Assessment 1
Statistics and probability theory

SS 2007

Prof. Dr. M.H. Faber

ETH Zurich

Thursday 3rd of May 2007
08:15 – 09:45

Surname: ....

Name: ....

Stud. Nr.: ....

Course of studies: ....

Date and duration:

Thursday, 3rd of. May 2007
Start: 8:15
End: 9.45
Duration: 90 minutes

Aids:

- No communication medium (e.g. cell phones, calculators with Bluetooth etc.)

Hints:

- Please control first, if you have received all the materials (listed under: Contents).

- Please place your Legi on your desk.
-
- Please write your **name on every sheet of paper,** at the bottom left side**.**
-
- Use only the provided sheets of paper.

- When you have finished, place **all materials** back in the envelope and raise your hand to call an assistant to collect it. You are allowed to leave till 9.15. If you finish later than this, wait quietly until the end time of the assessment (9.45).

- **Do not open** the paper fastener.

Contents

- General information and exercises (12 pages).
- 1 sheet of paper (checkered)

Part 1: „Multiple Choice"

In answering the following multiple choice questions it should be noted that for some of the questions several answers may be correct. Tick ALL the correct alternative(s) in every question as: ☑

1.1 90 Persons are questioned about their income. 55 of them earn 10000 CHF a year, 34 of them earn 20000 CHF and 1 person, director of a bank, earns 2000000 CHF a year. Which one(s) of the following statement(s) is(are) correct?
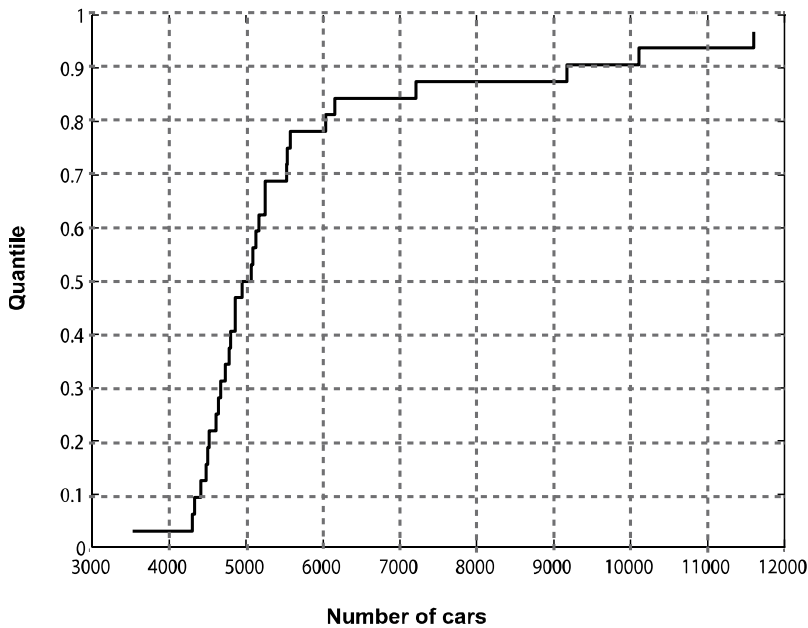
The mode is 10000 CHF. ☐

The mean is 33300 CHF. ☐

None of the above. ☐

1.2 The following figure shows a cumulative distribution plot of the daily traffic flow through the Gotthard tunnel in January 1997.

The median is equal to 5000 cars. ☐

The mode is equal to 5000 cars. ☐

The mean is equal to 5000 cars. ☐

1.3 Which one(s) of the following statements is(are) correct?

The coefficient of variation is a measure of the dispersion of a data set. ☐

The sample covariance is a measure of the correlation of two data sets. ☐

None of the above. ☐

1.4 Two sets of observed values, A and B, are being compared. The observed values are shown in the following table. Which statement(s) is(are) correct?

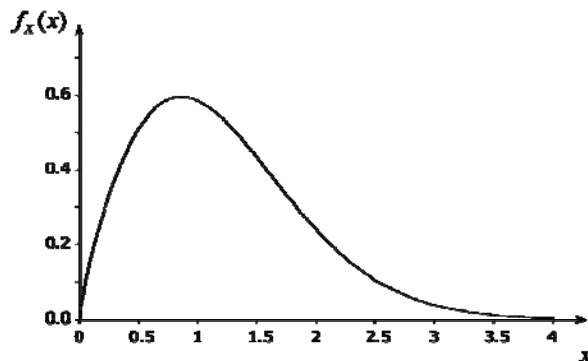| Set A	2	4	6	8
Set B	8	10	12	14

The sample variances of the two data sets are the same. ☐

The sample covariance of data set A is higher than the one of data set B. ☐

The coefficients of variation of the two data sets are the same. ☐

1.5 The probability density function of a continuous random variable X is shown in the following figure.

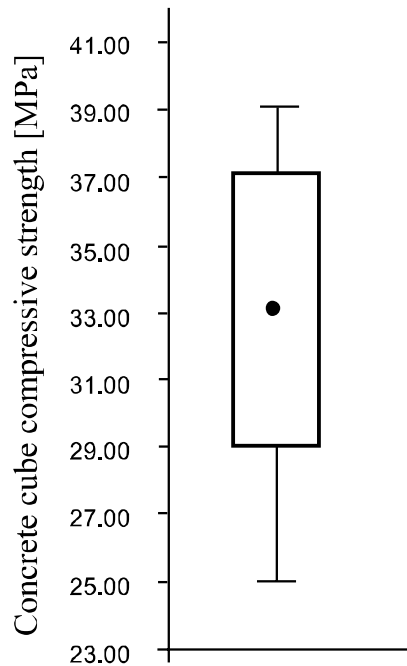Which of the following statement(s) is(are) correct?

The distribution of the random variable is symmetric. ☐

The distribution of the random variable is skewed to the right. ☐

The mode of this data set is larger than the mean value. ☐

1.6 In the following figure a Tukey Box Plot is shown. Which statement(s) is(are) correct?

The outside values of this data set are lying within the interquartile range. ☐

The 0.25 quantile is 25 MPa. ☐

50% of the data lie between 25 MPa and 39 MPa. ☐

50% of the data lie between 29 MPa and 37 MPa. ☐

1.7 The 0.25 quantile of a data set corresponds to a value of the data set for which:

25% of the data values in the data set are larger. ☐

75% of the data values in the data set are larger. ☐

1.8 Which of the following statement(s) is(are) correct?

The sample correlation coefficient is defined in the interval $0 \le r_{xy} \le 1$. ☐

A perfect correlation between two data sets exists if r_{xy} is equal to 0. ☐

A perfect correlation between two data exists if r_{xy} is equal to 1. ☐

1.9 An easy and cheap on-site test method to determine whether the reinforcement in a concrete beam is corroded or not has been developed by an engineer. Based on experiments conducted in the laboratory, the engineer found that the test accurately predicts the occurrence of corrosion with a probability of 0.75. In order to account for differences in laboratory and actual on-site conditions, he however decides that the prediction probability of the test method should be set lower at 0.6. The assignment of this probability is based on which of the following definitions of probability:

Frequentistic ☐

Classical ☐

Bayesian ☐

None of the above ☐

1.10 The water supply for a city C comes from two sources A and B as shown in the figure below. Water is transported by pipelines 1, 2, 3 and 4. If required, either one of the two sources, by itself, is sufficient to supply the water for the city.
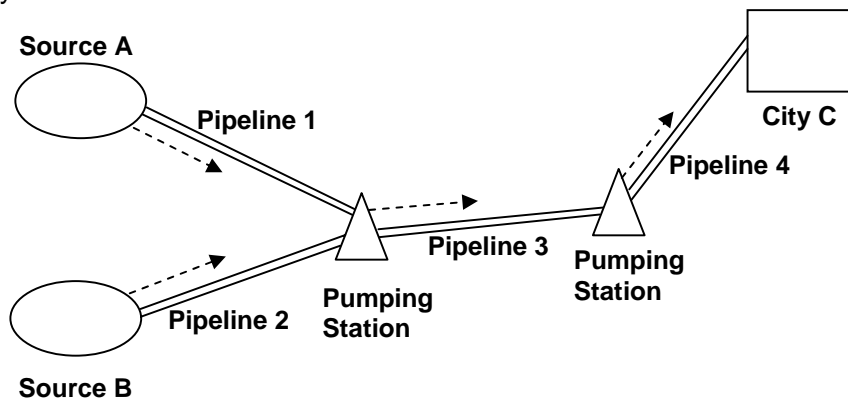
The failure of a pipeline will lead to a shortage of water in the city. If events E_1, E_2, E_3 and E_4 denote the failure of pipelines 1, 2, 3 and 4 respectively, the event that there is no shortage of water in city C can be written in the following way(s):

$\{\bar{E}_1 \cup \bar{E}_2\} \cap \{\bar{E}_3 \cap \bar{E}_4\}$ ☐

$\overline{\{E_1 \cap E_2\} \cup \{E_3 \cup E_4\}}$ ☐

$\overline{E_1 \cup E_2 \cup E_3 \cup E_4}$ ☐

$\overline{\{E_1 \cap E_2\}} \cap \overline{\{E_3 \cup E_4\}}$ ☐

1.11 Two events A and B are mutually exclusive. Which of the following expressions is(are) correct?

$P(A|B) = P(A)$ ☐

$P(A|B) = 0$ ☐

$P(A \cup B) = P(A) + P(B)$ ☐

$P(A \cup B) = 1$ ☐

1.12 An experiment has four possible mutually exclusive outcomes : A, B, C and D. Which of the following assignments of probability values is(are) NOT correct?

$P(A) = 0.38, P(B) = 0.19, P(C) = 0.11, P(D) = 0.32$ ☐

$P(A) = 0.34, P(B) = 0.29, P(C) = 0.28, P(D) = 0.19$ ☐

$P(A) = 0.3, P(B) = 0.3, P(C) = 0.28, P(D) = 0.19$ ☐

1.13 In a class of 225 civil engineering graduate students, 134 are enrolled for a course in statistics (event A), 87 are enrolled for a course in mechanics (event B) and 28 are enrolled in both of these courses. The number of students who are not enrolled in either course is:

4 ☐

32 ☐

16 ☐

1.14 If A and B are two independent events, which of the following expressions is(are) correct?

$P(\bar{A} \cap \bar{B}) = P(\bar{A})P(\bar{B})$ ☐

$P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B)$ ☐

$P(\bar{A} \cap \bar{B}) = 1 - P(A)P(B)$ ☐

1.15 From past experience, the probability that a painter A can do a painting work is 60%. The probability that a painter B can do a painting work is 80%.The probability that either one of the two painters can do a painting work is 90%. A and B are not independent. What is the probability that painter B can do a painting work given that painter A cannot do the same painting work?

0.45 ☐

0.30 ☐

0.90 ☐

0.75 ☐

1.16 It costs 60 CHF to test a certain component used in a machine and determine whether the component is defective before installing it. However if the component is installed without testing and later found to be defective, it costs 1200 CHF to repair the resulting damage to the machine. It is more profitable to install the component without testing if it is known that:

1% of all components produced are defective ☐

2% of all components produced are defective ☐

4% of all components produced are defective ☐

6% of all components produced are defective ☐

1.17 Probability distribution functions may be defined in terms of their moments. If X is a discrete random variable which one(s) of the following is(are) correct?

The second moment of X corresponds to its mean value, σ_X^2. ☐

The second central moment of X corresponds to its variance, σ_X^2. ☐

The first moment of X corresponds to its mean value, μ_X. ☐

1.18 The coefficient of variation $CoV[X]$ of a random variable X is a descriptor of:

the variability of the random variable around its expected value. ☐

the variability of the random variable around the median. ☐

1.19 The cumulative distribution function of a continuous random variable X is illustrated in the following diagram.
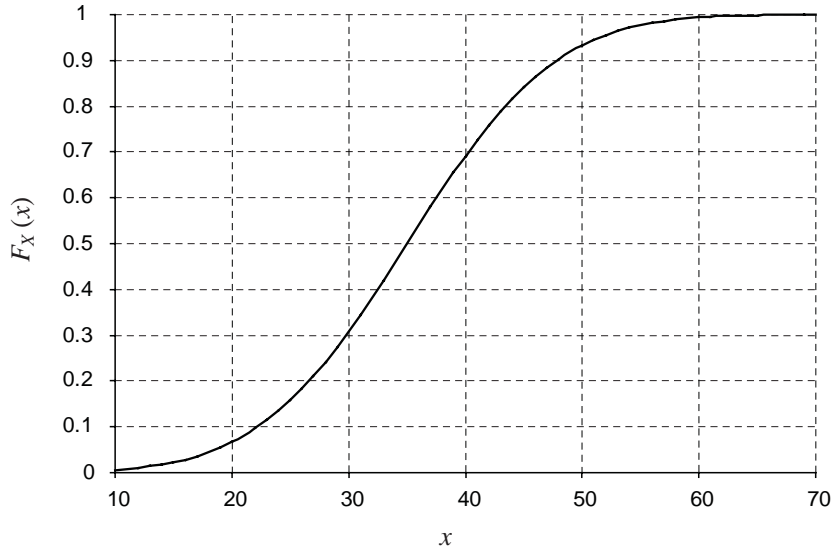
The probability of X exceeding the value of 30 is approximately equal to:

$P(X > 30) = 0.7$ ☐

$P(X > 30) = 0.3$ ☐

1.20 X is a random variable with mean value equal to $\mu_X = 18$ and standard deviation equal to $\sigma_X = 1$. The random variable X is now standardized into the random variable Z. Then Z:

has a mean value equal to $\mu_Z = 0$ and standard deviation equal to $\sigma_Z = 1$ ☐

has a mean value equal to $\mu_Z = 1$ and standard deviation equal to $\sigma_Y = 0$ ☐

1.21 A structure is designed to withstand wind speeds with a return period of 40 years. The probability that the design wind speed will be exceeded for the first time on the eleventh year after the structure is given in service is equal to:

0.02 ☐

$9.3 \cdot 10^{-17}$ ☐

1.22 The cost associated with the occurrence of an event A is a function of a random variable X with mean value $\mu_X = 100$ CHF and standard deviation $\sigma_X = 10$ CHF. Their relation can be written as: $C_A = a + bX + cX^2$, where a, b and c are constants and equal to 10, 0.5 and 1 respectively.

The expected cost of event A is equal to:

$E[C_A] = 10060$ CHF. ☐

$E[C_A] = 10160$ CHF. ☐

$E[C_A] = 160$ CHF. ☐

1.23 The random variables X and Y are correlated with a correlation coefficient equal to $\rho_{XY} = 0.25$. The mean values of the random variables are $\mu_X = 100$ and $\mu_Y = 80$, while their standard deviations are $\sigma_X = 10$ and $\sigma_Y = 8$. The variance of the function $D = 2X - Y$ is equal to:

$Var[D] = 424$. ☐

$Var[D] = 464$. ☐

$Var[D] = 374$. ☐

Part 2: „Exercise - Moments of a random variable "

1) A discrete random variable X is represented by the following probability mass function (see also Figure 1):

$$p_X(x) = \begin{cases} 0.3 & , x = 1 \\ 0.4 & , x = 2 \\ 0.2 & , x = 3 \\ 0.1 & , x = 4 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the mean and the standard deviation of the random variable X.

2) A continuous random variable Y is represented by the following probability density function (see also Figure 1).

$$f_Y(y) = \begin{cases} 0.3 & , 0 < y \le 1 \\ 0.4 & , 1 < y \le 2 \\ 0.2 & , 2 < y \le 3 \\ 0.1 & , 3 < y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

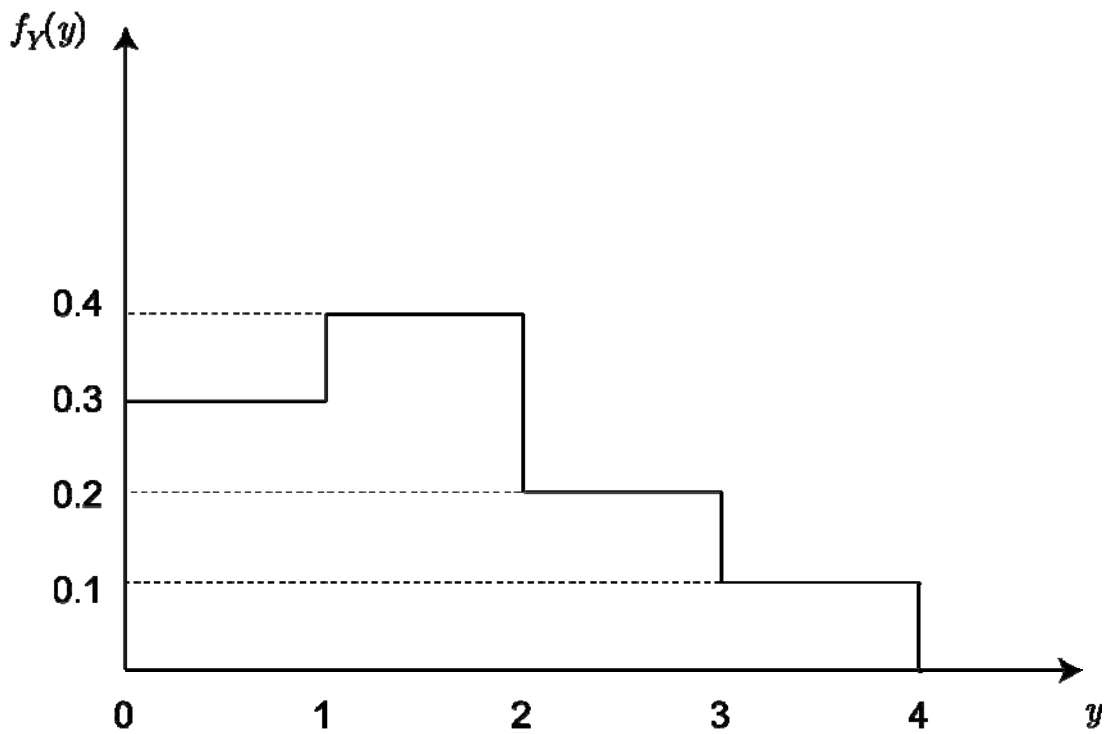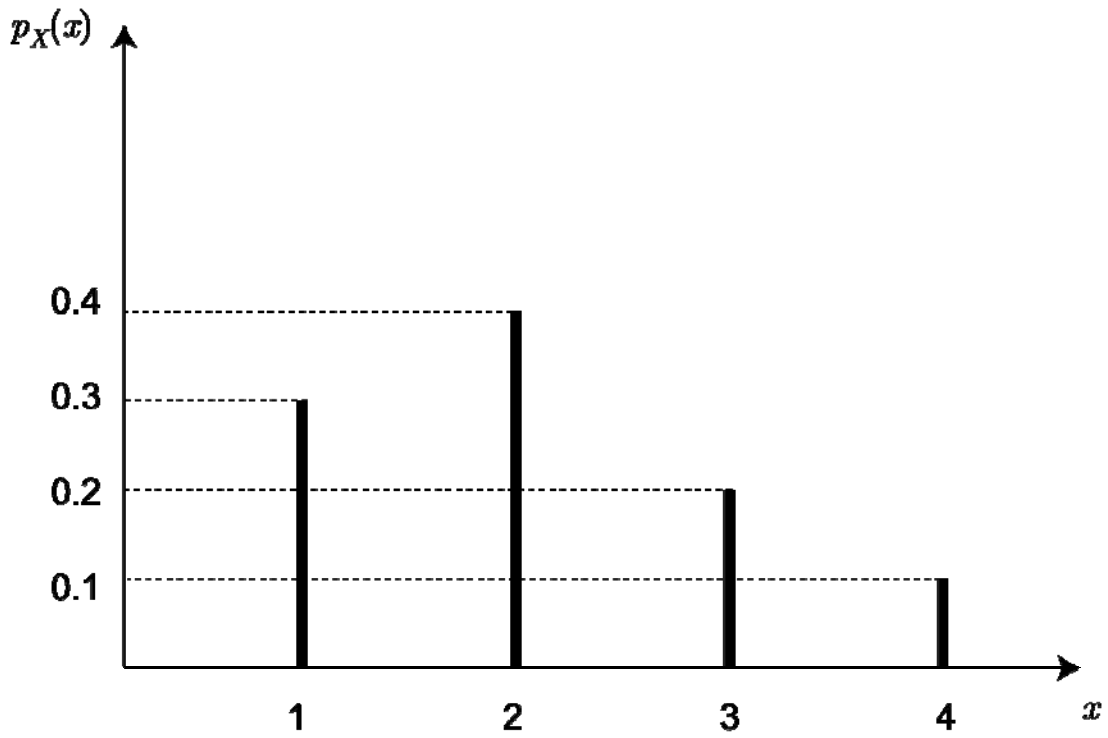Calculate the mean and the standard deviation of the random variable Y.

Surname, Name:

Figure 1: Probability mass function of X **(top) and probability density function of** Y **(bottom).**