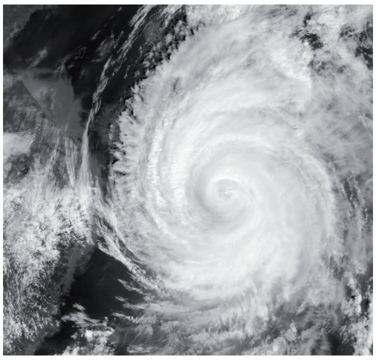
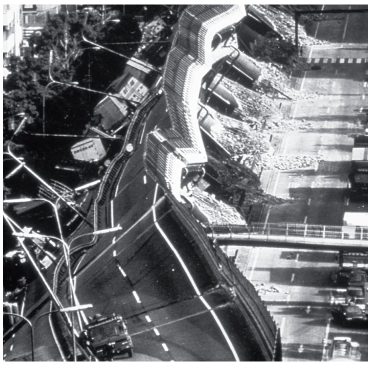
Statistics and Probability Theory

Solutions of the Tutorial Exercises






Lecture Notes

Prof. Dr. M. H. Faber

SS 2007

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

EXERCISE TUTORIAL 1- SOLUTION:

Exercise 1.1 - Solution

The probability remains the same because the probability is specified in the same way for the total period.

Exercise 1.2 - Solution

Based on the definition of risk it is:

$R_1 = P_1 C_1 = 0.1 \cdot 100 = 10$

and in the same way

$R_2 = 5$ and $R_3 = 20$

Therefore event 3 is associated with the higher risk.

Exercise 1.3 - Solution

Mean death risk Per year and per 100000 persons	
Overall	
110 100 300 800 2000 5000	25 years old 35 years old 45 years old 55 years old 65 years old 75 years old
Work risks	
100 90 50 15 10 5	Wood cutting, wood transport Forest enterprise Worker on a construction site Chemistry industry Mechanical factory Office work
Miscellaneous risks	
400 300 150 100 20 10 10 5 3 1 1 1 0.2 0.1	20 cigarettes a day 1 bottle of wine per day Motor bicycling Wing aircraft as a hobby Driving a car (20-24 years) Pedestrian, Houseworker 10000 km by personal car Hiking 10000 km on the highway Plane crash per flight Fire in a building 10000 km by train Death due to earthquake Death due to lightning

Table 1.3.1: Mean death risk (source: *Sicherheit und Zuverlässigkeit im Bauwesen,* Schneider J.).

Based on Table 1.1 the activity with the higher risk is: smoking 20 cigarettes a day.

Exercise 1.4- Solution

In the analysis of data, correlations can be determined by different measurements. However in the mentioned measurements there is no direct connection between the number of storks and the number of births.

Exercise 1.5- Solution

Events:

E_1: A failure of the bridge at mid span under the action of an abnormal load

E_2: A failure of the bridge under the action of an abnormal load

Event E_1 is a subset of event E_2. Therefore a failure of the bridge is more probable as a result of the action of an abnormal load.

Exercise 1.6- Solution

The probability of an event occurring lies in general between 0 (event will not occur) and 1 (event will definitely occur). Therefore it is not possible to speak about 1000% of safety.

Exercise 1.7- Solution

Probability of:

no Avalanche occurs

1 Avalanche occurs

2 Avalanches occur

…

25 Avalanches occur

(2 … 25 Avalanches occur: to be found)

…

+ ________________________________

The sum of all probabilities is 1.

Therefore the probabilities that no avalanche occurs and that only 1 avalanche occurs need to be determined and be subtracted from the sum of all probabilities.

The probability of occurrence of an avalanche at one summit is:

$$P_j(avalanche) = \frac{1}{40} = 0.025 \text{ ,where } j = 1,2,\ldots.n$$

The probability that no avalanche occurs at one summit is:

$P_j(no\ \ avalanche)= 1-\frac{1}{40}=0.975$, where $j=1,2,\ldots.n$

The probability of an avalanche only at one summit and at no other summit is calculated as:

$$P(avalance\, only\ on\ j^{th}\ summit) = P(avalanche)\cdot(1-P(avalanche))^{24} = 0.025\cdot 0.975^{24} = 0.0136$$

Then the probability that no avalanche occurs at any summit (event A) is calculated as:

$$P(A) = (1-P_1(avalanche))\cdot (1-P_2(avalanche))\cdots(1-P_{25}(avalanche)) = 0.975^{25} = 0.531$$

Therefore, the probability that only one avalanche occurs in 25 summits (event B) is

$$P(B) = \sum_{j=1}^{25} P(avalance\, only\ on\ j^{th}\ summit) = 25\cdot 0.0136 = 0.340$$

The probability that at least two avalanches occur (event C) can be calculated as:

$$P(C) = 1 - P(A) - P(B) = 1 - 0.531 - 0.340 = 0.129$$

Exercise 1.8 - Solution

Let us assume that we have 1000 reinforcement bars (rebars). According to tests, 1% of these rebars are corroded; that is 10 of the rebars are corroded while 990 are not corroded. Also it is known that the test method will indicate all the corroded rebars. Therefore the test will definitely detect 10 corroded rebars. However there is a 10% probability that the test will indicate that the rebars are corroded although they are not, i.e. there may be an observation of 99 corroded rebars while they are not actually corroded!

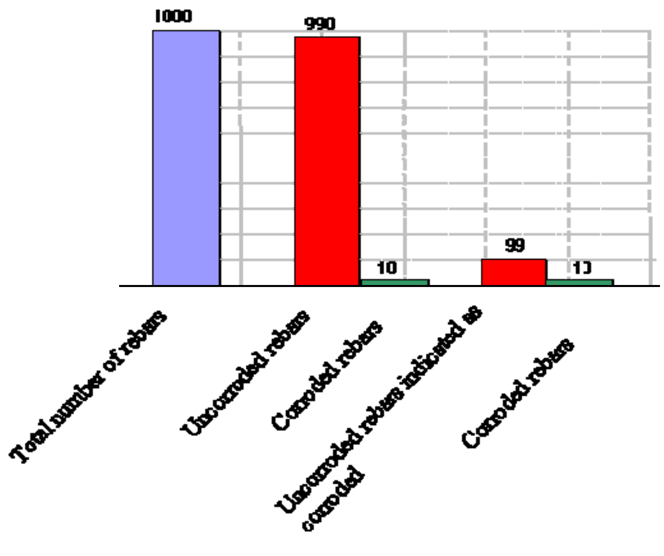


Figure 1.8.1: Bar diagram.

From the 1000 rebars, 99+10=109 are indicated as being corroded. However, only 10 rebars are really corroded. Therefore the probability that corrosion is present provided that the test indicates corrosion is: $P(corrosion) = \frac{10}{10+99} = 0.0917$

EXERCISE TUTORIAL 2- SOLUTION:

Exercise 2.1 - Solution

a. i) and ii) are sensible, but iii) and iv) are not. Probabilities cannot be separated and complementary events describe quantities and not the probability.

b.

i) Event A and/or B occur; mathematical: Quantity.

ii) Event B does not occur and event C occurs; mathematical: Quantity.

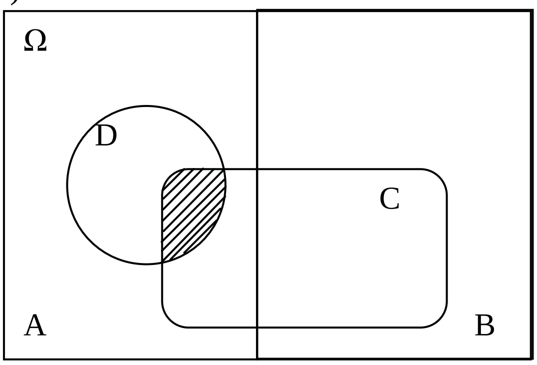
iii) The probability of event A to occur; mathematical: Number between 0 and 1.

iv) The probability that events A and B and C will occur and/or the complementary events will not occur; mathematical: Number between 0 and 1.

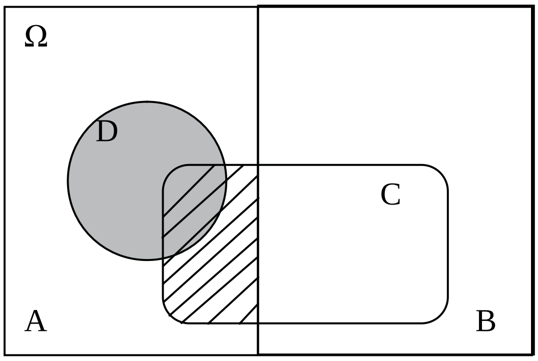
v) An empty sample space; mathematical: Empty set.

c.

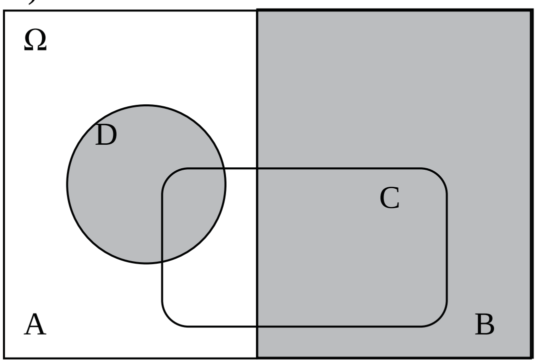
i)



ii)



iii)



Exercise 2.2 - Solution

Case 1) It is:

$A = \{2,4,6\},\ B = \{3,6\},\ A \cap B = \{6\}$

$$P(A)=\frac{1}{2} \qquad P(B)=\frac{1}{3}$$

$$P(A\cap B)=\frac{1}{6}=P(A)\cdot P(B)$$ Events A and B are independent.

Case 2) It is:

$A=\{2,4,6\},\ B=\{2,3,5\},\ A\ \cap\ B=\{2\}$

$$P(A)=\frac{1}{2} \qquad P(C)=\frac{1}{2}$$

$$P(A\cap C)=\frac{1}{6}\neq P(A)\cdot P(C)$$ Events A and C are not independent.

Exercise 2.3 - Solution

The probability that a vehicle is moving in one direction is:

Direction 1-Event A_1: $$P(A_1)=\frac{n_1}{(n_1+n_2)}=\frac{50}{(50+200)}=0.2$$

Direction 2-Event A_2: $$P(A_2)=\frac{n_2}{(n_1+n_2)}=\frac{200}{(50+200)}=0.8$$

It can be seen that there is a higher probability of a vehicle moving in direction 2.

The probability that a vehicle will turn to the secondary road is:

Vehicles from direction 1 $$P(B\mid A_1)=\frac{m_1}{n_1}=\frac{25}{50}=0.5$$

Vehicles from direction 2 $$P(B\mid A_2)=\frac{m_2}{n_2}=\frac{40}{200}=0.2$$

The probability that a vehicle from either direction will turn to the secondary road is:

$$P(B)=P(A_1)\cdot P(B\mid A_1)+P(A_2)\cdot P(B\mid A_2)=0.2\cdot 0.5+0.8\cdot 0.2=0.26$$

Exercise 2.4 - Solution

A = Equipment of Institute A (IAC)

B = Equipment of Institute B (IHW)

D = Inaccurate equipment

Probability that the device was provided from one or the other institute:

$P(A)=0.2$ $P(B)=0.8$

Probability of using an inaccurate device:

$P(D \mid A) = 0.05 \qquad P(D \mid B) = 0.02$

Probability of measuring with a device from an institute given there are some inaccurate devices:

$$P(A \mid D) = \frac{P(A) \cdot P(D \mid A)}{P(A) \cdot P(D \mid A) + P(B) \cdot P(D \mid B)} = \frac{0.2 \cdot 0.05}{0.2 \cdot 0.05 + 0.8 \cdot 0.02} = 0.385$$

Exercise 2.5 - Solution

The following events are identified:

K: Reinforcement is corroded

I: Indication of corrosion

$P(K) = 0.01 \qquad P(\bar{K}) = 0.99$

$P(I \mid K) = 1.00 \qquad P(\bar{I} \mid K) = 0.00$

$P(I \mid \bar{K}) = 0.10 \qquad P(\bar{I} \mid \bar{K}) = 0.90$

We can write this into a table:

True state	Indication	
	K	$\bar{K}$
K	1.00	0
$\bar{K}$	0.10	0.90

$$P(K \mid I) = \frac{P(I \mid K) \cdot P(K)}{P(I \mid K) \cdot P(K) + P(I \mid \bar{K}) \cdot P(\bar{K})} = \frac{1.00 \cdot 0.01}{1.00 \cdot 0.01 + 0.10 \cdot 0.99} = 0.0917$$

Exercise 2.6 - Solution

The annual failure probability is calculated as:

$$P(F_1 \cup F_2) = P(F_1) + P(F_2) - P(F_1 \cap F_2) = 0.04 + 0.08 - 0.04 \times 0.08 = 0.1168\,.$$

EXERCISE TUTORIAL 3- SOLUTION:

Exercise 3.1 – Solution

In order to plot the required graphical representations the data ordered in ascending form are used (Table 3.1.1). Based on the rule suggested by Benjamin and Cornell (see lecture notes Equation C.8) the number of intervals to be used is 6. Table 3.1.2 shows the summary of the observed traffic flow data.

The maximum observation in direction 2 is 35852 and the minimum observation is 24846. The length of the interval may be thus chosen as:

$$\frac{35852-24846}{6}=1834\approx 2000.$$

The intervals are chosen as:

$24500-26500$, Midpoint=25500

$26500-28500$, Midpoint=27500

$28500-30500$, Midpoint=29500

$30500-32500$, Midpoint=31500

$32500-34500$, Midpoint=33500

$34500-36500$, Midpoint=35500

	Interval (Number of cars $*10^3$)	Interval Midpoint (Number of cars $*10^3$)	Number of observations	Frequency %	Cumulative frequency
Direction 1	24.5-26.5	25.5	3	10.000	0.100
	26.5-28.5	27.5	1	3.333	0.133
	28.5-30.5	29.5	3	10.000	0.233
	30.5-32.5	31.5	3	10.000	0.333
	32.5-34.5	33.5	16	53.333	0.867
	34.5-36.5	35.5	4	13.333	1.000
	Interval (Number of cars $*10^3$)	**Interval Midpoint (Number of cars $*10^3$)**	**Number of observations**	**Frequency %**	**Cumulative frequency**
Direction 2	17.5-20.0	18.75	3	10.000	0.100
	20.0-22.5	21.25	2	6.667	0.167
	22.5-25.0	23.75	4	13.333	0.300
	25.0-27.5	26.25	2	6.667	0.367
	27.5-30.0	28.75	8	26.667	0.633
	30.0-32.5	31.25	11	36.667	1.000

Table 3.1.2 Summary of the observed traffic flow.

Figures 3.1.1 and 3.1.2 show the frequency distributions and cumulative distribution diagrams for the traffic flow data. Even though one could use the values of the cumulative frequency of Table 3.1.2 to make the cumulative frequency plots, the quantiles of the observations (Table 3.1.3) are instead used. That is, as mentioned in the script (section C.3) due to the fact that the observations are known. The cumulative frequencies in Table 3.1.2 would be used if only the intervals of the observations where known. However try to plot the cumulative frequencies as an exercise for yourself using the interval and the cumulative frequencies of Table 3.1.2.

Number (i)	**Direction 1**		**Direction 2**	
	ordered	**Quantile =** $\frac{i}{n+1}$	**ordered**	**Quantile =** $\frac{i}{n+1}$
1	24846	0.0323	17805	0.0323
2	24862	0.0645	18123	0.0645
3	25365	0.0968	19735	0.0968
4	28252	0.1290	20903	0.1290
5	29224	0.1613	21145	0.1613
6	29976	0.1935	22762	0.1935
7	30035	0.2258	22828	0.2258
8	30613	0.2581	23141	0.2581
9	32158	0.2903	24609	0.2903
10	32472	0.3226	26525	0.3226
11	32618	0.3548	26846	0.3548
12	32962	0.3871	27746	0.3871
13	33091	0.4194	28117	0.4194
14	33197	0.4516	28858	0.4516
15	33198	0.4839	28877	0.4839
16	33245	0.5161	29080	0.5161
17	33380	0.5484	29586	0.5484
18	33406	0.5806	29965	0.5806
19	33788	0.6129	29994	0.6129
20	33888	0.6452	30263	0.6452
21	33937	0.6774	30313	0.6774
22	34007	0.7097	30366	0.7097
23	34013	0.7419	30629	0.7419
24	34076	0.7742	30680	0.7742
25	34425	0.8065	30788	0.8065
26	34455	0.8387	30958	0.8387
27	34576	0.8710	31074	0.8710
28	35237	0.9032	31405	0.9032
29	35843	0.9355	31994	0.9355
30	35852	0.9677	32384	0.9677

Table 3.1.3 Quantile values of the traffic flow observations.

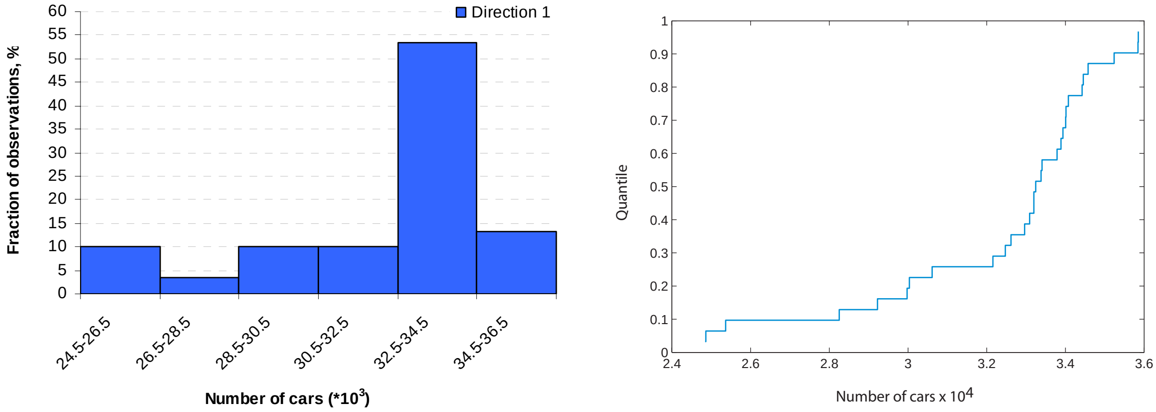


Figure 3.1.1: **Frequency distribution and cumulative distribution plot of the observed traffic flow in Rosengartenstrasse (direction 1).**

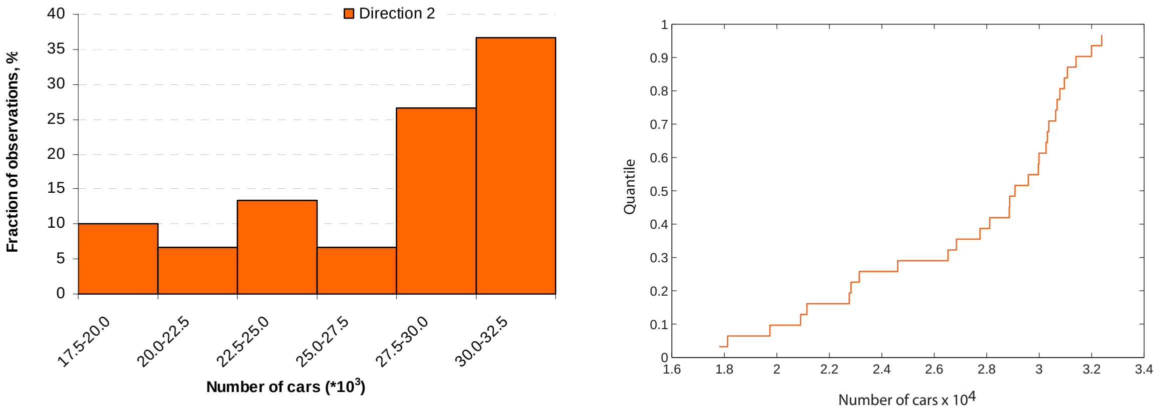


Figure 3.1.2 **Frequency distribution and cumulative distribution plot of the observed traffic flow in Rosengartenstrasse (direction 2).**

It can be seen from the frequency distributions that the traffic flow in direction 2 is lower than in direction 1. In direction 1 the highest frequency is observed within the range of 32500 and 34500 cars per day while the highest frequency for direction 2 is observed in the range of 30000 and 32500 cars per day. Additionally it is seen that both distributions are skewed to the left.

Plotting the cumulative distributions in the same scale, Figure 3.1.3, we have a direct comparison of the two data sets. It can be seen that the cumulative distribution plot of direction 1 is shifted significantly to the right, thus the traffic flow in this direction is higher than in direction 2.

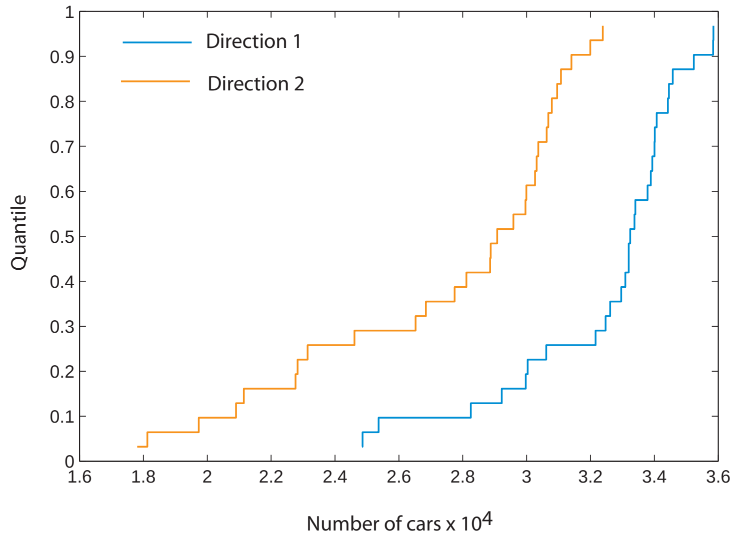


Figure 3.1.3: Cumulative distribution plot of the observed traffic flow in Rosengartenstrasse (direction 1 and 2).

However it can be seen from the plotted histograms that much information is lost due to the chosen number of intervals. Following the solution is provided for another number of intervals.

	Interval (Number of cars *10^3)	Interval Midpoint (Number of cars *10^3)	Number of observations	Frequency %	Cumulative frequency
Direction 1	24.50-25.75	25.125	3	10.000	0.100
	25.75-27.00	26.375	0	0.000	0.100
	27.00-28.25	27.625	0	0.000	0.100
	28.25-29.50	28.875	2	6.667	0.167
	29.50-30.75	30.125	3	10.000	0.267
	30.75-32.00	31.375	0	0.000	0.267
	32.00-33.25	32.625	8	26.667	0.533
	33.25-34.50	33.875	10	33.333	0.867
	34.50-35.75	35.125	2	6.667	0.933
	35.75-37.00	36.25	2	6.667	1.000
	Interval (Number of cars *10^3)	**Interval Midpoint (Number of cars *10^3)**	**Number of observations**	**Frequency %**	**Cumulative frequency**
Direction 2	17.5-19.0	18.25	2	6.667	0.067
	19.0-20.5	19.75	1	3.333	0.100
	20.5-22.0	21.25	2	6.667	0.167
	22.0-23.5	22.75	3	10.000	0.267
	23.5-25.0	24.25	1	3.333	0.300
	25.0-26.5	25.75	0	0.000	0.300
	26.5-28.0	27.25	3	10.000	0.400
	28.0-29.5	28.75	4	13.333	0.533
	29.5-31.0	30.25	10	33.333	0.867
	31.0-32.5	31.75	4	13.333	1.000

Table 3.1.3: Summary of the observed traffic flow.

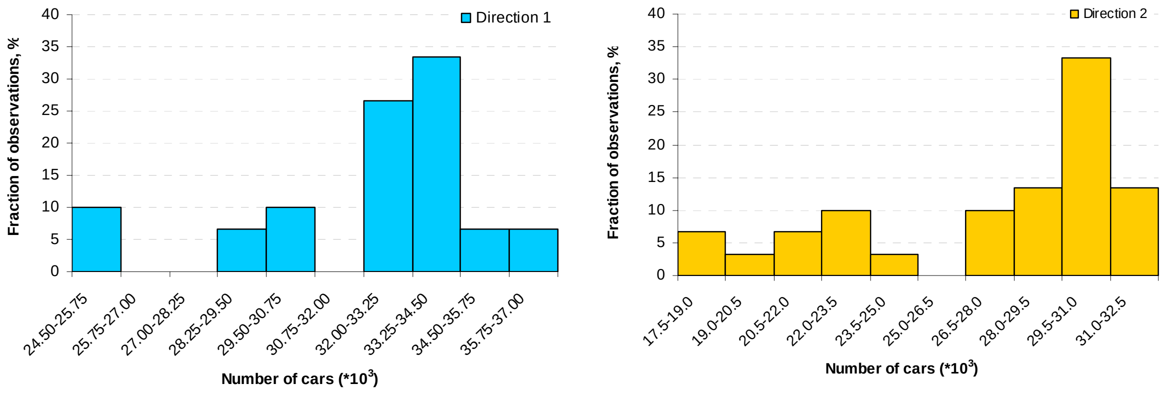


Figure 3.1.4: Frequency distribution plots of the observed traffic flow in Rosengartenstrasse (direction 1 and 2).

From Figures 3.1.4 it is seen that a larger number of intervals enables to view more clearly the features of the distributions.

Exercise 3.2 - Solution

In order to plot the Tukey box plot five main features are required as shown in Table C.8 in the lecture notes. These are:

- the lower quartile
- the lower adjacent value
- the median
- the upper adjacent value
- the upper quartile

Consider the data of traffic flow in direction 1. Based on Equation C.10 from the lecture notes a value v is required such that

$$v = nQ_v + Q_v$$

Therefore for the lower quartile (i.e. the 0.25 quartile) it is:

$$v = 30 \cdot 0.25 + 0.25 = 7.75$$

v has a non integer value. The value is splitted to its integer part $k = 7$ and the fractional part $p = 0.75$.Therefore x_v^o is:

$$x_v^o = (1-p)x_7^o + px_{7+1}^o = (1-0.75)\cdot 30035 + 0.75 \cdot 30613 = 30468.5 \approx 30469 \text{ cars}$$

In the same way for the upper quartile it is:

$$v = 30 \cdot 0.75 + 0.75 = 23.25$$

Thus with the help of Table 3.1.1 it is:

$$x_v^o = (1-p)x_{23}^o + px_{23+1}^o = (1-0.25)\cdot 34013 + 0.25 \cdot 34076 = 34028.75 \approx 34029 \text{ cars}$$

In order to calculate the median it is:

$$\nu = 30 * 0.5 + 0.5 = 15.5$$

$$x_\nu^o = (1-p)x_{15} + px_{15+1} = (1-0.5)*33198 + 0.5*33245 = 33221.5 \approx 33222 \text{ cars}$$

To evaluate the adjacent values the interquartile range is required:

$$r = Q_{0.75} - Q_{0.25} = 34029 - 30469 = 3560$$

The lower adjacent value is the smallest observation that is greater than or equal to the lower quartile minus *1.5r.* It is:

$$Q_{0.25} - 1.5r = 30469 - 1.5 \cdot 3560 = 25129$$

Thus from Table 3.1.1 the lower adjacent value is 25365.

In the same way the upper adjacent value is found as:

$$Q_{0.75} + 1.5r = 34029 + 1.5 \cdot 3560 = 39369$$

Therefore form Table 3.1.1 the upper adjacent value is a value less than or equal to 39369, that is 35852 which actually coincides with the higher value of the data set.

Table 3.2.1 summarizes the above features showing also the outside values of both data sets. It can be seen that in direction 2 there are no outside values.

Statistic	**Direction 1**	**Direction 2**
Lower adjacent value	25365	17805
Lower quartile	30469	23063
Median	33222	28979
Upper quartile	34029	30642
Upper adjacent value	35852	32384
Outside values	24846	
	24862	

Table 3.2.1: Statistics for the Tukey box plot for the traffic flow data in Rosengartenstrasse (direction 1 and 2).

Figure 3.2.1 shows the Tukey box plots for both directions. It can be seen that all main features of the distribution of the data set for direction 1 are much higher than the corresponding ones for direction 2. It can be also observed that the data are not symmetrical and the upper tales are shorter than the lower ones. The median is shifted to the upper part of the box plot in both directions and that shows that the distributions are skewed to the left.

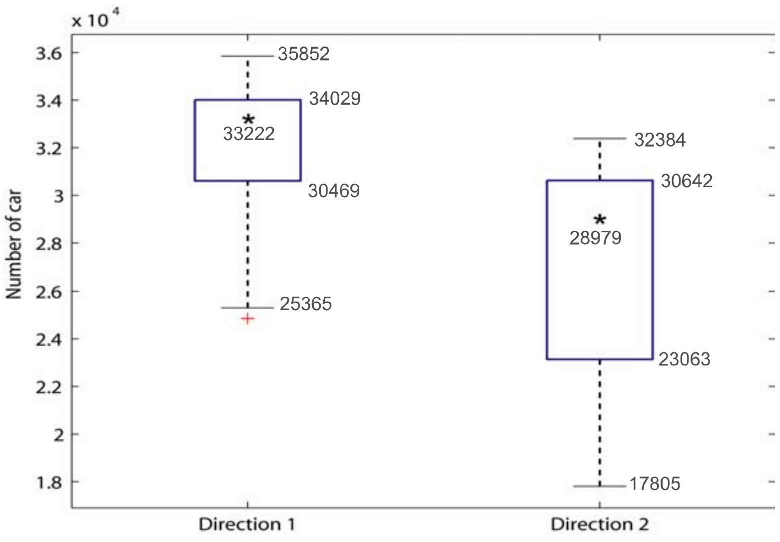


Figure 3.2.1: **Tukey box plots of the traffic flow data in Rosengartenstrasse (direction 1 and 2).**

Exercise 3.3 - Solution

In order to make the Q-Q plot the first thing is to examine the number of observations in each data set. Examining the number of observations within the two data sets it is seen that both have 30 observations. Therefore the Q-Q plot is simply a plot of the observations of the one data set against the observations of the other data set, Figure 3.3.1. To plot the Tukey mean difference plot the information of Table 3.3.1 is required.

Direction 2	**Direction 1**	y_i-x_i	$(y_i+x_i)/2$
17805	24846	7041	21325.5
18123	24862	6739	21492.5
19735	25365	5630	22550.0
20903	28252	7349	24577.5
21145	29224	8079	25184.5
22762	29976	7214	26369.0
22828	30035	7207	26431.5
23141	30613	7472	26877.0
24609	32158	7549	28383.5
26525	32472	5947	29498.5
26846	32618	5772	29732.0
27746	32962	5216	30354.0
28117	33091	4974	30604.0
28858	33197	4339	31027.5
28877	33198	4321	31037.5
29080	33245	4165	31162.5
29586	33380	3794	31483.0
29965	33406	3441	31685.5
29994	33788	3794	31891.0

30263	33888	3625	32075.5
30313	33937	3624	32125.0
30366	34007	3641	32186.5
30629	34013	3384	32321.0
30680	34076	3396	32378.0
30788	34425	3637	32606.5
30958	34455	3497	32706.5
31074	34576	3502	32825.0
31405	35237	3832	33321.0
31994	35843	3849	33918.5
32384	35852	3468	34118.0

Table 3.3.1: Values for the Tukey mean-difference plot of the traffic flow data in Rosengartenstrasse.

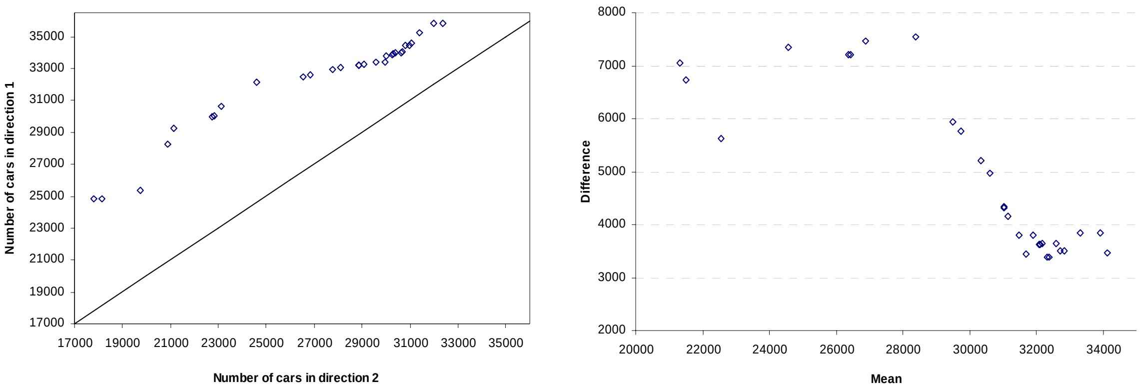


Figure 3.3.1: Q-Q plot of the traffic flow data and Tukey mean-difference plot.

It can be seen from Figure 3.3.1 that the data lie far from the symmetry line in the Q-Q plot and are concentrated on the side of direction 1. From the Tukey mean-difference plot it is seen that for a large part of the data sets the traffic flow in direction 1 is about 3500 cars per day higher than in direction 2.

Exercise 3.5- Solution

Mean of the number of the newcomers : $\bar{x}$ =2161

Mean of the number of the total students: $\bar{y}$ =13147

Standard deviation of the number of the newcomers: s_X =1337.

Standard deviation of the number of the total students: s_Y =8801.

Total number of observations: n =6.

Coefficient of correlation of the numbers of the newcomers and the total students: $r_{XY} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \bar{x})(y_i - \bar{y})}{s_X s_Y} = \frac{11604968}{1337 \cdot 8801} = 0.99$.

	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y - \bar{y})^2$	$(x_i - \bar{x})(y - \bar{y})$
A	3970	24273	1868.83	11126.5	3492538	123799002	20793574
B	732	5883	-1369.17	-7263.5	1874617	52758432	9944942
C	499	2847	-1602.17	-10299.5	2566938	106079700	16501516
D	1300	5358	-801.17	-7788.5	641868	60660732	6239887
E	3463	23442	1361.83	10295.5	1854590	105997320	14020755
F	2643	17076	541.83	3929.5	293583	15440970	2129134
Σ	12607	78879	-	-	10724135	464736158	69629808
Σ / n	2161.17	13146.5	-	-	1787356	77456026	11604968
$\sqrt{\Sigma / n}$	-	-	-	-	1336.92	8800.91	-

Exercise 3.6- Solution

The relationships between the height of the station and the maximum temperatures, and the height of the station and the minimum temperatures in May are obtained in Figure 3.6.1.

Let x_i, y_i and z_i ($i = 1, 2, ..., 10$) represent the height of the i^{th} station, maximum temperature and minimum temperature at the i^{th} station respectively. Using the calculation sheet the following descriptive statistics are obtained:

Mean values:

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1379, \qquad \bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = 13.7, \qquad \bar{z} = \frac{1}{10}\sum_{i=1}^{10} z_i = 4.36$$

Standard deviations:

$$s_x = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})^2} = 834, \; s_y = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(y_i - \bar{y})^2} = 1.99, \; s_z = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(z_i - \bar{z})^2} = 3.69$$

Covariances:

$$s_{xy} = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})\cdot(y_i - \bar{y}) = -1513, \qquad s_{xz} = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})\cdot(z_i - \bar{z}) = -2887$$

Correlation coefficients:

$$\rho_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = -0.91, \qquad \rho_{xz} = \frac{s_{xz}}{s_x \cdot s_z} = -0.94$$

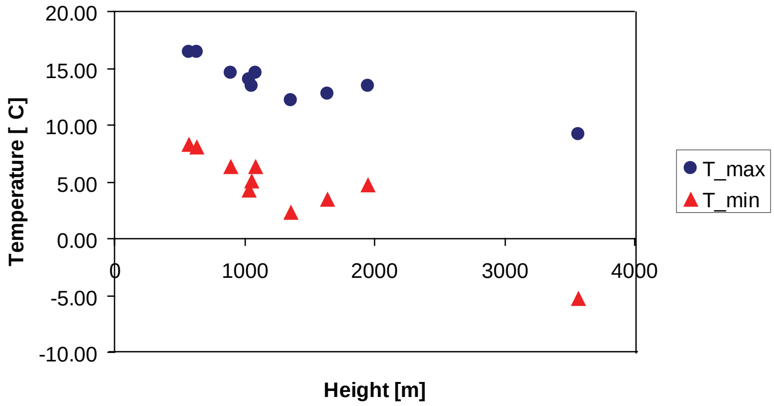


Figure 3.6.1: The relationship between height of station and maximum/minimum temperatures.

x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
Height [m]	T_{max} [°C]					
1355	12.2	-24.1	580.81	-1.5	2.25	36.15
890	14.6	-489.1	239218.81	0.9	0.81	-440.19
1950	13.4	570.9	325926.81	-0.3	0.09	-171.27
1040	14	-339.1	114988.81	0.3	0.09	-101.73
1085	14.6	-294.1	86494.81	0.9	0.81	-264.69
1055	13.4	-324.1	105040.81	-0.3	0.09	97.23
574	16.4	-805.1	648186.01	2.7	7.29	-2173.77
3572	9.2	2192.9	4808810.4	-4.5	20.25	-9868.05
632	16.4	-747.1	558158.41	2.7	7.29	-2017.17
1638	12.8	258.9	67029.21	-0.9	0.81	-233.01

Table 3.6.2: Calculation sheet for Height – T_{max} relation.

x_i	z_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$z_i - \bar{z}$	$(z_i - \bar{z})^2$	$(x_i - \bar{x})(z_i - \bar{z})$
Height [m]	T_{min} [°C]					
1355	2.3	-24.1	580.81	-2.06	4.2436	49.646
890	6.3	-489.1	239218.81	1.94	3.7636	-948.854
1950	4.7	570.9	325926.81	0.34	0.1156	194.106
1040	4.3	-339.1	114988.81	-0.06	0.0036	20.346
1085	6.3	-294.1	86494.81	1.94	3.7636	-570.554
1055	5.1	-324.1	105040.81	0.74	0.5476	-239.834
574	8.3	-805.1	648186.01	3.94	15.5236	-3172.094
3572	-5.3	2192.9	4808810.4	-9.66	93.3156	-21183.414
632	8.1	-747.1	558158.41	3.74	13.9876	-2794.154
1638	3.5	258.9	67029.21	-0.86	0.7396	-222.654

Table 3.6.3: Calculation sheet for Height – T_{min} relation.

Exercise 3.7:

The relative and cumulative frequencies are obtained in Table 3.7.1. The histogram is shown in Figure 3.7.1 and 3.7.2.

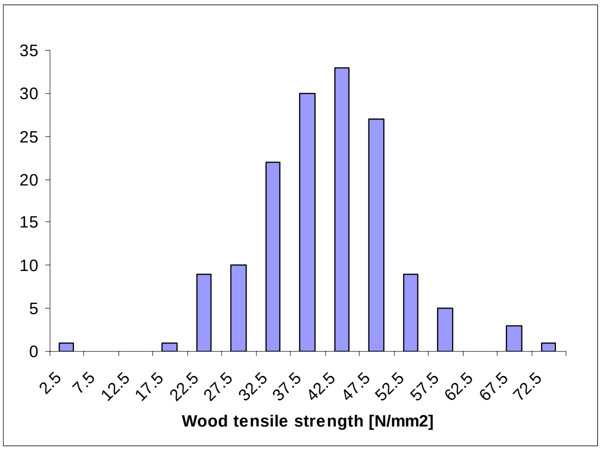


Figure 3.7.1: Histogram.

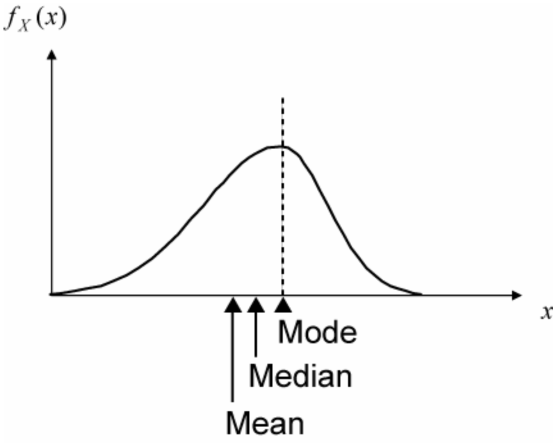
a. The probability that the tensile strength lies between 20 and 25 [N/mm^2] is obtained from Table 3.7.2 as $P[A] = \frac{n_k}{n} = \frac{9}{151} = 0.06$

b. $$P[B] = \frac{\sum_{i=1}^{5} n_i}{n} = \frac{(1+0+0+1+9)}{151} = \frac{11}{151} = 0.062$$

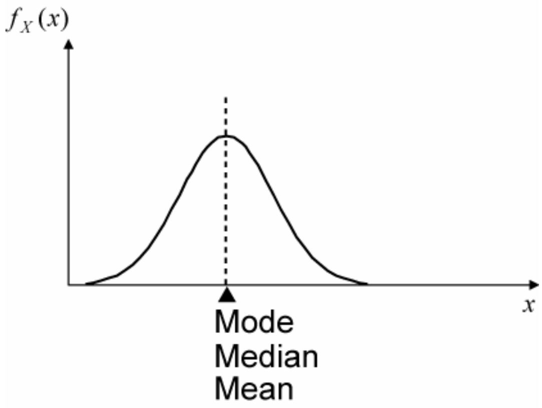
Upper limit [N/mm2]	Class center [N/mm2]	Abs. frequency n_i	Rel. frequency	Cumulative frequency
5	2.5	1	0.007	0.007
10	7.5	0	0.000	0.007
15	12.5	0	0.000	0.007
20	17.5	1	0.007	0.013
25	22.5	9	0.060	0.073
30	27.5	10	0.066	0.139
35	32.5	22	0.146	0.285
40	37.5	30	0.199	0.483
45	42.5	33	0.219	0.702
50	47.5	27	0.179	0.881
55	52.5	9	0.060	0.940
60	57.5	5	0.033	0.974
65	62.5	0	0.000	0.974
70	67.5	3	0.020	0.993
75	72.5	1	0.007	1.000

Table 3.7.2: Relative and cumulative frequencies of wood tensile strength.

Exercise 3.8:



Left skewed distribution



Symmetrical distribution

EXERCISE TUTORIAL 4- SOLUTION:

Exercise 4.1 – Solution

a. The integration of the probability density function over the entire support must be one.

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \Rightarrow c \cdot \int_{0}^{60} x \cdot (15 - \frac{x}{4})\, dx = c \cdot \left[\frac{15}{2} \cdot x^2 - \frac{1}{12} \cdot x^3 \right]_0^{60} = 1$$

$$\Rightarrow c \cdot (27000 - 18000) = 1 \Rightarrow c = \frac{1}{9000}$$

b.

$$\int_{-\infty}^{x} c \cdot y \cdot (15 - \frac{y}{4})\, dy = \frac{1}{9000} \cdot \left[\frac{15}{2} \cdot y^2 - \frac{1}{12} \cdot y^3 \right]_0^{x} = \frac{1}{9000} \cdot \left(\frac{15}{2} \cdot x^2 - \frac{1}{12} \cdot x^3 \right)$$

$$F_X(x) = \begin{cases} 0 & \text{x} < 0 \\ \frac{1}{9000} \cdot \left(\frac{15}{2} \cdot x^2 - \frac{1}{12} \cdot x^3 \right) & 0 \le \text{x} \le 60 \\ 1 & 60 < \text{x} \end{cases}$$

c. Let a be a number between 0 and 60, then:

$$P(X \le a) = \frac{1}{9000} \cdot \int_{0}^{a} x \cdot (15 - \frac{x}{4})\, dx = \frac{1}{9000} \cdot \left[\frac{15}{2} \cdot x^2 - \frac{1}{12} \cdot x^3 \right]_0^{a} = 0.9$$

$$\frac{1}{9000} \cdot \left(\frac{15}{2} \cdot a^2 - \frac{1}{12} \cdot a^3 \right) \equiv 0.9 \Rightarrow \frac{a^3}{12} - \frac{15}{2} \cdot a^2 + 8100 = 0$$

$$\Rightarrow a^3 - 90 \cdot a^2 + 97200 = 0 \Rightarrow a = 48.30$$

So the values of 30.00 CHF and 40.00 CHF do not exceed the 90% quantile.

d. Considering the symmetry of the probability density function of X, the mean value is obtained as: $(0+60)/2 = 30.$ Or, the mean value is obtained also as follows:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \frac{1}{9000} \cdot \int_{0}^{60} x^2 \cdot (15 - \frac{x}{4})\, dx = \frac{1}{9000} \cdot \left[5 \cdot x^3 - \frac{1}{16} \cdot x^4 \right]_0^{60} \Rightarrow$$

$$E(X) = \frac{1}{9000} \cdot (1080000 - 810000) = \frac{270000}{9000} = 30$$

Exercise 4.2 - Solution

a. The probability density function is obtained as:

$$f_X(x) = \begin{cases} 0 & x < a \\ h \cdot \dfrac{(x-a)}{(b-a)} & a \le x < b \\ h & b \le x < c \\ h \cdot \dfrac{(x-d)}{(c-d)} & c \le x < d \\ 0 & d \le x \end{cases}$$

Integration of the probability density function gives the cumulative distribution function as follows:

$$F_X(x) = \int_{-\infty}^{\infty} f_X(x) dx$$

$$F_X(x) = \begin{cases} 0 & x < a \\ h \cdot \dfrac{(x-a)^2}{2 \cdot (b-a)} + C_1 & a \le x < b \\ h \cdot x + C_2 & b \le x < c \\ h \cdot \dfrac{(x-d)^2}{2 \cdot (c-d)} + C_3 & c \le x < d \\ 0 & d \le x \end{cases}$$

Integration takes place over the continuing terms.

By $x = a$ $\quad 0 = h \cdot \dfrac{(a-a)^2}{2 \cdot (b-a)} + C_1 \Rightarrow C_1 = 0$

By $x = b$ $\quad h \cdot \dfrac{(b-a)^2}{2 \cdot (b-a)} = h \cdot b + C_2 \Rightarrow C_2 = -\dfrac{(a+b)}{2} \cdot h$

By $x = c$ $\quad h \cdot \dfrac{(x-d)^2}{2 \cdot (c-d)} + C_3 = h \cdot x + C_2 \Rightarrow h \cdot \dfrac{(c-d)^2}{2 \cdot (c-d)} + C_3 = h \cdot c - \dfrac{(a+b)}{2} \cdot h$

$$\Rightarrow C_3 = \left(\frac{(c+d) - (a+b)}{2} \right) \cdot h$$

Finally the cumulative distribution function becomes:

$$F_X(x) = \begin{cases} 0 & (x \le a) \\ h\dfrac{(x-a)^2}{2(b-a)} & (a < x \le b) \\ hx - \dfrac{(a+b)}{2}h & (b < x \le c) \\ h\dfrac{(x-d)^2}{2(c-d)} + \dfrac{(c+d)-(a+b)}{2}h & (c < x \le d) \\ 1 & (d \le x) \end{cases}$$

b. The mode value is the value at which lies the maximum of the density function. In the existing case, no distinct maximum is available. Instead of a mode value, an area is indicated from *b* to *c*.

The parameter *h* may be estimated by evaluating the value of $F_X(x)$ at $x = 6$:

$\int_{-\infty}^{\infty} f_X(x)dx = 1$ e.g. area under the density function = 1

Thus it is: $\dfrac{(d-a)+(c-b)}{2} \cdot h = 1 \Rightarrow \dfrac{(6-1)+(3-2)}{2} \cdot h = 1 \Rightarrow 3 \cdot h = 1 \Rightarrow h = \dfrac{1}{3}$

c. For a=1, *b*=2, *c*=3, *d*=6 and *h*=1/3 the probability density function gets the following form:

$$f_X(x) = \begin{cases} 0 & x < 1 \\ \dfrac{(x-1)}{3} & 1 \le x < 2 \\ \dfrac{1}{3} & 2 \le x < 3 \\ -\dfrac{(x-6)}{9} & 3 \le x < 5 \\ 0 & 5 \le x \end{cases}$$

The mean value is then calculated as follows

$$\mu_x = E\left[x\right] = \int_{-\infty}^{\infty} x \cdot f_x(x) \cdot dx = \int_1^2 \frac{x \cdot (x-1)}{3} dx + \int_2^3 \frac{x}{3} \cdot dx + \int_3^6 \frac{-x \cdot (x-6)}{9} dx$$

$$= \left[\frac{x^3}{9} - \frac{x^2}{6}\right]_1^2 + \left[\frac{x^2}{6}\right]_2^3 - \left[\frac{x^3}{27} - \frac{x^2}{3}\right]_3^6 = \frac{28}{9}$$

c. Using the parameters a =1, b=2, c=3, d=6 and h=1/3 and by integration of the probability density function it is:

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{(x-1)^2}{6} & 1 \le x < 2 \\ \frac{x}{3} - \frac{1}{2} & 2 \le x < 3 \\ -\frac{(x-6)^2}{18} + 1 & 3 \le x < 6 \\ 0 & 6 \le x \end{cases}$$

It is easy to find that $F_X(3) = 0.5$. Therefore, the median is 3.

e. The median can be determined graphically through the illustration of the probability density function. It is that x value, for which the area under the density function is half the total area, Figure 4.2.2.

Area $\quad A_1 = (2-1) \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$

Area $\quad A_2 = (3-2) \cdot \frac{1}{3} = \frac{1}{3}$

Area $\quad A_3 = (6-3) \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}$

Since $A_1 + A_2 = A_3$ the median lies at x=3.

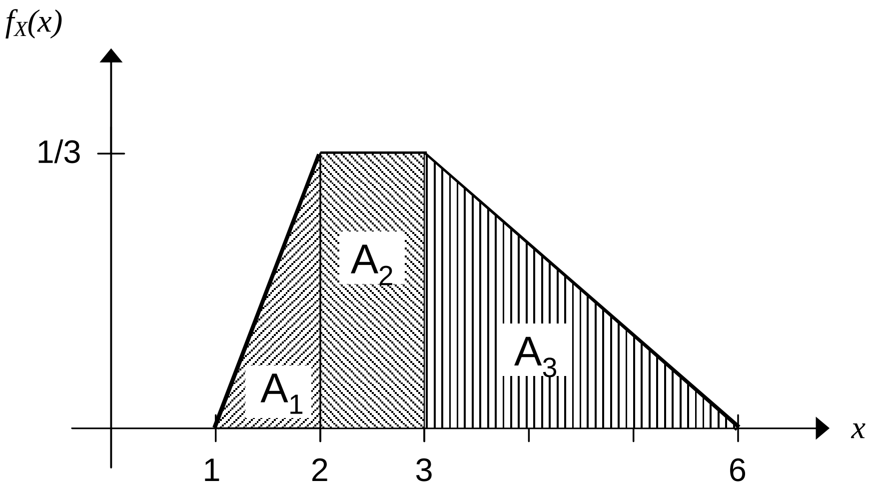


Figure 4.2.2: **Determination of the median.**

The graphical interpretation of the mean value is the center of gravity of the shape of the probability density function. That means that moments are necessary for the estimation of the mean values. The mean value lies where moments of the corresponding areas are in equilibrium Figure 4.2.3. Therefore in a graphical solution the areas A_i and the associated lever arms d_i should be estimated to evaluate *x*. It is useful to know that: $\sum_{i=1}^{5} A_i \cdot d_i = 0$

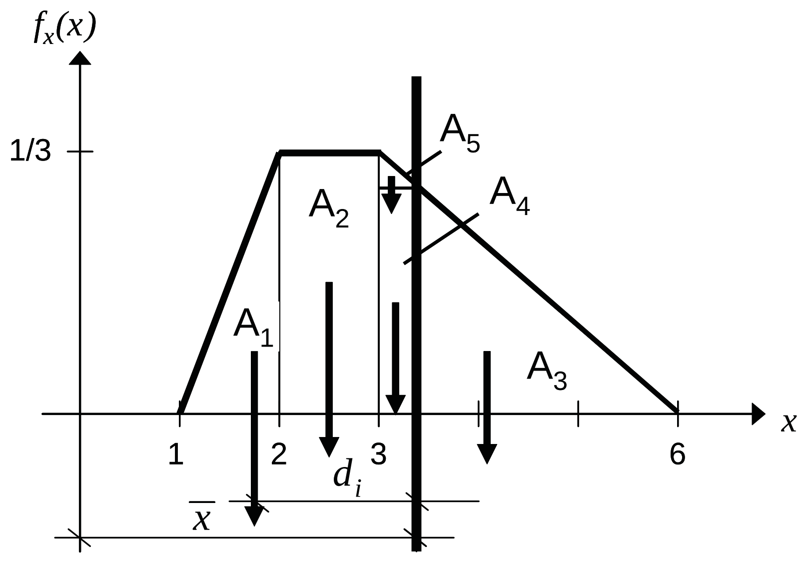


Figure 4.2.3: **Estimation of the mean values.**

EXERCISE TUTORIAL 5- SOLUTION:

Exercise 5.1 – Solution

a. The expected values of *X* and *Y* are:

$$E(X)=\int_{-\infty}^{\infty} x\cdot f_X(x)dx=\int_{-1}^{1}\frac{1}{2}xdx=\left[\frac{1}{4}x^2\right]_{-1}^{1}=0$$

$$E(Y)=\int_{-\infty}^{\infty} y\cdot f_Y(y)dy=\frac{3}{4}\cdot\int_{0}^{2} y\cdot\left(2y-y^2\right)dy=\frac{3}{4}\cdot\left[\frac{2}{3}\cdot y^3-\frac{y^4}{4}\right]_0^2=1$$

The expected value of $6X-4Y+2$ is obtained as follows:

$$E(6X-4Y+2)=6E(X)-4E(Y)+2=-2$$

b. The variances of *X* and *Y* are obtained as follows:

$$Var(X)=E\left[X^2\right]-\left(E\left[X\right]\right)^2 \qquad Var(Y)=E\left[Y^2\right]-\left(E\left[Y\right]\right)^2$$

$$E\left[X^2\right]=\int_{-\infty}^{\infty} x^2 f_X(x)dx=\int_{-1}^{1}\frac{1}{2}x^2dx=\left[\frac{x^3}{6}\right]_{-1}^{1}=\frac{1}{3}$$

$$E\left[Y^2\right]=\int_{-\infty}^{\infty} y^2 f_Y(y)\,dy=\frac{3}{4}\cdot\int_{0}^{2} y^2\cdot\left(2y-y^2\right)dy=\frac{3}{4}\left[\frac{y^4}{2}-\frac{y^5}{5}\right]_0^2=\frac{6}{5}$$

$$Var(X)=\frac{1}{3} \qquad Var(Y)=\frac{1}{5}$$

So the covariance of $Cov(6X;4Y)$ is then obtained as follows:

$$Cov(X;Y)=\rho_{X,Y}\cdot\sqrt{Var(X)\cdot Var(Y)}=\sqrt{\frac{1}{15}\cdot\frac{1}{3}\cdot\frac{1}{5}}=\sqrt{\frac{1}{225}}$$

$$Cov(6X;4Y)=6\cdot4\cdot Cov(X;Y)=6\cdot4\cdot\sqrt{\frac{1}{45}}=24\cdot\sqrt{\frac{1}{45}}$$

c.

$$Var(6X-4Y+2)=Var(6X)+Var(4Y)-2\cdot Cov(6X;4Y)=$$

$$6^2\cdot Var(X)+4^2\cdot Var(Y)-2\cdot Cov(6X;4Y)=36\cdot\frac{1}{3}+16\cdot\frac{1}{5}-2\cdot24\cdot\sqrt{\frac{1}{45}}\cong 8.04$$

d.

$$E(6X^2-4Y^2)=6E(X^2)-4E(Y^2)=6E(X^2)-4(Var(Y)+\left[E(Y)\right]^2)=$$

$$6\cdot\frac{1}{3}-4\cdot(\frac{1}{5}+1^2)=-\frac{14}{5}$$

Exercise 5.2 - Solution

a. $P[N_U = N_G] = 0.2910 + 0.3580 + 0.1135 + 0.0505 = 0.813$.

b. The probability of interest is represented by a conditional probability:

$$P[N_U \mid N_G = 2] = \frac{P[N_U \cap (N_G = 2)]}{P[N_G = 2]}.$$

The conditional probabilities are obtained as:

$$P[N_U = 0 \mid N_G = 2] = \frac{P[(N_U = 0) \cap (N_G = 2)]}{P[N_G = 2]} = \frac{0.01}{0.1785} = 0.056$$

$$P[N_U = 1 \mid N_G = 2] = \frac{P[(N_U = 1) \cap (N_G = 2)]}{P[N_G = 2]} = \frac{0.025}{0.1785} = 0.1401$$

$$P[N_U = 2 \mid N_G = 2] = \frac{P[(N_U = 2) \cap (N_G = 2)]}{P[N_G = 2]} = \frac{0.1135}{0.1785} = 0.6359$$

$$P[N_U = 3 \mid N_G = 2] = \frac{P[(N_U = 3) \cap (N_G = 2)]}{P[N_G = 2]} = \frac{0.03}{0.1785} = 0.1681.$$

EXERCISE TUTORIAL 6- SOLUTION:

Exercise 6.1 – Solution

a. For S_n it is:

$$E[S_n] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \mu_X = 50$$

$$V[S_n] = V\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} V[X_i] = \sum_{i=1}^{n} \sigma_X^2 = 200$$

And thus, based on the central limit theorem, S_{50} is Normal distributed with: $N(50, 200)$

For $\bar{X}_n$ it is:

$$E\left[\bar{X}_n\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{50} 50 = 1$$

$$V\left[\bar{X}_n\right] = \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{50^2} 200 = 0.08$$

And thus $\bar{X}_{50}$ is Normal distributed with: $N(1, 0.08)$.

b. X_1 is Normal distributed with $N(1, 2^2)$. A new random variable Z is introduced such as $Z = \frac{X_1 - 1}{2}$ with $N(0,1)$. Then it is:

$$P(E[X_1] - 1 \le X_1 \le E[X_1] + 1) = P(0 \le X_1 \le 2) = P\left(-\frac{1}{2} \le Z \le \frac{1}{2}\right)$$

$$P_Z\left(\frac{1}{2}\right) - P_Z\left(-\frac{1}{2}\right) = P_Z\left(\frac{1}{2}\right) - \left(1 - P_Z\left(\frac{1}{2}\right)\right) = P_Z\left(\frac{1}{2}\right) + P_Z\left(\frac{1}{2}\right) - 1 = 2 \cdot \Phi\left(\frac{1}{2}\right) - 1$$

$$\cong 2 \cdot 0.692 - 1 = 0.384$$

c. For $n = 50$ it is $Z = \frac{S_n - 50}{\sqrt{200}}$. Then:

$$P(E[S_n] - 1 \le S_n \le E[S_n] + 1) = P(49 \le S_n \le 51) = P(-0.07 \le Z \le 0.07)$$

$$= 2 \cdot \Phi(0.07) - 1 \cong 2 \cdot 0.53 - 1 = 0.06$$

d. For $n = 50$ it is $Z = \frac{\bar{X}_{50} - 1}{\sqrt{0.08}}$. Then:

$$P\left(E\left[\bar{X}_n\right] - 1 \le \bar{X}_n \le E\left[\bar{X}_n\right] + 1\right) = P\left(0 \le \bar{X}_{50} \le 2\right) = P(-3.5 \le Z \le 3.5)$$

$$= 2 \cdot \Phi(3.5) - 1 \cong 2 \cdot 0.999 - 1 = 0.998$$

Exercise 6.2 – Solution

Consider the following events:

Event *H*: overflow in a given year

Event *K*: no overflow in a given year

$$P(H) = \frac{1}{1000} = 0.001 = p_1$$
$$P(K) = 1 - 0.001 = 0.999 = \overline{p}_1$$

n in the following corresponds to the 10 year period.

a. The event of overflow in a given year during a 10 year period may be described by a geometric distribution:

$$P(H_{overflow,1}) = (p_1) \cdot (1 - p_1)^{n-1} = (0.001) \cdot (0.999)^9 = 0.000991$$

b. According to the Binomial distribution it is (the frequency of occurrence implies no difference in the solution):

$$P(H_{overflow,2}) = \frac{10!}{2! \cdot (10-2)!}(p_1)^2 \cdot (\overline{p}_1)^{10-2} = 45 \cdot (0.001)^2 \cdot (0.999)^8 = 0.000045$$

$$P(H_{overflow,0}) = \frac{10!}{0! \cdot (10-0)!}(p_1)^0 \cdot (\overline{p}_1)^{10-0} = (0.001)^0 \cdot (0.999)^{10} = 0.99004$$

c. The probabilities of the events of “no overflow” and “overflow once” need to be estimated:

$$P(H_{overflow.,1}) = \frac{10!}{1! \cdot (10-1)!}(p_1)^1 \cdot (\overline{p}_1)^{10-1} = 10 \cdot (0.001)^1 \cdot (0.999)^9 = 0.00991$$

$$P(H_{overflow,0}) = \frac{10!}{0! \cdot (10-0)!}(p_1)^0 \cdot (\overline{p}_1)^{10-0} = (0.001)^0 \cdot (0.999)^{10} = 0.99004$$

$$P(H_{\max,1}) = P(H_{overflow,0}) + P(H_{overflow.,1}) = 0.99004 + 0.00991 = 0.99995$$

d. According to the Binomial distribution it is (the frequency of occurrence implies no difference in our solution):

$$P(H_{overflow,10}) = \frac{100!}{10! \cdot (100-10)!}(p_1)^{10} \cdot (\overline{p}_1)^{100-10} = 1.6 \cdot 10^{-17}$$

e. For the case that the number of the considered years is high (m=100) and the yearly probability of overflow is small (p_1=0.001), the Poisson distribution can be used.

$$\mu_y = m \cdot p_1 = 100 \cdot 0.001 = 0.1$$

$$P(H_{overflow,10}) = \frac{{\mu_y}^y}{y!} \cdot e^{-\mu_y} = \frac{0.1^{10}}{10!} \cdot e^{-0.1} = 2.5 \cdot 10^{-17}$$

f. From the Poison distribution it is:

$$P(H_{\max,10}) = P(H_{overflow,0}) + P(H_{overflow,1}) + ... + P(H_{overflow.,10})$$

$$= \frac{0.1^0 \cdot e^{-0.1}}{0!} + \frac{0.1^1 \cdot e^{-0.1}}{1!} + ... + \frac{0.1^{10} \cdot e^{-0.1}}{10!} = 0.999 \cong 1$$

g. Using the Binomial distribution it is:

$$P(H_{overflow,0}) = \frac{1000!}{0! \cdot (1000-0)!} (p_1)^0 \cdot (\overline{p}_1)^{1000-0} = (0.001)^0 \cdot (0.999)^{1000} = 0.368$$

And the required probability is the probability of the complementary event:

$$P(H_{overflow,\geq 1}) = 1 - 0.368 = 0.632$$

Using the Poisson distribution instead it is:

$$\mu_y = m \cdot p_1 = 1000 \cdot 0.001 = 1$$

$$P(H_{overflow,\geq 1}) = 1 - \frac{1^0}{0!} \cdot e^{-1} = 0.632$$

EXERCISE TUTORIAL 7

Exercise 7.1 - Solution

a. The mean occurrence rate of a rainfall in the first 5 months of a year is obtained as:

$$\nu = \int_0^3 \frac{2\cdot t}{3}\,dt + \int_3^5 2\;dt = 7\,.$$

Therefore the probability that 3 or more rainfalls occur in the first 5 month is:

$$P[X \geq 3] = 1 - P[X \leq 2] = 1 - (\frac{7^0}{0!}e^{-7} + \frac{7^1}{1!}\cdot e^{-7} + \frac{7^2}{2!}\cdot e^{-7}) = 0.97\,.$$

where X is the number of rainfall in the first 5 months.

b. Let Y, Z be the numbers of occurrence of rainfall during the 8th, 9th and the 10th month and during the 11th, 12th and the 13th month respectively. The mean occurrence rates in each period are obtained as:

$$\nu_Y = \frac{1}{3}\int_7^{10} (13 - t)\,dt = 4.5$$

$$\nu_Z = \frac{1}{3}\int_{10}^{13} (13 - t)\,dt = 1.5\,.$$

The probability of interest is:

$$P[Y \leq 1 \text{ and } Z \leq 1] = (\frac{4.5^0}{0!}e^{-4.5} + \frac{4.5^1}{1!}\cdot e^{-4.5})\cdot(\frac{1.5^0}{0!}e^{-1.5} + \frac{1.5^1}{1!}\cdot e^{-1.5}) = 0.034\,.$$

Exercise 7.2 - Solution

a. Let A represent the event which corresponds to a return period of 475 years. The probability that A occurs in a year $P_A(1)$ is:

$$P_A(1) = \frac{1}{475}\,.$$

The probability that A occurs in 50 years $P_A(50)$ can be calculated as:

$$P_A(50) = 1 - (1 - P_A(1))^{50} = 1 - \left(1 - \frac{1}{475}\right)^{50} = 0.1\,.$$

b. The probability that A occurs within the next 475 years, $P_A(475)$ can be calculated as: $P_A(475) = 1 - (1 - P_A(1))^{475} = 1 - \left(1 - \frac{1}{475}\right)^{475} = 0.633\,.$

Exercise 7.3 – Solution

a.

$$P[yearly\ \max \geq 15.000] = 1 - F_X(x = 15.000) = 1 - e^{-e^{-\alpha(15.000-u)}}$$

$$\alpha = \frac{\pi}{\sigma_x\sqrt{6}} = \frac{\pi}{3.000\sqrt{6}} = 4.2752\ 10^{-4}$$

$$u = \mu_x - \frac{0,57722}{\alpha} = 10.000 - \frac{0,57722}{4.2752\ 10^{-4}} = 8649,809$$

$$1 - F_X(x = 15.000) = 1 - e^{-e^{-4.2752\ 14^{-4}\ (15.000-8649.81)}} = 1 - e^{-e^{-2.715}} = 1 - 0.9359 = 0.0641$$

The probability that the annual maximum discharge will exceed 15.000 m^3/s is 0.0641.

b.

$$1 - \frac{1}{100} = F_X(x) = e^{-e^{-\alpha(x-u)}} = 0.99 \Leftrightarrow \ln(-\ln(0.99)) = -\alpha(x-u) \Leftrightarrow \frac{\ln(-\ln(0.99))}{-\alpha} + u = x$$

.

$$\Leftrightarrow \frac{\ln(-\ln(0.99))}{-4,2752\ 10^{-4}} + 8649.809 = x \Leftrightarrow 10760.08 + 8649.809 = x \Leftrightarrow 19409.889 = x$$

The discharge that corresponds to a return period T of 100 years is $19410\ m^3/s$.

c.

$$F_Y(y) = P[Y \leq y] = [F_X(x)]^{20} =$$

$$F_Y(y) = \left(e^{-e^{-\alpha(x-u)}}\right)^{20}$$

$$F_Y(y) = e^{-e^{-20\alpha(x-u)}}$$

d.

$$1 - F_Y(15000) = 1 - e^{-e^{-20\ 4,2756\ 10^{-4}(15000-8649,81)}} = 1 - e^{-1,324} = 1 - 0,266$$

$$1 - F_Y(15000) = 0,734$$

The probability that the 20-year-maximum discharge will exceed 15.000 m^3/s is 0.734.

EXERCISE TUTORIAL 8 - SOLUTION

Exercise 8.1 - Solution

a. From the Pythagorean Theorem it follows that:

$$f^2 + a^2 + b^2 = d^2$$

Therefore, the error in d propagates according to $\varepsilon_d = \sqrt{\varepsilon_f^2 + \varepsilon_a^2 + \varepsilon_b^2}$.

Then, $\frac{\varepsilon_d}{\sigma_\varepsilon} = \sqrt{(\frac{\varepsilon_f}{\sigma_\varepsilon})^2 + (\frac{\varepsilon_a}{\sigma_\varepsilon})^2 + (\frac{\varepsilon_b}{\sigma_\varepsilon})^2}$ is Chi-distributed with three degrees of freedom.

The probability density function of $Z = \frac{\varepsilon_d}{\sigma_\varepsilon}$ is:

$$f_Z(z) = \frac{z^{(3-1)}}{2^{3/2-1}\Gamma(3/2)} e^{(-z^2/2)}$$

Therefore, the probability density function of ε_d is obtained as:

$$f_{\varepsilon_d}(\varepsilon_d) = \frac{1}{2\sqrt{2}}(\frac{\varepsilon_d}{\sigma_e})^2 \frac{1}{\sqrt{\pi}/2} e^{(-(\frac{\varepsilon_d}{\sigma_e})^2/2)} \left|\frac{dz}{d\varepsilon_d}\right| = \frac{1}{\sqrt{2\pi}}(\frac{\varepsilon_d}{\sigma_\varepsilon})^2 e^{(-(\frac{\varepsilon_d}{\sigma_e})^2/2)} \frac{1}{\sigma_\varepsilon}$$

b. The error in c propagates according to $\varepsilon_c = \sqrt{\varepsilon_a^2 + \varepsilon_b^2}$.

$Y = \frac{\varepsilon_c}{\sigma_c} = \sqrt{(\frac{\varepsilon_a}{\sigma_\varepsilon})^2 + (\frac{\varepsilon_b}{\sigma_\varepsilon})^2}$ is Chi-distributed with two degrees of freedom and the probability density function is:

$$f_Y(y) = \frac{y^{(2-1)}}{2^{2/2-1}\Gamma(2/2)} e^{(-y^2/2)} = ye^{-\frac{y^2}{2}} .$$

The probability that the error in c exceeds $2.4\sigma_\varepsilon$ is obtained as:

$$P(\varepsilon_c \geq 2.4\sigma_\varepsilon) = P(\frac{\varepsilon_c}{\sigma_\varepsilon} \geq 2.4) = P(Y \geq 2.4) = \int_{2.4}^{\infty} ye^{-\frac{y^2}{2}} dy = 0.056 .$$

Exercise 8.2 - Solution

1. Formulate the null and alternate hypotheses:

The null hypothesis H_0 is formulated as the true mean μ being equal to 30 MPa. The alternate hypothesis H_1 is then simply given by $\mu \neq 30$ MPa.

$H_0: \quad \mu = 30 \ MPa \qquad\qquad H_1: \quad \mu \neq 30 \ MPa$

(It may be strange to assume that the null hypothesis is $\mu = 30 \ MPa$, since $\mu > 30 MPa$ is also acceptable if $\mu = 30 \ MPa$ is acceptable. However, in this exercise, for simplicity the null hypothesis and the alternative hypothesis are taken as above.)

2. Formulate an operating rule:

The operating rule is given as: $P(30 - \Delta \leq \overline{X} \leq 30 + \Delta) = 1 - \alpha$

3. Choose the level of significance α:

$\alpha = 10\%$.

4. Determine the condition of sampling (what kind of and how many data?):

15 samples of the compressive strength are taken at one day from the concrete production.

5. Do the calculations:

$$P(30 - \Delta \leq \overline{X} \leq 30 + \Delta) = 1 - \alpha \Rightarrow$$
$$P(30 - \Delta \leq \overline{X} \leq 30 + \Delta) = 0.9 \Rightarrow 2\Phi(\frac{30 + \Delta - 30}{\sigma_x / \sqrt{n}}) - 1 = 0.9 \Rightarrow \Phi(\frac{\Delta}{\sqrt{16.36/15}}) = 0.95$$

where $\overline{X}$ is the sample statistic. From the probability table for the standard Normal distribution (Annex T, Table T.1), it is:

$$\frac{\Delta}{\sqrt{16.36/15}} = 1.645 \Rightarrow \Delta = 1.72 .$$

Thus, if the sample mean $\overline{x}$ from the 15 samples lies within the interval:

$\left[28.28 MPa \leq \overline{x} \leq 31.72 MPa \right]$ then the null hypothesis cannot be rejected at the 10% significance level.

6. Obtain the sample mean:

The sample mean is equal to 32.25 MPa.

7. Judge the null hypothesis H_0

Since 32.25 MPa is outside the interval, the null hypothesis is rejected.

In the same procedure, the interval for accepting the null hypothesis at the 1% significance level is obtained as $\left[27.31MPa \le \overline{X} \le 32.69MPa\right]$. Since the sample mean (32.25MPa) is within the interval, the null hypothesis cannot be rejected.

Exercise 8.3 - Solution

1. Formulate the null and alternate hypotheses:

$H_0 : \mu = 23.7 \qquad H_1 : \mu \ne 23.7$

2. Formulate an operating rule:

The operating rule is given as: $\mu - 1.96\frac{\sigma}{\sqrt{n}} \le \overline{x} \le \mu + 1.96\frac{\sigma}{\sqrt{n}}$

where σ is the standard deviation of the traveling time and n is the number of samples.

3. Choose the level of significance α:

$\alpha = 5\%$.

4. Determine the condition of sampling (what kind of and how many data?)

$n = 13$ samples of the traveling time.

5. Do the calculations:

$$\mu - 1.96\frac{\sigma}{\sqrt{n}} \le \overline{x} \le \mu + 1.96\frac{\sigma}{\sqrt{n}} \Leftrightarrow 23.7 - 1.96\frac{3}{\sqrt{13}} \le \overline{x} \le 23.7 + 1.96\frac{3}{\sqrt{13}}$$
$$\Leftrightarrow 22.07 \le \overline{x} \le 25.33$$

6. Obtain the sample mean.

$\overline{x} = 22.3$ minutes.

7. Judge the hypothesis H_0.

The sample mean is in the interval $\left[22.07 \le \overline{x} \le 25.33\right]$. Therefore the null hypothesis cannot be rejected at the 5% significance level.

Exercise 8.4 - Solution

a.

1. Formulate the null and alternate hypotheses:

$H_0: \quad \mu_X = 40 \; hour/week$

$H_1: \quad \mu_X \neq 40 \; hour/week$

2. Formulate an operating rule:

The operating rule is stated as: The null hypothesis cannot be rejected at the α significance level if the following is satisfied:

$-k_{\alpha/2} < \dfrac{\overline{X}-\mu}{\sigma_X \frac{1}{\sqrt{n}}} < k_{\alpha/2}$, where n is the number of measurements

3. Choose the level of significance α:

$\alpha = 0.05$

4. Determine the condition of sampling (what kind of and how many data?)

Weekly working hours of 9 workers.

5. Do the calculations:

$$-k_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma_X \frac{1}{\sqrt{n}}} < k_{\alpha/2} \Leftrightarrow -\Phi^{-1}\left(1-\frac{0.05}{2}\right) < \frac{\overline{X}-40}{\sqrt{9.5}\frac{1}{\sqrt{9}}} < \Phi^{-1}\left(1-\frac{0.05}{2}\right) \Leftrightarrow$$

$$-1.96 < \frac{\overline{X}-40}{\sqrt{9.5}\frac{1}{\sqrt{9}}} < 1.96 \Leftrightarrow 37.99 \; hours < \overline{X} < 42.01 \; hours$$

The null hypothesis cannot be rejected if the sample mean of the weekly working hour of the 9 workers lies between 37.99 hours and 42.01 hours.

6. Obtain the sample mean.

The sample mean is:

$$\frac{1}{9}\cdot(39+41+40+42+43+40+39+37+43) = 40.33 \text{ hour/week}$$

7. Judge the null hypothesis H_0.

The sample mean lies within the interval $[37.99; 42.01]$ and hence the null hypothesis cannot be rejected at the 5% significance level.

b.

1. Formulate the null and alternate hypotheses:

$H_0: \quad \mu_X \le \mu_Y$

$H_1: \quad \mu_X > \mu_Y$

2. Formulate an operating rule:

The null hypothesis cannot be rejected at the α significance level if the following condition is satisfied: $\overline{X} - \overline{Y} \le \Delta$ where Δ is a critical value to be determined in the following.

3. Choose the level of significance α:

$\alpha = 0.05$.

4. Determine the condition of sampling (what kind of and how many data?)

The weekly working hours of 9 workers before and after the installation of the new rule respectively.

5. Do the calculations:

The critical value Δ is obtained as:

$$P\left[\overline{X} - \overline{Y} \le \Delta\right] = 1 - 0.05 \Rightarrow \Phi\left(\frac{\Delta - \mu_{\overline{X}-\overline{Y}}}{\sigma_{\overline{X}-\overline{Y}}}\right) = 0.95 \Rightarrow \Phi\left(\frac{\Delta - 0}{\sqrt{\frac{\sigma_X^2}{k} + \frac{\sigma_Y^2}{l}}}\right) = 0.95 \Rightarrow$$

$$\Phi\left(\frac{\Delta}{\sqrt{\frac{9.5}{9} + \frac{9.5}{9}}}\right) = 0.95 \Rightarrow \Delta = 2.39$$

The null hypothesis cannot be rejected at the 5% significance level if the difference of the mean values of the random variables X and Y is smaller or equal to 2.39.

6. Obtain the sample mean difference.

$\overline{x} = 40.33$

$\overline{y} = 39.33$

$z = \overline{x} - \overline{y} = 40.33 - 39.33 = 1.00$ hours.

7. Judge the null hypothesis H_0.

Since it is $\overline{x} - \overline{y} = 1.00 < 2.39$, the null hypothesis cannot be rejected at the 5% significance level.

Exercise 8.5 - Solution

a and b. The probability density function and the cumulative distribution functions are

$$f_X(x) = \begin{cases} \frac{2}{10000^2} x & 0 \le x \le 10000 \\ 0 & otherwise \end{cases} \quad (8.5.1)$$

$$F_X(x) = \begin{cases} 0 & 0 \le x \\ \left(\frac{x}{10000}\right)^2 & 0 < x \le 10000 \\ 1 & x > 10000 \end{cases} \quad (8.5.2)$$

Taking the square root of both sides of Equation (8.5.1), a linear relationship between the square root of $F_X(x)$ and x is obtained:

$$F_X(x) = \left(\frac{x}{10000}\right)^2 \Leftrightarrow \sqrt{F_X(x)} = \frac{x}{10000} \quad (8.5.3)$$

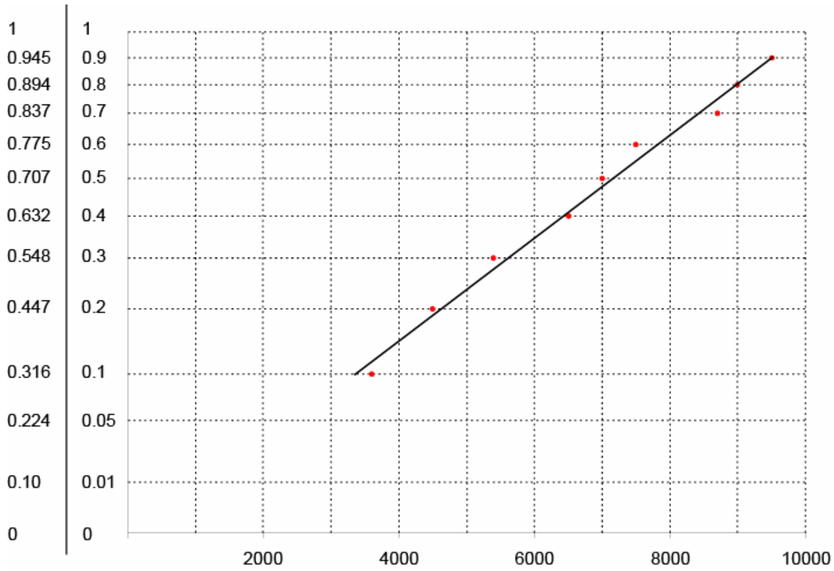
For values of the cumulative distribution function in the interval $[0;1]$ the following table is obtained:

$\sqrt{F_X(x)}$	$F_X(x)$
0	0
0.31	0.1
0.45	0.2
0.55	0.3
0.63	0.4
0.71	0.5
0.77	0.6
0.84	0.7
0.89	0.8
0.94	0.9
1.0	1.0

With the help of the above table the probability paper is created by rescaling the y-axis. Plot the data in the probability paper. If the data fit on a straight line the data follow the triangular distribution. The cumulative distribution function used to plot the data is obtained in the following table.

i	No. of cars	$F_X(x_i^o) = \frac{i}{N+1}$
1	3600	0.1
2	4500	0.2
3	5400	0.3
4	6500	0.4
5	7000	0.5
6	7500	0.6
7	8700	0.7
8	9000	0.8
9	9500	0.9

It is seen that the data fit well on a straight line and hence the hypothesis of the triangular distribution can be accepted.



Exercise 8.6 - Solution

a. The cumulative distribution function of the exponential distribution is written as:

$$F_T(t) = 1 - \exp(-\lambda t)$$

The complementary cumulative distribution function is expressed by:

$$G_T(t) = 1 - F_T(t) = \exp(-\lambda t)$$

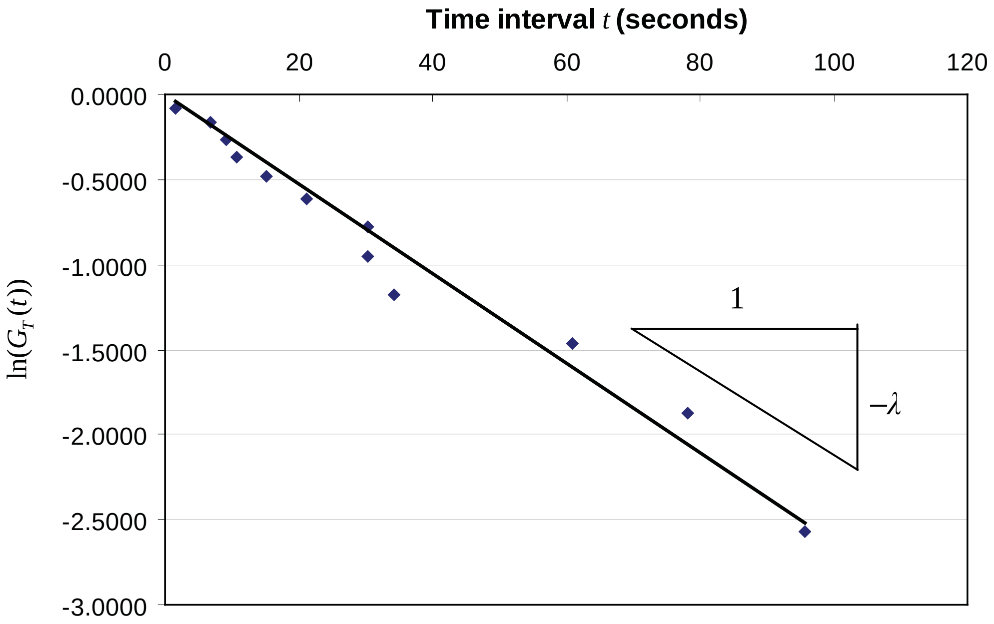
By taking the natural logarithm the following relation is obtained:

$$\ln(G_T(t)) = -\lambda t$$

The pair $(t, \ln G(t))$ represents a straight line with a slope of $-\lambda$. The assumption that the time interval is exponentially distributed is checked by plotting the data in the probability paper. To do so the following calculation sheet is used:

i	Time interval (seconds)	$F_T(t_i^o)$	$1-F_T(t_i^o)$	$\ln(G_T(t))$
1	1.52	0.077	0.923	-0.080
2	6.84	0.154	0.846	-0.167
3	9.12	0.231	0.769	-0.262
4	10.64	0.308	0.692	-0.368
5	15.2	0.385	0.615	-0.486
6	21.28	0.462	0.538	-0.619
7	30.4	0.538	0.462	-0.773
8	30.4	0.615	0.385	-0.956
9	34.2	0.692	0.308	-1.179
10	60.8	0.769	0.231	-1.466
11	78.28	0.846	0.154	-1.872
12	95.76	0.923	0.077	-2.565

The probability paper with the plotted data is shown in the following figure: The data fit well on a straight line and hence it is reasonable to say that the time interval of car arrivals is exponentially distributed.



b. The sample mean is obtained as:

$$\bar{x} = \frac{1}{12}(1.52 + 6.84 + 9.12 + 10.64 + 15.2 + 21.28 + 30.4 + 30.4 + 34.2 + 60.8 + 78.28 + 95.76)$$
$$= 32.87 \text{ seconds}$$

Since the negative slope of the line in the above figure corresponds to the parameter λ, the parameter is estimated from the figure as 0.026. The mean value is then estimated from the following relation:

$$\frac{1}{\hat{\lambda}} = \frac{1}{0.026} = 38.1 \text{ seconds}$$

EXERCISE TUTORIAL 9 - SOLUTION

Exercise 9.1 - Solution

a. The likelihood function is written as:

$$L(\mu,\sigma)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

where x_i is the i^{th} observation of concrete compressive strength. The log likelihood function is written as:

$$l=\ln(L)=\sum_{i=1}^{n}\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right)=-n\ln\left(\sqrt{2\pi}\right)-n\ln(\sigma)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

b. The estimators with the maximum likelihood method are obtained by solving the following equations simultaneously.

$$\frac{\partial l(\mu,\sigma)}{\partial\mu}=0 \text{ and } \frac{\partial l(\mu,\sigma)}{\partial\sigma}=0$$

$$\frac{\partial l}{\partial\mu}=\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)=0$$

$$\Leftrightarrow\sum_{i=1}^{n}x_i-n\mu=0$$

$$\Leftrightarrow\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\frac{\partial l}{\partial\sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2=0$$

$$\Leftrightarrow$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

By substituting the numbers of x_i,

$$\hat{\mu}=\frac{1}{20}(24.4+27.6+...+39.7)=\frac{1}{20}\times653.3=32.67$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2=4.04.$$

c. Analytical moments are obtained as:

$$\lambda_1 = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu$$

$$\lambda_2 = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sigma^2 + \mu^2$$

Sample moments are obtained from the data as:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = 32.67 \quad \text{and} \qquad m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{21674.6}{20} = 1083.7$$

By equating the analytical moments and the sample moments,

$$(\lambda_1 =)\mu = 32.67 (= m_1)$$

$$(\lambda_2 =)\mu^2 + \sigma^2 = 1083.7 (= m_2)$$

The estimates are thus $\hat{\mu} = 32.67$ and $\hat{\sigma} = 4.04$.

Exercise 9.2 – Solution

a. The likelihood function is written as:

$$L = \prod_{i=1}^{n} \lambda \exp(-\lambda x)$$

and the log likelihood function is written as:

$$l = \ln(L) = \sum_{i=1}^{n} \left(\ln\left(\lambda \exp(-\lambda x_i)\right)\right) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

The maximum likelihood estimator is obtained as:

$$\frac{dl}{d\lambda} = 0$$

$$\Leftrightarrow \frac{dl}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\Leftrightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{20}{653.3} = 0.031$$

b. Whereas it is almost always possible to estimate the parameters of distributions by means of the maximum likelihood method or the method of moment, it does not necessarily mean that the distribution drawn with the estimated parameters fits the data well, see Figure 9.2.1.

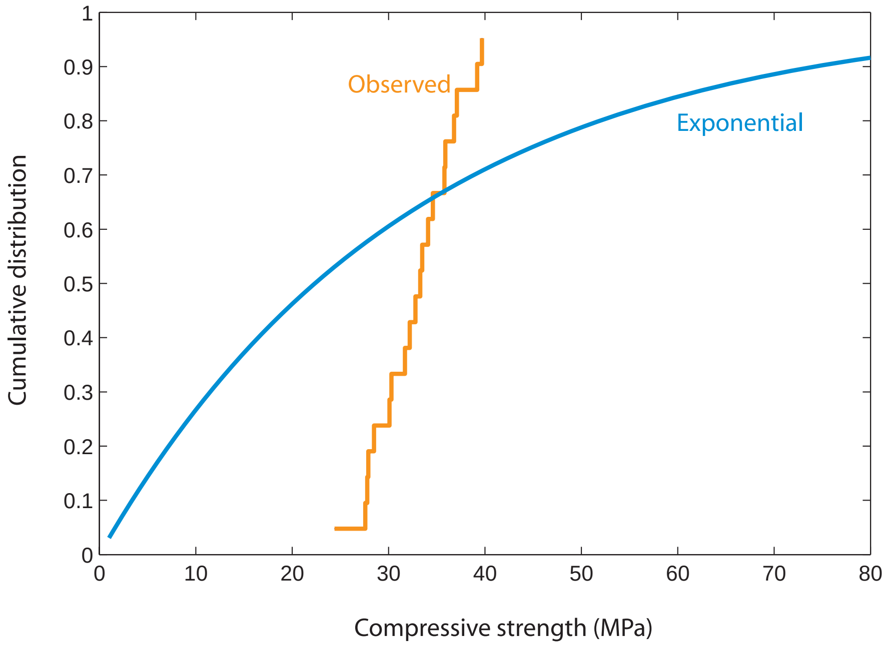


Figure 9.2.1: Cumulative distribution function and observed cumulative distribution.

EXERCISE TUTORIAL 10 - SOLUTION

Exercise 10.1 – Solution

a.

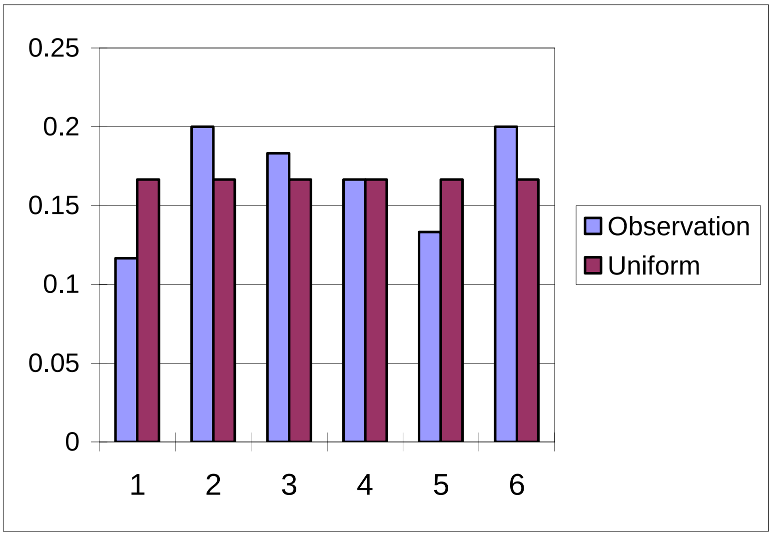


Figure 10.1.1: Histogram of observations and uniform mass probability.

b. $$P[N_{o,j}=10,\ j=1,2,3,4,5,6]=\frac{60!}{(10!)^6}\left(\frac{1}{6}\right)^{60}=0.0000745\,.$$

Remark that the probability that the observations of the resulting side distribute uniformly is very small even if the dice is symmetric.

c. The null hypothesis that the dice is symmetric is expressed as: $p(x_j)=1/6,\ (j=1,2,3,4,5,6)$.

The sample statistic is:

$\varepsilon_m^2=\sum_{j=1}^{6}\frac{(N_{o,j}-N_{p,j})^2}{N_{p,j}}$, where $N_{p,j}=np(x_j)$, with n being the number of total trials and $N_{p,j}$ the number of outcomes of side j.

The operating rule, i.e. the critical value Δ to which the sample statistic shall be compared in order to judge the null hypothesis, is: $P(\varepsilon_m^2\geq\Delta)=\alpha$, stating that the null hypothesis shall be rejected, at the α% significance level, if the sample statistic is larger or equal to the critical value. The sample statistic follows the Chi-square distribution with 6-1=5 degrees of freedom. At the 5% significant level, the null hypothesis shall be rejected if the sample statistic is larger than 11.07, see the probability table for the Chi-square distribution (Annex T, Table T.3).

Since the sample statistic is obtained as $\varepsilon_m^2 = 2.20 \le \Delta = 11.07$ from the observations, see Table 10.1.2, the null hypothesis that the dice is symmetric cannot be rejected at the 5% significance level.

Side	$N_{o,j}$	$p(x_j)$	$N_{p,j} = np(x_j)$	ε_m^2
1	7	9	10	9/10
2	12	4	10	4/10
3	11	1	10	1/10
4	10	0	10	0/10
5	8	4	10	4/10
6	12	4	10	4/10
Sum	60			= 2.20

Table 10.1.2: Calculation sheet for the χ^2 - goodness of fit test.

Exercise 10.2 – Solution

a. The parameters are estimated as:

$$\hat{\mu} = m_1 = 32.67$$

$$\hat{\sigma} = \sqrt{m_2 - m_1^2} = \sqrt{1083.4 - 32.67^2} = 4.04 .$$

b. The sample statistic for the χ^2 goodness-of-fit test is given as:

$\varepsilon_m^2 = \sum_{j=1}^{k} \frac{(N_{o,j} - N_{p,j})^2}{N_{p,j}}$, where $N_{p,j} = np(x_j)$, with *n* being the number of total trials and $N_{p,j}$ the number of outcomes within a certain interval.

The operating rule, i.e. the critical value Δ to which the sample statistic shall be compared in order to judge the null hypothesis, is: $P(\varepsilon_m^2 \ge \Delta) = \alpha$, stating that the null hypothesis shall be rejected, at the α% significance level, if the sample statistic is larger or equal to the critical value.

The sample statistic follows the Chi-square distribution with 4-1-2=1 degree of freedom. At the 5% significant level, the null hypothesis shall be rejected if the sample statistic is larger than 3.84, see the probability table for the Chi-square distribution (Annex T, Table T.3).

Since the sample statistic is obtained as $\varepsilon_m^2 = 0.163 < \Delta = 3.84$ from the observations, see Table 10.2.2, the null hypothesis that the dice is symmetric cannot be rejected at the 5% significance level.

Interval	$N_{o,j}$	$p(x_j)$	$N_{p,j} = np(x_j)$	ε_m^2
-30	5	$\Phi\left(\frac{30-32.67}{4.04}\right) = 0.254$	5.08	0.001
30-33	5	$\Phi\left(\frac{33-32.67}{4.04}\right) - \Phi\left(\frac{30-32.67}{4.04}\right) = 0.278$	5.56	0.06
33-36	6	$\Phi\left(\frac{36-32.67}{4.04}\right) - \Phi\left(\frac{33-32.67}{4.04}\right) = 0.263$	5.26	0.10
36-	4	$1-\Phi\left(\frac{36-32.67}{4.04}\right) = 0.205$	4.10	0.002
Sum	20			= 0.163

Table 10.2.3: Calculation sheet for the χ^2 - goodness of fit test.

Exercise 10.3:

a.

i	x_i	$F_o(x_i^o) = \frac{i}{n}$	$F_p(x_i^o)$	$\lvert F_o(x_i^o) - F_p(x_i^o)\rvert$
1	6.33	0.05	0.121891	0.071891
2	6.85	0.1	0.258987	0.158987
3	7.17	0.15	0.369813	0.219813
4	7.41	0.2	0.464813	0.264813
5	7.57	0.25	0.526882	0.276882
6	7.81	0.3	0.62297	0.32297
7	7.86	0.35	0.641923	0.291923
8	7.90	0.4	0.657023	0.257023
9	7.96	0.45	0.678091	0.228091
10	8.06	0.5	0.712015	0.212015
11	8.11	0.55	0.73037	0.18037
12	8.13	0.6	0.734124	0.134124
13	8.17	0.65	0.750047	0.100047
14	8.29	0.7	0.784557	0.084557
15	8.33	0.75	0.795964	0.045964
16	8.73	0.8	0.889861	0.089861
17	9.07	0.85	0.941416	0.091416
18	9.19	0.9	0.954406	0.054406
19	9.19	0.95	0.954574	0.004574
20	10.18	1	0.996354	0.003646

Table 10.3.2: Calculation sheet for the Kolmogorov-Smirnov goodness of fit test.

The sample statistic for the Kolmogorov-Smirnov goodness of fit test is calculated with the help of Table 10.3.2 and it is: $\varepsilon_{max} = \max_{i=1}^{n}\left[\left|F_o(x_i^o) - F_p(x_i^o)\right|\right] = 0.32297$.

The operating rule, i.e. the critical value Δ to which the sample statistic shall be compared in order to judge the null hypothesis, is: $P(\varepsilon_{max} \geq \Delta) = \alpha$, stating that the null hypothesis shall be rejected, at the α % significance level, if the sample statistic is larger or equal to the critical value.

At the 1% significant level and *n*=20, the null hypothesis shall be rejected if the sample statistic is larger than 0.352, (Annex T, Table T.4). Since the sample statistic is obtained as $\varepsilon_{max} = 0.322 \leq \Delta = 0.352$ from the observations, see Table 10.3.2, the null hypothesis cannot be rejected at the 1% significance level.

b.

The sample statistic for the Kolmogorov-Smirnov goodness of fit test is calculated with the help of Table 10.3.3 and it is: $\varepsilon_{max} = \max_{i=1}^{n}\left[\left|F_o(x_i^o) - F_p(x_i^o)\right|\right] = 0.28297$.

The operating rule, i.e. the critical value Δ to which the sample statistic shall be compared in order to judge the null hypothesis, is: $P(\varepsilon_{max} \geq \Delta) = \alpha$, stating that the null hypothesis shall be rejected, at the α % significance level, if the sample statistic is larger or equal to the critical value.

At the 1% significant level and *n*=50, the null hypothesis shall be rejected if the sample statistic is larger than 0.352, (Annex T, Table T.4). Since the sample statistic is obtained as $\varepsilon_{max} = 0.28297 > \Delta = 0.231$ from the observations, see Table 10.3.2, the null hypothesis shall be rejected at the 1% significance level.

i	x_i	$F_o(x_i^o) = \frac{i}{n}$	$F_p(x_i^o)$	$\left|F_o(x_i^o) - F_p(x_i^o)\right|$
1	5.83	0.02	0.047393	0.027393
2	6.33	0.04	0.121891	0.081891
3	6.40	0.06	0.134778	0.074778
4	6.41	0.08	0.137037	0.057037
5	6.56	0.1	0.173362	0.073362
6	6.66	0.12	0.201526	0.081526
7	6.80	0.14	0.241197	0.101197
8	6.85	0.16	0.258987	0.098987
9	6.94	0.18	0.288944	0.108944
10	7.08	0.2	0.336548	0.136548
11	7.17	0.22	0.369813	0.149813
12	7.19	0.24	0.380149	0.140149

13	7.31	0.26	0.423959	0.163959
14	7.41	0.28	0.464813	0.184813
15	7.57	0.3	0.526882	0.226882
16	7.60	0.32	0.539873	0.219873
17	7.81	0.34	0.62297	0.28297
18	7.84	0.36	0.634307	0.274307
19	7.86	0.38	0.641923	0.261923
20	7.90	0.4	0.657023	0.257023
21	7.94	0.42	0.670326	0.250326
22	7.96	0.44	0.678091	0.238091
23	7.98	0.46	0.684461	0.224461
24	8.06	0.48	0.712015	0.232015
25	8.11	0.5	0.73037	0.23037
26	8.13	0.52	0.734124	0.214124
27	8.17	0.54	0.750047	0.210047
28	8.22	0.56	0.764028	0.204028
29	8.26	0.58	0.775566	0.195566
30	8.29	0.6	0.784557	0.184557
31	8.29	0.62	0.786522	0.166522
32	8.33	0.64	0.795964	0.155964
33	8.53	0.66	0.8482	0.1882
34	8.57	0.68	0.857948	0.177948
35	8.67	0.7	0.878718	0.178718
36	8.69	0.72	0.882976	0.162976
37	8.71	0.74	0.887226	0.147226
38	8.71	0.76	0.887688	0.127688
39	8.73	0.78	0.889861	0.109861
40	8.82	0.8	0.90585	0.10585
41	8.86	0.82	0.912768	0.092768
42	9.07	0.84	0.941416	0.101416
43	9.19	0.86	0.954406	0.094406
44	9.19	0.88	0.954566	0.074566
45	9.19	0.9	0.954574	0.054574
46	9.25	0.92	0.960285	0.040285
47	9.29	0.94	0.963293	0.023293
48	9.42	0.96	0.972263	0.012263
49	9.62	0.98	0.983147	0.003147
50	10.18	1	0.996354	0.003646

Table 10.3.3: Calculation sheet for the Kolmogorov-Smirnov goodness of fit test.

c. The sample statistic for the Kolmogorov-Smirnov goodness of fit test is calculated with the help of Table 10.3.2 and it is: $\varepsilon_{max} = \max_{i=1}^{n}\left[\left|F_o(x_i^o) - F_p(x_i^o)\right|\right] = 0.32297$.

The operating rule, i.e. the critical value Δ to which the sample statistic shall be compared in order to judge the null hypothesis, is: $P(\varepsilon_{max} \geq \Delta) = \alpha$, stating that the null hypothesis shall be rejected, at the α% significance level, if the sample statistic is larger or equal to the critical value. At the 5% significant level and *n*=20, the null hypothesis shall be rejected if the sample statistic is larger than 0.294, (Annex T, Table T.4). Since the sample statistic is obtained as $\varepsilon_{max} = 0.32297 > \Delta = 0.294$ from the observations the null hypothesis shall be rejected at the 5% significance level.

d. The sample statistic for the Kolmogorov-Smirnov goodness of fit test is calculated with the help of Table 10.3.2 and it is: $\varepsilon_{max} = \max_{i=1}^{n}\left[\left|F_o(x_i^o) - F_p(x_i^o)\right|\right] = 0.28297$.

The operating rule, i.e. the critical value Δ to which the sample statistic shall be compared in order to judge the null hypothesis, is: $P(\varepsilon_{max} \geq \Delta) = \alpha$, stating that the null hypothesis shall be rejected, at the α% significance level, if the sample statistic is larger or equal to the critical value. At the 5% significant level and *n*=50, the null hypothesis shall be rejected if the sample statistic is larger than 0.192, (Annex T, Table T.4). Since the sample statistic is obtained as $\varepsilon_{max} = 0.28297 > \Delta = 0.192$ from the observations the null hypothesis shall be rejected at the 5% significance level.

Exercise 10.4:

a. The first and the second sample moments are:

$m_1 = 26.41$

$m_2 = 747.55$

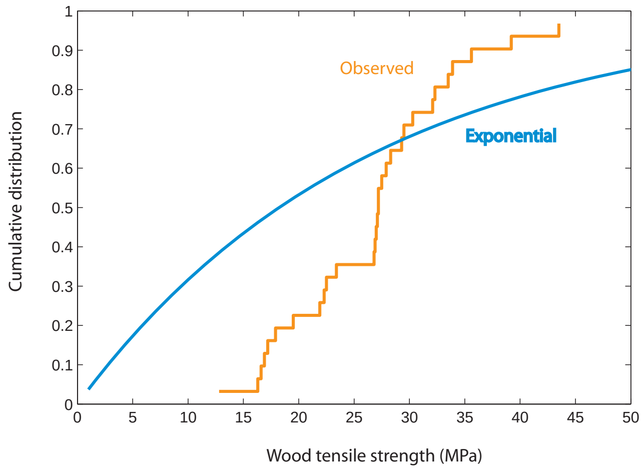
The exponential distribution has the following cumulative distribution function:

$F_X(x) = 1 - \exp(-\lambda x), \qquad x > 0$

The first analytical moment μ_1 is: $\mu_1 = \dfrac{1}{\lambda}$

Equating m_1 and μ_1, the parameter λ is estimated as: $m_1 = \mu_1 \Leftrightarrow \hat{\lambda} = \dfrac{1}{m_1} = 0.038$

b. The cumulative distribution function is shown in the following figure. Remark that the model of the exponential distribution is quite poor, although it is possible to estimate the parameter in the exponential distribution with the method of moment.



c. The sample statistic for the χ^2 – goodness of fit test is: $\varepsilon_m^2 = \sum_{j=1}^{k} \frac{(N_{o,j} - N_{p,j})^2}{N_{p,j}}$, where $N_{p,j} = np(x_j)$, with n being the number of total measurements and $N_{p,j}$ the number of measurements within a certain interval. The operating rule, i.e. the critical value Δ to which the sample statistic shall be compared in order to judge the null hypothesis, is: $P(\varepsilon_m^2 \geq \Delta) = \alpha$, stating that the null hypothesis shall be rejected, at the α% significance level, if the sample statistic is larger or equal to the critical value.

The sample statistic follows the Chi-square distribution with 4-1-1=2 degrees of freedom. At the 10% significant level, the null hypothesis shall be rejected if the sample statistic is larger than 4.6, see the probability table for the Chi-square distribution (Annex T, Table T.3). Since the sample statistic is obtained as $\varepsilon_m^2 = 43.55 > \Delta = 4.6$ from the observations, see Table 10.4.3, the null hypothesis that the shall be rejected at the 10% significance level.

Interval	$N_{o,j}$	$p(x_j)$	$N_{p,j} = np(x_j)$	ε_m^2
-20	7	0.49	14.7	4.03
20-25	4	0.08	2.4	1.07
25-30	11	0.07	2.1	37.72
30-	8	0.36	10.8	0.73
Sum	30			43.55

Table 10.4.3: Calculation sheet for the χ^2 - goodness of fit test.

c. The sample statistic for the Kolmogorov-Smirnov goodness of fit test is calculated with the help of Table 10.4.4 and it is: $\varepsilon_{max} = \max_{i=1}^{n}\left[\left|F_o(x_i^o) - F_p(x_i^o)\right|\right] = 0.412$.

The operating rule, i.e. the critical value Δ to which the sample statistic shall be compared in order to judge the null hypothesis, is: $P(\varepsilon_{max} \geq \Delta) = \alpha$, stating that the null hypothesis shall be rejected, at the α% significance level, if the sample statistic is larger or equal to the critical value.

At the 10% significance level and n=30, the null hypothesis shall be rejected if the sample statistic is larger than 0.22, (Annex T, Table T.4). Since the sample statistic is obtained as $\varepsilon_{max} = 0.412 > \Delta = 0.22$ from the observations the null hypothesis shall be rejected at the 10% significance level.

i	x_i	$F_o(x_i^o) = \frac{i}{n}$	$F_p(x_i^o)$	$\lvert F_o(x_i^o) - F_p(x_i^o) \rvert$
1.0	12.8	0.033	0.401	0.367
2.0	16.3	0.067	0.479	**0.412**
3.0	16.6	0.100	0.485	0.385
4.0	16.9	0.133	0.491	0.358
5.0	17.2	0.167	0.497	0.331
6.0	17.9	0.200	0.511	0.311
7.0	19.5	0.233	0.542	0.308
8.0	21.9	0.267	0.584	0.317
9.0	22.3	0.300	0.590	0.290
10.0	22.5	0.333	0.593	0.260
11.0	23.4	0.367	0.608	0.241
12.0	26.8	0.400	0.658	0.258
13.0	26.9	0.433	0.659	0.226
14.0	27.0	0.467	0.660	0.194
15.0	27.1	0.500	0.662	0.162
16.0	27.2	0.533	0.663	0.130
17.0	27.2	0.567	0.663	0.096
18.0	27.5	0.600	0.667	0.067
19.0	27.9	0.633	0.672	0.039
20.0	28.3	0.667	0.678	0.011
21.0	29.3	0.700	0.690	0.010
22.0	29.5	0.733	0.693	0.041
23.0	30.3	0.767	0.702	0.064
24.0	32.1	0.800	0.723	0.077
25.0	32.3	0.833	0.725	0.108
26.0	33.5	0.867	0.738	0.129
27.0	33.9	0.900	0.742	0.158
28.0	35.6	0.933	0.759	0.174
29.0	39.2	0.967	0.792	0.175
30.0	43.5	1.000	0.824	0.176

Table 10.4.4: Calculation sheet for the Kolmogorov-Smirnov goodness of fit test.

EXERCISE TUTORIAL 11 - SOLUTION

Exercise 11.2 – Solution

a. Based on the information provided the following event tree is constructed for carrying out the prior analysis:

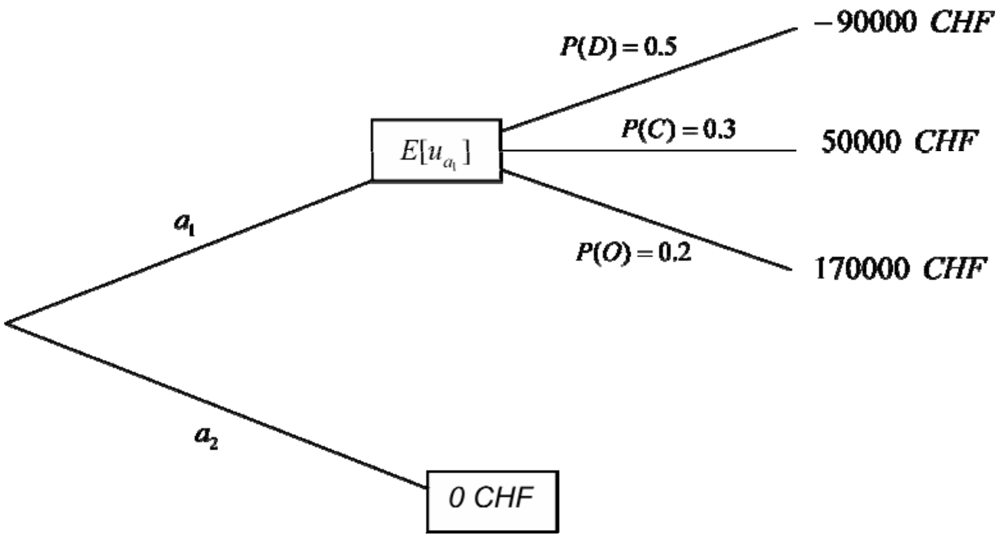


Figure 11.2.1: Event tree for carrying out the prior decision analysis.

The benefit associated with the opening of the borehole, a-priori, is estimated as follows:

$$E'[u_{a_1}] = P'[D]\cdot(-90000) + P'[D]\cdot(15000) + P'[O]\cdot(170000)$$
$$= 0.5\cdot(-90000) + 0.3\cdot(15000) + 0.2\cdot(170000)$$
$$= 4000\ CHF$$

Hence the action that gives the larger utility (larger expected benefit in terms of cost) is action a_1,

$$E'[u] = \max\left\{E'\left[u_{a_1}\right];\ E'\left[u_{a_2}\right]\right\} = \max\{4000;\ 0\} = 4000\ CHF$$

and hence a-priori the engineer would decide to open up the borehole.

b. and c.

The event tree is now extended to include the cases of performing a test, a_{11}, or not performing a test, a_{12}. The following probabilities can readily be estimated:

In case that the test is carried out the probability of receiving the indication that the well is dry is:

$$P'(I_D) = P(I_D \mid D)\cdot P'(D) + P(I_C \mid C)\cdot P'(C) + P(I_O \mid O)\cdot P'(O)$$
$$= 0.6\cdot 0.5 + 0.3\cdot 0.3 + 0.1\cdot 0.2 = 0.41$$

The probabilities of the states of the well are updated given the above indication:

$$P''(D \mid I_D) = \frac{P(I_D \mid D)\cdot P'(D)}{P'(I_D)} = \frac{0.6\cdot 0.5}{0.41} = \frac{0.3}{0.41} = 0.732$$

$$P''(C \mid I_D) = \frac{P(I_D \mid C)\cdot P'(C)}{P'(I_D)} = \frac{0.3\cdot 0.3}{0.41} = \frac{0.09}{0.41} = 0.220$$

$$P''(O \mid I_D) = \frac{P(I_D \mid O)\cdot P'(O)}{P'(I_D)} = \frac{0.1\cdot 0.2}{0.41} = \frac{0.02}{0.41} = 0.048$$

Similarly for the other two possible outcomes of the test it is:

$$P'(I_C) = P(I_C \mid D)\cdot P'(D) + P(I_C \mid C)\cdot P'(C) + P(I_C \mid O)\cdot P'(O)$$
$$= 0.1\cdot 0.5 + 0.3\cdot 0.3 + 0.5\cdot 0.2 = 0.24$$

$$P''(D \mid I_C) = \frac{P(I_C \mid D)\cdot P'(D)}{P'(I_C)} = \frac{0.1\cdot 0.5}{0.24} = \frac{0.05}{0.24} = 0.208$$

$$P''(C \mid I_C) = \frac{P(I_C \mid C)\cdot P'(C)}{P'(I_C)} = \frac{0.3\cdot 0.3}{0.24} = \frac{0.09}{0.24} = 0.375$$

$$P''(O \mid I_C) = \frac{P(I_C \mid O)\cdot P'(O)}{P'(I_C)} = \frac{0.5\cdot 0.2}{0.24} = \frac{0.1}{0.24} = 0.417$$

$$P'(I_O) = P(I_O \mid D)\cdot P'(D) + P(I_O \mid C)\cdot P'(C) + P(I_O \mid O)\cdot P'(O)$$
$$= 0.3\cdot 0.5 + 0.4\cdot 0.3 + 0.4\cdot 0.2 = 0.35$$

$$P''(D \mid I_O) = \frac{P(I_O \mid D)\cdot P'(D)}{P'(I_O)} = \frac{0.3\cdot 0.5}{0.35} = \frac{0.15}{0.35} = 0.429$$

$$P''(C \mid I_O) = \frac{P(I_O \mid C)\cdot P'(C)}{P'(I_O)} = \frac{0.4\cdot 0.3}{0.35} = \frac{0.12}{0.35} = 0.343$$

$$P''(O \mid I_O) = \frac{P(I_O \mid O)\cdot P'(O)}{P'(I_O)} = \frac{0.4\cdot 0.2}{0.35} = \frac{0.08}{0.35} = 0.228$$

The expected utility can be written:

$$E[u] = \sum_{i=1}^{n} P'[I_i]E''[u \mid I_i] = \sum_{i=1}^{n} P'[I_i] \max_{j=1,\dots,m}\left\{E''[u(a_j) \mid I_i]\right\}$$

Where n is the number of different possible experiment findings and m is the number of different decision alternatives. So it is:

$$E''[u|I_D] = \max\{P''[D|I_D](-90000) + P''[C|I_D](50000) + P''[O|I_D](170000); 0]\} =$$
$$= \max\{0.732(-90000)+0.220(50000)+0.048(170000); 0\} = \max\{-46720; 0\} =$$
$$=0\ CHF$$

Similarly:

$$E''[u|I_C] = \max\{P''[D|I_C](-90000) + P''[C|I_C](50000) + P''[O|I_C](170000); 0]\} =$$
$$= \max\{0.208(-90000)+0.375(50000)+0.417(170000); 0\} = \max\{-46720; 0\} =$$
$$=70920\ CHF$$

$$E''[u|I_C] = \max\{P''[D|I_O](-90000) + P''[C|I_O](50000) + P''[O|I_O](170000); 0]\} =$$
$$= \max\{0.429(-90000)+0.343(50000)+0.228(170000); 0\} = \max\{17300; 0\} =$$
$$=17300\ CHF$$

And the expected utility considering the costs of the test is:

$$E[u] = \{E''[u \mid I_D]\cdot P'(I_D) + E''[u \mid I_C]\cdot P'(I_C) + E''[u \mid I_O]\cdot P'(I_O)\} - 10000 =$$
$$= \{(0)\cdot 0.41 + (70920)\cdot 0.24 + (17300)\cdot 0.35\} - 10000 =$$
$$= 23076 - 10000 = 13076\ CHF$$

Hence if this is compared to the case of not carrying out the test it can be seen that the utility is higher in the case that the test is carried out.

The benefit associated from opening up the borehole is then equal to 13076 CHF.

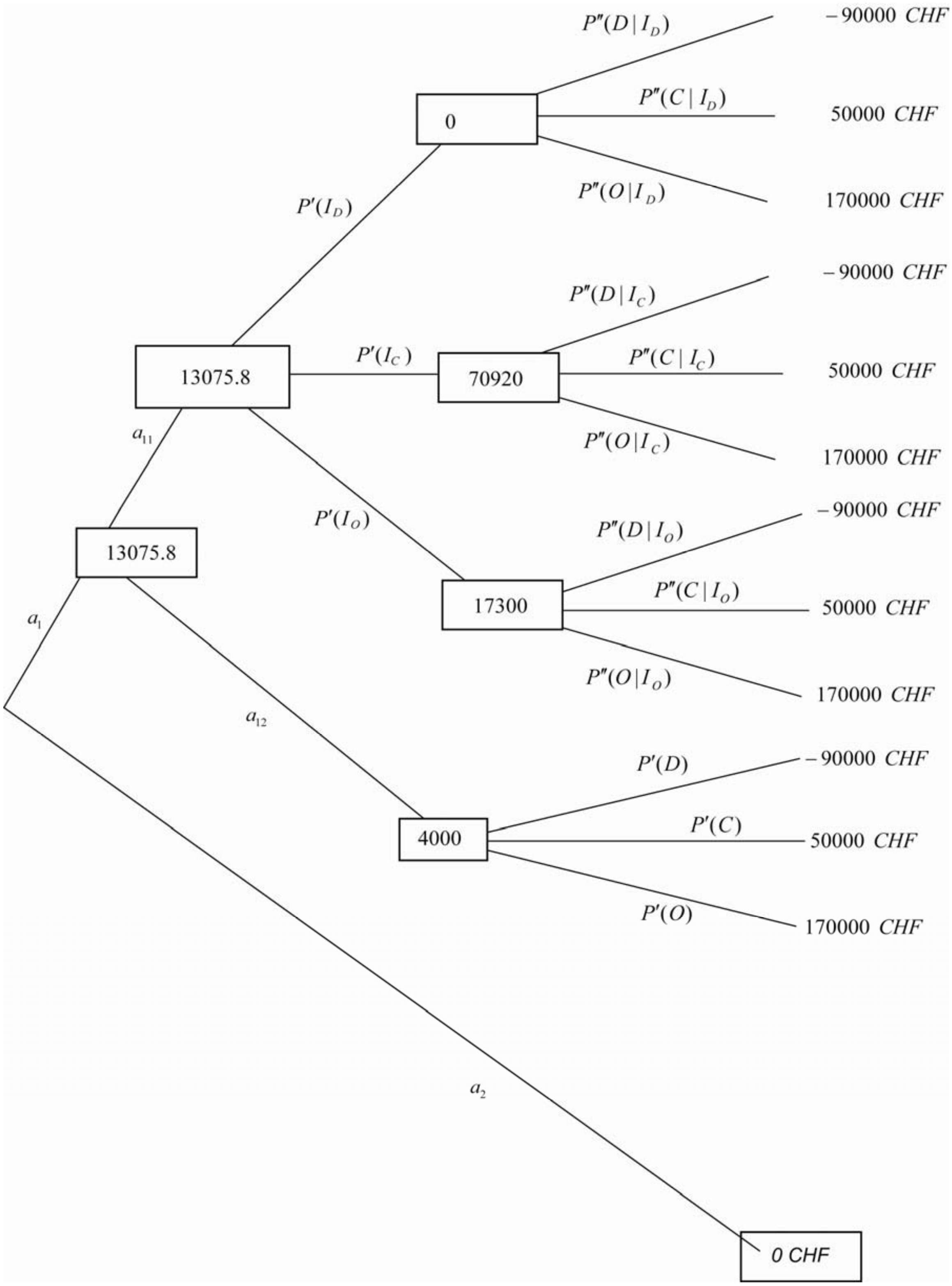


Figure 11.2.2: Event tree for carrying out the pre-posterior decision analysis.

Exercise 11.3 – Solution

a. The company should consider the following two choices:

A_1: Develop a well locally.

A_2: Construct a pipeline.

The capacity of the well is associated with uncertainty and there are two cases:

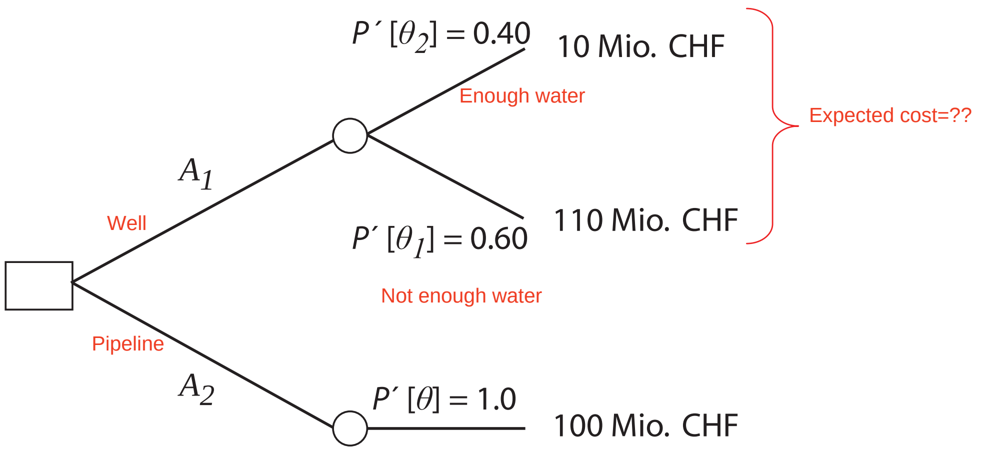
θ_1: Capacity less than100 kl.

θ_2: Capacity greater than 100 kl.

Based on experience the prior probabilities of the above mentioned cases are:

$P'[\theta_1] = 0.60$

$P'[\theta_2] = 0.40$

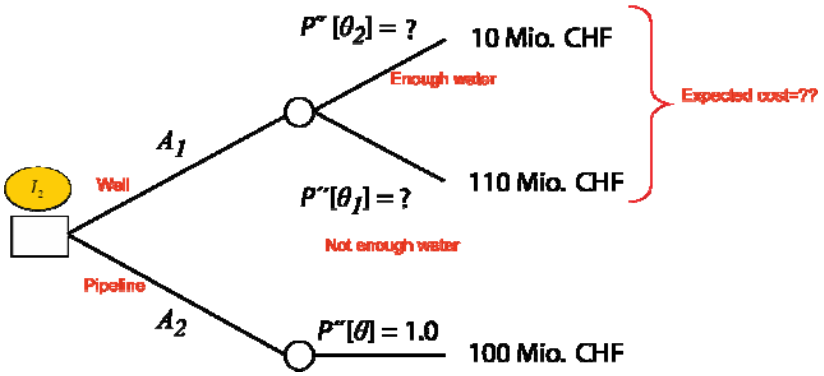


The minimized expected cost is:

$$E'[C] = \min \ \{P'[\theta_1]\cdot 10 + P'[\theta_2]\cdot(100+10), \ \ 100\} =$$
$$\min \ \{0.4\cdot 10 + 0.6\cdot 110, \ 100\} = 70 \quad \text{Mio. CHF}$$

Action A_1 poses less expected cost and hence the company should develop a well locally.

b.



According to the Bayes' theorem:

Likelihood

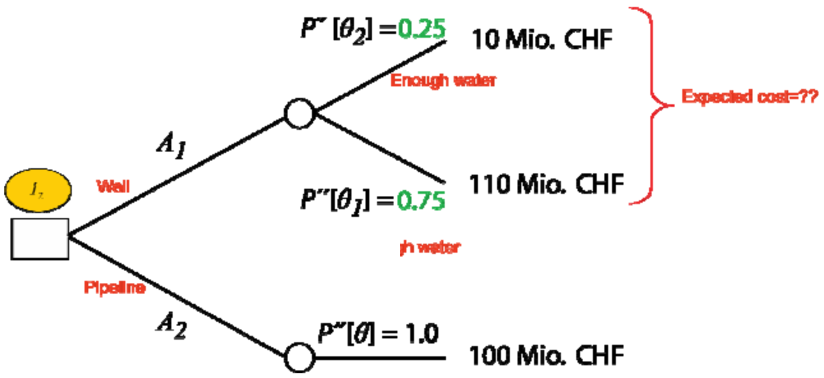
$$\underline{P(X \mid E)} = \frac{P(X \cap E)}{P(E)} = \frac{\underline{P(E \mid X)}}{P(E \mid X_1)P(X_1) + \cdots P(E \mid X_n)P(X_n)}\underline{P(X)}$$

Posterior prob. Prior prob.

Given the indicator I_2:

$$P''(\theta_1 \mid I_2) = \frac{P(I_2 \mid \theta_1)}{P(I_2 \mid \theta_1)P(\theta_1) + P(I_2 \mid \theta_2)P(\theta_2)} P'(\theta_1) = \frac{0.2}{0.2 \cdot 0.6 + 0.1 \cdot 0.4} \cdot 0.6 = 0.75$$

$$P''(\theta_2 \mid I_2) = \frac{P(I_2 \mid \theta_2)}{P(I_2 \mid \theta_1)P(\theta_1) + P(I_2 \mid \theta_2)P(\theta_2)} P'(\theta_2) \frac{0.1}{0.2 \cdot 0.6 + 0.1 \cdot 0.4} \cdot 0.4 = 0.25$$



The minimized expected cost is:

$$E''[C \mid I_2] = \min \ \{P''[\theta_1] \cdot 10 + P''[\theta_2] \cdot (100+10), 100\} =$$
$$\min \ \{0.25 \cdot 10 + 0.75 \cdot 110,\ 100\} = 85 \quad \text{Mio. CHF}$$

Action A_1 poses less expected cost and hence the company should develop a well locally.

c.

Before all, it has to be decided whether a test well should be constructed or not.

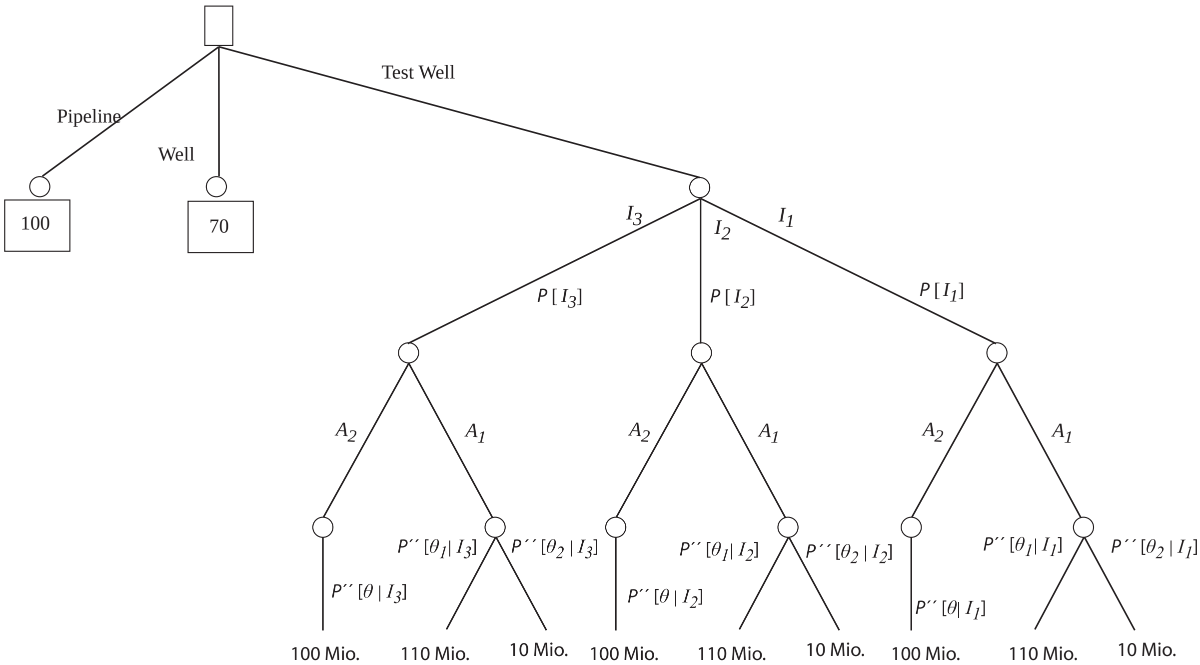
There are three decision alternatives:

A_1: Develop a well locally

A_2: Construct a pipeline

A_3: Develop a test well before develop/construct a well/pipeline.

If a test well is developed, three possible results can be obtained, namely the indicator I_1, I_2 and I_3. After the development of a test well the company can decide whether the well should be developed locally (A_1) or construct a pipeline (A_2).



How large are the probabilities that the results of a test well are I_1, I_2 and I_3 respectively? The probabilities can be calculated by:

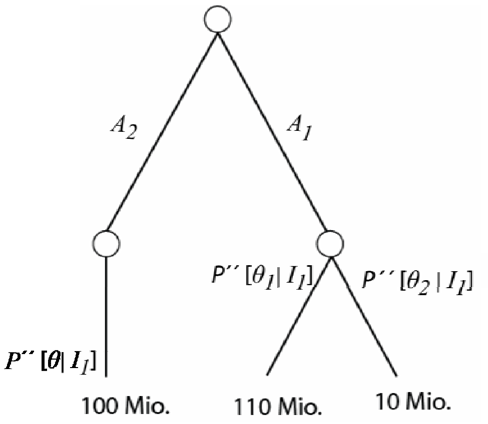
$$P[I_1] = P[I_1 | \theta_1] \cdot P[\theta_1] + P[I_1 | \theta_2] \cdot P[\theta_2] = 0.1 \cdot 0.6 + 0.8 \cdot 0.4 = 0.38$$

$$P[I_2] = P[I_2 | \theta_1] \cdot P[\theta_1] + P[I_2 | \theta_2] \cdot P[\theta_2] = 0.2 \cdot 0.6 + 0.1 \cdot 0.4 = 0.16$$

$$P[I_3] = P[I_3 | \theta_1] \cdot P[\theta_1] + P[I_3 | \theta_2] \cdot P[\theta_2] = 0.7 \cdot 0.6 + 0.1 \cdot 0.4 = 0.46$$

The posterior analysis for I_1, I_2 and I_3 is performed. For this purpose, the company needs to know the probabilities of the state Θ for each indicator.

Given the indicator I_1, the posterior probabilities can be calculated by:



$$P''(\theta_1 | I_1) = \frac{P(I_1 | \theta_1) \cdot P'(\theta_1)}{P'(I_1)} = \frac{0.1 \cdot 0.6}{0.38} = \frac{0.06}{0.38} = 0.158$$

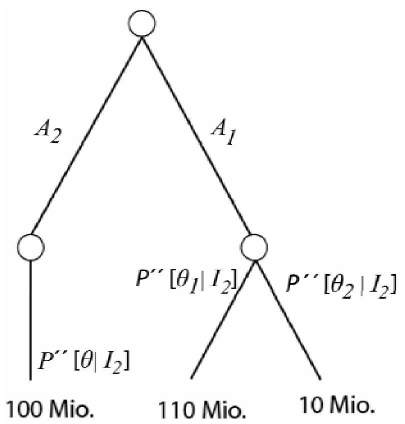
$$P''(\theta_2 | I_1) = \frac{P(I_1 | \theta_2) \cdot P'(\theta_2)}{P'(I_1)} = \frac{0.8 \cdot 0.4}{0.38} = \frac{0.32}{0.38} = 0.842$$

The minimized expected cost is:

$$E''[C | I_1] = \min \; \{P''[\theta_2 | I_1] \cdot 10 + P''[\theta_1 | I_1] \cdot (100 + 10), 100\} =$$
$$\min \; \{0.842 \cdot 10 + 0.158 \cdot 110, \, 100\} = 26 \quad \text{Mio. CHF}$$

Therefore, given I_1, action A_1 should be chosen.

The posterior analysis for the indicator I_2 is already done in part



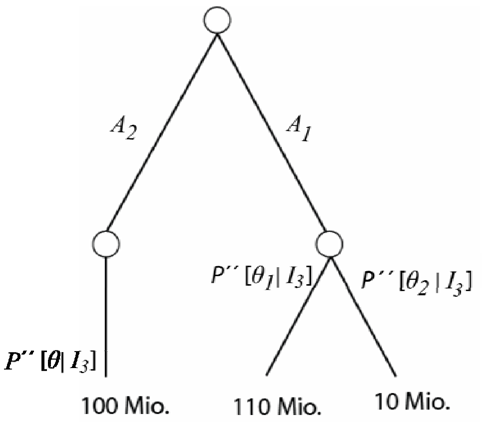
$$P''(\theta_1 | I_2) = \frac{P(I_2 | \theta_1) \cdot P'(\theta_1)}{P'(I_2)} = \frac{0.2 \cdot 0.6}{0.16} = \frac{0.12}{0.16} = 0.75$$

$$P''(\theta_2 \mid I_2) = \frac{P(I_2 \mid \theta_2) \cdot P'(\theta_2)}{P'(I_2)} = \frac{0.1 \cdot 0.4}{0.16} = \frac{0.04}{0.16} = 0.25$$

$$E''[C \mid I_2] = \min \ \{P''[\theta_1 \mid I_2] \cdot (10) + P''[\theta_2 \mid I_2] \cdot (100+10), \ \ 100\} =$$
$$\min \ \{0.25 \cdot 10 + 0.75 \cdot 110, 100\} = 85 \ \ \text{Mio. CHF}$$

Given the result of the test well I_2, action A_1 should be chosen.

Posterior analysis for the indication I_3 is perfomed as:



$$P''(\theta_1 \mid I_3) = \frac{P(I_3 \mid \theta_1) \cdot P'(\theta_1)}{P'(I_3)} = \frac{0.7 \cdot 0.6}{0.46} = \frac{0.42}{0.46} = 0.913$$

$$P''(\theta_2 \mid I_3) = \frac{P(I_3 \mid \theta_2) \cdot P'(\theta_2)}{P'(I_3)} = \frac{0.1 \cdot 0.4}{0.46} = \frac{0.04}{0.46} = 0.087$$

The minimized expected cost is:

$$E''[C \mid I_3] = \min \ \{P''[\theta_1 \mid I_3] \cdot (10) + P''[\theta_2 \mid I_3] \cdot (100+10), 100\} =$$
$$\min \ \{0.087 \cdot 10 + 0.913 \cdot 110, 100\} = 100 \ \ \text{Mio. CHF}$$

Given the result of the test well I_3, alternative A_2 should be chosen.

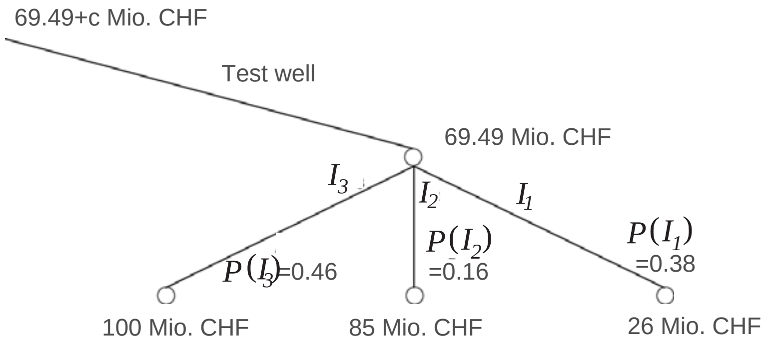
By multiplying the expected costs associated with each decision with the probabilities that each indication is obtained, the expected cost when the test well is obtained as:

The minimized expected cost is:

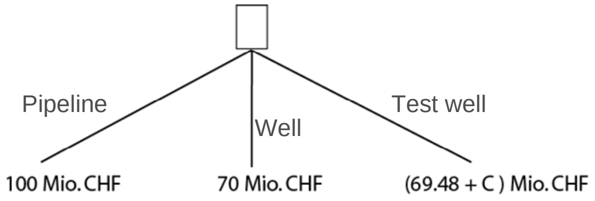
$$E''[C] = E''[C \mid I_1] \cdot P(I_1) + E''[C \mid I_2] \cdot P(I_2) + E''[C \mid I_3] \cdot P(I_3)$$
$$= 26 \cdot 0.38 + 85 \cdot 0.16 + 100 \cdot 0.46 = 69.49$$

Since the development of a test well requires a cost c, the cost has to be added to the expected cost:

$E''[C] = 69.49 + c$



Finally, the following decision tree is obtained:



The minimum cost for the test well that allows the company to construct the test well is:

$$69.48 + c \leq \min(A_1, A_2) = 70$$
$$\Leftrightarrow c \leq 0.52 \ \text{Mio.CHF}$$

Since the test well costs 1 Million CHF, the best decision is to develop a local well without constructing a test well.