EXERCISE TUTORIAL 1

Exercise 1.1

In spite of a small seismic activity, the risk of a large earthquake with significant consequences always exists. A large earthquake may occur once in 1000 years. In a given region, 300 years have passed without a significant earthquake occurring. How large is the probability that a significant earthquake will occur in this region, in the current year?

The probability has increased ☐

The probability has remained the same ☐

The probability has decreased ☐

Exercise 1.2

Considering an activity with only one event with potential consequences, the risk is that probability that this event will occur multiplied with the consequences given the event occurs.

Which of the following events is associated with the highest risk?

Event	**1**	**2**	**3**
Event probability	10%	1%	20%
Consequences	100 SFr	500 SFr	100 SFr
Risk			

Exercise 1.3

Following a number of different activities is given, which involve death as a possible consequence. Which is the riskiest one?

Crossing a bridge ☐

Smoking 20 cigarettes per day ☐

Traveling 100000 km by train ☐

Exercise 1.4

In a region, an investigation was carried out of the number of storks and births. It was figured out that when the number of storks is high then the amount of births is also high and vice versa. The statistics indicate that these events – the number of births and the number of storks are correlated. What do you think?

It has been proved statistically that the storks bring the children ☐

There is no direct connection between the two events so we cannot speak about correlation ☐

The statistical analysis has shown that the stork is a protected bird ☐

Exercise 1.5

A reinforced concrete bridge shows large cracks at mid span. As a result water can reach the reinforcement and eventually corrosion will initiate. What is more probable?

A failure of the bridge at mid span under the action of an abnormal load ☐

A failure of the bridge under the action of an abnormal load ☐

Exercise 1.6

Engineer Meier is "1000%" certain that the pedestrian bridge constructed by him is capable to withstand the load resulting from the bike racers taking part in the "Tour de Suisse". Which statement is correct?

Mr. Meier has made a wrong evaluation. A "200%" certainty would be enough ☐

If Mr. Meier made no miscalculations, he is right ☐

There is neither 1000% certainty nor absolute safety in civil engineering ☐

Exercise 1.7

In an Alp region, there are 25 very high summits. These are covered with snow over the entire year and each day there is the same probability of occurrence of an avalanche. This amounts to 1/40. How large is the probability in this region of at least two avalanches occurring at the same day? It is assumed that only one avalanche may occur on the same summit at the same day.

Exercise 1.8

A non destructive test method is carried out to determine whether the reinforcement of a structural component is corroded or not. From a number of past tests, it is known that the probability of the reinforcement being corroded is 1%. If the reinforcement is corroded, this will be indicated by the test. However there is also a 10% probability that the test will indicate that the reinforcement is corroded although this is not true (false indication).

How large is the probability that corrosion is present, if the non destructive test indicates corrosion?

EXERCISE TUTORIAL 2

Exercise 2.1

a. Which of the following expressions are meaningful in the way they are written?

i) $P\left[A \cup [B \cap C]\right]$ ii) $P[A] + P[B]$

iii) $P\left[\overline{A}\right] \cap P[B]$ iv) $\overline{P[B]}$

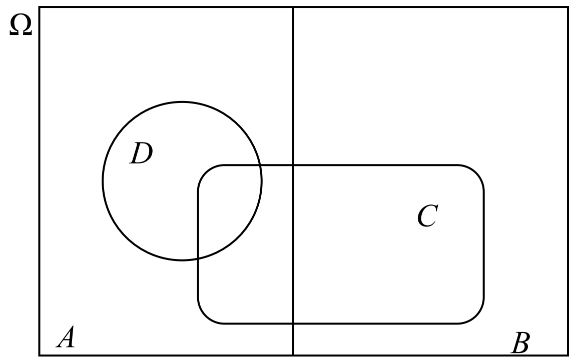
b. Assume *A, B* and *C* represent different events. Explain in words the meaning of the following expressions and what they represent in mathematical terms (i.e. numbers, vectors, functions, sets).

i) $A \cup B$ ii) $\overline{B} \cap C$ iii) $P[A]$

iv) $P\left[[A \cap B \cap C] \cup \left[\overline{A} \cap \overline{B} \cap \overline{C}\right]\right]$ v) $\varnothing$

c. Using the diagram provided below show the following events.

i) $C \cap D$ ii) $[D\backslash C] \cup [C \cap A]$ iii) $B \cup D$



Exercise 2.2:

We are throwing an ideal dice and considering the following events:

Case 1). *A*: "An even number comes"

B: "A number dividable by 3 comes"

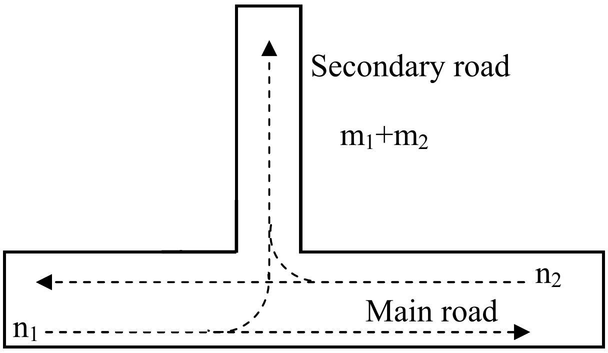
Case 2) *A*: "An even number comes"

B: "A prime number comes"

For each case, calculate the probability that both events (*A* and *B*) occur simultaneously.

Exercise 2.3:

The observation of the traffic flow at a crossing, see the following figure, shows that $n_1 = 50$ vehicles move on the main road in direction 1. From those $m_1 = 25$ vehicles turn to the secondary road. $n_2 = 200$ vehicles move on the main load in direction 2 and $m_2 = 40$ vehicles turn to the secondary road. How large is the probability that a vehicle moving on the main road will turn to the secondary road?



Exercise 2.4:

Measurements are to be carried out with measurement devices. Since a large number of devices is required, 20% of them will be provided by IAC (Institute for the Atmosphere and Climate) and 80% will be provided by IHW (Institute for Hydraulics and Water management). 5% of the devices provided by IAC do not fulfill the required accuracy, while 2% of the devices provided by IHW do not fulfill the required accuracy.

A student carried out a measurement using a device without knowing from which institute the device was provided. Thereby, she found the inaccuracy involved in the measurement. How large is the probability that the measurement was carried out with a device provided by IAC?

Exercise 2.5:

In exercise tutorial 1, exercise 1.8 was given as:

A non destructive test method is carried out to determine whether the reinforcement of a structural component is corroded or not. From a number of past tests, it is known that the probability of the reinforcement being corroded is 1%. If the reinforcement is corroded, this will be indicated by the test. However there is also a 10% probability that the test will indicate that the reinforcement is corroded although this is not true (false indication).

How large is the probability that corrosion is present, if the non destructive test indicates corrosion? Calculate the probability using the Bayes' theorem.

Exercise 2.6:

The failure of a building in the city of Tokyo may be caused by two independent events:

F_1: A big earthquake.

F_2: A strong typhoon.

The annual probabilities of occurrence of the above events are:

$P(F_1) = 0.04$

$P(F_2) = 0.08$

Calculate the annual failure probability for the building.

Exercise 2.7 (Group Exercise):

Due to the increasing demand on drinking and processing water, the groundwater discharge flow has to be discussed. The hazard of long-term ground-lowering is analysed, whereby it is assumed that the ground-lowering depends on the thickness of the clay layer, h. The thickness of the clay layer is classified in the following:

C_1: $0 \le h \le 20\,cm$, C_2: $20\,cm < h \le 40\,cm$, C_3: $40\,cm < h$

Based on experience a geologist estimates the following prior probabilities that the thickness of the clay layer at a site belongs to one of the above cases:

$P(C_1) = 0.2$, $P(C_2) = 0.47$, and $P(C_3) = 0.33$

A geo-electrical test may be useful to update the prior probability on the ground category, although the test result may not always be correct. From past experience, the probabilities of the correct/false indication of the test are known as are listed in the (uncompleted) table below:

Category of thickness of clay layer C_i	Indication of the category of the thickness of the clay layer		
	$I = C_1$	$I = C_2$	$I = C_3$
C_1	0.84		0.03
C_2	0	0.77	
C_3		0.02	0.89

Table 2.7.1: Probability of indication of the category of the thickness of the clay layer.

a. Complete the table.

b. A geo-electrical test was carried out and indicated as C_3 the thickness of the clay layer. What is the probability that the thickness of the clay layer belongs to C_1, C_2, C_3?

EXERCISE TUTORIAL 3

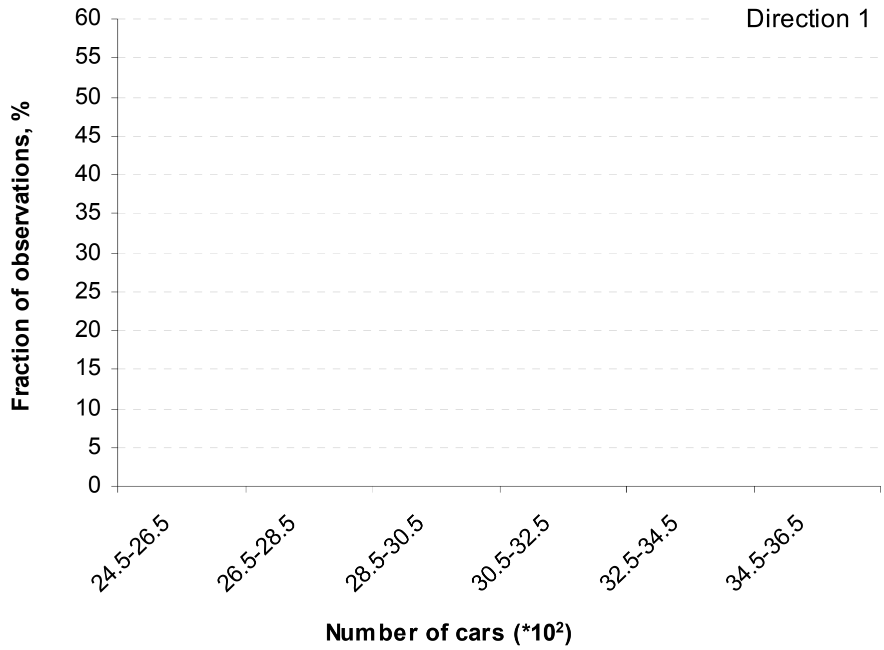
Exercise 3.1

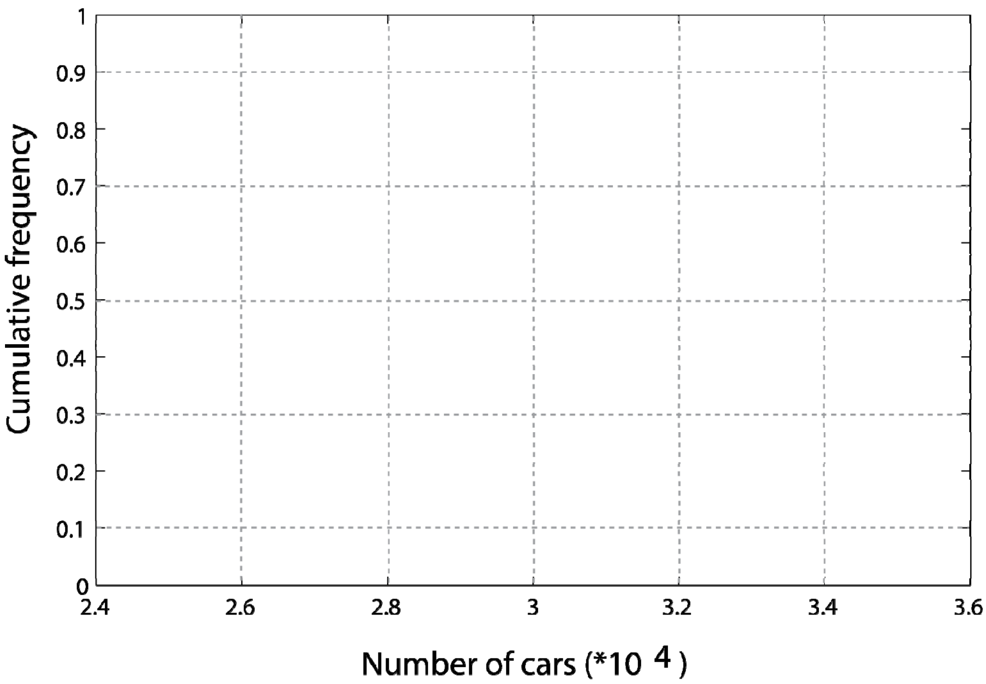
Two sets of data are provided, each of which represents the daily traffic flow in Rosengartenstrasse in Zurich during the month of April 2001 (Table 3.1.1). Direction 1 corresponds to driving towards Bucheggplatz, while direction 2 corresponds to driving towards Escher Wyss Platz.

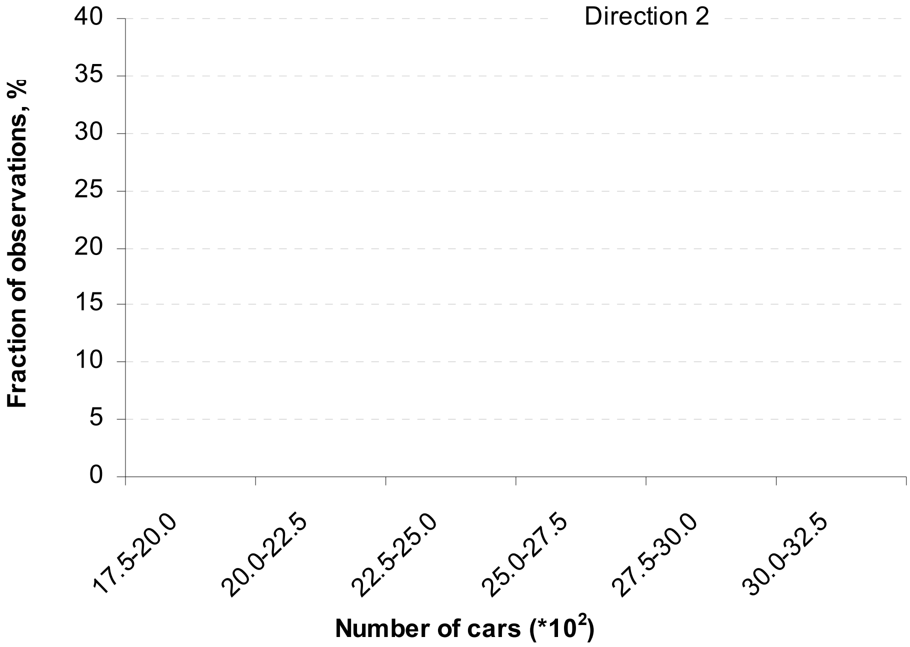
Date	**Direction 1**		**Direction 2**	
	unordered	**ordered**	**unordered**	**ordered**
01.04.2001	32618	24846	24609	17805
02.04.2001	33380	24862	29965	18123
03.04.2001	34007	25365	30629	19735
04.04.2001	33888	28252	30263	20903
05.04.2001	35237	29224	31405	21145
06.04.2001	35843	29976	31994	22762
07.04.2001	33197	30035	26846	22828
08.04.2001	30035	30613	22762	23141
09.04.2001	32158	32158	30366	24609
10.04.2001	33406	32472	29994	26525
11.04.2001	34576	32618	30958	26846
12.04.2001	34013	32962	30680	27746
13.04.2001	24846	33091	19735	28117
14.04.2001	28252	33197	21145	28858
15.04.2001	25365	33198	17805	28877
16.04.2001	24862	33245	18123	29080
17.04.2001	32472	33380	28117	29586
18.04.2001	33245	33406	28858	29965
19.04.2001	33788	33788	29080	29994
20.04.2001	34076	33888	30313	30263
21.04.2001	29976	33937	23141	30313
22.04.2001	29224	34007	20903	30366
23.04.2001	32962	34013	27746	30629
24.04.2001	33937	34076	29586	30680
25.04.2001	33198	34425	30788	30788
26.04.2001	34455	34455	31074	30958
27.04.2001	35852	34576	32384	31074
28.04.2001	33091	35237	26525	31405
29.04.2001	30613	35843	22828	31994
30.04.2001	34425	35852	28877	32384

Table 3.1.1: Daily traffic flow through Rosengartenstrasse, Zurich-April 2001.

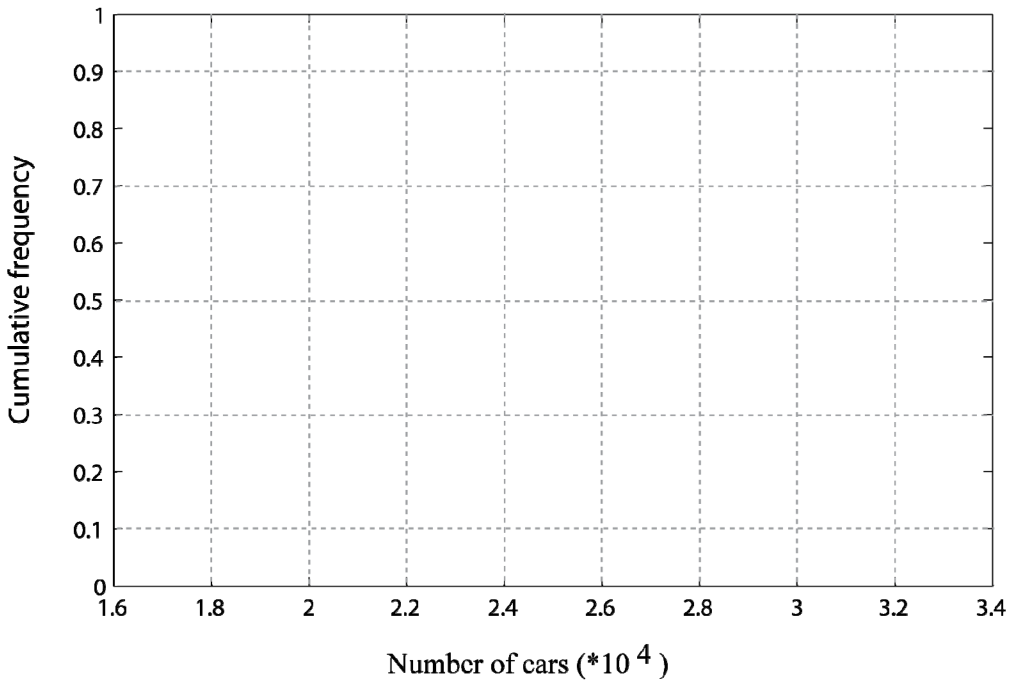
Provide frequency distributions and cumulative frequency distributions of the observed data. What is your first impression of the data? Try to make a comparison between the two directions.







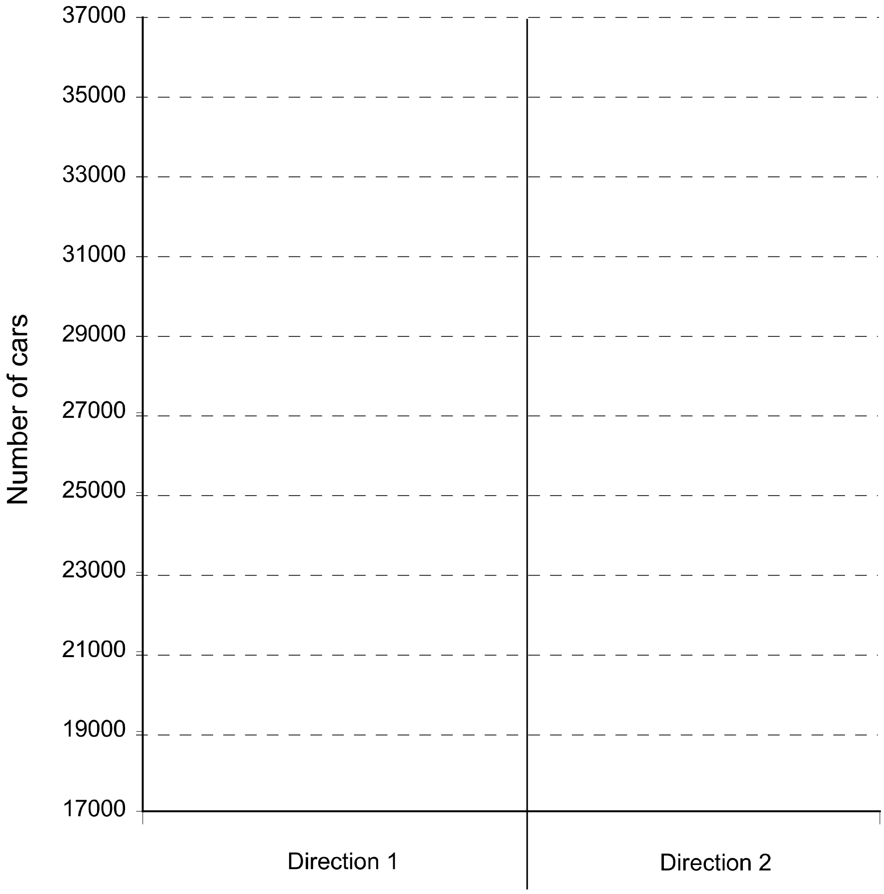
Direction 2
Fraction of observations, %
40
35
30
25
20
15
10
5
0
17.5-20.0
20.0-22.5
22.5-25.0
25.0-27.5
27.5-30.0
30.0-32.5
Number of cars (*10^2)



Cumulative frequency
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1.6
1.8
2
2.2
2.4
2.6
2.8
3
3.2
3.4
Number of cars (*10 4)

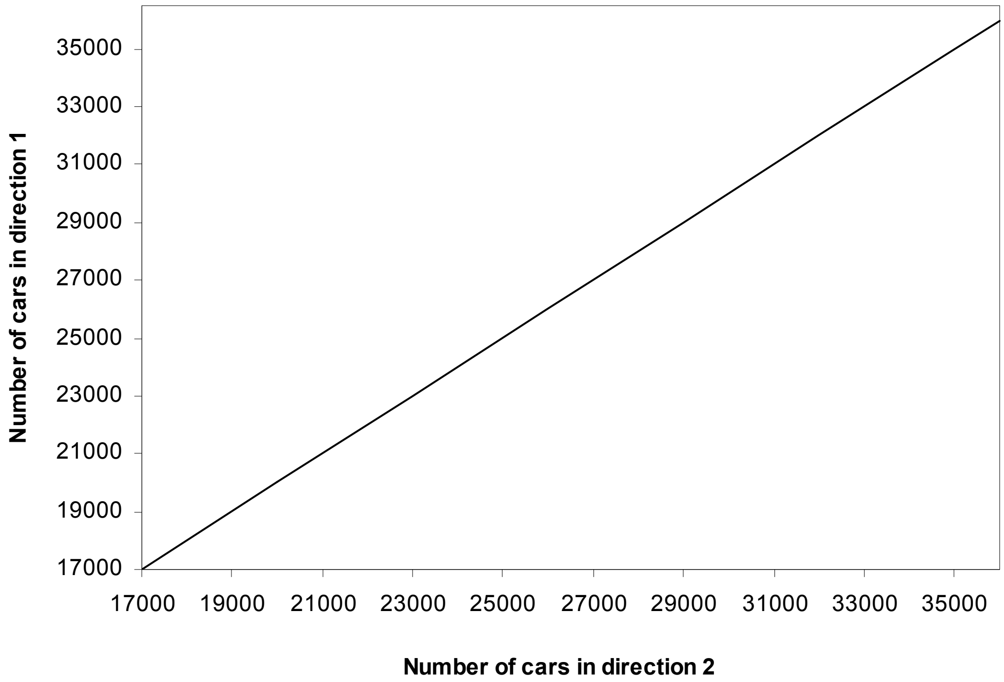
Exercise 3.2

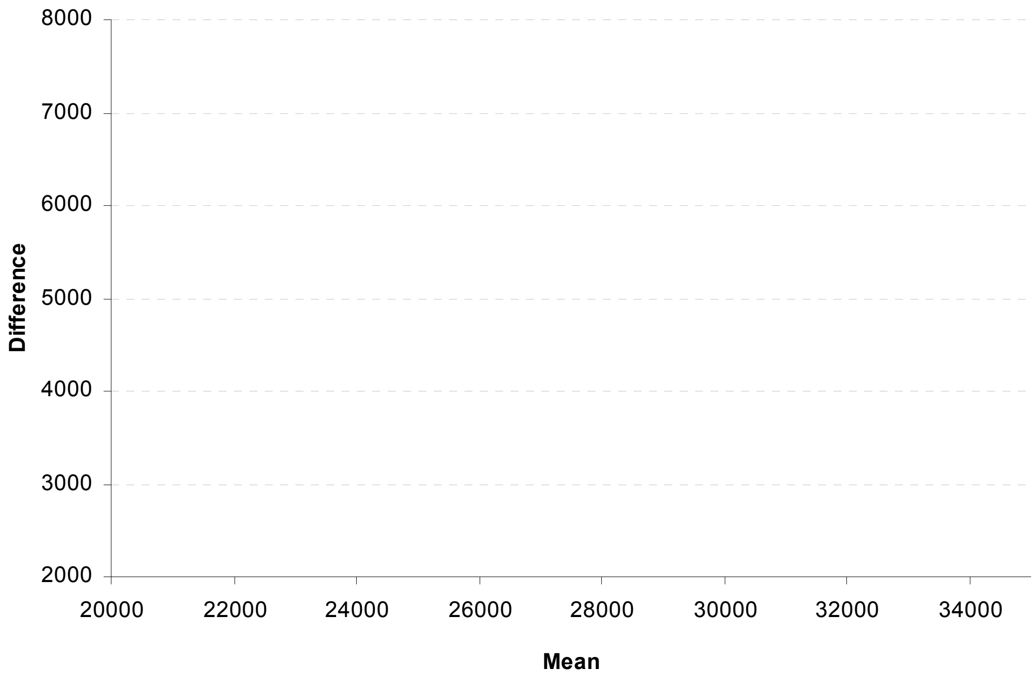
Use a Tukey box plot to provide a summary of the main features of the distribution of each data set of Table 3.1.1. Plot the Tukey box plots on the same graph so that you are able to compare these features. Do you observe any symmetry in the data sets?



Exercise 3.3

Make a Q-Q plot (Quantile-Quantile plot) to compare the two data sets of Table 3.1.1. What do you observe in regard to the traffic flows in directions 1 and 2? Provide an approximate value of the difference in the daily traffic flow between the two directions using a Tukey mean-difference plot.





Exercise 3.4 (Group exercise)

Resistivity measurements help to predict the possible corrosion of bridge structures. During a general bridge inspection the data shown in Table 3.4.1 were obtained from resistivity measurements along the two bridge lanes (direction 1 and 2):

a. Draw two box plots for the data provided in Table 3.4.1 (direction 1 and direction 2). Show the main features of the box plots and write their values next to the corresponding points on the diagrams. Plot also the outside values, if any.

b. Tukey box plot is a helpful tool for assessing the symmetry of data sets. Discuss the symmetry/skewness of the resistivity data for both lanes.

c. Choose a suitable number of intervals and plot the histogram for the resistivity data of direction 1.

Measurement No. (*i*)	Direction 1 Resistivity (kOhm)	Direction 2 Resistivity (kOhm)
1	20.2	3.8
2	20.4	5.6
3	22.1	6.5
4	23.8	7.1
5	24.3	7.9
6	24.7	8.2
7	25.3	9.1
8	25.6	9.3
9	25.7	9.6
10	25.9	9.8
11	26.2	10.3
12	26.7	10.9
13	26.9	11.1
14	27.3	11.7
15	27.6	12.2
16	27.6	12.6
17	27.8	12.9
18	27.9	13.8
19	28.3	13.9
20	28.7	14.5
21	28.9	15
22	28.9	15.4
23	29.3	17.1
24	29.4	17.8
25	29.9	23.4

Table 3.4.1 Resistivity measurements.

Exercise 3.5

The data sets in Table 3.5.1 show the number of newcomers to the university and the number of total students at the university. Estimate the correlation of these numbers using the following calculation sheet.

	Univ. A	Univ. B	Univ. C	Univ. D	Univ. E	Univ. F
Newcomer	3970	732	499	1300	3463	2643
Total students	24273	5883	2847	5358	23442	17076

Table 3.5.1: Number of newcomers to the university and number of total students at the university.

Calculation sheet

	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
A	3970	24273					
B	732	5883					
C	499	2847					
D	1300	5358					
E	3463	23442					
F	2643	17076					
Σ			-	-			
Σ / n			-	-			
$\sqrt{\Sigma / n}$	-	-	-	-			-

Exercise 3.6:

Table 3.6.1 shows observations taken at weather stations located at different heights. Plot the relations between the temperatures (T_{max} and T_{min}) and the heights of the weather stations in order to see the correlation between the temperatures and the heights. Then, calculate the correlation coefficient of the observed data – Height-T_{max} and Height-T_{min} .

Station	Höhe [m ü.M.]	T_{max} - Mai [°C]	T_{min} Mai [°C]
Adelboden	1355	12.20	2.30
Chateau-d'Oex	890	14.60	6.30
Grimsel	1950	13.40	4.70
Grindelwald	1040	14.00	4.30
Gstaad	1085	14.60	6.30
Guttannen	1055	13.40	5.10
Interlaken	574	16.40	8.30
Jungfraujoch	3572	9.20	-5.30
Meiringen	632	16.40	8.10
Mürren	1638	12.80	3.50

Table 3.6.1: Maximum and minimum temperatures observed at weather stations located at different heights through Switzerland.

Exercise 3.7:

An experiment was carried out to measure the tensile strength of wood. The critical strength is assumed equal to the load at which the wood sample broke. In Table 3.7.1 the measured strengths have been classified in intervals of 5 N/mm^2. The wood which is assumed to follow the same distribution as the tested wood is used for the construction of a building. The probability of failure or the reliability of the wood material can be estimated based on the measurement results. First draw the histogram and the cumulative frequency and:

a. Estimate the probability that the tensile strength of wood lies between 20-25 N/mm^2.

b. Estimate the probability of failure of the wood in the case where 25 N/mm^2 is loaded to the wood.

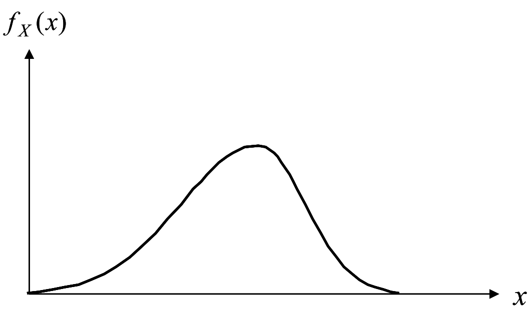
class upper limit [N/mm2]	class center [N/mm2]	absolute frequency	rel. frequency	cum. rel. frequency
5	2.5	1		
10	7.5	0		
15	12.5	0		
20	17.5	1		
25	22.5	9		
30	27.5	10		
35	32.5	22		
40	37.5	30		
45	42.5	33		
50	47.5	27		
55	52.5	9		
60	57.5	5		
65	62.5	0		
70	67.5	3		
75	72.5	1		

Table 3.7.1: Classified measured wood tensile strengths.

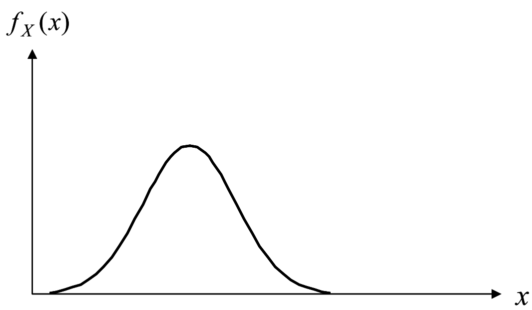
Exercise 3.8:

Identify and write down the skewness features (right-skewed or left-skewed) of the distributions shown in the following figure. Then, mark the median, the mean and the mode for each distribution.

_______________ Distribution



_______________ Distribution



EXERCISE TUTORIAL 4

Exercise 4.1:

The monthly expense X [CHF] for water consumption including sewage fee for a 2-persons household may be considered as a random variable with the following density function:

$$f_X(x) = \begin{cases} c \cdot x \cdot (15 - \frac{x}{4}) & \text{for } 0 \leq x \leq 60 \\ 0 & \text{otherwise} \end{cases}$$

a. Which value of c should be chosen?

b. Describe the cumulative distribution function $F_X(x)$ of the random variable X.

c. Which of the following four values, 30.00 CHF, 40.00 CHF, 50.00 CHF and 60.00 CHF does not exceed the 90%-quantile of the monthly expense?

d. How large is the mean monthly expense for water consumption including sewage fee for a 2-persons household?

Exercise 4.2:

The probability density function of a random variable is shown in Figure 4.2.1.
Let $a = 1$, $b = 2$, $c = 3$ and $d = 6$.

a. Describe the cumulative distribution function and the probability density function.

b. Define the parameter h and identify the mode.

c. Calculate the mean value.

d. Calculate the value of the median.

e. Obtain graphically the median. Discuss how the mean value may be evaluated graphically.

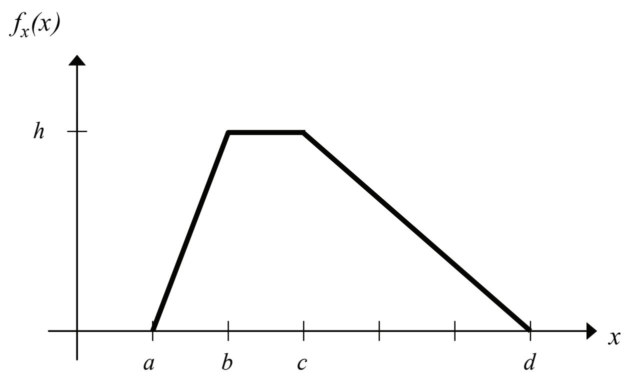


Figure 4.2.1: Probability density function.

Exercise 4.3 (Group exercise):

The probability density function of a random variable X is shown in Figure 4.3.1. In the interval [0, 4] the function is linear and in the interval [4, 12] the function is parabolic which is tangent to x-axis at point Q.

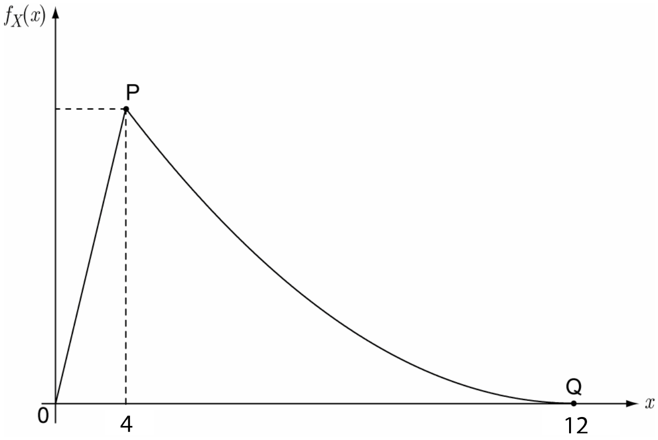


Figure 4.3.1: Probability density function.

a. Determine the coordinate of point P(x,y) and then describe the probability density function.

b. Describe and draw the cumulative distribution function of X with some characteristic numbers in the figure.

c. Calculate the mean value of X.

d. Calculate $P[X>4]$.

EXERCISE TUTORIAL 5

Exercise 5.1

The marginal probability density functions of a two dimensional random variable $Z = \left(X, Y\right)^T$ are defined as:

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_Y(y) = \begin{cases} \frac{3}{4} \cdot (2y - y^2) & \text{for } 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}.$$

The correlation coefficient ρ_{XY} between X and Y equals to $\sqrt{\frac{1}{3}}$.

a. Calculate the expected value of $6X - 4Y + 2$.

b. Calculate the covariance $Cov(6X; 4Y)$.

c. Calculate the variance of $6X - 4Y + 2$.

d. Calculate the expected value of $6X^2 - 4Y^2$.

Exercise 5.2

Wind loads are considered in the design of buildings in a region. In order to specify the design wind load, wind speed is measured over a long period with a reliable measuring device (it is herein called “accurate device”). The number of days when the observation of wind speed exceeds a given threshold (60 km/h) is counted in each year.

However, in previous years, the measurement has been undertaken with a less accurate device (it is herein called “less accurate device”). The correspondence between the measurement with the accurate device and the measurement with the less accurate device is of interest. Therefore, the joint probability of the measured wind speeds with both devices is established by calibration in the following next few years. Table 5.2.1 shows the joint probability of the numbers of the days when measured wind speed exceeds the threshold with the accurate device and with the less accurate device. N_U represents the number of the days when the wind speed measured with the accurate device exceeds the threshold, and N_G represents the number of the days when the wind speed measured with the less accurate device exceeds the threshold.

	$N_U = 0$	$N_U = 1$	$N_U = 2$	$N_U = 3$	$P(N_G)$
$N_G = 0$	0.2910	0.0600	0.0000	0.0000	0.3510
$N_G = 1$	0.0400	0.3580	0.0100	0.0000	0.4080
$N_G = 2$	0.0100	0.0250	0.1135	0.0300	0.1785
$N_G = 3$	0.0005	0.0015	0.0100	0.0505	0.0625
$P(N_U)$	0.3415	0.4445	0.1335	0.0805	$\sum = 1.00$

Table 5.2.1: Joint probability of N_G and N_U.

a. Calculate the probability that the number of days at which the wind speed, measured by each device, exceeds the threshold coincides.

b. Assume that the accurate device always measures the exact wind speed. What are the probabilities that the wind speed which exceeds the threshold prevails 0, 1, 2 and 3 time(s) in a year when the wind speed measured with the less accurate device exceeds the threshold twice?

Exercise 5.3 (Group exercise)

Highway bridges may require maintenance in their life time. The duration where no maintenance is required, T, is assumed exponentially distributed with the mean value of 10 years. The maintenance activity takes some time, which is represented by S. The time S is also assumed exponentially distributed with the mean value of 1/12 year.

a. Assuming that T and S are independent, obtain the distribution of the time between subsequent maintenance activities are initiated, Z, i.e., $Z = T + S$.

b. How large is the probability $P(Z \leq 5)$?

c. Assume that two bridges in a highway system are opened and the times until the bridges require the maintenance are represented by T_1 and T_2, which are independent identically distributed as T. How large is the probability that in the next 5 years no maintenance is required for the two bridges?

EXERCISE TUTORIAL 6

Exercise 6.1

Let random variables $\{X_i\}_{1\le i\le 50}$ be independent, identically Normal distributed with the mean value of 1 and the standard deviation of 2. The following are defined:

$$S_n = X_1 + X_2 + ... + X_n$$

and

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + ... + X_n) = \frac{S_n}{n}$$

where n is equal to 50.

a. Calculate the mean and the variance of S_n and $\bar{X}_n$.
b. Calculate $P(E[X_1]-1 \le X_1 \le E[X_1]+1)$.
c. Calculate $P(E[S_n]-1 \le S_n \le E[S_n]+1)$.
d. Calculate $P(E[\bar{X}_n]-1 \le \bar{X}_n \le E[\bar{X}_n]+1)$.

Exercise 6.2

A dike is designed to withstand a "1000-year return period flood". How large is the probability that water will overflow the dike in the following cases:

a. During a 10 year period, for first time in a given year?
b. During a 10 year period, twice?
c. Will not overflow during a 10 year period?
d. During a 10 year period, at most once?
e. During a 100 year period, 10 times?
f. During a 1000 year period, once or more often?

It is assumed that flood occurs once a year.

Exercise 6.3 (Group exercise)

An environmental planning engineering company obtains a project in return for a project proposal with the success rate of 27%.

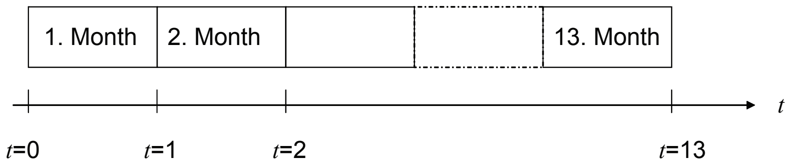
Assume that you have taken over this company and you need to make the business plan for the forthcoming years.

a. How large is the probability that the company will have at least one success after 12 project proposals?
b. How large is the probability that only the last of 10 project proposals is accepted?
c. How large is the probability that at most 2 out of 13 project proposals are accepted?

EXERCISE TUTORIAL 7

Exercise 7.1

The occurrence of rainfall in an area in a year may be described by a non-homogeneous Poisson process with the intensity, namely, the mean rate of occurrence of rainfall per unit time, $\lambda(t)$, where t is defined in the interval $[0;13]$ and describes the time in a monthly unit (i.e., 4 weeks).



The intensity is defined as follows:

$$\lambda(t) = \begin{cases} \dfrac{2t}{3} & (0 \le t \le 3) \\ 2 & (3 < t \le 7) \\ \dfrac{13-t}{3} & (7 < t \le 13) \end{cases}$$

a. Calculate the probability that in the first 5 months of a year, three or more rainfalls occur.

b. Calculate the probability that a rainfall occurs at most once during the 8th, 9th and 10th month and at most once during the last 3 months of a year.

Hint: For a non-homogeneous Poisson process, the intensity $\lambda(t)$ varies with time. The mean occurrence rate ν for any time interval (t_1, t_2) of the Poisson process can be described by:

$$\nu = \int_{t_1}^{t_2} \lambda(t)\, dt$$

Exercise 7.2

An earthquake hazard map is often represented in terms of peak ground acceleration and a return period of 475 years is adopted in the map in many countries.

a. Show that the event with a return period of 475 years corresponds to the event whose occurrence probability is 10% in 50 years, under the assumption that an event follows a homogeneous Poisson process.

b. What is the probability that an earthquake with a return period of 475 years will occur within the next 475 years?

Hint: Assume that the occurrence of the earthquake follows a homogeneous Poisson process.

Exercise 7.3

The annual maximum discharge X of a particular river is assumed to follow the Gumbel distribution with mean μ_X=10000 m^3/s and standard deviation σ_X =3.000 m^3/s.

a. Calculate the probability that the annual maximum discharge will exceed 15000 m^3/s.

b. What is the discharge that corresponds to a return period T of 100 years?

c. Find an expression for the cumulative distribution function of the river's maximum discharge over the 20 year lifetime of an anticipated flood-control project. Assume that the individual annual maxima are independent random variables.

d. What is the probability that the 20-year-maximum discharge will exceed 15000 m^3/s?

Hint: The Gumbel distribution function is expressed as:

$$-\infty < x < \infty$$

$$F_X(x) = \exp\left(-\exp\left(-\alpha(x-u)\right)\right)$$

$$\mu_X = u + \frac{0.577216}{\alpha}$$

$$\sigma_X = \frac{\pi}{\alpha\sqrt{6}}$$

μ_X – Mean

σ_X – Standard deviation

u – Parameter of the distribution

α – Parameter of the distribution

Exercise 7.4 (Group exercise)

Diesel engines are used, among others, for electrical power generation. The operational time T of a diesel engine until a breakdown, is assumed to follow an Exponential distribution with mean $\mu_T = 24$ months. Normally such an engine is inspected every 6 months and in case that a default is observed this is fully repaired. It is assumed herein that a default is a serious damage that leads to breakdown if the engine is not repaired.

a. Calculate the probability that such an engine will need repair before the first inspection.

b. Assume that the first inspection has been carried out and no repair was required. Calculate the probability that the diesel engine will operate normally until the next scheduled inspection.

c. Calculate the probability that the diesel engine will fail between the first and the second inspection.

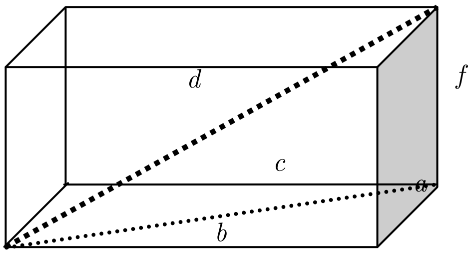
d. A nuclear power plant owns 6 such diesel engines. The operational lives t_1, t_2,…, t_6 of the diesel engines are assumed statistically independent. What is the probability that at most 1 engine will need repair at the first scheduled inspection?

e. It is a requirement that the probability of repair at each scheduled inspection is not more than 60%. The operational lives t_1, t_2,…, t_6 of the diesel engines are assumed statistically independent. What should be the inspection interval?

EXERCISE TUTORIAL 8

Exercise 8.1

Consider a cubic shaped box illustrated in the following figure. Measurements have been performed on a, b and f. It is assumed that the measurements involve the error ε that is assumed to be Normal distributed, unbiased and with standard deviation σ_ε.



a. Obtain the probability density function and cumulative distribution function of the error in d when this is assessed using the above measurements a, b and f.

b. If the assessment of c is made in the same way, how large is the probability that the error in c is larger than 2.4 σ_ε?

Exercise 8.2

In order to check the quality of the concrete production in a construction site, the compressive strength of the concrete produced were measured. Based on previous experience the compressive strength is assumed to follow the Normal distribution and its variance is assumed known and equal to 16.36 (MPa^2). Acceptance criterion of the concrete production is that the mean of the population of the produced concrete is equal to $30 \pm \Delta\ (MPa)$. 15 samples have been tested and the results are shown in Table 8.2.1. Should the quality of the concrete production be accepted? Test the hypothesis at significance levels of 10% and 1%.

Sample number (i)	Compressive Strength (MPa)	Sample number (i)	Compressive Strength (MPa)
1	24.4	9	30.3
2	26.5	10	39.7
3	27.8	11	38.4
4	29.2	12	33.3
5	39.2	13	33.5
6	37.8	14	28.1
7	35.1	15	34.6
8	30.8	-	-

Table 8.2.1: Measured concrete compressive strength.

Exercise 8.3

A student living in Baden has read in a report that the mean driving time from Baden to ETH Hoenggerberg is 23.7 minutes with a standard deviation of 3 minutes. In the following 13 days he made a note of his travel time and obtained the sample mean of 22.3 minutes.

Assuming that the travel time is Normal distributed, testify the correctness o of the report at a significance level of 5%. (Assume that the standard deviation is known as being equal to 3 minutes.)

Exercise 8.4

In a laboratory, 30 measurements are taken to control the water quality every day. Each measurement result is assumed to follow the Normal distribution with a mean of $\mu = 23\ ng/ml$ and a standard deviation of $\sigma = 4.3\ ng/ml$.

a. How large is the probability that a measurement result is less than $23 ng/ml$? How large is the probability that a measurement result lies in the interval $[19.5\ ng/ml; 20.5 ng/ml]$?

b. How large is the probability of the daily mean being less than $20 ng/ml$?

c. The lab bought a new instrument for measurements and decided to carry out a number of measurements in order to calibrate the new instrument with the old instrument. 15 measurements were carried out and the result was summarized as: the sample mean is $\bar{x} = 19 ng/ml$ and the sample standard deviation $s = 5\ ng/ml$. Test at the 5% significance level if the measurement result from the new instrument belongs to the same population of the measurement result from the old instrument.

Exercise 8.5

The weekly working hours in the building industry were decided to be reduced by 2 hours. On a construction site it is to be tested whether this new rule is applied or not since the labor union insists that the workers work the same hours as before the reduction. 9 workers were selected randomly and their weekly working hours were measured before (X) and after (Y) the reduction of the working hours. It is assumed that X and Y are Normal distributed. The variance of the weekly working hours before and after the reduction is assumed to be $\sigma_X^2 = \sigma_Y^2 = 9.5$ hours^2. The results are shown in the following table:

a. Can it be said, based on the data, that the mean of the working hours before the reduction is 40 hours/week at the 5% significance level.

b. Test the claim of the labor union at the 5% significance level.

Number of workers (i)	Working time before reduction (hours)	Working time after reduction (hours)
1	38	38
2	41	39
3	40	41
4	42	39
5	43	40
6	40	40
7	39	39
8	37	38
9	43	40

Table 8.5.1: Working hours per week.

Exercise 8.6

Table 8.3 provides a number N of data on the daily traffic flow in Rosengartenstrasse in Zurich.

Day (i)	Number of cars
1	3600
2	4500
3	5400
4	6500
5	7000
6	7500
7	8700
8	9000
9	9500

Table 8.6.1: Number of cars.

a. Construct the probability paper for the triangular distribution $f(x) = 2/10000^2 x$ for $0 < x < 10000$.

b. Check if the daily traffic flow is triangularly distributed with the help of the probability paper.

Hint: Use the following cumulative distribution function: $F(x_i^o) = \dfrac{i}{N+1}$

Exercise 8.7 (Group Exercise)

To rebuild a car park, the arrival times of cars were measured. The time interval between arriving cars are shown in Table 8.7.1.

a. Check graphically, if the time interval of car arrivals can be represented by an exponential distribution.

b. Calculate the mean value of the time interval of car arrivals. Under the assumption that the time interval is exponentially distributed, determine the mean value of the time interval from the graph produced in part (a.).

i	Time interval (seconds)
1	1.52
2	6.84
3	9.12
4	10.64
5	15.2
6	21.28
7	30.4
8	30.4
9	34.2
10	60.8
11	78.28
12	95.76

Table 8.7.1: Time interval of arrival.

Hint: Use the following cumulative distribution function: $F(x_i^o) = \dfrac{i}{N+1}$

EXERCISE TUTORIAL 9

Exercise 9.1:

In order to model the concrete compressive strength of a certain concrete production, 20 samples were measured and the result of the measurements is shown in Table 9.1.1. It is assumed that the population of the samples follows the Normal distribution $N(\mu, \sigma^2)$.

a. Describe the likelihood function.

b. Estimate the unknown parameters (μ,σ) with the maximum likelihood method.

c. Estimate the unknown parameters with the method of moment.

No. of sample (i)	Compressive strength (MPa)	No. of sample (i)	Compressive strength (MPa)
1	24.4	11	33.3
2	27.6	12	33.5
3	27.8	13	34.1
4	27.9	14	34.6
5	28.5	15	35.8
6	30.1	16	35.9
7	30.3	17	36.8
8	31.7	18	37.1
9	32.2	19	39.2
10	32.8	20	39.7

Table 9.1.1: Compressive strength of 20 samples.

Exercise 9.2:

What happens if the Exponential distribution is assumed instead of the Normal distribution in exercise 9.1?

a. Estimate the parameters of the Exponential distribution with the maximum likelihood method.

b. Draw the cumulative distribution function with the estimated parameter, together with the observed cumulative distribution.

Exercise 9.3 (Group Exercise):

It is known that the data shown in Table 9.3.1 are the realizations of a random variable X characterized by the cumulative distribution function $F_X(x) = x^{\alpha}$, $0 \le x \le 1$ with unknown parameter α. Estimate the parameter α in the following method.

a. Estimate the unknown parameter α with the *method of moment*.

b. Estimate the unknown parameter α with the *maximum likelihood method*.

c. Draw the cumulative distribution function with the estimated parameter in (b.) and the observed cumulative distribution.

No. of data	Realization
1	0.22
2	0.97
3	0.92
4	0.59
5	0.39
6	0.74
7	0.81
8	0.86
9	0.39
10	0.67

Table 9.3.1: Realizations of a random variable.

EXERCISE TUTORIAL 10

Exercise 10.1:

A dice is suspected to be asymmetric, resulting in the inhomogeneity of probability that each side of a dice comes out. In order to judge this suspicion statistically, 60 trials were made and the result is shown in Table 10.1.1.

a. Draw the relative frequency histogram, and compare with the uniform probability mass function under the assumption that the dice is symmetric.

b. What is the probability that each side of the dice comes out 10 times respectively in 60 trials when the dice is symmetric?

c. Set the symmetry property of the dice as the null hypothesis and test the hypothesis with the χ^2 test at the 5% significance level.

Side of the dice	No. of realizations
1	7
2	12
3	11
4	10
5	8
6	12
Sum:	60

Table 10.1.1: Result of trials.

Exercise 10.2:

For the estimation of the concrete compressive strength of a certain concrete production, 20 samples were measured and the result is shown in Table 10.2.1. It is assumed that the concrete compressive strength follows the Normal distribution.

a. Estimate the unknown parameters of the distribution with the method of moments.

b. Test the goodness of fit for the distribution with estimated parameters with the χ^2 test at the 5% significance level. Adopt the intervals in Table 10.2.2.

No. of sample (i)	Compressive strength (MPa)	No. of sample (i)	Compressive strength (MPa)
1	24.4	11	33.3
2	27.6	12	33.5
3	27.8	13	34.1
4	27.9	14	34.6
5	28.5	15	35.8
6	30.1	16	35.9
7	30.3	17	36.8
8	31.7	18	37.1
9	32.2	19	39.2
10	32.8	20	39.7

Table 10.2.1: Compressive strength of 20 samples.

Interval	$N_{o,j}$	$p(x_j)$	$N_{p,j} = np(x_j)$	ε_m^2
-30				
30-33				
33-36				
36-				
Sum				

Table 10.2.2: Calculation sheet for the χ^2 - goodness of fit test.

Exercise 10.3:

In order to understand the role of the number of samples and the significance level in the hypothesis testing, the following experiment is made.

50 realizations are simulated which follow the Normal distribution with the mean of 8 and the standard deviation of 1, see Table 10.3.1. The null hypothesis is that these realizations follow the Normal distribution with a mean of 7.5 and a standard deviation of 1 – therefore, it may be expected that the null hypothesis is rejected. Test the null hypothesis with the following given conditions with the Kolmogorov-Smirnov test.

a. Test the null hypothesis at the 1% significance level, using only the first 20 realizations.

b. Test the null hypothesis at the 1% significance level, using 50 realizations.

c. Test the null hypothesis at the 5% significance level, using only the first 20 realizations.

d. Test the null hypothesis at the 5% significance level, using 50 realizations.

No.	Realization	No.	Realization	No	Realization	No.	Realization	No.	Realization
1	7.57	11	7.81	21	8.29	31	7.60	41	6.40
2	6.33	12	8.73	22	6.66	32	8.69	42	8.26
3	8.13	13	7.41	23	8.71	33	8.82	43	6.94
4	8.29	14	10.18	24	9.62	34	8.71	44	9.42
5	6.85	15	7.86	25	7.31	35	9.29	45	7.19
6	9.19	16	8.11	26	8.86	36	8.67	46	8.53
7	9.19	17	9.07	27	9.25	37	9.19	47	8.22
8	7.96	18	8.06	28	6.41	38	6.80	48	7.08
9	8.33	19	7.90	29	6.56	39	7.98	49	5.83
10	8.17	20	7.17	30	8.57	40	7.84	50	7.94

Table 10.3.1: Realizations.

Exercise 10.4:

The strength of 30 wood samples has been measured and the results are shown in Table 10.4.1. The strength is assumed to follow an exponential distribution.

a. Estimate the parameters of the exponential distribution using the method of moments.

b. Draw the cumulative distribution function with the estimated parameters, together with the observed cumulative distribution.

c. Test the goodness of fit for the exponential distribution with the χ^2 test at the 10% significance level. Adopt the intervals in Table 10.4.2.

d. Test the goodness of fit for the exponential distribution with the Kolmogorov-Smirnov test at the 10% siginificance level. The parameter of the exponential distribution is assumed as $\lambda = 0.04$.

No.	Strength (MPa)	No.	Strength (MPa)	No.	Strength (MPa)
1	12.8	11	23.4	21	29.3
2	16.3	12	26.8	22	29.5
3	16.6	13	26.9	23	30.3
4	16.9	14	27	24	32.1
5	17.2	15	27.1	25	32.3
6	17.9	16	27.2	26	33.5
7	19.5	17	27.2	27	33.9
8	21.9	18	27.5	28	35.6
9	22.3	19	27.9	29	39.2
10	22.5	20	28.3	30	43.5

Table 10.4.1: Strength of wood samples.

Interval	$N_{o,j}$	$p(x_j)$	$N_{p,j} = np(x_j)$	ε_m^2
-20				
20-25				
25-30				
30-				
Sum				

Table 10.4.2: Calculation sheet for the χ^2 - goodness of fit test.

Exercise 10.5 (Group Exercise):

The strength of 30 wood samples has been measured and the result is shown in Table 10.5. The strength is assumed to follow a Weibull distribution (with zero lower bound).

a. Estimate the parameters of the Weibull distribution using the method of moments.

b. Draw the cumulative distribution function with the estimated parameters, together with the observed cumulative distribution.

c. Test the goodness of fit for the Weibull distribution with the test at the 10% significance level. Adopt the intervals in Table 10.4.2.

d. Test the goodness of fit for the Weibull distribution with the Kolmogorov-Smirnov test at the 10% siginificance level. The parameters of the Weibull distribution are given by $k = 4.00$ and $u = 28$ from past experience.

Hint: Cumulative distribution function of the Weibull distribution is written as:

$$F_X(x) = 1 - \exp\left\{ -\left(\frac{x}{u}\right)^k \right\}, x > 0$$

The mean and standard deviation of Weibull distribution are given respectively as:

$$\mu = u\Gamma\left(1 + \frac{1}{k}\right), \qquad \sigma = u\sqrt{\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)}$$

EXERCISE TUTORIAL 11

Exercise 11.1 (Group Exercise):

After heavy snowfall, you need to decide whether to clean up a roof from the snow or not. There is though also the possibility to carry out firstly a test to check the melting status of the snow on the roof. The test basically provides an indication, I, of the wetness of the snow on the roof.

Therefore the decision problem is focused on either clean up the roof, do not clean up the roof or carry out first a test. In the following some information is provided to enable in the decision making.

The cost associated with carrying out a test is equal to 1000 CHF. The clean up of the roof can be made from the local fire department. This option is associated with a cost equal to 4000 CHF. In the case of collapse of the roof to the snow load the associated cost is equal to 1.000.000 CHF.

The probability of collapse of the roof has been estimated using First Order Reliability Methods (FORM). If the snow is dry, SD, the probability of collapse is: $P_f(SD) = 10^{-3}$. If the snow is wet, SW, the probability of collapse is: $P_f(SW) = 6.2 \cdot 10^{-3}$. In case where there is no snow, SN, on the roof the probability of collapse is equal to: $P_f(SN) = 5 \cdot 10^{-4}$.

The drawback of carrying out a test is that the test can provide a correct indication of the wetness, W, of the snow only in 75% of situations, $P(I_W | W) = 0.75$. You have a lot of experience in the assessment of snow wetness and based on this experience you believe that the probability that the snow on the roof is wet is equal to 0.6, $P(SW) = 0.6$.

The decision tree is provided in Figure 11.1.1. Estimate the branch probabilities (where necessary since some have been already provided) and fill in the boxes the costs associated with the different events. Which decision is the most beneficial one in terms of cost?

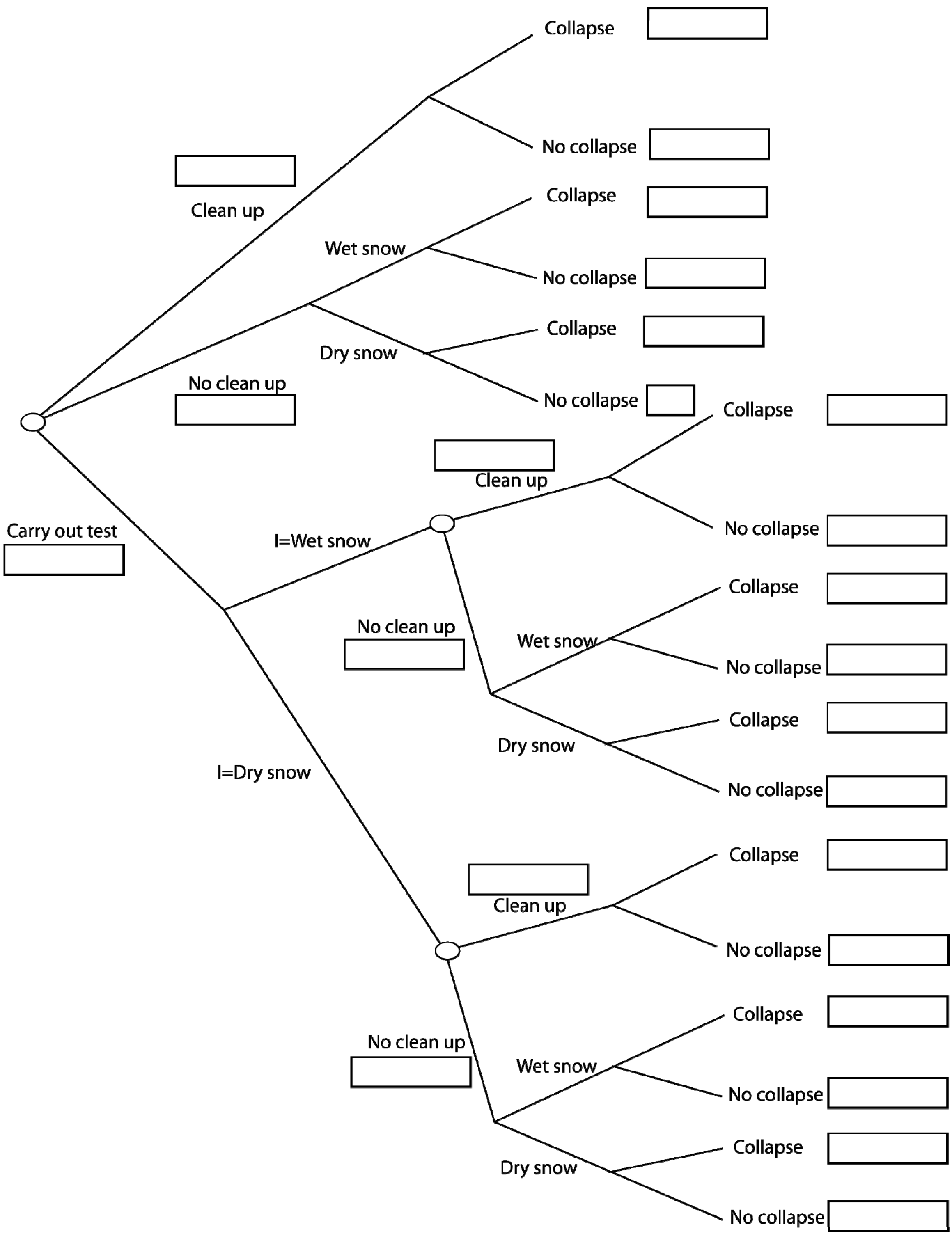


Figure 11.1.1: Event tree.

Exercise 11.2:

An oil producer needs to make the decision of opening up, a_1, or not, a_2, a borehole at a location where there is an old oil well. However, he is not sure of whether such an action is advantageous or not. The well maybe dry, D, it may be contaminated, C, (i.e. will have oil but and other substances also) or it may contain oil, O. Those events have a –priori probabilities equal to 50%, 30% and 20% respectively.

The benefit associated with the decision of opening up the borehole are estimated as the difference between the cost of opening the borehole and the cost benefit from the condition of the well (i.e. the income, if any, gained from the oil in the well). If the well is dry the benefit associated with opening up the borehole is -90.000 CHF. If the well is contaminated the benefit is equal to 50000 CHF and if the well has oil the benefit is equal to 170000 CHF.

a. Carry out a prior decision analysis (using an appropriate event tree) and determine whether it is beneficial to open up the borehole.

b. In order to receive some additional information regarding the state of the well the engineer can carry out a probe test that will cost 10000 CHF. The test can provide the following indications about the state of the well, WS :

I_D : the well is dry

I_C : the well is contaminated

I_O : the well has oil

The information provided from the test is uncertain. The probabilities associated with a correct or false indication from the test about the state of the well given the true state of the well are given in the following table, $P(I_{WS}|WS)$

	State of the well		
Indication	D: dry	C: contaminated	O: oil
I_D	0.6	0.3	0.1
I_C	0.1	0.3	0.5
I_O	0.3	0.4	0.4

Table 11.2.1: Likelihood of the true capacity of the well given the probe test results.

Construct an appropriate decision tree and estimate whether it is beneficial to carry out a test (pre-posterior analysis).

Exercise 11.3:

A company is about to invest in the construction of a factory in a desert area. In order to maintain the machines and the workers, $100kl$ of water is needed in a day. There are two options: one option is to develop a well locally at the site (action A_1) while the other option is to construct a pipeline to get the water from another location (action A_2). The development of a well locally costs 10 million CHF whereas the construction of a pipeline it costs 100 million CHF. The problem is that it is not certain that the construction of a well locally will provide with sufficient water. In case that the construction of a well locally will not provide sufficient water, a pipeline will need to be constructed. Answer the questions following the instructions.

a. (Prior analysis) Based on the experience from similar geological conditions, it is assumed that the probability that a local well will be able to supply enough water is 40%. With this information, which action should the company support A_1 or A_2?

The capacity of the well can be diagnosed by developing a small test well. It costs 1 million CHF to develop the small test well. However, the information obtained from the test well is only indicative, which means it cannot be known for sure whether there is sufficient water or not. The likelihood of the actual water capacity of the local well given a test result is described in Table 11.3.1:

	True capacity of the local well	
Indicators	θ_1 :less than 100 kl	θ_2 :larger than 100 kl
I_1 : Capacity >105 kl	0.1	0.8
I_2 : 95< Capacity <105	0.2	0.1
I_3 : Capacity < 95	0.7	0.1

Table 11.3.1: Likelihood of the true capacity of the well given the test results.

b. (Posterior analysis) Now a test well has been made and the indication of I_2 is obtained. Should the well be developed?

c. (Pre-posterior analysis) Before of all, should have a test well been developed?