** of the random variable X, given the experimental knowledge that the random variable Y takes the value y (with $p_Y(y) > 0$) is defined as

$$P_{X|Y}(x|y) := P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P_{X,Y}(x,y)}{P_y(y)}. \tag{24}$$

It is useful to compute the joint PMF sequentially with $P_{X,Y}(x,y) = P_Y(y)P_{X|Y}(x|y)$.

Task: Compute the conditional PMF $P_{X|Y}(x|y)$ for the random variables X and Y describing the result of the first, respectively the second roll of a fair dice. Assume that the second roll of the fair dice was a 4 ($y = 4$). Use the joint PMF of the random variables X and Y from task (1) in section 2.2 and the related marginal PMFs.

The **conditional expectation** of X given a value y of Y is defined by

$$\mathbb{E}\{X|Y = y\} = \sum_x x P_{X|Y}(x|y). \tag{25}$$

2.4 Independence

Two random variables X and Y are independent, if the following holds

$$\forall x, y : P_{X,Y}(x,y) = P_X(x)P_Y(y). \tag{26}$$

Then also $\mathbb{E}\{XY\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}$ and $\mathbb{E}\{g(X)h(Y)\} = \mathbb{E}\{g(X)\}\mathbb{E}\{h(Y)\}$ for any function g and h. Interesting to note is that the variance of the sum of two **independent** random variables X and Y is the sum of the variance of each random variable: $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$.

Task: Check whether the random variables X and Y each describing the result of rolling a fair dice are independent.

3 Continuous Random Variables

After a long day in the casino, Max decides to go home. He hops on his bike and start riding home. At one of his favourite spots he stops to determine the correct GPS position $\mathbf{p} = \begin{bmatrix} x & y \end{bmatrix}^\top$. He's only interested in two dimensions, namely longitude and latitude. Unfortunately he notices, that the signal is a bit noisy. Assume that the noise is additive, i.i.d. and has zero mean.

$$\mathbf{p}_{\text{measurement}} = \begin{bmatrix} x + \epsilon_x \\ y + \epsilon_y \end{bmatrix}$$

A random variable X which can take an infinite number of values is a continuous random variable, e.g. the noise of the x-value of the GPS position. The concepts introduced for discrete random variables also exist for continuous random variables: The continuous random variable can be described in terms of a nonnegative **probability density function (PDF)**. The PDF in the continuous case is defined as

$$P(X \in B) = \int_B p_X(x)\,\mathrm{d}x \tag{27}$$

for every subset $B \subseteq \Omega$. For example if $B = [a, b]$ with $a, b, X \in \mathbb{R}$, then $P(a \leq X \leq b) = \int_a^b p_X(x)\mathrm{d}x$. This probability can be interpreted as the area below the graph of the PDF. A PDF of a continuous random variable also has to satisfy the axioms of a probability law: The probability $P(X \in B)$ is nonnegative, as $p_x(x) \geq 0$ for every x. The probability of an event containing the whole sample space is normalized to 1:

$$\int_{-\infty}^{\infty} p_X(x)\,\mathrm{d}x = P(-\infty < X < \infty) = 1. \tag{28}$$

To get more accurate information on his position, Max decides to observe the data from the GPS receiver every 5 seconds and to compute the expected position from the recorded data.

Task: Open the Jupyter notebook `statistics.ipynb`, load the dataset `position.npy` and plot the position data.

Expected Value and Variance For any function $g(X)$ it holds the expected value rule

$$\mathbb{E}\{g(X)\} = \int_{-\infty}^{\infty} g(x) p_X(x)\,\mathrm{d}x. \tag{29}$$

If $g(X) = X$, then the expected value is

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x p_X(x)\,\mathrm{d}x. \tag{30}$$

The variance is defined as $\sigma_X^2 = \mathbb{E}\{(X - \mathbb{E}\{X\})^2\}$, in analogy with the discrete case.

Task: Compute the expectation value of Max's position. Plot it in a figure with the data. (Hint: The data provides only a finite number of outcomes, so only finitely many values x and y of the continuous random variables X and Y are nonzero. The outcomes are equally likely. Therefore you can interpret the expectation value as the mean of the data.)

Task: Then compute the variance and the standard deviation of the measurement noise of the x- and y-data. Useful Python commands are `np.var()` and `np.std()`.

Normal Distribution A random variable with a **Gaussian distribution**, described by the following PDF

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{31}$$

is said to be **normally distributed**, where $\mu = \mathbb{E}\{X\}$ and $\sigma^2 = \mathrm{var}(X)$. Normality is preserved by a linear transformation ($Y = aX + b$). The expected value and the variance of Y can be calculated to be $\mathbb{E}\{Y\} = a\mu + b$ and $\mathrm{var}(Y) = a^2\sigma^2$.

Task: Given a random variable X with a normal propability density $p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ with μ the mean and σ^2 the variance of X, please write down the PDF $p_Z(z)$ of the random variable $Z = 5X + 3$, using μ and σ.

3.1 Multiple Continuous Random Variables

The intuitive interpretation as well as the main properties of joint, marginal and conditional PDF parallel the discrete case. Two jointly continuous random variables X and Y, that are associated to a single experiment, can be described by a **joint PDF** $p_{X,Y}$. It has to be nonnegative and satisfy

$$P((X, Y) \in B) = \iint_{(x,y)\in B} p_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{32}$$

for every subset B of a two-dimensional space. Let B be the entire two-dimensional space, we obtain the normalization property

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1. \tag{33}$$

The **marginal PDF** can be computed from the joint PDF as follows

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x, y)\,\mathrm{d}y \tag{34}$$

and analogously for $p_Y(y)$.

Expected Value, Variance and Standard Deviation The expectation rule defined for continuous variables also holds for multiple random variables X and Y and some function $g(X, Y)$.

$$\mathbb{E}\{g(X,Y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{35}$$

Given a linear function $g(X, Y) = aX + bY$ the ecpected value results in $\mathbb{E}\{aX + bY\} = a\mathbb{E}\{X\} + b\mathbb{E}\{Y\}$.

Conditioning For any fixed y (event $Y = y$) with $p_Y(y) > 0$, the **conditional PDF** of X is defined by $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$. The probability of $(X \in A)$ given $Y = y$ can be calculated as

$$P(X \in A|Y = y) = \int_A p_{X|Y}(x|y)\mathrm{d}x. \tag{36}$$

The conditional version of the expected value rule $\mathbb{E}\{g(X)|Y = y\} = \int_{-\infty}^{\infty} g(x)p_{X|Y}(x|y)\,\mathrm{d}x$ remains valid.

Independence In full analogy with discrete random variables, the continuous random variables X and Y are independent, if $p_{X,Y}(x, y) = p_X(x)p_Y(y)$, for all x, y. If this holds, also the random variables $g(X)$ and $h(Y)$ are independent, for any functions g and h. For the expected value it holds that $\mathbb{E}\{XY\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}$ and, more generally, $\mathbb{E}\{g(X)h(Y)\} = \mathbb{E}\{g(X)\}\mathbb{E}\{h(Y)\}$. And for the variance of the sum of X and Y we have $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$.

3.2 Covariance and Correlation

Let X and Y be two continuous random variables with joint PDF $f_{X,Y}(x, y)$. In addition to the variances, one is also interested in the so-called **covariance** $\sigma(X, Y)$ between X and Y. It is a measure of how much the two random variables change together. The covariance is defined as the expectation of $(X - \mu_X)(Y - \mu_Y)$

$$\sigma(X,Y) = \mathbb{E}\{(X - \mu_X)(Y - \mu_Y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)p_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{37}$$

Task: Compute the covariance of the scalar valued variable $Y = aX + b$ with constants $a \in \mathbb{R}$ and $b \in \mathbb{R}$ and the random variable $X \in \mathbb{R}$ with mean $\mathbb{E}\{X\} = \mu_X$.

The physical unit of the covariance is the product of the units of the two random variables. To obtain a unitless quantity, one often divides the covariance by the two standard deviations, which yields the **correlation** $\rho(X, Y)$ between X and Y:

$$\rho(X,Y) = \frac{\sigma(X,Y)}{\sigma_X\sigma_Y}. \tag{38}$$

The correlation is a measure of linear relation between the random variables. One can show that, if two random variables are correlated, the correlation is always in the interval [-1, 1]. If the correlation is zero, one says that the two variables are uncorrelated. If two variables are statistically independent, then they are always uncorrelated, but the converse is not true in general.

Before Max finally goes home he would like to know, how much the noise n_x and n_y changes together. *Task: Please help him and compute the covariance and the correlation of the noise (random variables X and Y). In the Jupyter notebook you already find code to compute the covariance matrix. Is the noise correlated? Give the covariance matrix*

$$\mathrm{cov}(X,Y) = \begin{pmatrix} \sigma_X^2 & \sigma(X,Y) \\ \sigma(X,Y) & \sigma_Y^2 \end{pmatrix} = \dots$$

This matrix can be used to deform a unit circle to visualize the spread of the distribution of the random variables and the relation between them. As you have learned in the last tutorial, such an ellipse is called **confidence ellipsoid**.

Task: The covariance matrix to can be used to plot a confidence ellipsoid for the random variables X and Y describing the noise n_x and n_y, respectively. Execute the rest of the Jupyter notebook and have a look at the result.