

Exercises Tutorial 5

Statistics and Probability Theory

Prof. Dr. Michael Havbro Faber Swiss Federal Institute of Technology Zurich ETHZ

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

$$f_{X}(x) = \begin{cases} c \cdot x \cdot (60 - x) & \text{for } 0 \le x \le 60 \\ 0 & \text{otherwise} \end{cases} \xrightarrow{\text{Change to}} f_{X}(x) = \begin{cases} c \cdot x \cdot (15 - \frac{x}{4}) & \text{for } 0 \le x \le 60 \\ 0 & \text{otherwise} \end{cases}$$
Exercise 4 Solution

- 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?

Solution 4.1 a. Which value of *c* should be chosen?



Solution 4.1 b. Describe the cumulative distribution function $F_X(x)$ of the random variable X.

Cumulative distribution function

$$F_X(x) = \int_{-\infty}^{x} f_X(t) dt \qquad \qquad f_X(x) = \begin{cases} c \cdot x \cdot (60 - x) & \text{for } 0 \le x \le 60 \\ 0 & \text{otherwise} \end{cases}$$

Work in the intervals where the pdf is defined

Case of $0 \le x \le 60$

$$F_X(x) = \int_0^x f_X(t) dt = \int_0^x \frac{1}{36000} \cdot t(60 - t) dt = \frac{1}{36000} \left[\frac{60}{2} t^2 - \frac{1}{3} t^3 \right]_0^x = \frac{1}{36000} \left(30x^2 - \frac{1}{3}x^3 \right)$$

Case of 60 < x

$$F_X(x) = 1$$

$$F_{X}(x) = \begin{cases} \frac{1}{36000} \cdot \left(\frac{60}{2} \cdot x^{2} - \frac{1}{3} \cdot x^{3}\right) & 0 \le x \le 60\\ 1 & 60 < x \end{cases}$$



0

Solution 4.1 c. Which of the following four values, 30.00 CHF, 40.00 CHF, 50 CHF and 60 CHF does not exceed the 90%-quantile of the monthly expense?

 $P(X \leq \alpha) = F_{Y}(x) = 0.9$

First we need to find the value corresponding to the 90% quantile

60

??

$$P(X \le \alpha) = \frac{1}{36000} \cdot \int_{0}^{\alpha} t(60-t) dt \Rightarrow 0.9 = \frac{1}{36000} \cdot \int_{0}^{\alpha} t(60-t) dt \Rightarrow$$

$$\alpha = 48.30$$
So the values of 30 and 40 CHF do not exceed the 90%- quantile of the monthly expense

Solution 4.1 d. How large is the mean monthly expense for water consumption including sewage fee for a 2-persons household?

Mean---First moment

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$



$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx =$$

$$\frac{1}{36000} \int_{0}^{60} x^2 (60 - x) dx = \frac{1}{36000} \left[20x^3 - \frac{1}{4}x^4 \right]_{0}^{60}$$

$$= \frac{1}{36000} (4320000 - 3240000) = \frac{1080000}{36000} = 30$$

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

$$f_{X}(x) = \begin{cases} \frac{1}{2} & \text{for } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad f_{Y}(y) = \begin{cases} \frac{3}{4} \cdot (2y - y^{2}) & \text{for } 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

The correlation coefficient between X and Y equals to $\sqrt{\frac{1}{15}}$

- a. Calculate the expected value of 6X 4Y + 2
- b. Calculate the covariance of Cov[6X, 4Y]
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

Generally: (script Equations D.16 to D.18)

Expectation operator

```
E[c] = c

E[cX] = cE[X]

E[a+bX] = a+bE[X]

E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]
```

Variance operator

$$Var[c] = 0$$
$$Var[cX] = c^{2}Var[X]$$
$$Var[a+bX] = b^{2}Var[X]$$

- a. Calculate the expected value of 6X 4Y + 2b. Calculate the covariance of Cov(6X; 4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

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

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad f_Y(y) = \begin{cases} \frac{3}{4} \cdot (2y - y^2) & \text{for } 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

The correlation coefficient between X and Y equals to $\sqrt{\frac{1}{15}}$

Steps:

• Which is the first thing to do??

Exercise 5.1

- a. Calculate the expected value of X 4Y + 2
- b. Calculate the covariance of Cov(6X; 4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

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



Steps:

Which is the first thing to do??

E[a+bX] = a+bE[X]

 $E[6X - 4Y + 2] = \dots$

Find the expected value and variance of *X* and *Y*!

What do these measures express???

- a. Calculate the expected value of 6X 4Y + 2
- b. Calculate the covariance of Cov(6X;4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

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



Steps:

1. Find the expected value and variance of X and Y! (Script Equations D.5-D11)

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

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

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

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

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

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



Steps:

1. Find the expected value and variance of X and Y! (Script Equations D.5-D11)

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

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

- a. Calculate the expected value of 6X 4Y + 2
- b. Calculate the covariance of Cov(6X;4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

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



Steps:

2. After calculating the expected value and variance of both variables use the properties of the respective operator: (*Script Equations D.16-D18*)

a. E[6X - 4Y + 2]

E[a+bX] = a+bE[X]

 $E[6X - 4Y + 2] = \dots$

- a. Calculate the expected value of 6X 4Y + 2
- b. Calculate the covariance of Cov(6X;4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

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



Steps:

2. After calculating the expected value and variance of both variables use the properties of the respective operator: (*Script Equations D.16-D.18*)

b. *Cov*[6*X*,4*Y*]

$$Var[a+bX] = b^{2}Var[X]$$
Watch this!!!
Cov[6X,4Y] = 6.4 Cov[X,Y]

Script Equation D.23

 $Cov[X,Y] = \rho_{XY} \cdot \sqrt{Var[X] \cdot Var[Y]}$

- a. Calculate the expected value of 6X 4Y + 2
- b. Calculate the covariance of Cov(6X;4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

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



Steps:

2. After calculating the expected value and variance of both variables use the properties of the respective operator: (*Script Equations D.16-D.18*)

c. $Var[6X - 4Y + 2] = Var[6X] + Var[4Y] - 2 \cdot Cov[6X, 4Y]$

Watch this!!!

WHY??

- a. Calculate the expected value of 6X 4Y + 2
- b. Calculate the covariance of Cov(6X;4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

Steps:

Exercise 5.1

2. After calculating the expected value and variance of both variables use the properties of the respective operator: (*Script Equations D.16-D.18*)

C.
$$Var[6X - 4Y + 2] = Var[6X] + Var[4Y] - 2 \cdot Cov[6X, 4Y]$$

Watch this!!! WHY??

Script Equation D.25

$$Y = a_0 + \sum_{i=1}^n a_i X_i$$

$$E[Y] = a_0 + \sum_{i=1}^n a_i E[X_i]$$

$$Var[Y] = \sum_{i=1}^{n} a_{i}^{2} Var[X_{i}] + 2\sum_{\substack{i,j=1\\i\neq j}}^{n} a_{i}a_{j}C_{X_{i}X_{j}}$$

- a. Calculate the expected value of 6X 4Y + 2
- b. Calculate the covariance of Cov(6X;4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

Steps:

Exercise 5.1

2. After calculating the expected value and variance of both variables use the properties of the respective operator: (*Script Equations D.16-D.18*)

c.
$$Var[6X - 4Y + 2] = Var[6X] + Var[4Y] - 2 \cdot Cov[6X, 4Y]$$

Watch this!!! WHY??

Script Equation D.25

$$Y = a_0 + \sum_{i=1}^n a_i X_i$$

$$E[Y] = a_0 + \sum_{i=1}^n a_i E[X_i]$$

$$Var[Y] = \sum_{i=1}^{n} a_{i}^{2} Var[X_{i}] + 2\sum_{\substack{i, j=1 \ i \neq j}}^{n} a_{i}a_{j}C_{X_{i}X_{j}}$$

 $Var[6X - 4Y + 2] = Var[6X] + Var[4Y] - 2 \cdot Cov[6X, 4Y]$



 $Var[6X - 4Y + 2] = \left[6^{2} Var[X] \right] + \left[(-4)^{2} Var[Y] \right] + 2 \left[6 \left[-4 \right] Cov[X, Y] \right]$

- a. Calculate the expected value of 6X 4Y + 2
- b. Calculate the covariance of Cov(6X;4Y)
- c. Calculate the variance of 6X 4Y + 2
- d. Calculate the expected value of $6X^2 4Y^2$

Steps:

Exercise 5.1

2. After calculating the expected value and variance of both variables use the properties of the respective operator: (*Script Equations D.16-D.18*)

d.
$$E[6X^2 - 4Y^2] = 6E[X^2] - 4E[Y^2]$$

$$E\left[X^{2}\right] = \int_{-\infty}^{\infty} x^{2} f_{X}(x) dx$$
$$f_{X}(x) = \begin{cases} \frac{1}{2} & \text{for } -1 \le x \le 1\\\\ 0 & \text{otherwise} \end{cases}$$

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	$\Sigma = 1.00$

Simplify!!

What is given??

Measurements of wind speed are taken with an accurate and a less accurate device.

The joint probability of the measurements by both devices of the number of days in which the wind speed exceeds a threshold.

 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.

Simplify!!

What is given??

Measurements of wind speed are taken with an accurate and a less accurate device.

The joint probability of the measurements by both devices of the number of days in which the wind speed exceeds a threshold.

 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.

What is required??

- 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 5.2 a.

 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	$\Sigma = 1.00$

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

 $P[N_{U} = N_{G}] = 0.2910 + 0.3580 + 0.1135 + 0.0505 = 0.813$



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

Probability that $N_U < N_G$

Probability that $N_U > N_G$

	$N_U = 0$	$N_{U} = 1$	$N_{U} = 2$	$N_{U} = 3$	$P(N_G)$
$N_G = 0$	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	$\Sigma = 1.00$

Exercise 5.2 b.

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?

Conditional probability....Baye's rule

		$N_U = 0$	$N_{U} = 1$	$N_{U} = 2$	$N_{U} = 3$	$P(N_G)$
1	$\mathbf{V}_G = 0$	0.2910	0.0600	0.0000	0.0000	0.3510
1	$V_G = 1$	0.0400	0.3580	0.0100	0.0000	0.4080
1	$V_G = 2$	0.0100	0.0250	0.1135	0.0300	0.1785
1	$V_{G} = 3$	0.0005	0.0015	0.0100	0.0505	0.0625
1	$P(N_U)$	0.3415	0.4445	0.1335	0.0805	$\Sigma = 1.00$

How can we write this???

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 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=S+T.
- b. How large is the probability $P(Z \le 5)$?
- 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?

Example with discrete random variables

Consider two dices. The outcome of throwing each one is described by the discrete random variables *X* and *Y*.

We are looking for the probability that the sum of the numbers that will come out when we will throw the dices is equal to e.g. 10.

The sum is a random variable itself which can be described as: Z=X+Y

The probability that the sum is equal to 10 is: ???

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Example with discrete random variables

Consider two dices. The outcome of throwing each one is described by the discrete random variables *X* and *Y*.

We are looking for the probability that the sum of the numbers that will come out when we will throw the dices is equal to e.g. 10.

The sum is a random variable itself which can be described as: Z=X+Y

The probability that the sum is equal to 10 is:

In the case of continuous random variables: Z=X+Y

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

What is given?

Time where no maintenance is required, T: exponentially distributed $\mu_T = 10$ years

Time of maintenance S: exponentially distributed $\mu_s = \frac{1}{12}$ years

What is given?

Time where no maintenance is required, T: exponentially distributed $\mu_T = 10$ years

Time of maintenance S: exponentially distributed $\mu_s = \frac{1}{12}$ years

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



What is given?

Time where no maintenance is required, T: exponentially distributed $\mu_T = 10$ years

Time of maintenance S: exponentially distributed $\mu_s = \frac{1}{12}$ years

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



In the case of continuous random variables: Z=X+Y

What is given?

Time where no maintenance is required, T: exponentially distributed $\mu_T = 10$ years

Time of maintenance S: exponentially distributed $\mu_s = \frac{1}{12}$ years

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

Z is defined within the interval [0;z]

$$f_{Z}(z) = \int_{0}^{z} f_{T}(t) f_{S}(z-t) dt = \dots$$

$$f_T(t) = \frac{1}{\mu_T} \cdot e^{\frac{-t}{\mu_T}} \qquad f_S(s) = \frac{1}{\mu_S} \cdot e^{\frac{-s}{\mu_S}}$$

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

 $P(Z \le 5) = F_Z(z) = \dots$

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

 $P(Z \le 5) = F_Z(z) = \dots$

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?

How can we express the problem??