Probability Theory and Mathematical Statistics Handouts

Gábor Kende
Renáta Németh
Probability Theory and Mathematical Statistics Handouts
by Gábor Kende and Renáta Németh
Publication date 2011.
Table of Contents

Introduction .................................................................................................................. vi
1. Experiments, events, operations with events, probabilities (theory) .......................... 1
   1. Events .................................................................................................................. 1
   2. Operations on events .............................................................................................. 1
   3. Some more definitions: .......................................................................................... 2
   4. Probability, some basic rules ................................................................................ 3
   5. Conditional probabilities ....................................................................................... 3
   6. The total probability formula and Bayes’s formula ................................................ 3
2. Probabilities – introductory exercises ...................................................................... 5
   1. (A) SIMPLE EXERCISES ................................................................................ 5
   2. (B) MULTIPLICATION RULE; CONDITIONAL PROBABILITIES ..................... 6
   3. (C) ADDITION RULE ....................................................................................... 7
   4. (D) INDEPENDENCE AND CONDITIONAL PROBABILITIES .......................... 9
   5. (E) MULTIPLICATION RULE AND ADDITION RULE COMBINED (a preliminary exercise for the total probability formula and Bayes’s formula) .................. 10
3. Independence of events, questions (exercises) ....................................................... 12
4. Total probability formula & Baye's formula (exercises) ........................................... 16
5. Variables – simple exercises – distributions, expected values (E.V.) & standard deviations (SD) (exercises) .......................................................... 20
6. Variables (univariate): distribution, density function, distribution function, expected value, variance and covariance, standard deviation (theory) .............. 22
   1. Random variables .................................................................................................. 22
   2. Distribution, density function, cumulative distribution function ............................ 22
   3. Expected value (E.V.) ......................................................................................... 23
      3.1. The expected value of a random variable X with finite range is: .................... 23
      3.2. The expected value of a discrete variable X with infinite range is ................. 23
      3.3. The expected value of continuous variables: ...................................... 23
      3.4. Some properties of the expected value: ......................................................... 24
   4. The variance and the standard deviation (S.D.) ..................................................... 26
   5. Expected values and standard deviations of sample sums and sample means (sampling with replacement): the Square Root Law ................................................. 27
      5.1. Two auxiliary statements: .............................................................................. 27
      5.2. The Square Root Law: ................................................................. 27
   6. An upper bound for the proportion or the probability of large deviations: the Chebyshev inequality .................................................................................. 28
   7. Standard errors for the sample sum and the sample mean, drawing without replacement : the "correction factor" (finite correction) .............................................. 29
   8. Expected values, standard errors, variances and covariances (exercises) ............. 31
   9. Random variables: distribution, cumulative distribution function, density function, expected value (exercises) ................................................................. 36
10. Expected value and standard deviation of continuous variables; other exercises (exercises) ............................................................... 42
11. Roulette/1: expected values, standard deviations (exercises) ............................... 49
12. Using the normal table (exercises) ......................................................................... 51
13. Normal approximation (exercises) ........................................................................ 54
14. Roulette/2: normal approximation (exercises) ....................................................... 63
15. Some kinds of distributions (theory) ....................................................................... 66
    1. Discrete distributions ........................................................................................... 66
    2. Continuous distributions .................................................................................... 68
16. Some kinds of distributions – expected values, standard deviations, probabilities (questions, exercises) ................................................................. 70
17. Law of large numbers (LLN) and central limit theorem (CLT) (theory) ................. 72
    1. The law of large numbers (LLN) ........................................................................ 72
    2. Central limit theorem (CLT) ................................................................................ 72
18. Hypothesis testing/1 – z test, introductory exercises (exercises) ............................ 74
19. Hypothesis testing/2 – one-sample z test (exercises) ............................................. 78
20. Hypothesis testing/3 – decision rules; one- and two-tailed tests; questions (exercises) ...... 80
21. Hypothesis testing/4 – e.g. probabilities of errors (exercises) ................................. 83
22. Bivariate concepts (theory) ......................................................................................... 87
   1. Joint distribution, marginal distributions, conditional distributions (discrete variables) .... 87
   2. Joint cumulative distribution function, joint density, marginal densities, conditional densities (joint distribution; continuous variables) ........................................................................ 88
   3. Conditional expected value ......................................................................................... 90
   4. Conditional variance: ................................................................................................. 91
   5. The expected value of a function of a random vector variable .................................... 92
   6. Covariance .................................................................................................................. 93
   7. Variance in a direction; the covariance matrix ............................................................... 96
   8. Supplement – variance, covariance, correlation, and geometry .................................... 96
   9. Bivariate joint normal distributions ........................................................................... 97
10. A) Steiner's identity (theory) ....................................................................................... 100
23. Bivariate distributions (exercises) ............................................................................... 102
24. Estimation: concepts (theory) ...................................................................................... 110
   1. Estimations ................................................................................................................ 110
   2. Characterizing a single estimation: bias, standard error and root mean square error, variance and mean square error ................................................................................... 110
   3. Characterizing estimation series: asymptotic unbiasedness and consistency ............... 111
   4. Confidence intervals ................................................................................................... 112
25. Estimations (exercises) ............................................................................................... 114
List of Examples

1.1. 5 chips in a box, 3 yellow, 2 green. Two chips are drawn at random, consecutively, without replacement. ........................................................................................................... 4
6.1. ..................................................................................................................................... 27
6.2. ..................................................................................................................................... 28
6.3. ..................................................................................................................................... 29
6.4. ..................................................................................................................................... 30
15.1. .................................................................................................................................... 66
15.2. .................................................................................................................................... 66
15.3. .................................................................................................................................... 66
15.4. .................................................................................................................................... 67
15.5. .................................................................................................................................... 67
15.6. .................................................................................................................................... 68
22.1. .................................................................................................................................... 90
Introduction

This material is compiled for the Probability Theory and Mathematical Statistics course of the Economics BA at the Faculty of the Social Sciences of Eötvös Loránd University. It consists mainly of exercises matched closely to the theoretical subject matter of the course. For a few topics brief theoretical summaries are also provided. These are, for the most part, only background drafts, discussing matters to be found in almost every probability or statistics textbook. There is, though, a longer handout explaining basic concepts about bivariate distributions and correlations. This topic can not be easily found elsewhere with a like selection of material and relatively easy presentation.

In form the tradition of giving out printed materials at the beginning or end of classes is followed: the material consists of handouts.

A kisebb betűs szedés az elméleti handoutkon háttérként szolgáló, a vizsgán nem kért bizonyításokat, indoklásokat, és kiegészítő anyagot jelöl; a feladatok között pedig korábbi feladatok többé-kevésbé változatlan ismétlését, amit azután rendes szedéssel követ "az igazi" kérdés.

Smaller print on theoretical pages indicates material for background knowledge, not to be met with on exams; among exercises it indicates exercises repeated in a more or less unchanged form to be followed by the 'real' question.

Exercises signed with an asterisk (*) are more difficult.
Chapter 1. Experiments, events, operations with events, probabilities (theory)

1. Events

Events are "things" we attribute probabilities to. (These are what we map probabilities to. They constitute the domain of probability seen as a function).

Special events:

an event sure to occur is called the certain event and denoted \( \Omega \)
an event impossible to occur is called the impossible event and denoted \( \emptyset \).

2. Operations on events

\( \neg A \): occurs if A does not occur

(denoted also: \( \neg A \), \( A' \), read as "not A" or "A complement");

\( A \cup B \): occurs if at least one of A and B occurs

(denoted also: \( A + B \), \( A \cup B \) ; read as "A or B");

\( A \cap B \): occurs if A and B both occur

(denoted also: \( A \land B \), \( A \& B \), \( A \cap B \) read as "A and B", "A intersection B", or simply "AB");

Assuming an experiment and some events to start from we also assume these operations executable; their results also belong to the set of "all events". [This set of "all events" plus the operations together make up an event algebra.]

Basic identities of event operations:

\[
\begin{align*}
A \cup A &= A \\
A \cup \emptyset &= A \\
\neg (A \cup B) &= \neg A \cap \neg B \\
\neg (A \cap B) &= \neg A \cup \neg B \\
(A \cup B) \cap C &= A \cup (B \cap C) \\
(A \cap B) \cup C &= A \cap (B \cup C) \\
\neg \Omega &= \emptyset \\
\neg \emptyset &= \Omega \\
\neg \neg A &= A \\
\neg \Omega &= \emptyset \\
\neg \emptyset &= \Omega \\
\neg (\neg A) &= A
\end{align*}
\]
3. Some more definitions:

Events A and B are mutually exclusive, if \( A \cap B = \emptyset \) that is, if it is impossible for A and B to occur simultaneously.

Events \( A_i, A_2, \ldots, A_n \) are pairwise mutually exclusive, if for all \( i,j \): \( A_i \cap A_j = \emptyset \).

A and \( \overline{A} \) are, for example, mutually exclusive.

A set of events \( A_i, A_2, \ldots, A_n \) is a partition, if

1. \( A_i, A_2, \ldots, A_n \) are pairwise mutually exclusive and
2. \( \bigcup_{i=1}^n A_i = \Omega \)

Example: consider an elementary-school logic set consisting of plastic tokens, in form squares, circles and triangles, in colour red, blue, yellow and green, in size little and big, which can be either full or with a hole in them (that is, a full single set consists of \( 3 \times 4 \times 2 \times 2 = 48 \) items). Assume an experiment of choosing one from among the 48 items at random; then

- \( A = \) (choosing a red one) and \( B = \) (choosing a yellow one) are mutually exclusive;
- \( A = \) (choosing a little circle) and \( B = \) (choosing a big square) are mutually exclusive;
- \( A = \) (choosing a little circle) and \( B = \) (choosing a big square) and \( C = \) (choosing a big blue circle) are pairwise mutually exclusive;
- Examples for partitions:
  - \( A = \) red, \( B = \) blue, \( C = \) yellow, \( D = \) green
  - \( A = \) red, \( B = \) not red
  - \( A = \) red, \( B = \) not red but angular, \( C = \) not red and not angular.

Exercises:

1. experiment = one roll with a dice
   - give an example of two mutually exclusive events together constituting a partition;
   - give an example of two mutually exclusive events not constituting a partition;
   - give an example of 4 pairwise mutually exclusive events that together constitute a partition;
   - give an example of 4 pairwise mutually exclusive events that do not constitute a partition;
   - give an example of 6 events constituting a partition.

2. experiment = 2 tosses with a coin
   - give examples of partitions consisting of two events
   - give an example of a partition consisting of three events
   - give an example of a partition consisting of four events
   - (*) give an example of a partition consisting of one event.
4. Probability, some basic rules

The probability of an event $A$ is denoted $P(A)$.

$0 \leq P(A) \leq 1$ (probabilities can be between 0 and 1)

$P(\Omega)=1$ (the probability of something that is sure to occur equals 1)

- If events $A$ and $B$ are mutually exclusive then $P(A \cup B) = P(A) + P(B)$
  (this is called the addition rule)

- Consequence(1) (explain): $P(\emptyset) = 0$ (the probability of something impossible equals 0)

- Consequence(2) (explain): $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive events, then

$$P\left( \bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} P(A_i) \quad (1.1)$$

Remark: it does not follow from $P(A)=0$ that $A=\emptyset$

example: toss a fair coin until the first heads. Denote $A$ the event that we never get a heads. Evidently, $A$ is not impossible ($A \neq \emptyset$) at the same time, $P(A)<1/2$, $P(A)<1/4$, $P(A)<1/8$, etc, so, for any $n \in \mathbb{N}$ holds that $P(A)<\left(\frac{1}{2}\right)^n$; but then $P(A)$ cannot be a positive number – so $P(A)$ must be 0.

Likewise, it does not follow from $P(A)=1$ that $A=\Omega$

example: the event "sooner or later we succeed in tossing a heads" in the previous example.

5. Conditional probabilities

The conditional probability of event $B$ given event $A$ shows in what percent of the cases when $A$ occurs $B$ also occurs. (We zoom into $A$, considering now this to be 100%.) Denoted by (read "$P B$ given $A$"). Its definition, according to the above explanation, is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (1.2)$$

Multiplying both sides with $P(A)$ we get

$$P(A \cap B) = P(A) \cdot P(B|A). \quad (1.3)$$

the so-called multiplication rule.

6. The total probability formula and Bayes’s formula

a. With the combination of the addition rule and the multiplication rule we get the so-called total probability formula:

let $A_1, A_2, \ldots, A_n$ be a partition, $B$ an event, then

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i) \quad (1.4)$$

[proof.approx.: $B = B \cap \Omega = B \cap \left( \bigcup_{i=1}^{n} A_i \right) = \bigcup_{i=1}^{n} (B \cap A_i) = (B \cap A_1) \cup \ldots \cup (B \cap A_n)$ /constructing $B$ from pairwise mutually exclusive events/, so $P(B) = \sum_{i=1}^{n} P(B \cap A_i)$ (addition rule) and this, according to the multiplication rule, $\sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)$]

b. Its twin brother is Bayes’s formula:
let \( A_1, A_2, \ldots, A_n \) be a partition, \( B \) an event, then

\[
P(A_j | B) = \frac{P(A_j \cap B)}{\sum P(A_i \cap B)}
\]

(1.5)

Proof: according to the definition of conditional probability:

\[
P(A_j | B) = \frac{P(A_j \cap B)}{P(B)}
\]

What remains is rewriting the numerator according to the multiplication rule, the denominator according to the formula of the full probability.

**Example 1.1.** 5 chips in a box, 3 yellow, 2 green. Two chips are drawn at random, consecutively, without replacement.

a. Find the chance that the second draw is green. Solution:

\[
\]

\[
\]

\[
= \frac{3}{5} \cdot \frac{2}{4} + \frac{2}{5} \cdot \frac{2}{4} = \frac{6}{20} + \frac{4}{20} = \frac{1}{4}
\]

(1.6)

(1.7)

(1.8)

b. Find the chance that the first draw was green, given that the second draw is green. (That is, in what percent of these games-with-two-draws in which the second draw is green, was the first draw green, too.) Solution:

\[
P(1. = G | 2. = G) = \frac{P(1. = G \land 2. = G)}{P(2. = G)} = \frac{\frac{3}{5} \cdot \frac{2}{4}}{\frac{6}{20}} = \frac{2}{8/20} = \frac{1}{4}
\]

(1.9)

**Readings**


Chapter 2. Probabilities – introductory exercises

1. (A) SIMPLE EXERCISES

1.) There are ten tokens in a box, 3 blue, 7 yellow (the tokens are well mixed, one cannot see in the box). Experiment=one draw from the box (reaching in the box & taking out one token).

• what is the probability of drawing a blue token?
• what is the probability of drawing a yellow token?
• what is the probability of drawing a black token?
• what is the probability of drawing a coloured token?

2.) We draw one card from the top of a well-shuffled deck of cards. (A deck of cards consists of 4 suits, that is hearts, diamonds, clubs & spades. Each suit consists of 13 ranks that is 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King and an Ace. There are altogether 52 cards to a deck. Hearts are signed with a red heart, diamonds with a red diamond, spades with a black spade-formed leaf, clubs with a black three-lobed leaf.)

• what is the probability of drawing the queen of clubs?
• what is the probability of drawing one of the clubs?
• what is the probability of drawing one of the kings?
• what is the probability of drawing a black card?
• what is the probability of drawing one of the major cards? (Calling the Jacks, Queens, Kings & Aces of all suits major.)

3.) Experiment=one roll with a fair dice.

• what is the probability of getting an ace?
• what is the probability of getting an even number?
• what is the probability of getting an ace or a deuce?
• what is the probability of getting at least 5?
• what is the probability of getting a number divisible by 3?

3’) Experiment=one roll with a dice loaded so that the probability of getting an ace equals 0.5, while the probability of getting one of the remaining five numbers equals 0.1 each.

• what is the probability of getting an ace?
• what is the probability of getting an even number?
• what is the probability of getting an ace or a deuce?
• what is the probability of getting at least 5?
• what is the probability of getting a number divisible by 3?

4.) There are 10 tokens in a box, 3 black, 7 white. The tokens are numbered: two of the blacks have 10 written on them, one of the blacks has 20; four of the whites have 10 written on them, three of the whites have 20. One token is drawn at random from the box.
• what is the probability of drawing the black 20?
• what is the probability of drawing a white 10?
• what is the probability of drawing a white token?
• what is the probability of drawing a 20?
• what is the probability of drawing a white 20?

5.) One card is drawn from a deck of cards that has been tampered with but we don't know what cards the deck consists of. Yet we know that, drawing one card at random, the probability of drawing a spade equals 15 percent, the probability of drawing a club equals 25 percent. Does it follow that the probability of drawing a black card equals 15+25=40 percent?

6.) We draw one card from a deck of cards that has been tampered with we but don't know what cards the deck consists of. If a spade or a king is drawn, we gain 1 $. Is is known for sure that, drawing one card at random, the probability of drawing a spade equals 25 percent, the probability of drawing a king equals 5 percent.

a. Does it follow that the probability of gaining 1 $ equals 25%+5%=30%? b) If it does not follow, can we tell the probability of gaining 1 $?

b. If it does not follow, can we tell the probability of gaining 1 $?

c. If it cannot be told, what do we know of this probability? It cannot be less than ______ . It cannot be more than _______. (Fill in the spaces.)

7.) We draw from a deck of cards that has been tampered with we but don't know what cards the deck consists of. If a spade or a king is drawn, we gain 1 $. We know for sure that, drawing one card at random, the probability of drawing a spade equals 25 percent, the probability of drawing a king equals 5 percent. We also know that the probability of drawing a king of spades equals 3%.

a. Does it follow that the probability of gaining 1 $ equals 25%+5%=30%? 

b. If it does not follow – can we tell the probability of gaining 1 $?

c. If it cannot be told, what do we know of this probability? It cannot be less than ______ . It cannot be more than _______. (Fill in the spaces.)

8.) Three rolls with a fair dice.

a. what is the probability of getting at least one ace?

b. what is the probability of getting three aces?

c. what is the probability of not getting at least one ace?

2. (B) MULTIPLICATION RULE; CONDITIONAL PROBABILITIES

1.) Two draws without replacement from a well-shuffled deck of cards.

a. what is the probability of the first card being a king and the second card being a queen?

b. what is the probability of the first card being a king?

c. what is the probability of the second card being a queen given that the first card is a king?

d. *what is the probability of the first card being a king given that the second card is a queen?

2.) Two draws without replacement from a well-shuffled deck of cards.
a. what is the probability of the first card being a spade and the second card being a hearts?
b. what is the probability of the first card being a spade?
c. what is the probability of the second card being a hearts given the first card is a spade?
d. * what is the probability of the first card being a spade given the second card is a hearts?

3.) A fair coin is tossed twice. What is the probability of getting heads both times?

4.) A fair coin is tossed twice. What is the probability of getting first a head and then a tail?

5.) A fair coin is tossed three times. What is the probability of getting all heads?

6.) A coin is tossed twice. The coin is loaded so that, with one throw, the probability of getting a head equals 0.6.
   a. what is the probability of getting all heads?
   b. what is the probability of getting all tails?

7.) A coin is tossed three times. The coin is loaded so that, with one throw, the probability of getting a head equals 0.6.
   a. what is the probability of getting all heads?
   b. what is the probability of getting all tails?

8.) Three marbles in a box, a green, a blue and a yellow. Two draws from the box. What is the probability of drawing green both times, if the draws are
   a. with replacement?
   b. without replacement?

9.) Three draws without replacement from a well-shuffled deck of cards.
   a. what is the probability that all three will be hearts?
   b. what is the probability that all three will be kings?
   c. what is the probability of the first being a king, the second a queen and the third a jack?

10.) Three rolls with a fair dice
   a. what is the probability of all three being sixes?
   b. what is the probability of all three being less than six (1...5)?
   c. what is the probability of neither being a six?
   d. what is the probability of getting at least one six from the three rolls?

11.) 20 marbles in a box, 10 marked with 'H', the other 10 marked with 'T'. Two draws from the box, with replacement. What is the probability of getting all 'H's ?

12.) 20 marbles in a box, ten marked with an 'H', the other ten marked with a 'T'. Three draws from the box, with replacement. What is the probability of getting all 'H's ?

(Are there, among the former ones, exercises closely related to the last two?)

3. (C) ADDITION RULE
1.) There were fifty children at a party where cookies and ice-cream were served. Of the fifty 12 ate cookies, 17 ate ice-cream. True or false: 12+17=29 children ate either cookies or ice-cream. Explain.

2.) Two cards are dealt from the top of a well-shuffled deck. You can choose:
   i. you get one dollar if the first card is an ace or the second card is an ace
   ii. you get one dollar if at least one ace shows.
      Which is better? Or are they the same? Explain.

3.) We are going to roll with two dice. The chance that the first shows 1 is 1/6. The chance that the second shows 2 is 1/6. True or false: the chance that the first shows 1 or the second shows 2 is 1/6+1/6=2/6. Explain.

4) Ten cards in a box, the cards are numbered (1..10). Five draws at random from the box, with replacement. True or false: the chance that at least one 7 shows equals 5/10. Explain.

5) A number is drawn at random from a box. There is a 20% chance for it to be 10 or less. There is a 10% chance for it to be 50 or more. True or false: there is a 70% chance for it to be between 10 and 50 (endpoints excluded). Explain.

6) Five cards in a box, 2 hearts, 3 spades. Two draws at random from the box. What is the probability of drawing at least one hearts, if the draws are
   a. with replacement?
   b. *) without replacement?

7) Five cards in a box, 2 hearts, 3 spades. Two draws at random from the box. What is the probability of drawing a hearts first time or drawing a spade second time, if the draws are
   a. with replacement?
   b. *) without replacement?

8) Five cards in a box, 1 King and 4 Tens. We are going to draw two cards at random, consecutively, without replacement. We would like to know the chance that there will be at least one King among the draws. Our reasoning goes like this:
   • The event we would like to know the chance of is "the first draw will be a king OR the second draw will be a king"
   • there is 1 king among the 5 cards so the chance that the "first draw will be a king" equals 1/5
   • each of the 5 cards has an equal chance to be the card drawn the second time, so the chance that the "second draw will be a king" equals 1/5
   • so the chance that "the first draw will be a king OR the second draw will be a king" equals 1/5+1/5.
      Is it right? Is it wrong? Is anything missing?

9) Five cards in a box, 2 Kings and 3 Tens. We are going to draw two cards at random, consecutively, without replacement. We would like to know the chance that there will be at least one King among the draws. Our reasoning goes like this:
   • The event we would like to know the chance of is "the first draw will be a king OR the second draw will be a king"
   • there are 2 kings among the 5 cards so the chance that the "first draw will be a king" equals 2/5
• each of the 5 cards has an equal chance to be the card drawn the second time, so the chance that the "second draw will be a king" equals 2/5

• so the chance that "the first draw will be a king OR the second draw will be a king" equals 2/5 + 2/5 = 4/5.

Is it right? Is it wrong? Or is anything missing?

4. (D) INDEPENDENCE AND CONDITIONAL PROBABILITIES

1.) One roll with a fair dice –

A bets that it will be an even number;

B bets that it will be 4 or more.

a. what is A’s chance for winning?

b. The die has been rolled. A does not yet know the outcome, but sees from B’s face that B has won. What is A’s chance for winning now?

c. Is it good news for A that B has won? is it bad news? or is it all the same?

(A may take his bet when the die has been rolled so, of course, ha cannot see the roll. But, seeing B’s face, he knows whether B has won.)

d. When deciding if he is to take his bet upon the just-happened roll he does not know the outcome of, is it good for A to take into account whether B has won?

    When is it better for him to bet: if B has won? or if B has not won? or is it all the same?

e. Is knowing whether B has won of any help for A in finding his chance for winning?

f. Is A’s success independent of B’s success? (in the probabilistic sense)

2.) (Previous exercise, roles partially reversed.)

One roll with a fair dice –

A bets that it will be an even number;

B bets that it will be 4 or more.

a. what is B’s chance for winning?

b. The die has been rolled. B does not yet know the outcome, but sees from A’s face that A has won. What is B’s chance for winning now?

c. Is it good news for B that A has won? or is it bad news? or is it all the same?

Now it is B who can wait with his decision until after the die is rolled.

d. When deciding if he is to take his bet upon the just-happened roll he does not know the outcome of, is it good for B to take into account whether A has won?

    When is it better for him to bet: if A has won? or if A has not won? or is it all the same?

e. Is knowing whether A has won of any help for B in finding his chance for winning?

f. Is B’s success independent of A’s success?

3.) A game: 10 cards, numbered 1..10, in a box. One draw from the box,
Anne bets that the draw will be an even number – or she does not bet.

Basil bets that the draw will be a number divisible by 3 – or he does not bet.

They repeat this game for a number of times.

– in what percent of the games will Anne win, approximately?

Anne may decide whether to bet when, the draw having been made, she sees if Basil has won.

– does betting only when Basil wins improve Anne’s chances?

– does betting only when Basil loses improve Anne’s chances?

In what percent of the games Basil wins will Anne win?

In what percent of the games Basil loses will Anne win?

Is knowing whether Basil has won of any help for Anne in estimating her chances for winning?

Denote A the event that Anne wins in a game (that is, an even number is drawn);

denote B the event that Basil wins in a game (that is, a number divisible by 3 is drawn);

– are events A and B independent?

5. (E) MULTIPLICATION RULE AND ADDITION RULE COMBINED (a preliminary exercise for the total probability formula and Bayes’s formula)

- There are two boxes, box I. with 50 silver marbles and 50 golden marbles in it, and box II. with 90 silver marbles and 10 golden marbles in it.

The game (that is, the experiment) consists of three stages:

– stage 1 (selecting a box): the Master of Ceremonies rolls a fair dice and selects a box for the players to draw from; if the result is 1, 2, 3 or 4 it will be box I. If the result is 5 or 6 it will be box II. The boxes look identical and the players can not see the rolling of the dice so they do not know which box they are drawing from.

– stage 2 (first draw): player A draws one marble at random from the box selected in stage 1. (The result is noted down, the marble drawn put back in the box, the marbles in the box are thoroughly mixed.)

– stage 3 (second draw): player B draws one marble at random from the box selected in stage 1. (The result is noted down, the marble put back, the marbles in the box thoroughly mixed.)

(Assume that drawing a golden marble is better than drawing a silver marble.)

Imagine observing 3000 such games. Approximately,

a) in how many of these will the players draw from box I.? ( from box II.?)

b) in how many of these will A draw a golden marble?

c) in how many of these will B draw a golden marble?

d) in how many of these will A draw a silver marble?

a’) find the chance that the players draw from box I. ( from box II.)

b’) find the chance that A draws a golden marble.
c’) find the chance that B draws a golden marble.

d’) find the chance that A draws a silver marble.

e) how many of the 3000 games will both players draw golden marbles in, approximately?

e’) find the chance that both players draw golden marbles.

Approximately, of the 3000 games,

f) in how many will A get a silver, B a golden marble?

g) in how many will A get a golden, B a silver marble?

h) in how many will both A and B get silver marbles?

Find the chance that

f’) A will get a silver, B a golden marble.

g’) A will get a golden, B a silver marble.

h’) both A and B will get silver marbles.

i) in what percent of the games in which A has got a golden marble does B get a golden marble, too?

j) in what percent of the games in which A has got a silver marble does B get a golden marble?

k) in what percent of the games in which A has got a golden marble does B get a silver marble?

l) in what percent of the games in which A has got a silver marble does B get a silver marble, too?

i’) find the chance that B draws a golden marble, given that A has drawn a golden marble.

j’) find the chance that B draws a golden marble, given that A has drawn a silver marble.

k’) find the chance that B draws a silver marble, given that A has drawn a golden marble.

k’) find the chance that B draws a silver marble, given that A has drawn a silver marble.

m) Is it good news for B if A draws a golden marble? is it bad news? is it all the same?

Has any of the above exercises been connected with DeMorgan’s laws?

Readings

Chapter 3. Independence of events, questions (exercises)

1) One roll with a fair dice,

   a) A=(getting an even number)  B=(getting a number divisible by 3)
   b) A=(getting an even number)  B=(getting 1 or 6)
   c) A=(getting an even number)  B=(getting at least 4)
   d) A=(getting a number divisible by 3) B=(getting at least 4)
   e) A=(getting a number divisible by 3) B=(getting 1 or 6)
   f) A=(getting an even number)  B=((getting a number divisible by 5)
   g) A=(getting 1 or 2) B=(getting 1 or 6)
   h) A=(getting 1, 2 or 3) B=(getting 1 or 6)

Are A and B independent?

2) One roll with a die loaded so that the chance of getting a 6 is 50 percent, the chance of getting 1,2,3,4 or 5 is 10% each. Are A and B independent?

   a) A=(getting an even number)  B=(getting a number divisible by 3)
   b) A=(getting an even number)  B=(getting 1 or 6)

3) Two rolls with a fair dice. Are A and B independent?

   a) A=(getting an even number first)  B=(the sum of the two numbers is even)
   b) A=(getting an even number first)  B=(the product of the two numbers is even)
   c) A=(getting 1 first) B=(the sum of the two numbers equals 7)
   d) A=(getting 1 first) B=(the sum of the two numbers is 5 or less)
   e) A=(getting 1 first) B=(the sum of the two numbers is between 5 and 7, endpoints included)
   f) A=(getting 1 first) B=(1 shows at least once)

4) Two rolls with a die loaded so that the chance of getting a 6 is 50 percent, the chance of getting 1,2,3,4 or 5 is 10% each. Are A and B independent events, if

   a) A=(getting an even number first)  B=(the sum of the two numbers is even)
Independence of events, questions
(exercises)

b) A=(getting an even number first) B=(the product of the two numbers is even)
c) A=(getting 1 first) B=(the sum of the two numbers equals 7)
d) A=(getting 1 first) B=(the sum of the two numbers is 5 or less)
e) A=(getting 1 first) B=(the sum of the two numbers is between 5 and 7, endpoints included)
f) A=(getting 1 first) B=(1 shows at least once)

5) Two cards are dealt from the top of a well-shuffled deck (without replacement).

1/52 is the chance that the first card is the king of spades;
1/52 is the chance that the second card is the king of spades.

Right or wrong: 1/52 x 1/52 is the chance that the first card is the king of spades AND the second card is the king of spades. (Explain briefly.)

6) Two cards are dealt from the top of a well-shuffled deck (without replacement).

4/52=1/13 is the chance that the first card is a king
4/52=1/13 is the chance that the second card is a king.

Right or wrong: 1/13 x 1/13 is the chance that the first card is a king AND the second card is a king. (Explain briefly.)

7) Two cards are dealt from the top of a well-shuffled deck (with replacement).

1/52 is the chance that the first card is the king of spades;
1/52 is the chance that the second card is the king of spades.

Right or wrong: 1/52 x 1/52 is the chance that the first card is the king of spades AND the second card is the king of spades. (Explain briefly.)

8) Two cards are dealt from the top of a well-shuffled deck (with replacement).

4/52=1/13 is the chance that the first card is a king
4/52=1/13 is the chance that the second card is a king.

Right or wrong: 1/13 x 1/13 is the chance that the first card is a king AND the second card is a king. (Explain briefly.)

9) Jól megkevert franciakártya-pakliból húzunk egy lapot. Független-e a B eseménytől az A esemény, hogyha

a) A=(it is red) B=(the card has got a number /2..10/ on it)
b) A=(it is red) B=(it is a hearts)
c) A=(the card has got a number /2..10/ on it) B=(it is a king)
d) A=(it is red) B=(it is a hearts or a spade)
e) A=(it is a 2 or a 4) B=(it is a 4 or a 6)
Independence of events, questions
(exercises)

(The red suits are the hearts and the spades)

10) A card is dealt from a deck containing only six cards:
   the king of spades, the king of clubs and the king of hearts,
   the queen of hearts, the queen of diamonds and the queen of spades.
   1/2 is the chance that the card is a king;
   1/3 is the chance that the card is a spade.
   Is it true: 1/2 x 1/3 is the chance that the card is the king of spades?

11) A card is dealt from a deck containing only six cards:
   the king of spades, the king of clubs and the king of hearts,
   the queen of hearts, the queen of diamonds and the queen of spades.
   1/2 is the chance that the card is a king;
   1/6 is the chance that the card is a diamond.
   Is it true: 1/2 x 1/6 is the chance that the card is the king of diamonds?

12) A card is dealt from a deck containing only six cards:
   the king of spades, the king of clubs and the king of hearts,
   the queen of hearts, the queen of diamonds and the queen of spades.
   a) are suit and rank independent? (suits are: clubs, spades, diamonds and hearts; ranks are the 13 cards in a suit:
      2,3,...10,J,Q,K,A) (see Freedman-Pisani-Purves:Statistics, p.230-233)
   b) are A and B independent events?
      A=(it is a king)    B=(it is a club)
   c) are A and B independent events?
      A=(it is a king)    B=(it is a spade)
   d) are A and B independent events?
      A=(it is a king)    B=(it is red)
   e) are A and B independent events?
      A=(it is a king)    B=(it is a hearts)

12') A card is dealt from a deck containing only six cards:
   the king of spades, the king of clubs and the king of hearts,
   the queen of spades, the queen of clubs and the queen of hearts.
   a) are suit and rank independent?
   b) are A and B independent events?
      A=(it is a king)    B=(it is a club)
   c) are A and B independent events?
Independence of events, questions
(exercises)

A=(it is a king)  B=(it is a spade)
d) are A and B independent events?

A=(it is a king)  B=(it is red)
e) are A and B independent events?

A=(it is a king)  B=(it is hearts)

13) Two rolls with a fair dice. Denote A that the first number is even; B that the second number is even; C that the sum of the numbers is even. True or false:
a) P(AB)=P(A)P(B)
b) P(AC)=P(A)P(C)
c) P(BC)=P(B)P(C)
d) P(ABC)=P(A)P(B)P(C)

14) 3 tosses with a coin. Denote A that at least one head shows and at least one tail shows; B that more than one heads show. Are A and B independent?

15) One roll with two dice (a blue and a green one). Denote A that the sum of the numbers is even; B that a 2 shows on the blue die; C that at least one 2 shows.
a) are A and B independent?
b) are A and C independent?

16) At an oral exam the questions are written on 25 seemingly identical slips of paper. Of the 25 questions 22 are "good", 3 are "wicked". Two students, X and Y enter and draw one question each, at random, without replacement. Which of them is to draw first is decided by tossing a coin. Denote A that X draws first; B that X draws a "good" question; C that Y draws a "good" question.
a) are A and B independent?
b) are B and C independent?
c) Assume the coin is loaded so that X's chance to draw first is p, Y's chance to draw first is (1−p). Are A and B independent now?

17) Ally has got three coins in his pocket, one fair and two loaded so that the chance of getting a heads is 0.80 with them. He chooses one of the coins randomly then tosses it twice. Denote H₁ that the first toss is a heads; H₂ that the second toss is a heads. Are H₁ and H₂ independent?

18) Ten cards, numbered 1..10, in a box. Two draws at random from the box, without replacement. Denote Aᵢ (i=1..10) that the ith draw happens to be the card with an i on it. Are Aᵢ and Aᵢ independent?

Readings

Chapter 4. Total probability formula & Bayes's formula (exercises)

1) Two cards are dealt, without replacement, from a well-shuffled deck. What is the chance that the second card will be a king?

2) Three cards are dealt, without replacement, from a well-shuffled deck. What is the chance that the third card will be a king?

3) Three doormen guard a discotheque. A checks 20% of the guests, B checks 30% of the guests and C checks the remaining 50%. A has a 75% chance of noticing if the membership card of the guest he is checking is invalid, B has a 50% and C has only a 20% chance. (The guests cannot choose, they get randomly to the three doormen.)

– X has an invalid membership card – what are his chances of getting in? (He tries only once, in this exercise.)

4) In Trolland 80% of the yellow-haired and 50% of the dark-haired inhabitants are clever. 40% of the inhabitants are yellow-haired and 60% are dark-haired.

a) What percent of all the inhabitants are clever in Trolland?

b) The names of all the trolls of Trolland are thrown into a big hat, one name is drawn at random. What is the chance of a clever troll's name being drawn?

c) The names of all Trolland trolls being thrown into a big hat, one name has been drawn at random. The owner of this name has been examined and found clever. What is the chance of him being yellow-haired?

5) 10 litres of a 15 volume percent solution of alcohol (A) and 20 litres of a 10 volume percent solution of alcohol (B) are mixed.

– what alcohol percentage is the mixed solution?

– what percentage of the alcoholic content of the mixed solution comes from solution A?

6) 70% of the population of Amazony are women, 30% men. 20% of the population of Trolland are women, 20% men. The two states unite. 40% of its population come from Amazony, 60% come from Trolland.

– what is the percentage of women in the new state?

– what percent of the women in the new state come from Amazony?

7) The first question of the maths exam will be a multiple choice question. The chance that Ally knows the right answer to it is 0.20. Not knowing the right answer, he chooses one of the options at random.

a) find the chance that Ally checks the right answer

b) find the chance that he knew the right answer given he has checked the right answer.

8) The first question of the maths exam was a multiple choice question. Some students knew the right answer. Those not knowing it chose one of the options at random. Of the 600 students taking the exam, 300 checked the right answer. How many knew the right answer, approximately?

9) The questions are written on 25 seemingly identical slips of paper at an oral exam. Only 3 of the 25 questions are "good". Enter Celia, then Demetrius, they draw one question each. (They keep their questions until exit.) Find the chance that

a) Celia draws a good question.

b) Demetrius draws a good question.
c) they both draw good questions.

10) A new contagious disease has emerged in Trolland, the black-disease. It is well-curable during the first two years after contagion – but the symptoms manifest themselves only after the third year. A new screening protocol might help; its precision is such that it signals the existence of the contagion in 95% of those having the virus, and signals contagion only at 3% of those not having the virus. Assume that at present 2% of the Troll population has got the infection. All the inhabitants of Trolland (that is 5 million Trolls) are to be screened.

a) How many trolls will be found having the virus in the screening, approximately?

Among those found having the virus there will be trolls having the virus – and, because the screening process is not error-free, also trolls indeed not having the virus.

b) What percent of those found having the virus really have the virus?

b') what percent of those found having the virus do not have the virus? (Of those sent to further medical examinations, which might prove to be a bit unnerving, what percent have got here erroneously, only because of the unavoidable classification errors of the screening process?)

11) There are 3 red, 2 blue and 4 green marbles in a box. Three draws are made at random, one by one, without replacement.

a) find the chance that the second draw is blue, given the first draw is red;

b) find the chance that the first draw was red (you missed the drawing), given the second draw is blue (you have seen this draw).

12) This exercise is about families with two children; denote bg the families where the first child is a boy, the second child is a girl. Assume that the 4 possible combinations (bb, bg, gb, gg) have equal chances. Given one of the children is a boy, find the chance that the other child is a boy, too.

13) Given the first child is a boy, find the chance that the other child is a boy, too.

14)* A deck is shuffled and then dealt evenly among four players. Peter is one of the players.

a) Find the chance of these events:

\[ A = (\text{Peters's topmost card is the ace of hearts}) \]
\[ B = (\text{Peters's topmost card is an ace}) \]
\[ C = (\text{Peter has got the ace of hearts}) \]
\[ D = (\text{Peter has got at least one of the aces}) \]
\[ E = (\text{each of the players has got an ace}) \]
\[ F = (\text{Peter has got two of the aces [not more, not less]}) \]

b) Find the value of these conditional probabilities:

\[ P(\text{C | A})P(\text{A | C})P(\text{B | D})P(\text{A | D})P(\text{C | E}) \]
\[ P(\text{F | A})P(\text{F | D})P(\text{F | C}) \]

15) There are 3 red and 7 blue marbles in satchel (A), 6 red and 4 blue marbles in satchel (B), 9 red and 1 blue marbles in satchel (C). From which satchel to draw one marble is decided upon by rolling a dice. With a roll of 1 satchel (A) is selected, with a roll of 2 or 3 satchel (B), with a roll of 4,5 or 6 satchel (C) is selected. Find the chance that the roll was a 1 given a red marble is drawn.

16) There are 5 red, 3 blue and 2 yellow marbles in a box. Two marbles are drawn, consecutively. Find the chance that the first marble is red given the two marbles are of the same colour.
17) There are 2 golden rings in the first of three little two-drawer boxes, a golden and a silver ring in the second box, and two silver rings in the third box (a ring in each drawer). One of the three boxes is selected at random, one of its drawers is selected, then we pull it open. There is a golden ring in it. Find the chance that it is the first box. (Bertrand's box paradox.)

18.) There are coloured plastic chips in three, seemingly identical boxes, 1 red, 9 blue in the first box, 5 red and 5 blue in the second box and 9 red and 1 blue in the third box. One of the boxes is selected at random, then two chips are drawn from the box, with replacement.

a) find the chance that the first chip is red;

b) find the chance that the second chip is red;

c) find the chance that both chips are red;

d) are the two draws independent?

19.) (Three prisoners/1) There live in a jail three prisoners sentenced to death. One of them will be executed next morning. Which of them will be the one is decided upon by a draw giving each a 1/3 chance. The decision is already known to the jailer, but not known to the prisoners yet. They are not even allowed to get information concerning themselves. Mr.X wants to know more, though. He tells the jailer that, as at least one of the other two prisoners (Mr.Y and Mr.Z) certainly stays alive, it is not against the rules if the jailer tells him the name of one not to be executed next morning (a name not of himself). The jailer agrees and tells him that Mr.Y will not be executed. Mr.X becomes very sad. Up to now, he says, he had a 2/3 chance not to be the one to be executed, but from now on his chance has diminished to a low 1/2. Is he right?

19') (Three prisoners/2) There live in a jail three prisoners. Which of them will be the one tomorrow to shovel coal early in the morning in the cold rain is decided upon by a draw done by the administration the previous evening, giving each a 1/3 chance. The decision is already known to the jailer, but not known to the prisoners yet. They are not even allowed to get information concerning themselves. (They will get this information at the 4 o’clock reveille). Mr.X wants to know more, though. He tells the jailer that, as at least one of the other two prisoners (Mr.Y and Mr.Z) may certainly sleep till later, it is not against the rules if the jailer tells him the name of one not to shovel coal early in the morning (a name not of himself). The jailer agrees and tells him that Mr.Y is to sleep till later. Mr.X becomes very sad. Up to now, he says, he had a 2/3 chance not to be the one to be shovelling coal in the rain, but from now on his chance has diminished to a low 1/2. Is he right?

20) Alex has got three coins in his pocket, one fair and two loaded such that the chance of getting a heads is 0.90 with them each. He chooses one of the coins randomly (giving 1/3 chance to each) then tosses it twice (or three times, in the last three questions). Find the values of these, unconditional and conditional, probabilities:

a) P(the first toss is heads)

b) P(the fair coin is drawn and the first toss is heads)

c) P(the fair coin is drawn given the first toss is heads)

d) P(the fair coin is drawn given the first toss is tails)

c') P(a loaded coin is drawn given the first toss is heads)

d') P(a loaded coin is drawn given the first toss is tails)

e) P(the second toss is heads given the first toss is heads)

f) P(the second toss is heads given the first toss is tails)

e') P(the second toss is tails given the first toss is heads)

f') P(the second toss is tails given the first toss is tails)

g) P(a loaded coin is drawn given both draws are heads)

h) P(the fair coin is drawn given one was heads and one tails of the two tosses)
i) \( P(\text{the third toss will be heads given the first and second tosses both were heads}) \)

j) \( P(\text{the first three tosses are all heads}) \)

k) \( P(\text{two are heads, one is tails of the first three tosses}) \)

21) (continues exercise 20) Denote \( H_1 \) that the first toss is a head; denote \( H_2 \) that the second toss is a head. Are \( H_1 \) and \( H_2 \) independent? Explain briefly.

**Readings**


Chapter 5. Variables – simple exercises – distributions, expected values (E.V.) & standard deviations (SD) (exercises)

1) One roll with a fair dice. What are the possible values of the rolls? – find the chance of each.

2) Two rolls with a fair dice; (a) what are the possible values of the sum of the rolls? – find the chance of each; (b) what are the possible values of the product of the rolls? – find the chance of each.

3) There are three numbered cards in a box, one 0, one 1 and one 10. One card is drawn (this is the experiment);\(X:=\)the number drawn. Find the distribution of\(X\) [that is: what are its possible values? and what are the chances of each?].

4) There are eight numbered cards in a box, five 0s, two 1s and one 10. Experiment=one draw, \(X:=\)the number drawn. Find the distribution of \(X\).

5) There are three numbered cards in a box, one 0, one 1 and one 10. Experiment=two draws, without replacement.

a) \(X:=\)the sum of the numbers drawn. Find the distribution of \(X\).

b) \(Y:=\)the second draw. Find the distribution of \(Y\).

6) There are eight numbered cards in a box, five 0s, two 1s and one 10. Experiment=two draws, without replacement.

a) \(X:=\)the sum of the numbers drawn. Find the distribution of \(X\).

b) \(Y:=\)the second draw. Find the distribution of \(Y\).

7) as 5) and 6), but with replacement

8) as (a) of 5)–7), but with the products to be examined instead of the sums.

9) A fair coin is tossed twice; denote \(X\) "the number of heads tossed". What are the possible values of \(X\)? – find the chance of each. [That is, find the distribution of \(X\)]

10) Playing with a fair coin, you get 1 $ from the bank if a head is tossed; you get 10 $ from the bank if a tail is tossed. Observing two tosses, what are the possible values of your net gain? [that is, of the sum of the gains] – find the chance of each.

11) Playing with a fair coin, you get nothing from the bank if a head is tossed; you get 1 $ from the bank if a tail is tossed. Observing two tosses, what are the possible values of your net gain? – find the chance of each.

12) Playing with a fair coin, you get 1 $ from the bank if a head is tossed; but you pay 1 $ for the bank if a tail is tossed. Observing two tosses, what are the possible values of your net gain? – find the chance of each.

13) Playing with a fair coin, you get 2 $ from the bank if a head is tossed; but you pay 1 $ for the bank if a tail is tossed. Observing two tosses, what are the possible values of your net gain? – find the chance of each.

14) As exercises 9–13, but for three tosses instead of two.

15) As exercises 9–13, but tossing a coin loaded such that the chance of tossing a head is 0.60, the chance of tossing a tail is 0.40.
Variables – simple exercises –
distributions, expected values
(E.V.) & standard deviations (SD)
(exercises)

16) A fair coin is tossed three times; the variable is the number of heads tossed – what are its possible values? – find the chance of each.

17) There are 10 blue and 10 green marbles in a box. Three are drawn at random, with replacement. X:="the number of blue marbles drawn". What are its possible values? – find the chance of each.

18) (Same as 17 but without replacement.)

19) A coin loaded such that the chance of tossing a head is 0.60, is tossed three times; X:="the number of heads tossed". What are its possible values? – find the chance of each.

20) There are 6 blue and 4 green marbles in a box; three draws are made at random, with replacement. X:="the number of blue marbles drawn". What are its possible values? – find the chance of each.

21) (Same as 20, but without replacement.)

22) A fair coin is tossed until a head shows. X:=number of tosses until the first head. What are its possible values? – find the chance of each.

23) A fair dice is tossed until a six shows. X:=number of tosses until the first six. What are its possible values? – find the chance of each.

24) A loaded coin with a probability of heads=0.60 is tossed until a head shows. X:=number of tosses until the first head. What are its possible values? – find the chance of each.

25) There are ten numbered cards in a box, 7 ones, 2 twos and 1 nine. Three draws are made from the box at random with replacement. X:="the sum of the draws. Find the distribution of X. Make a bar chart. (A bar chart is a graph showing the possible values of X on the x-axis, their probabilities with bars proportionate in height to the probabilities.)

26) Find the expected values (E.V.) of the random variables (distributions) in exercises 1–21 and 25.

27) Find the standard deviations (S.D.) (also standard error, S.E.) of the random variables in exercises 1–21 and 25.

Readings

Chapter 6. Variables (univariate): distribution, density function, distribution function, expected value, variance and covariance, standard deviation (theory)

1. Random variables

When the outcome of a chance experiment is a number it is called a random variable. (If the outcome is a single number, it is a univariate random variable (this handout is about univariate variables), if the outcome consists of several numbers – that is, it is a vector –, it is a multivariate random variable – see handouts (22)-(23).)

2. Distribution, density function, cumulative distribution function

The distribution of a random variable – what values it takes and with what probabilities – can be described in several ways. If its range is a finite set (finite discrete distribution) then the values can be enumerated in a table giving the probabilities also (a distribution table).

An infinite discrete variable has an infinite range consisting of discrete values. Its values can be enumerated, so its distribution can be described the above way, enumerating the values and their respective probabilities. (For example X:=number of rolls with a fair dice until a six shows.)

- Continuous distributions are those that can be given with a density function. The function \( f_x(x) \) is the density function (d.f.; also: probability density function [p.d.f.] or density) of the random variable \( X \) if for sets \( A \subseteq \mathbb{R} \) the probability of \( X \) falling in \( A \) can be given with the integral of \( f_x(x) \) on \( A \):

\[
P(X \in A) = \int_{A} f_x(x)dx
\]  

(6.1)

- The cumulative distribution function (c.d.f.) \( F_x(x) \) of a univariate random variable can be given as:

\[
F_x(x) = P(X < x)
\]  

(6.2)

so \( F_x(x) \) gives the probability of the variable being less than \( x \). It gathers, that is it cumulates, the probabilities pertaining to values less than \( x \). Though less graphic than the density function, it has the advantage of being universal: all univariate random variables have cumulative distribution functions.

Some properties of density functions: if \( f_x(x) \) is a density then

a. \( f_x(x) \geq 0 \)

b. \( \int_{-\infty}^{\infty} f_x(x)dx = 1 \)

Some properties of cumulative distribution functions: if \( F_x(x) \) is a cumulative distribution function then

a. \( \lim_{x \to -\infty} F_x(x) = 0 \)

b. \( \lim_{x \to \infty} F_x(x) = 1 \)

\( ^{\text{a actually only for sufficiently nice subsets of } \mathbb{R} – \text{ intervals, unions of a finite or countably infinite set of intervals, their complements, etc., called Borel sets. (See measure theory)}.} \)
Variables (univariate); distribution, density function, distribution function, expected value, variance and covariance, standard deviation (theory)

c. $F_X(x)$ is nondecreasing, and
d. $F_X(x)$ is continuous on the left for all $x \in \mathbb{R}$

Connection between densities and cumulative distribution functions: if the random variable $X$ is continuous (that is, has a density function) then

$$F_X(x) = f_X(x) \quad \text{and} \quad \int_{-\infty}^x f_X = F_X(x) \quad \text{(6.3)}$$

3. Expected value (E.V.)

3.1. The expected value of a random variable $X$ with finite range is:

$$E(X) := \sum_{i=1}^{N} P(X = x_i) \cdot x_i \quad \text{(6.4)}$$

Making a number of experiments on a random variable $X$ we would like to know the average of its observed values, approximately. For example, when gambling, denote $X$ the value of your net gain in one game, in dollars (with signs: a negative value signifies a loss) – the expected value of $X$ then shows your average gain, per game.

Take $N$ experiments with $X$. Denote $X_i (i=1...n)$ the values of $X$ and $p_i (i=1..n)$ their respective probabilities. Then of the $N$ experiments there will be approximately $N \cdot p_i$ cases when the value of $X$ (that is, your gain) equals $x_i$. The sum of your gains from the $N$ games will be, approximately, $\sum_{i=1}^{N} x_i \cdot N \cdot p_i$; approximately, and the average of the gains from the $N$ games (the expected value of $X$) will be its $1/N$th part, that is

$$E[X] = \frac{1}{N} \sum_{i=1}^{N} x_i \cdot N \cdot p_i = \sum_{i=1}^{N} x_i \cdot p_i \quad \text{(6.5)}$$

A bit more exactly:

divide the above infinite sum into two, one part summing over positive values of $X$, the other part summing over the negative values:

$$S_+ = \sum_{x_i > 0} P(X = x_i) \cdot x_i \quad \text{and} \quad S_- = \sum_{x_i < 0} P(X = x_i) \cdot x_i$$

If both partial sums are finite, the expected value is computed according to the above definition (and the sequence and grouping of the numbers to be added might be arbitrary).

If one of the partial sums is finite & the other infinite then the expected value is plus or minus infinity. (Eg. if $S_+$ is infinite $E(X) = +\infty$.)

If both partial sums are infinite then the distribution has no expected value (its expected value is undefined).

(The reason: if an infinite sum $\sum_{x_i} x_i$ is such that $x_i \to \emptyset$, and both $S_+$ and $S_-$ are unbounded then the original sum can be reordered such that it converge to zero; can be reordered such that it converge to $+\infty$ or to $-\infty$; or to any given $x_0 \in \mathbb{R}$. Therefore it is better considering these sums undefined.)

3.3. The expected value of continuous variables:

\[ mostly \ continuous. \ Where \ it \ is \ discontinuous \ it \ has \ a \ leap \ because \ of \ being \ nondecreasing. \ At \ these \ leaps \ it \ is \ continuous \ on \ the \ left (explain). \ The \ leaps \ signify \ single \ values \ with \ positive \ probabilities. \]
if variable $X$ has a density function $f_X(x)$ then

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

Why:

- similarly to introducing $\hat{\bar{X}}$, the range of $X$ is divided into intervals $I_i (i = 1, \ldots, n)$. One point $x_i$ is selected from each interval $I_i (x_i \in I_i)$, then the continuous $X$ is approximated by the discrete variable $X'$ such that if $X \in I_i$ then $X' := x_i$. (That is, instead of $x$-values in $I_i$, we take the single value of $x_i$ close to them.)

This way $X'$ is close to $X$, so

$$E(X) \approx E(X') = \sum_i x_i P(X' = x_i) = \sum_i x_i P(X \in I_i) = \sum_i (x_i \int_{I_i} f_X(x)) =$$

$$= \sum_i (\int_{I_i} x f_X(x) dx) \approx \sum_i (\int_{I_i} f_X(x) dx) = \int_{-\infty}^{\infty} f_X(x) dx$$

Sometimes the expected value of a function $g(x)$ (e.g., the square or the logarithm etc.) of a random variable $X$ with a known density function $f_X(x)$ is needed. The formula is:

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

Why: the range of $X$ is divided into intervals $I_i (i = 1, \ldots, n)$. One point is selected from each interval ($x_i \in I_i$), then the continuous $X$ is approximated by the discrete variable $X'$ such that if $X \in I_i$ then $X' := x_i$. This way $g(X')$ is near to $g(x)$ (if $g(x)$ is continuous), so

$$E(g(X)) \approx E(g(X')) = \sum_i g(x_i) P(X' = x_i) = \sum_i g(x_i) P(X \in I_i) =$$

$$\approx \sum_i g(x_i) \int_{I_i} f_X(x) dx \approx \sum_i \int_{I_i} g(x) f_X(x) dx =$$

$$\approx \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx$$

(Though for the above explanation the continuousness of $g(x)$ is necessary at two points (that is, at the approximations `≈') the formula holds for non-continuous functions $g(x)$ as well. The reasoning needs higher mathematics.)

Especially, the expected value of the square of a variable is often needed. This can be computed

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) dx$$

Remark: like in the discrete case, qualifications hold if at least one of the 'partial sums' $\int_{x_0}^{x} x \cdot f_X(x) dx$ and $\int_{x_0}^{\infty} x \cdot f_X(x) dx$ is not finite. If exactly one of them is infinite the E.V. is defined plus or minus infinity; if both are infinite no E.V. is defined.

### 3.4. Some properties of the expected value:

(a) $E(c) = c$ \quad /with $c$ denoting a variable that does not vary, its value being the constant $c$/

(b) $E(c+X) = c + E(X)$

(c) $E(cX) = c \cdot E(X)$

(d) $E(X+Y) = E(X) + E(Y)$ \quad /$X$ and $Y$ random variables with finite expected values $E(X)$ and $E(Y)$/  

(e) if $X$ and $Y$ are independent random variables with finite expected values $E(X)$ and $E(Y)$ then $E(XY) = E(X) \cdot E(Y)$
Variables (univariate): distribution, density function, distribution function, expected value, variance and covariance, standard deviation (theory)

If the random variables \( X \) and \( Y \) with finite expected values \( E(X) \) and \( E(Y) \) are dependent then their co-dependence might be described with the difference between \( E(XY) \) and \( E(X)E(Y) \) – that is, with the covariance of \( X \) and \( Y \): 
\[
\text{cov}(X,Y) := E(XY) - E(X)E(Y)
\]

(f) another formula for \( \text{cov}(X,Y) = E((X-E(X))(Y-E(Y))) \)

Why it is called co-variance: the quantity on the right-hand side of (f) is positive if the deviations of \( X \) and \( Y \) are more or less simultaneous from their respective expected values upwards, and are more or less simultaneous downwards. (In this case the factors of the product are either both positive, the product being positive also, or are both negative, the product being positive again.) While, if the deviations of \( X \) and \( Y \) are more or less simultaneous but are in opposite directions – that is, \( Y \) is big when \( X \) is small and \( Y \) is small when \( X \) is big – then, mostly, one factor of the product will be positive, the other negative the product being typically negative, so its expected value will be negative, too.

explanation of (d), for variables with discrete & finite distributions

denote 
\[
p_{i,j} := P(X = x_i \land Y = y_j)
\]

\[
p_{i,\cdot} := \sum_j p_{i,j} := P(X = x_i)
\]

\[
p_{\cdot,j} := \sum_i p_{i,j} := P(Y = y_j)
\]

/ probabilities defining the distribution of \( X \),

\[
p_{i,\cdot} := \sum_j p_{i,j} := P(Y = y_j)
\]

/ probabilities defining the distribution of \( Y \),

\[
E(X + Y) = \sum_{i,j} p_{i,j} (x_i + y_j) = \sum_i (p_{i,x_i} + p_{i,y_i}) = \sum_i p_{i,x_i} + \sum_j p_{j,y_j}
\]

explanation of (e), for variables with discrete & finite distributions

– if \( X \) and \( Y \) are independent then 
\[
P(X = x_i \land Y = y_j) = P(X = x_i)P(Y = y_j) \]

denote 
\[
p_i := P(X = x_i), \; p_j := P(Y = y_j)
\]

then 
\[
E(XY) = \sum_{i,j} P(X = x_i \land Y = y_j) = \sum_{i,j} p_{i,x_i} \cdot p_{j,y_j} = \sum_{i,j} p_{i,x_i} \cdot p_{j,y_j} = \sum_i \left( p_{i,x_i} \sum_j p_{j,y_j} \right) = \sum_i \left( p_{i,x_i} \cdot E(Y) \right) = E(Y) \cdot \sum_i p_{i,x_i} = E(Y) \cdot E(X)
\]

/1st transformation: by definition of expected value / 2nd: independence / 3rd: regrouping / 4th: summing in 2 steps / 5th: factoring out non-varying terms from the inner sum / 6th: by definition of \( E(Y) \) / 7th. factoring out the multiplier / 8th: by definition of \( E(X) \)

explanation of (f) / equivalence of the formulas: 
\[
\]

– the 1st transformation is, schematically, \((a-b)(c-d)=ac-ad-bc+bd\) (with \(a=X, b=E(X)\) etc),

– the second is \(E(X+Y)=E(X)+E(Y)\), with the sum of four products inside the left-hand parentheses,
Variables (univariate): distribution, density function, distribution function, expected value, variance and covariance, standard deviation (theory)

– the 3rd is the $E(cX) = cE(X)$ transformation applied to the second and third products /with $c = E(X)$ then with $c = E(Y)$ ./

4. The variance and the standard deviation (S.D.)

The variance of an $X$ random variable with finite expected value $E(X)$ is the expected value of the square of its deviation from its expected value, that is:

$$\text{var}(X) = E((X - E(X))^2).$$

(6.13)

Abbr.: $D^2(X)$ ill. var($X$).

The standard deviation is the square root of the variance.

$$D(X) := \sqrt{\text{var}(X)}$$

(6.14)

Abbr.: $D(X)$ or S.D. (Called also standard error, S.E.)

From a users’ standpoint, the standard deviation shows what size the deviations of $X$ from the expected value of $X$ are about, within what radius the values of $X$ disperse around $E(X)$: the S.D. is the medium size of these deviations. (More exactly, the variance is the mean square error of $X$ around its average; the S.D. is the root mean square error of $X$ around the average.)

(Why the average is not used defining 'medium size' deviations: the average of the deviations always equals zero, so it is not very informative. Alternatively, the mean of the absolute values of the deviations could also be considered, but it is more difficult to handle mathematically and does not have those advantageous properties the variance and S.D have (see also square root law). )

Important to remember: the S.D. is a medium size deviation. There exist deviations smaller and deviations larger than it is.

Alternative formula for the variance: the variance is sometimes more easily computed with the formula.

$$\text{var}(X) = E(X^2) - (E(X))^2$$

(6.15)

(the equivalence of the formulae: denote for a while $m$ the expected value of $X$ (that is, $E(X)$ ); then

\[ \text{var}(X) = E((X - m)^2) = E(X^2 - 2mE(X) + m^2) = E(X^2) - 2mE(X) + m^2 = E(X^2) - (E(X))^2 \]

– the 2nd step is the transformation $(a-b)^2 = a^2 - 2ab + b^2$ within the $E(\ldots)$ parentheses,
– at the 3rd step rules $E(X+Y)=E(X)+E(Y)$ and $E(c)=c$ have been applied,
– at the 4th and 6th steps $m=E(X)$ is used.)

Some properties of the variance and the standard deviation:

\[ \text{var}(x)(c)=0 \quad \text{D}(c)=0 \]
\[ \text{var}(x)(c+X)=\text{var}(x)(X) \quad \text{D}(c+X)=\text{D}(X) \]
\[ \text{var}(x)(cX)=c^2 \cdot \text{var}(x)(X) \quad \text{D}(cX)=c \cdot \text{D}(X) \]

and, if $X$ and $Y$ are independent random variables with finite variances $\text{var}(X)$ and $\text{var}(Y)$, then

\[ \text{var}(x)(X+Y) = \text{var}(x)(X) + \text{var}(x)(Y) \quad \text{D}(X+Y) = \sqrt{\text{D}(X) + \text{D}(Y)}. \]

In general, that is, if the independence of $X$ and $Y$ is not known, the variance of the sum is
5. Expected values and standard deviations of sample sums and sample means (sampling with replacement): the Square Root Law

5.1. Two auxiliary statements:

(1) Let \(X_1, X_2, \ldots, X_n\) be independent random variables with finite expected values \(m_1, m_2, \ldots, m_n\) and finite standard deviations \(d_1, d_2, \ldots, d_n\). Denote \(X\) the sum of these variables \((\sum_{i=1}^{n} X_i)\). Then

\[
E(X) = \sum_{i=1}^{n} m_i, 
\text{var}(X) = \sum_{i=1}^{n} d_i^2 \\
D(X) = \sqrt{n} \cdot d
\]

(2) Consider an experiment: \(n\) draws from a box with numbers of average \(m\) and standard deviation \(d\) (either with or without replacement). Denote \(X\) the \(i\)th draw. Then

\[
E(X_i) = m, 
D(X_i) = d
\]

(The distribution of the random variable of the \(i\)th draw [a probability distribution] and the distribution of the numbers in the box [in the sense distributions are spoken of in descriptive statistics] are the same distributions.)

5.2. The Square Root Law:

Numbered cards in a box, with average \(m\) and standard deviation \(d\); \(n\) draws from the box, with replacement. Denote \(X\) the sum of the draws (that is, the sample sum). Denote \(Y\) the average of the draws (the sample mean). Then

\[
E(X) = n \cdot m, 
\text{var}(X) = n \cdot d^2 \\
D(X) = \sqrt{n} \cdot d
\]

furthermore

\[
E(Y) = m, 
\text{var}(Y) = \frac{d^2}{n}, 
D(Y) = \frac{d}{\sqrt{n}}
\]

(proof: \(X = \sum_{i=1}^{n} X_i\), where the random variables \(X_i\) denoting the \(i\)th draw are independent with expected values \(m\) and variances \(d_i^2\). (The distribution of every \(X_i\) is identical to the distribution of the box; the E.V. of \(X\) is equal to the average of the box; the variance of \(X\) is equal to the variance of the box.) Hence the statements concerning the sample sum (upper row). For the statements concerning the sample mean, \(Y=X/n\) is also used.)

Drawing a sample of \(n\) from a population with average \(m\) and standard deviation \(d\), with replacement, the random variable of the sample sum disperses around its expected value \(nm\) in a radius of \(\sqrt{n} \cdot d\), while the sample mean disperses around its expected value \(m\) in a radius of \(d/\sqrt{n}\).

Example 6.1.
the population average of the Troll population's income is to be measured in a sample survey (simple random sampling, size=100). Assume the average to be 1000 marks, the standard deviation to be 2000 marks in the population. Thus the sample mean, to be seen by the researchers, would be somewhere around 1000 marks, but its approximate error – that is, the standard error – would be at about 200 marks, very large. The sample size of 100 seems insufficient for a useful estimation.

Similarly,

- the medium error of the sample mean from samples of size 400 would be 100 marks
- from samples of size 1600 the medium error of the sample means would be 50 marks, and even
- from samples of size 10,000 the medium error of the sample means would still be around 20 marks.

The same equalities can be applied in estimating the rate of some feature in the population with a sample survey, eg. the rate of the left-handed. The sample statistics of interest would then be the number of left-handed people in the sample (frequency) and the rate of left-handed people in the sample (relative frequency). These can be considered as the sum and the mean of the sample, respectively, modelling the population of right-handed and left-handed this way:

- assume that the population consists of 3 million left-handed and 7 million right-handed people; sampling 100 from among them can be modelled by

(1) drawing 100 cards with replacement from a box of 10 million cards, of which 3 million cards have an "L" and 7 million have an "R" written on them – and then counting those with an "L" among the draws. But if

(2) a "1" is written on the backs of cards with the letter "L" and a "0" is written on the back of cards with the letter "R", then the count of cards with the letter "L" is equal to the sum of the draws.

It is because of this that the following statement is of any interest:

the standard deviation of [0/1] boxes: if a population consists of only 0s and 1s (p being the rate of the 1s and 1-p being the rate of 0s) then the standard deviation of this population equals \( \sqrt{p(1-p)} \). (prove it)

**Example 6.2.**

assume the population rate of some feature (eg. yellow hair) to be 20%. Researchers try to estimate this rate from a sample survey of size=100. What will the error of their estimate be, approximately?

(a) number of yellow-haired people in the sample (frequency). The researchers' experiment (sampling from the population and counting the yellow-haired) can be modelled as 100 draws from a [0/1] box with 20% ones and 80% zeros in the box, with replacement and then computing the sum of the draws. So the variable of interest is the sample sum. The E.V. of the variable is 20. For the S.E. of the variable the S.D. of the box is needed. That is \( d = \sqrt{0.20 \cdot 0.80} = 0.40 \). Therefore the standard error of the sample frequency is \( S.E.\text{sample} = d \cdot \sqrt{n} = 0.40 \cdot 10 = 4 \) the researchers' estimate will be around 20%, with an error of about 4%.

(b) rate of yellow-haired in the sample (relative frequency) : the researchers's experiment (sampling from the population and computing the rate of the yellow-haired in the sample) can be modelled as 100 draws from a [0/1] box with 20% ones and 80% zeros in the box, with replacement and then computing the mean of the draws. So the variable of interest is the sample mean. The E.V. of the sample mean is 0.20. For the S.E. of the sample mean the S.D. of the box is needed. That is \( d = \sqrt{0.20 \cdot 0.80} = 0.40 \). Therefore the standard error for the sample mean is \( S.E.\text{mean} = d \cdot \sqrt{n} = 0.40 \cdot \sqrt{10} = 0.40 \cdot 3.16 = 1.26 \), the estimate will be around 20%, with an error of about 4%.

**6. An upper bound for the proportion or the probability of large deviations: the Chebyshev-inequality**
Variables (univariate): distribution, density function, distribution function, expected value, variance and covariance, standard deviation (theory)

a. Chebyshev's inequality for random variables: the probability that a random variable deviates from its expected value with more than \( k \) times its standard deviation is less than \( 1/k^2 \).

b. Chebyshev's inequality for numeric populations: in a numeric population, the proportion of numbers deviating from the average of the population with more than \( k \) times its standard deviation is less than \( 1/k^2 \).

proof:

\[
D^2(X) = \sum_{|s_i|=k \cdot D} p_i \cdot d_i^2 = \sum_{|s_i|=k \cdot D} p_i \cdot d_i^2 + \sum_{|s_i|=k \cdot D} p_i \cdot d_i^2 \\
= \sum_{|s_i|=k \cdot D} p_i \cdot 0 + \sum_{|s_i|=k \cdot D} p_i \cdot (k \cdot D)^2 = 0 + P(|d| > k \cdot D(X)) \cdot (k \cdot D)^2
\]

- dividing by \((kD)^2\), we get

\[
\frac{1}{k^2} > P(|d| > k \cdot D)
\]

Example 6.3.

the average of the monthly income of the population in Trolland is 100 marks, the S.D. is 200 marks.

a. at most what percent of the population may have incomes above 900 marks?

Solution: the Chebyshev for numeric populations can be applied: what percent in a population can be further than 4 S.D.s (900–100=800=4*200) from the average, at most. The answer is: not more than \( 1/4^2 \) of the population – that is, not more than 6.25% of the population.

b. at most what percent can have incomes over 2100 marks?

Solution: the distance of 2100 from the average is 10 times the S.D. – at most 1/10^2 of the population can be this far from the average – at most 1% of the population can be this rich. Similarly, at most 1/900 of the population can have incomes over 6100 marks (distance=6000 marks=30 S.Ds, proportion is 1/302 at most), etc.

Remark: the Chebyshev inequalities, assuming nothing concerning the shape of the distribution, may be applied almost always, e.g. in cases when the normality of the distribution cannot be assumed. But therefore they cannot tell us what the probabilities of large deviations are. Only an upper bound for these probabilities is given: a number the probability is sure to be less than.

7. Standard errors for the sample sum and the sample mean, drawing without replacement: the "correction factor" (finite correction)

We have got formulas for the SE of the sample sum, the sample mean, the sample frequency and the sample relative frequency, drawing with replacement. Drawing without replacement

\[
S.E. \text{ without replacement} = K \cdot S.E. \text{ with replacement}
\]

where

\[
K = \sqrt{\frac{N-n}{N-1}}
\]

(N denotes the size of the population, \( n \) denotes the size of the sample.)

K is called "correction factor", the operation\(^1\) is called "finite correction"

\(^{\dagger}\)a numeric population is like a set of numbers with the difference that the population may contain a number in two or more copies while a set either contains the number (that is, in one copy) or not contains it.

\(^{\ddagger}\)but still with simple random sampling: all members of the population in one box, \( n \) draws from the box, now without replacement.
Example 6.4.

researchers, having selected a simple random sample of size \( n=1,000 \) from a population of size \( N=8,000,000 \), have got a sample average of 120,000 Ft and 4750 Ft for the SE of the average, but the S.E. has been computed with replacement. What is the S.E., taking into account that the drawing was without replacement?

Solution: the S.E. with replacement has to be multiplied by the correction factor, so

\[
S.E_{\text{without replacement}} = K \cdot S.E_{\text{with replacement}} = \sqrt{\frac{7,999,999}{7,999,999}} \cdot 4750 \approx 4749.7 \text{Ft}
\]

Exercises:

– the same, but with a population size of 800 and a sample size of 100.
– the same, but with a population size of 200 and a sample size of 100.

Comment briefly (when is the finite correction essential?)

Readings


Chapter 7. Expected values, standard errors, variances and covariances (exercises)

A:

1) You can choose: – in game A you win 12 dollars with a chance of 50% or win nothing with a chance of 50%;
– in game B you win 18 dollars with a chance of 30% or win nothing with a chance of 70%.

Which is better?

2) You can choose: – in game A you win 12 dollars with a chance of 50% or win nothing with a chance of 50%;
– in game B you win 20 dollars with a chance of 30% or win nothing with a chance of 70%.

Which is better?

3) You can choose: – in game A you may win 100 dollars with a chance of 30% – but win only 10 dollars with a chance of 70%;
– in game B you may win 250 dollars with a chance of 10% – but win only 12 dollars with a chance of 90%.

Which is better?

3') You can choose: – in game A you may win 100 dollars with a chance of 30% – but win only 10 dollars with a chance of 70%;
– in game B you may win 250 dollars with a chance of 10% – but win only 15 dollars with a chance of 90%.

Which is better?

4) You are at a decision point; decision A is thought to result in a gain of 100,000 dollars (with 10% probability); or it may end in a gain of only 5,000 dollars (with 70% probability) – and there is a 20% chance that it will end in a 40,000 dollars loss.

With decision B there is a 25% chance for a gain of 20,000 dollars; 50% chance for a gain of 1,200 dollars; and 25% chance that your gain will be zero (but no loss either).

Which is better?

5) In a game your chance for winning nothing equals 90% – and you have a 10% chance for winning 10 dollars.
   a) what is your expected gain in this game?
   b) what is the S.D. of the gains in this game?

   (Let the experiment be one such game, X denoting your net gain; (b) then asks for the S.E. of X.)

6) In a game your chance for winning nothing equals 99% – and you have a 1% chance for winning 100 dollars.
   a) what is your expected gain in this game?
   b) what is the S.D. of the gains in this game?

7) In a game your chance for winning nothing equals 50% – but you have a 50% chance for winning 2 dollars.
   a) what is your expected gain in this game?
   b) what is the S.D. of the gains in this game?
8) Which is the best of the games in exercises 5–7?

In what sense is it the best?

9) Which of the games in exercises 5–7 would you recommend

– to somebody who is anxious, avoiding risks?

– to somebody who likes risks?

10) The values of a variable $X$ are all positive. Can the S.D. (that shows the medium deviations from the average) be bigger than the average?

11) Ten million cards representing the heights of a ten millions population in a box (cards with numbers, one for each inhabitant of Midgetland with his height in centimeters on it). The population average in Midgetland is 120 cm, the S.D. of the heights in the population is 10 cm.

Experiment: 100 draws from the box, with replacement. Denote $X$ the sum of the draws.

a) what is the average of the box? what is the S.D. of the box?

b) $E(X) =$?

c) $\text{var}(X) =$?

d) $\text{S.E.}(X) =$?

11') Ten million cards representing the heights of a ten millions population in a box (cards with numbers, one for each inhabitant of Midgetland with his height in centimeters on it). The population average in Midgetland is 120 cm, the S.D. of the heights in the population is 10 cm.

Experiment: 100 draws from the box, with replacement. Denote $Y$ the mean of the draws.

a) what is the average of the box? what is the S.D. of the box?

b) $E(Y) =$?

c) $\text{var}(Y) =$?

d) $\text{S.E.}(Y) =$?

12) Box with numbers of average $m$ and standard deviation $d$. Experiment: $n$ draws from the box, with replacement. Denote $X$ the sum of the draws.

a) find the expected value of the random variable $X$;

b) find the variance of the random variable $X$;

c) find the standard error of the random variable $X$.

Denote $Y$ the mean of the draws.

a) find the expected value of the random variable $Y$;

b) find the variance of the random variable $Y$;

c) find the standard error of the random variable $Y$.

B:

1) Seven 1s, two 2s and one 9 in a box. Two draws are made, with replacement. Denote $X_1$ the first draw, denote $X_2$ the second draw.

a) find the expected value of $X_1$. 
b) find the expected value of $X_1$.

c) find the expected value of the sum of the draws.

d) find the expected value of the product of the draws.

e) $D^2(X_1)=?; \ D^2(X_2)=?$

f) $\text{cov}(X_1, X_2)=?$

g) $D^2(X_1+X_2)=?$

h) $D(X_1)=?; \ D(X_2)=?$

i) $D(X_1+X_2)=?$

2) Seven 1s, two 2s and one 9 in a box. Two draws are made, without replacement. Denote $X_1$ the first draw, denote $X_2$ the second draw.

a) find the expected value of $X_1$.

b) find the expected value of $X_2$.

c) find the expected value of the sum of the draws.

d) find the expected value of the product of the draws.

e) $D^2(X_1)=?; \ D^2(X_2)=?$

f) $\text{cov}(X_1, X_2)=?$

g) $D^2(X_1+X_2)=?$

h) $D(X_1)=?; \ D(X_2)=?$

i) $D(X_1+X_2)=?$

* * * *

3) The number 1 is glued onto two sides of an otherwise fair dice, the number 2 is glued onto two sides and the number 3 is glued onto two sides so that the identical numbers be opposite each other. (When a 1 shows, a 1 is below, etc.) Two rolls with this dice. Denote $X_1$ the first roll, denote $X_2$ the second roll.

a) find the expected value of $X_1$.

b) find the expected value of $X_2$.

c) find the expected value of the sum of the draws.

d) find the expected value of the product of the draws.

e) find the covariance of $X_1$ and $X_2$.

f) find the variances of $X_1$ and of $X_2$.

g) find the variance of the sum of the two rolls.

4) The number 1 is glued onto two sides of an otherwise fair dice, the number 2 is glued onto two sides and the number 3 is glued onto two sides such that the identical numbers be opposite each other. (When a 1 shows, a 1 is below, etc.) One roll with this dice. Denote $X_1$ the number shown, denote $X_2$ the number below.

a) find the expected value of $X_1$.

b) find the expected value of $X_2$. 
c) find the expected value of the sum $X_1 + X_2$.

d) find the expected value of the product $X_1 \times X_2$.

e) find the covariance of $X_1$ and $X_2$.

f) find the variances of $X_1$ and of $X_2$.

g) find the variance of the sum of the two rolls.

5) The number 1 is glued onto two sides of an otherwise fair dice, the number 2 is glued onto two sides and the number 3 is glued onto two sides so that '1's are opposite '3's and '2's are opposite '2's. (When a 1 shows, a 3 is below, etc.) Two rolls with this dice. Denote $X_1$ the first roll, denote $X_2$ the second roll.

a) find the expected value of $X_1$.

b) find the expected value of $X_2$.

c) find the expected value of the sum of the draws.

d) find the expected value of the product of the draws.

e) find the covariance of $X_1$ and $X_2$.

f) find the variances of $X_1$ and of $X_2$.

g) find the variance of the sum of the two rolls.

6) The number 1 is glued onto two sides of an otherwise fair dice, the number 2 is glued onto two sides and the number 3 is glued onto two sides so that '1's be opposite '3's and '2's are opposite '2's. (When a 1 shows, a 3 is below, etc.) One roll with this dice. Denote $X_1$ the number shown, denote $X_2$ the number below.

a) find the expected value of $X_1$.

b) find the expected value of $X_2$.

c) find the expected value of the sum $X_1 + X_2$.

d) find the expected value of the product $X_1 \times X_2$.

e) find the covariance of $X_1$ and $X_2$.

f) find the variances of $X_1$ and of $X_2$.

g) find the variance of the sum of the two rolls.

* * * *

7) Two draws, without replacement, from a box with 10 balls (3 of gold and 7 of iron). Denote $Y_1$ the number of golden balls among the draws. $E(Y_1)$=?

7') Three draws, without replacement, from a box with 10 balls (3 of gold and 7 of iron). Denote $Y_3$ the number of golden balls among the draws. $E(Y_3)$=?

8) Two draws, with replacement, from a box with 10 balls (3 of gold and 7 of iron). Denote $Y_1$ the number of golden balls among the draws. $E(Y_1)$=?

8') Three draws, with replacement, from a box with 10 balls (3 of gold and 7 of iron). Denote $Y_3$ the number of golden balls among the draws. $E(Y_3)$=?

C:

1) The variable $X$ is uniformly distributed over the interval $[0; \pi / 2]$.
/a) sketch the graph of the density function/

b) find the expected value of the random variable X;

c) find the expected value of the random variable sin X

d) find the expected value of the random variable $X^2$

e) find the expected value of the random variable $\sin^2 X$

2) The values of the random variable X are from the interval $[0; \pi/2]$. The density function $f_X(x) = \frac{8x}{\pi^2}$ here. (Elsewhere it equals zero.)

/a) sketch the graph of the density function/

b) find the expected value of the random variable X;

c) find the expected value of the random variable sin X

d) find the expected value of the random variable $X^2$

e) find the expected value of the random variable $\sin^2 X$

3) The values of the random variable X are from the interval $[0; \pi/2]$. Its density function is $f_X(x) = \frac{4}{\pi} - \frac{8x}{\pi^2}$ on this interval. Elsewhere it equals zero.

/a) sketch the graph of the density function/

b) find the expected value of the random variable X;

c) find the expected value of the random variable sin X

d) find the expected value of the random variable $X^2$

e) find the expected value of the random variable $\sin^2 X$

Compare the results of exercises 1-3 (points (b) among themselves, points (c) among themselves etc.). Explain the differences.

**Readings**


Chapter 8. The square root law; measurement errors (exercises)

1) Blister-packed oranges in a store are said to be of weight 1000 grams.

Assume the average weight in the population of these packs to be 1000 grams and the standard deviation of the weights to be 50 grams.

a) taking 25 packs, the total weight is going to be around ______ kgs, give or take ______ or so.

b) taking 100 packs, the total weight is going to be around ______ kgs, give or take ______ or so.

a') taking 25 packs, the mean weight of the packs is going to be around ______ kgs, give or take ______ grams or so.

b') taking 100 packs, the mean weight of the packs is going to be around ______ kgs, give or take ______ grams or so.

(The packs may be assumed to be selected with simple random sampling from the total population of orange packs.)

2) There are numbered cards in a box; the average of the numbers is 200, their standard deviation is 80. Draws are made from this box by the players, one card each, with replacement. One hundred draws are observed.

a) the sum of the draws will be around _____ give or take _____ or so.

b) the mean of the draws will be around _____ give or take _____ or so.

(Complete the sentences.)

2') Numbered cards in a box; the average of the numbers is 200, their standard deviation is 80. Draws are made from this box by the players, one card each, with replacement. Four hundred draws are observed.

a) the sum of the draws will be around _____ give or take _____ or so.

b) the mean of the draws will be around _____ give or take _____ or so.

(Complete the sentences.)

3) The rolls of a bakery weigh in the average 55 grams (with a standard deviation of 13 grams). Ten rolls are selected at random;

a) the total weight of the ten rolls is going to be

   (i) exactly 550 grams  (ii) around 550 grams  (choose one of the options)

b) the expected value of the total weight of the ten rolls is going to be

   (i) exactly 550 grams  (ii) around 550 grams  (choose one of the options)

c) the total weight of the ten rolls is going to be around _____ grams, give or take ______ grams or so.

(Complete the sentence.)
4) Agents of the Troll Consumer Protection Agency are checking the weights of the rolls in a store. We know – but the Agency do not know – that the average weight of these rolls is 55 grams (the S.D. of the weights being 13 grams). Ten rolls are going to be selected at random by the agents;

a) the mean weight of the 10 rolls is going to be
   (i) exactly 55 grams  (ii) around 55 grams  (choose one of the options)

b) the expected value of the mean weight of the 10 rolls is going to be
   (i) exactly 55 grams  (ii) around 55 grams  (choose one of the options)

c) the mean weight of the 10 rolls is going to be around ______ grams, give or take ______ grams or so.
   (Complete the sentence.)

5) Researchers are going to measure the average height of 18 year old boys in county X from a sample so the present data might be compared to data from 5 years earlier. According to the researchers' hypothesis the average difference might only be a few millimetres (those born later being higher), so the researchers think a measurement with a standard error of only 1 millimetre is needed.

a) the standard error of what is to be 1 millimetre? Choose one of the options below:
   the sample / the sample sum / the sample mean / the population mean / the population sum / the population

Assume the standard deviation of heights in the 18-year-old population to be 20 centimetres.

b) a sample of size 25 being selected – what size the standard error in the sample mean is going to be?

c) a sample of size 100 being selected – what size the standard error in the sample mean is going to be?

d) find the necessary sample size for the measurement error (the S.E. of the sample mean) to be 1 centimetres.

e) find the necessary sample size for the measurement error (the S.E. of the sample mean) to be 1 millimetres.

6) The population average of monthly incomes of those working is to be measured by a sample survey. What size the measurement error will be

a) using a sample of size 100?

b) using a sample of size 400?

(Assume the population average of monthly incomes of those working to be 130,000 HUF, the S.D. of these incomes to be 100,000 HUF.

– with a sample of size 100 the sample mean is going to be around ___ HUF give or take ___ or so.

– with a sample of size 400 the sample mean is going to be around ___ HUF give or take ___ or so.)

c) find the necessary sample size for the measurement error to be only 1000 HUF.

d) find the necessary sample size for the measurement error to be only 100 HUF.

7) There are numbered cards in a box, zeroes and ones. Find the standard deviation of the box (that is, of the numbers in the box) assuming that

a) there are two cards in the box, one 1 and one 0;

b) there are five cards in the box, one 1 and four 0s;

c) there are five cards in the box, four 1 and one 0s;

d) there are ten cards in the box, one 1 and nine 0s;
The square root law; measurement errors (exercises)

e) there are ten cards in the box, nine 1 and one 0s.

8) Numbered cards in a box, zeroes and ones. Find the standard deviation of the box assuming that
a) there are twenty cards in the box, ten 1s and ten 0s;
b) there are twenty cards in the box, four 1s and sixteen 0s;
c) there are twenty cards in the box, sixteen 1s and four 0s;
d) there are twenty cards in the box, two 1s and eighteen 0s;
e) there are twenty cards in the box, eighteen 1s and two 0s.

9) Numbered cards in a box, zeroes and ones. Find the standard deviation of the box assuming that
a) the proportion of ones in the box is p=0.5;
b) the proportion of ones in the box is p=0.2;
c) the proportion of ones in the box is p=0.8;
d) the proportion of ones in the box is p=0.1;
e) the proportion of ones in the box is p=0.9;
– or is more information needed?

10) Numbered cards in a box, zeroes and ones (a [0/1] box). 100 draws from the box, with replacement. The
variable is the sum of the draws. Find the standard error of this variable, assuming that
a) the proportion of ones in the box is p=0.5;
b) the proportion of ones in the box is p=0.2;
c) the proportion of ones in the box is p=0.8;
d) the proportion of ones in the box is p=0.1;
e) the proportion of ones in the box is p=0.9.
The sum of the draws will be around _______ give or take _______ or so.

Fill in the blanks for the boxes in a), b), ...e).

11) The proportion of the population in favour of the re-establishment of slavery is to be measured in a number
of cities with sample surveys.

a) Assume the population proportion of supporters to be 50% in the city of A. So, in A,
a1) the proportion of supporters in a sample of 100 would be around ____%, give or take ____% or so;
a2) the proportion of supporters in a sample of 400 would be around ____%, give or take ____% or so;
a3) the proportion of supporters in a sample of 900 would be around ____%, give or take ____% or so;
a4) the proportion of supporters in a sample of 1600 would be around ____%, give or take ____% or so.
b) Assume the population proportion of supporters to be only 40% in the city of B. So, in B,
b1) the proportion of supporters in a sample of 100 would be around ____%, give or take ____% or so;
b2) the proportion of supporters in a sample of 400 would be around ____%, give or take ____% or so;
b3) the proportion of supporters in a sample of 900 would be around ____%, give or take ____% or so;
b4) the proportion of supporters in a sample of 1600 would be around ____%, give or take ____% or so.
c) Further assume the population proportion of supporters to be only a low 20% in the city of C. So, in C,
c1) the proportion of supporters in a sample of 100 would be around ____%, give or take ____% or so;
c2) the proportion of supporters in a sample of 400 would be around ____%, give or take ____% or so;
c3) the proportion of supporters in a sample of 900 would be around ____%, give or take ____% or so;
c4) the proportion of supporters in a sample of 1600 would be around ____%, give or take ____% or so.
d) Now assume the population proportion of supporters to be 95% in the city of D. Thus, in D,
d1) the proportion of supporters in a sample of 100 would be around ____%, give or take ____% or so;
d2) the proportion of supporters in a sample of 400 would be around ____%, give or take ____% or so;
d3) the proportion of supporters in a sample of 900 would be around ____%, give or take ____% or so;
d4) the proportion of supporters in a sample of 1600 would be around ____%, give or take ____% or so;
d5) the proportion of supporters in a sample of 10,000 would be around ____%, give or take ____% or so.
e) assume the population proportion of supporters to be the low 5% in the city of E. Thus, in E,
e1) the proportion of supporters in a sample of 100 would be around ____%, give or take ____% or so;
e2) the proportion of supporters in a sample of 400 would be around ____%, give or take ____% or so;
e3) the proportion of supporters in a sample of 900 would be around ____%, give or take ____% or so;
e4) the proportion of supporters in a sample of 1600 would be around ____%, give or take ____% or so;
e5) the proportion of supporters in a sample of 10,000 would be around ____%, give or take ____% or so.
f) assume the population proportion of supporters to be 80% in the city of F. Thus, in F,
f1) the proportion of supporters in a sample of 100 would be around ____%, give or take ____% or so;
f2) the proportion of supporters in a sample of 400 would be around ____%, give or take ____% or so;
f3) the proportion of supporters in a sample of 900 would be around ____%, give or take ____% or so;
f4) the proportion of supporters in a sample of 1600 would be around ____%, give or take ____% or so;
g) describe the dependence of the standard error from the pair \( \{ p; (1-p) \} \) (p being the proportion of ones, 
\((1-p)\) being that of zeroes in the box.)

12) 200 metal balls in a box, some golden, the others black. The proportion of golden balls is 20%. One hundred draws are observed (drawing with replacement). Of the 100 draws the number of golds

a) will be 20 – but only approximately 20 because a difference of about ____ is quite probable.
b) will be exactly 20.

Which is correct? Complete, if a).

13) 100 metal balls in a box, some golden, the others black. (Players draw with replacement.) The Master of Ceremonies tells that twenty of the hundred balls are golden. An angry player says no more than ten might be golden of the balls.
a) if the Master of Ceremonies is right, of 25 draws to be observed there will be about ____ golden, give or take ____ or so.

b) if there is only 10 golden balls among the 100, of 25 draws to be observed there will be about ____ golden, give or take ____ or so.

What is your opinion: is 25 draws enough to settle the matter?

14) A gambling machine states that "about every tenth player gets the big gain". It means, as the small print explains, that in each game the player has a 10% chance to win the "big gain".

If the machine works as stated then of 100 games there will be about _____ games when the player gets the big gain, get or take ____ or so. (Fill the blanks.)

15) The population proportion of Brown Party (B.P.) supporters must be around 15-25 percent in Trolland. Researchers want to measure this proportion with a precision such that the standard error of the sample percentage be not more than 1%. Find the necessary sample size.

Leading questions:

a) find the necessary sample size for the standard error of the sample percentage to be exactly 1%, assuming the proportion of Brown Party supporters to be 15% in the population;

b) find the necessary sample size for the standard error of the sample percentage to be exactly 1%, assuming the proportion of Brown Party supporters to be 25% in the population.

Questions leading to the leading questions:

a) assuming the proportion of Brown Party supporters to be 15% in the population,

a1) the proportion of B.P. supporters in a sample of 100 would be around ____%, give or take ____% or so;

a2) the proportion of B.P. supporters in a sample of 400 would be around ____%, give or take ____% or so;

a3) the proportion of B.P. supporters in a sample of 900 would be around ____%, give or take ____% or so;

a4) the proportion of B.P. supporters in a sample of 1600 would be around ____%, give or take ____% or so.

b) assuming the proportion of Brown Party supporters to be 25% in the population,

a1) the proportion of B.P. supporters in a sample of 100 would be around ____%, give or take ____% or so;

a2) the proportion of B.P. supporters in a sample of 400 would be around ____%, give or take ____% or so;

a3) the proportion of B.P. supporters in a sample of 900 would be around ____%, give or take ____% or so;

a4) the proportion of B.P. supporters in a sample of 1600 would be around ____%, give or take ____% or so.

16) There is a new spread of t.b. present in Trolland. Researchers want to establish the proportion of the infection in the population fro from a sample research. (Assume that the people selected in the sample are accessible, willing to cooperate, and that the screening process is working error-free.) Last time the proportion of infected people proved to be 5%. The researchers think they need a measurement such that its' standard error be not more than 0.5%.

a) find the necessary sample size.

b) find the necessary sample size for a measurement with a standard error of only 0.1%. (A bigger or a smaller sample is needed? how much [times] bigger? how much [times] smaller?)

Sample in the above exercises always means simple random sample, drawing with replacement. It is like putting chips in a big box (for each member of the population, one chip) and then making n draws (n denoting the

\[ n \] meaning the sample percentage
The square root law; measurement errors (exercises)

sample size) successively and with replacement. (Hence it is not impossible for a member of the population to be selected in the sample twice.)

Readings

Chapter 9. Random variables: distribution, cumulative distribution function, density function, expected value (exercises)

1) Two draws, with replacement, from the box [1, 1, 2, 3];
   a) find the distribution for the sum of the draws (its possible values and the chances of each, in a distribution-table)
   b) find the distribution for the product of the draws.

2) The same, without replacement.

3) Three draws, with replacement, from the box [1, 5],
   a) find the distribution for the sum of the draws;
   b) find the distribution for the product of the draws.

   Compare the shapes of the distributions. What do you see?

4) One roll with a fair die. Let X= the value of the roll. Find the cumulative distribution function of X and sketch the graph of it.

5) The density function of variable X equals 1/10 on the interval [5, 15], and 0 elsewhere. Find the values of these probabilities
   a) P(X<10)       b) P(X<5)       c) P(X<8 | X<10)       d) P(X<10 | X>8)

   e) find the median value of X;
   f) find the expected value of X;
   g) find the first quartile of X.

6) The density function of variable X equals 1/10 on the interval [5, 15], and 0 elsewhere. Find the cumulative distribution function of X. Make a graph of it.

7) The density function of variable X

\[ f_X(x) = \begin{cases} 
0.5 x & \text{if } x \in [0; 1] \\
0.5 & \text{if } x \in [1; 2] \\
-0.5x + 1.5 & \text{if } x \in [2; 3] 
\end{cases} \]

   and equals 0 elsewhere.

   Find the cumulative distribution function of X. Make a graph of it.

8) Two tosses with a fair coin. Let X= the number of heads. Find the cumulative distribution function of X and sketch the graph of it.

9) The density function of variable X

1 that is, four cards, numbered 1, 1, 2 and 3, in a box
Random variables: distribution, cumulative distribution function, density function, expected value (exercises)

\[ f(x) = \begin{cases} 
0.04x & \text{on interval } [0, 5] \\
-0.04x + 0.4 & \text{on interval } [5, 10] \\
0 & \text{elsewhere.} 
\end{cases} \]

a. Check if it is a density function.

Find the values of these probabilities:

b. \( P(X<10) \)

c. \( P(X<5) \)

d. \( P(X<3) \)

e. \( P(X<3 \mid X<5) \)

f. \( P(X<5 \mid X>3) \)

g. find the median value of \( X \);

h. find the expected value of \( X \);

i. find the first quartile of \( X \).

j. Find the cumulative distribution function of \( X \) and sketch the graph of it.

10) On interval \([5, 10]\) the density function of \( X \) equals \( 1/10 \), on \([15, 20]\) it also equals \( 1/10 \), elsewhere it equals 0.

a. Find the cumulative distribution function of \( X \) and sketch the graph of it.

b. \( P(X<8) = ? \)

c. \( P(8<X<16) = ? \)

11) On interval \([5, 10]\) the density function of \( X \) equals \( 1/10 \), on \([15, 25]\) it equals \( 1/20 \), elsewhere it equals 0.

a. Find the cumulative distribution function of \( X \) and sketch the graph of it.

b. \( P(X<8) = ? \)

c. \( P(8<X<16) = ? \)

12) First a coin is tossed, then,

- if the toss is heads, a dice is rolled and \( X \) equals the value of the roll;
- if the toss is tails, a number is taken from a device providing numbers distributed \( U[1;6] \) and the value of \( X \) equals this number. Find the cumulative distribution function of \( X \).

13) Defining the cumulative distribution function with the formula \( F(x) := P(X<x) \) implies that

\[ \lim_{x \to -\infty} F(x) = 0 \quad \lim_{x \to \infty} F(x) = 1 \]

F is nondecreasing and F is continuous on the left.

\textit{that is, a continuous distribution whose density function } = 0.20 \textit{ on } [1;6], \textit{ and } = 0 \textit{ elsewhere.}
Random variables: distribution, cumulative distribution function, density function, expected value 
(exercises)

It might be defined differently for example as \( F_X(x) := P(X \leq x) \).

(Show that it yields the same information about the distribution as the original definition \( F_X(x) := P(X < x) \). Accordingly, this definition is also widely used.)

What changes in statements (a)–(d) would changing the definition of \( F_X(x) \) this way imply?

14) \( F_X(x) := P(X > x) \) is another possible definition of the cumulative distribution function.

(Show that it also yields the same information as that defined \( F_X(x) := P(X < x) \).)

Compared to the original definition, what changes in statements (a)–(d) would defining \( F_X(x) \) this way imply?

15) How could you read these probabilities from the cumulative distribution function:

\[
\begin{align*}
a) & \ P(X \leq a) \\
b) & \ P(X \geq a) \\
c) & \ P(X > a) \\
d) & \ P(a < X < b) \\
e) & \ P(a < X \leq b) \\
f) & \ P(a \leq X \leq b)
\end{align*}
\]

16) How could you determine these probabilities from the density function:

\[
\begin{align*}
a) & \ P(a < X < b) \\
b) & \ P(a < X \leq b) \\
c) & \ P(a \leq X \leq b)
\end{align*}
\]

17) We know the density function of \( X; \ Y=2X \). How does the density function of \( Y \) look like?

18) We know the cumulative distribution function of \( X; \ Y=2X \). How does the cumulative distribution function of \( Y \) look like?

19) The density function of variable \( X \)

\[
\begin{align*}
fs(x) &= 0.25 x & \text{on interval } [0, 2] \\
\quad & -0.25 x +1 & \text{on interval } [2, 4] \\
\quad & 0 & \text{elsewhere.}
\end{align*}
\]

Let \( Y := 2X \). Sketch the graphs of both densities \( fs(x) \) and \( fx(x) \) on a common chart.

20) The distribution of \( X \) is exponential with parameter \( \lambda = 5 \), that is, its density function \( f(t) = 0 \) if \( t < 0 \) and \( f(t) = \lambda \cdot e^{-\lambda t} \) if \( t \geq 0 \) (so here \( f(t) = 5 \cdot e^{-5t} \) if \( t \geq 0 \)).

a. Check whether it is indeed a density function.

b. Find the probabilities \( P(X > 10) \) and \( P(X > 110 \mid X > 100) \). What do you see?

c. Show that for any \( t_0 > 0 \) and \( t_1 > 0 \) the following memoryless property of the exponential distribution holds:

\[
P(X > t_0 + t_1 \mid (X > t_1))
\]

21) A fair coin is tossed twice. Let \( X \) = the number of heads. Find the expected value of \( X \), that is, \( E(X) \).

22) A fair coin is tossed three times. Let \( X \) = the number of heads. Find the expected value of \( X \).

23) Two draws are made, with replacement, from the box \( [0, 1] \). Let \( X \) = the sum of the draws. Find \( E(X) \).

24) Three draws are made, with replacement, from the box \( [0, 1] \). Let \( X \) = the sum of the draws. Find \( E(X) \).

25) Two draws are made, with replacement, from the box \( [0, 1] \). Let \( X \) = the product of the draws. Find \( E(X) \).

26) Three draws are made, with replacement, from the box \( [0, 1] \). Let \( X \) = the product of the draws. Find \( E(X) \).
Random variables: distribution, cumulative distribution function, density function, expected value
(exercises)

27) Two draws, without replacement, from the box \([0, 0, 1, 1]\). Let \(X=\)the sum of the draws. Find \(E(X)\).

28) Three draws, without replacement, from the box \([0, 0, 1, 1]\). Let \(X=\)the sum of the draws. Find \(E(X)\).

29) Two draws, without replacement, from the box \([0, 0, 1, 1]\). Let \(X=\)the product of the draws. Find \(E(X)\).

30) Three draws, without replacement, from the box \([0, 0, 1, 1]\). Let \(X=\)the product of the draws. Find \(E(X)\).

31) Two draws, without replacement, from the box \([0, 0, 1, 1]\). Let \(X_1=\)the first draw, let \(X_2=\)the second draw.

Let \(X=\)the sum of the draws. Find \(E(X_1), E(X_2), E(X)\), and \(\text{cov}(X_1,X_2)\).

32) Two draws, with replacement, from the box \([0, 0, 1, 1]\). Let \(X_1=\)the first draw, let \(X_2=\)the second draw.

Let \(X=\)the sum of the draws. Find \(E(X_1), E(X_2), E(X)\), and \(\text{cov}(X_1,X_2)\).

* * *

33*) (infinity/1) A fair die is rolled until a six shows. Let \(X=\)the number of rolls until the first six. (\(X=1\) if the first roll is already a six.) Find the expected value of \(X\).

34) (infinity/2, the Petersburg Paradox) A gambler follows a simple strategy he says is sure to win: he bets 1 dollar on heads and if he wins he ends the game. If he does not win, he doubles his stake. If he wins now, he ends the game (he has paid 3 dollars and won 4 dollars so his net gain equals 1 dollar.) If he does not win he doubles his stake again ... until he wins. Then he ends the game. (A fair coin is tossed.)

- What is your opinion: is this a good strategy? If you say yes – why have not a lot of people got rich this way?

Some more questions:

a) Find the chance of winning with this strategy. (That is, the chance that sooner or later the player is going to win.)

b) Find the expected value of the net gain of the player.

c) Denote \(Y\) the capital necessary for the winning. Find the expected value of \(Y\). (\(Y\) equals the sum of the bets the player had payed until he won.)

35*) (infinity/2, Petersburg Paradox, modified) The following game is played: the player tosses a coin until the first head shows. His gain depends strongly on the number of tosses needed.

If only 1 toss has been necessary he gains 1 dollar;

if 2 tosses have been necessary he loses 2 dollars;

if 3 tosses have been necessary he gains 4 dollars;

if 4 tosses have been necessary he loses 8 dollars; and so on. (The player does not have to pay in advance any money. He only wins – or loses.)

a) Find the chance that the the game is going to end sooner or later.

b) Find the expected value of the player's, signed, net gain.

(For exercises 33...36, see page 2 of handout (6))

Readings

Chapter 10. Expected value and standard deviation of continuous variables; other exercises (exercises)

1) X is uniformly distributed on [0;1] (denoted U[0;1]) – that is, its density function equals 1 on this interval and 0 elsewhere.

   a) Sketch the graph of the density function.
   b) Sketch the graph of the cumulative distribution function.
   c) Find E(X) (that is, the expected value of X).
   d) Find D(X) (that is, the standard deviation of X).
   e) Find var(X) (the variance).
   f) Find E(X^2)

2) X is uniformly distributed on [5;10] – that is, its density equals 0.20 on this interval and 0 elsewhere.
   a) Find E(X).
   b) Find D(X).
   c) Find var(X).

3) The density of X, \(f_X(x)=1+x\) on [\(-1;0]\); \(f_X(x)=1-x\) on [0;1]; and \(f_X(x)=0\) elsewhere.
   a. check if \(f_X(x)\) is a density function.
   b. E(X)=?
   c. var(X)=?; \(D(X)=?\)

4.) The density of X, \(f_X(x)=x-10\) on [10;11]; \(f_X(x)=12-x\) on [11;12]; and \(f_X(x)=0\) elsewhere
   a. check if \(f_X(x)\) is a density function.
   b. E(X)=?
   c. var(X)=?; \(D(X)=?\)

5.) The density of X, \(f_X(x)=x/18\) on [0, 6] and \(f_X(x)=0\) elsewhere.
   a. check if \(f_X(x)\) is a density function.
   b. E(X)=?

6.) The distribution of X is U[0;5]. Find its expected value and standard deviation.

7.) The density function of X, \(f_X(x)=0.3\) on [2;3]; \(f_X(x)=0.2\) on [1;2] and [3;4]; \(f_X(x)=0.15\) on [0;1] and [4;5]; and \(f_X(x)=0\) elsewhere Sketch the graph of the density function. Find the expected value and the standard deviation of X.
8.) The density function of $X$, $f_X(x)=0.5$ on $[2;3]$; $f_X(x)=0.2$ on $[1;2]$ and $[3;4]$; $f_X(x)=0.05$ on $[0;1]$ and $[4;5]$; and $f_X(x)=0$ elsewhere Sketch the graph of the density function. Find the expected value and the standard deviation of $X$.

9.) The density function of $X$, $f_X(x)=0.70$ a $[2;3]$; $f_X(x)=0.10$ on $[1;2]$ and $[3;4]$; $f_X(x)=0.05$ on $[0;1]$ and $[4;5]$; and $f_X(x)=0$ elsewhere Sketch the graph of the density function. Find the expected value and the standard deviation of $X$.

10.) The density function of $X$, $f_X(x)=0.15$ on $[1;2]$ and $[3;4]$; $f_X(x)=0.3$ on $[0;1]$ and $[4;5]$; and $f_X(x)=0$ elsewhere Sketch the graph of the density function. Find the expected value and the standard deviation of $X$.

11.) The density function of $X$, $f_X(x)=0.35$ a $[0;1]$ and $[4;5]$; and $f_X(x)=0$ elsewhere Sketch the graph of the density function. Find the expected value and the standard deviation of $X$.

Compare the EVs and SDs of exercises 6..11. Explain.

12.) The distribution of $X$ is exponential with parameter $\lambda=5$ (that is, its density equals zero if $t<0$ and $f_X(t)=5e^{-5t}$ if $t\geq0$. Compute its EV and SD.

13.) Let $f(x)=\frac{3}{x^2}$ if $x>1$, $f(x)=0$ elsewhere.
   a. Check if $f(x)$ is a density.
   b. $E(X)=?$
   c. $\text{var}(X)=?; \quad D(X)=?$

14.) Legyen $f(x)=\frac{2}{x^3}$ if $x>1$, $f(x)=0$ elsewhere.
   a. Check if $f(x)$ is a density.
   b. $E(X)=?$
   c. $\text{var}(X)=?; \quad D(X)=?$

15.) Legyen $f(x)=\frac{1}{x^2}$ if $x>1$, $f(x)=0$ elsewhere.
   a. Check if $f(x)$ is a density.
   b. $E(X)=?$
   c. $\text{var}(X)=?; \quad D(X)=?$

(For exercises 13..16, see page 2 and 3 of handout (6))

17.) The numbers $x_1$, $x_2$, ..., $x_n$ are all positive. Is it possible that the SD of these numbers is greater than the mean?

18) Ten cards in a box, eight 1s and two 10s. Two draws, with replacement. Denote $X_1$ the first draw, $X_2$ the second draw; denote $Y=X_1 + X_2$ the sum of the draws.
   a. $\text{var}(X_1)=? \quad D(X_1)=?$
Expected value and standard deviation of continuous variables;
other exercises (exercises)

b. \( \text{var}(X_i) = ? \quad \text{D}(X_i) = ? \)
c. \( \text{var}(Y) = ? \quad \text{D}(Y) = ? \)
d. \( \text{covar}(X_i, X_j) = ? \)

19) Ten cards in a box, eight 1s and two 10s. Two draws, without replacement. Denote \( X_1 \) the first draw, \( X_2 \) the second draw; denote \( Y = X_1 + X_2 \) the sum of the draws.

a. \( \text{var}(X_1) = ? \quad \text{D}(X_1) = ? \)

b. \( \text{var}(X_2) = ? \quad \text{D}(X_2) = ? \)
c. \( \text{var}(Y) = ? \quad \text{D}(Y) = ? \)
d. \( \text{covar}(X_1, X_2) = ? \)

18') Ten cards in a box, eight 1s and two 10s. Two draws, with replacement. Denote \( X_1 \) the first draw, \( X_2 \) the second draw; denote \( Y = X_1 - X_2 \) the difference of the draws.

a. \( \text{var}(Y) = ? \quad \text{D}(Y) = ? \)

b. \( \text{covar}(X_1, X_2) = ? \)

19') Ten cards in a box, eight 1s and two 10s. Two draws, without replacement. Denote \( X_1 \) the first draw, \( X_2 \) the second draw; denote \( Y = X_1 - X_2 \) the difference of the draws.

a. \( \text{var}(Y) = ? \quad \text{D}(Y) = ? \)

b. \( \text{covar}(X_1, X_2) = ? \)

20) Ten cards in a box, nine 1s and one 11. Two draws, without replacement. Denote \( X_1 \) the first draw, \( X_2 \) the second draw; denote \( Y = X_1 + X_2 \) the sum of the draws.

a. \( \text{E}(X_1) = ? \)

b. \( \text{E}(X_2) = ? \)

c. \( \text{var}(X_1) = ? \quad \text{D}(X_1) = ? \)

d. \( \text{var}(X_2) = ? \quad \text{D}(X_2) = ? \)

e. \( \text{var}(Y) = ? \quad \text{D}(Y) = ? \)

Readings


Chapter 11. Roulette/1: expected values, standard deviations (exercises)

1) Alex is going to play roulette 100 times on a European roulette, putting a dollar on the first dozen each time. (In European roulette the wheel has 37 slots, of these 36 numbered 1..36, and a 0.) If one of the numbers 1..12 comes up he wins, getting his dollar back together with winnings of 2 dollars. If he loses he loses his dollar.

a) what is to be expected at the end of the 100 games: is he to win? is he to lose? and about how much?

b) about how much will his actual net gain from the 100 games deviate from the expected value computed in (a)?

(That is, how close are the actual values going to be to the value computed in (a)? If the expected gain is positive, is it almost sure Alex wins somewhat or is the chance of his winning only around 50%? Are the actual net gains from 100 games almost always positive or are they sometimes positive sometimes negative, with the chance of each around 50%?)

1’) Alex is going to play roulette 100 times on a Nevada roulette, putting a dollar on the first dozen each time. (In Nevada roulette the wheel has 38 slots, of these 36 numbered 1..36, plus a 0 and a 00.) If one of the numbers 1..12 comes up he wins, getting his dollar back together with winnings of 2 dollars. If he loses he loses his dollar.

a) what is to be expected at the end of the 100 games: is he to win? is he to lose? and about how much?

b) about how much will his actual net gain from the 100 games deviate from the expected value computed in (a)?

2) We are going to play roulette putting 1 dollar on the first dozen 100 times. Which is it better to play:

a) European roulette?

b) Nevada roulette?

3) We are going to play roulette 1000 times on a European roulette, putting 1 dollar on the first row each time. If one of the numbers 1,2,3 comes up we win, getting our dollar back together with winnings of 11 dollars; otherwise we get nothing.

a) what is to be expected at the end of the 1000 games: are we to win? are we to lose? about how much?

b) about how much will our actual net gain from the 1000 games deviate from the expected value computed in (a)?

3’) We are going to play roulette 1000 times on a Nevada roulette, putting 1 dollar on the first row each time. If one of the numbers 1,2,3 comes up we win, getting our dollar back together with winnings of 11 dollars; otherwise we get nothing.

a) what is to be expected at the end of the 1000 games: are we to win? are we to lose? about how much?

b) about how much will our actual net gain from the 1000 games deviate from the expected value computed in (a)?

4) We are going to play roulette putting 1 dollar on the first row 1000 times. Which is it better to play:

a) European roulette?

b) Nevada roulette?
5) Considering the net gain on the long run, which is better:

a) putting 1 dollar on the first dozen a great number of times?

b) putting 1 dollar on the first row the same number of times?

6) We play Nevada roulette with 1000 dollars.

Which gives a better chance of coming out with a fair (say, at least 500 dollars) gain,

a) putting the 1000 dollars on red at once? (If one of the 18 red numbers comes up, we get our 1000 dollars back together with winnings of 1000 dollars; otherwise we get nothing.)

b) putting 1 dollar on red 1000 times? (Each time one of the 18 red numbers comes up we get our dollar back together with winnings of 1 dollar. If another number comes up we get nothing.)

Readings

Chapter 12. Using the normal table (exercises)

1) The distribution of the variable Z is normal with expected value=0 and standard error=1. Find the chance that Z is in the following sets:

- \{z < 1\}
- \{z < 1.5\}
- \{z < 2\}
- \{z < 3.40\} (-\infty ; 2.40)
- \{z \geq 1\}
- \{z \geq 1.5\}
- \{z \geq 3\} (3.4 ; +\infty)
- (-\infty ; -1)
- \{z < -2\}
- [1; 2]
- [2.5; 3.5]
- [-1; +1]
- [-2; +2]
- [2; +2]
- [-2; +2]
- [1; 2)
- [2; +2)
- [2.5; 3.5)
- [-1; 2)
- [3; 5)
- [2; 6)
- [2; 6]
- [3; 6]

- the chance is 95% that Z is less than _____ (fill in the blanks)
- the chance is 99% that Z is less than _____ (fill in the blanks)
- the chance is 95% that Z is greater than _____ (fill in the blanks)
- the chance is 99% that Z is greater than _____ (fill in the blanks)
- the chance is 95% that Z is within +/- _____ around 0 (fill in the blanks)
- the chance is 99% that Z is within +/- _____ around 0 (fill in the blanks)

2) The distribution of X is normal with expected value=4 and standard error=1. Find the chance that X is in the following sets:

- \{x < 5\}
- \{x < 5.5\}
- \{x < 6\}
- \{x < 7.40\} (-\infty ; 6.40)
- \{x \geq 5\}
- \{x \geq 5.5\}
- \{x \geq 7\} (7.4 ; +\infty)
- (-\infty ; 3)
- \{x < 2\}
- [5; 6)
- [6.5; 7.5)
- [3; 5)
- [2; 6)
- [2; 6]
- [1.5; 6.5]
- [3; 6]

- the chance is 95% that X is less than _____ (fill in the blanks)
- the chance is 99% that X is less than _____ (fill in the blanks)
- the chance is 95% that X is greater than _____ (fill in the blanks)
- the chance is 99% that X is greater than _____ (fill in the blanks)
- the chance is 95% that X is within +/- _____ around 4 (fill in the blanks)
- the chance is 99% that X is within +/- _____ around 4 (fill in the blanks)

3) The distribution of X is normal with expected value=100 and standard error=1. Find the chance that X is in the following sets:

- \{x < 101\}
- \{x < 101.5\}
- \{x < 102\}
- \{x < 103.40\} (-\infty ; 102.40)
Using the normal table (exercises)

\[
\begin{align*}
\{x \geq 101\} & \quad \{x \geq 101.5\} & \quad \{x \geq 103\} & \quad [103.4 ; +\infty) \\
(- \infty; 99) & \quad \{x < 98 \} \\
[101; 102) & \quad [102.5; 103.5) \\
[ 99; 101) & \quad [ 98; 102) & \quad [98; 102] & \quad [-97.5; 102.5] \\
[ 99; 102] & \\
\end{align*}
\]

- the chance is 95\% that \( X \) is less than \( \) (fill in the blanks)
- the chance is 99\% that \( X \) is less than \( \) (fill in the blanks)
- the chance is 95\% that \( X \) is greater than \( \) (fill in the blanks)
- the chance is 99\% that \( X \) is greater than \( \) (fill in the blanks)
- the chance is 95\% that \( X \) is within +/- \( \) around 100 (fill in the blanks)
- the chance is 99\% that \( X \) is within +/- \( \) around 100 (fill in the blanks)

4) The distribution of \( X \) is normal with expected value=0 and standard error=2. Find the chance that \( X \) is in the following sets:

\[
\begin{align*}
\{x < 2\} & \quad \{x < 3\} & \quad \{x < 4\} & \quad \{x < 6.80\} & \quad (- \infty ; 4.80) \\
\{x \geq 2\} & \quad \{x \geq 3.0\} & \quad \{x \geq 6.0\} & \quad [6.80 ; +\infty) \\
(- \infty; -2) & \quad \{x < -4 \} \\
[2; 4) & \quad [5.0; 7.0) \\
[-2; +2) & \quad [-4; +4) & \quad [-4; +4] & \quad [-5; +5] \\
[-2; +4] & \\
\end{align*}
\]

- the chance is 95\% that \( X \) is less than \( \) (fill in the blanks)
- the chance is 99\% that \( X \) is less than \( \) (fill in the blanks)
- the chance is 95\% that \( X \) is greater than \( \) (fill in the blanks)
- the chance is 99\% that \( X \) is greater than \( \) (fill in the blanks)
- the chance is 95\% that \( X \) is within +/- \( \) around 0 (fill in the blanks)
- the chance is 99\% that \( X \) is within +/- \( \) around 0 (fill in the blanks)

5) The distribution of \( X \) is normal with expected value=0 and standard error=5. Find the chance that \( X \) is in the following sets:

\[
\begin{align*}
\{x < 5\} & \quad \{x < 7.5\} & \quad \{x < 10\} & \quad \{x < 17\} & \quad (- \infty ; 12.0) \\
\{x \geq 5\} & \quad \{x \geq 7.5\} & \quad \{x \geq 15\} & \quad [17.0 ; +\infty) \\
(- \infty; -5.0) & \quad \{x < -10 \} \\
[5; 10) & \quad [12.5; 17.5) \\
[-5; +5) & \quad [-10; +10) & \quad [-10; +10] & \quad [-12.5; +12.5] \\
[-5; +10] & \\
\end{align*}
\]
Using the normal table (exercises)

- the chance is 95% that X is less than ______ (fill in the blanks)
- the chance is 99% that X is less than ______ (fill in the blanks)
- the chance is 95% that X is greater than ______ (fill in the blanks)
- the chance is 99% that X is greater than ______ (fill in the blanks)
- the chance is 95% that X is within +/- ______ around 0 (fill in the blanks)
- the chance is 99% that X is within +/- ______ around 0 (fill in the blanks)

6) The distribution of X is normal with expected value=100 and standard error=2. Find the chance that X is in the following sets:

\[
\begin{align*}
\{x < 102\} & \quad \{x < 103\} & \quad \{x < 104\} & \quad \{x < 106.80\} & \quad (-\infty ; 104.80) \\
\{x \geq 102\} & \quad \{x \geq 103.0\} & \quad \{x \geq 106.0\} & \quad [106.80 ; +\infty) \\
(-\infty ; 98) & \quad \{x < 96\} \\
[102; 104] & \quad [105.0; 107.0) \\
[98; 102] & \quad [96; 104] & \quad [96; 104] & \quad [95; 105] & \quad [98 ; 104]
\end{align*}
\]

- the chance is 95% that X is less than ______ (fill in the blanks)
- the chance is 99% that X is less than ______ (fill in the blanks)
- the chance is 95% that X is greater than ______ (fill in the blanks)
- the chance is 99% that X is greater than ______ (fill in the blanks)
- the chance is 95% that X is within +/- ______ around 0 (fill in the blanks)
- the chance is 99% that X is within +/- ______ around 0 (fill in the blanks)

7) The distribution of the heights of the Trolland royal guards is normal with mean=180 cm and SD=10 cm.

a) what percentage of Trolland Royal Guards is shorter than 170 cm?

b) what percentage is higher than 204 cm?

c) "99 percent of the Guards' heights is within ______ cm around 180 cm." (fill in the blanks)

8) RedEyes ice creams are packed in tiny plastic boxes. The net weights of these packs are normally distributed with mean=40 grams and SD=8 grams. One pack is randomly selected. Find the chance that its weight is more than 60 grams.

Readings

Chapter 13. Normal approximation (exercises)

Exercise set "A":

1) A poll is planned by researchers who want to know the percentage of the ten millions population in favor of a new legislation. The sample size is going to be 1000. How precise is this measurement going to be?

Assume that the population proportion of those in favor of the proposal is 25%.

a) "The number of those in favor of the proposal in the sample is going to be________." (Fill in the blanks. Options: "exactly 250" / "around 250")

b) True or false: "The number of those in favor of the proposal in the sample is a random variable."

c) True or false: "The number of those in favor of the proposal in the sample is a variable with a hypergeometric distribution as the draws are without replacement."

d) True or false: "The number of those in favor of the proposal in the sample can be approximated with a binomial distribution because, though draws are without replacement, the size of the sample is tiny compared to the size of the population."

e) True or false: "The number of those in favor of the proposal in the sample can be approximated with a binomial distribution with parameters n=_____ and p=_____." (Fill in the blanks)

f) "The expected value of this binomial distribution is _______ and the SD is _______. This means that the number of those in favor of the proposal in the sample is going to be around _______, give or take _______ or so." (Fill in the blanks)

g) "The number of those in favor of the proposal in the sample can be approximated with a normal distribution because ________." (Fill in the blanks; options: "the sample size is great" / "the distribution of the population is close to normal").

h) "The distribution of the number of those in favor of the proposal in the sample is close to a normal distribution with expected value EV=_______ and standard deviation SD=_______." (Fill in the blanks)

i') "...so there is a 95% chance that the measured percentage of the supporters is going to be inside ±____% around ____%." (Fill in the blanks)

i") "...so there is a 99% chance that the measured percentage of the supporters is going to be inside ±____% around ____%." (Fill in the blanks)

2) A sample survey is planned by researchers wanting to know the average income of the population 25..45 years old. The sample size is going to be 1600. How precise is this measurement going to be?

Assume that the average of the incomes is 12,000 francs and the SD of these incomes is 15,000 francs in the target population.

a) true or false: "The mean of incomes in the sample is a random variable."

b) true or false: "The distribution of the sample mean of incomes is close to normal because the distribution of incomes in the population is close to normal."

c) true or false: "The distribution of the sample mean of incomes is close to normal because the sample size is big enough."

d) "The mean of incomes in the sample can be described with a normal distribution of expected value EV=_______ and standard deviation SD=_______." (Fill in the blanks)

So:
e) "There is a 95% chance that the difference between the sample mean and the population mean is going to be less than ______ francs." (That is, the chance that the measurement error exceeds this limit is only 5%). (Fill in the blanks)

f) "There is a 99% chance that the difference between the sample mean and the population mean is going to be less than ______ francs." (That is, the chance that the measurement error exceeds this limit is only 1%). (Fill in the blanks)

3) A sample measurement is planned by junior researchers wanting to know the average weight of ostrichs' eggs sold in XYZ hypermarkets. As ostrichs' eggs are expensive, the sample is to be small. It has been decided that the sample size is going to be 25. What precision is to be expected from this small-scale measurement?

Assume that the mean of the weights of ostrichs' eggs in the population is 8 kgs with an SD of 1 kg.

a) true or false: "The mean of the weights of eggs in the sample is a random variable."

b) true or false: "The distribution of the sample mean of the weights of eggs is close to normal because the distribution of weights in the population is close to normal."

c) true or false: "The distribution of the sample mean of the weights is close to normal because the sample size is big."

d) "The mean of the weights in the sample can be described with a normal distribution of expected value EV=_______ and standard deviation SD=_______." (Fill in the blanks)

So:

e) "There is a 95% chance that the difference between the sample mean of weights of eggs and the population mean of weights of eggs is going to be less than ______ kilograms." (That is, the chance that the measurement error exceeds this limit is only 5%). (Fill in the blanks)

f) "There is a 99% chance that the difference between the sample mean and the population mean is going to be less than ______ kilograms." (That is, the chance that the measurement error exceeds this limit is only 1%). (Fill in the blanks)

Exercise set "B":

1) We have got a big box of potatoes of variable kinds and sizes. The mean of the weights of potatoes in the box is 100 grams, the SD of these weights is 50 grams. We would like to know (a) what percentage of the potatoes in the box is heavier than 250 grams, (b) what percentage is heavier than 500 grams. Should normal approximation be applied here?

2) Almost nothing is known about the distribution of incomes of the citizens of Trolland. What is known from the Tax Office is that the average of these incomes is 100 golden francs and the SD of incomes is 2,000 golden francs. The population of Trolland is ten million. Rumour has it that a high proportion of Troll citizens are very rich. Two questions are to be answered:

a) whether it is possible that more than 1,000 citizens have incomes over 100,000 golden francs?

b) finding the highest possible number of Troll citizens with incomes over 10,000 golden francs.

Should normal approximation be applied here?

3) Loafs of bread at hypermarket ABC have weights with average=1000 grams and SD=30 grams.

a) Find the highest possible percentage of loafs with weights under 940 grams or over 1060 grams (that is, with a more than 60 grams deviation in weight from the nominal).

b) Find the highest possible percentage of loafs with weights differing more than 150 grams from the nominal 1000 grams.

c) Find the highest possible percentage of loafs with weights under 700 grams or over 1300 grams.
d*) Find the highest possible percentage of loafs with weights under 700 grams.

Does normal approximation apply?

4) A shipment of 16 loaves is selected randomly from the loaves in the previous exercise.

Denote X the total weight of the shipment.

a.) E(X)=? var(X)=? D(X)=?

b) Find the approximate chance of the total weight deviating more than 240 grams from the expected 16 kilograms.

c) Find the approximate chance of the total weight deviating more than 360 grams from the expected 16 kilograms.

d) Find the approximate chance of the total weight being under 15 kilograms.

Denote Y the mean of the weights of loaves in the shipment. E(Y)=? D(Y)=?

e) Find the approximate chance of the mean weight of loaves in the shipment deviating more than 30 grams from the expected 1000 grams.

f) Find the approximate chance of the mean weight of loaves in the shipment deviating more than 50 grams from the expected 1000 grams.

g) Find the approximate chance of the mean weight of loaves in the shipment deviating more than 15 grams from the expected 1000 grams. (...the chance that the mean is between 985 grams and 1015 grams.)

h) Find the approximate chance of the mean weight of loaves in the shipment deviating more than 22.5 grams from the expected 1000 grams.

Does normal approximation apply here?

5) A shipment of one hundred loaves is selected randomly from the loaves in the above two exercises. Denote X the total weight of the shipment, denote Y the mean of the weights of loaves in the shipment. E(X)=? D(X)=?

a) This means that the total weight of the shipment is around ____ kgs, give or take ____kgs or so. (Fill in the blanks.)

Should normal approximation be applied in answering questions b..e?

b) Find the approximate chance of the total weight deviating more than 240 grams from the expected 100 kilograms.

c) Find the approximate chance of the total weight deviating more than 360 grams from the expected 100 kilograms.

d) Find the approximate chance of the mean weight of loaves in the shipment deviating more than 30 grams from the expected 1000 grams.

e) Find the approximate chance of the mean weight of loaves in the shipment deviating more than 50 grams from the expected 1000 grams.

f) Find the approximate chance of the total weight being under 99 kilograms.

g) What is the chance that the difference between the mean of weights in the sample and the expected value exceeds 10 grams?

6) The population average of incomes in the Lowlands is 100 golden francs, the SD of these incomes is 200 golden francs.
We would like to know the percentage of citizens

a) with incomes over 500 golden francs;

b) with incomes over 1000 golden francs;

c) with incomes over 2000 golden francs.

Should normal approximation be applied?

7) [continues the previous exercise] A random sample of size n=400 is selected from the citizens of the Lowlands. Denote X the mean of incomes in the sample. E(X)=? D(X)=? We would like to know the approximate change that the mean of incomes in the sample

a) exceeds 500 golden francs;

b) exceeds 1000 golden francs;

c) exceeds 2000 golden francs.

Should normal approximation be applied?

Apply the normal approximation at exercises 8-12. (Is it all right? Explain.)

8) 60% of the Troll population is male, 40% female. A random sample of size 1000 is selected.

a) Find the chance that the percentage of women in the sample is below 35% or over 45%.

b) The chance is 99% that the proportion of women in the sample is between ____ % and ____%. (Fill in the blanks.)

9) 50% of the Troll population has brown hair, 50% has yellow hair. A random sample of size 1000 is selected.

a) Find the chance that the percentage of yellow-haired trolls in the sample is below 45% or over 55%.

b) There is a 99% chance that the proportion of yellow-haired trolls in the sample is between ____ % and ____%. (Fill in the blanks.)

c) How would you change your answer to (a) taking into account that the sampling is without replacement? (Population size to be assumed 100,000.)

10) 90% of the Troll population is bright, 10% brighter. A random sample of size 1000 is selected.

a) Find the chance that the percentage of brighter trolls in the sample is below 5% or over 15%.

b) The chance is 99% that the proportion of brighter trolls in the sample is between ____ % and ____%. (Fill in the blanks.)

c) How would you change your answer to (a) taking into account that the sampling is without replacement? (Population size to be assumed 100,000.)

11) A die has been rolled 6000 times and, instead of the expected 1000, there have been only 710 aces. Can the claim that the die is fair be refuted?

b) and in the less striking example of 925 aces out of the 6000 rolls?

12) A die has been rolled 600 times and, instead of the expected 100, there have been only 71 aces. Can the claim that the die is fair be refuted?

b) and in the less striking example of 92 aces out of the 600 rolls?

Exercise set "C": 


1) There are exactly 50% men and 50% women in a big city. A random sample of size 1600 is selected from among them. Variable to be observed=number of women in the sample.

a) Find the chance that there will be 820 or less women in the sample.

b) Find the chance that there will be 840 or less women in the sample.

c) Find the chance that there will be 780 or less women in the sample.

d) Find the chance that there will be 760 or less women in the sample.

e) Find the chance that the number of women in the sample is more than 780 but less than 820.

f) Find the chance that the number of women in the sample is more than 760 but less than 840.

g) It is almost sure (has a 95% chance) that there will be no more than ____ women in the sample.

h) It is almost sure (has a 99% chance) that there will be no more than ____ women in the sample.

i) It is almost sure (has a 99.9% chance) that there will be no more than ____ women in the sample.

j) It is quite unprobable (has a 5% chance) that there will be less than ____ women in the sample.

k) It is quite unprobable (has a 1% chance) that there will be less than ____ women in the sample.

l) It is quite unprobable (has a 0.1% chance) that there will be less than ____ women in the sample.

m) It is almost sure (has a 95% chance) that the number of women will be between 800±____ in the sample.

n) It is almost sure (has a 99% chance) that the number of women will be between 800±____ in the sample.

Fill in the blanks.

1') The same, but now sampling without replacement. (Population size to be assumed 1,000,000.)

2) There are exactly 50% men and 50% women in a big city. A random sample of size 400 is selected from among them. Variable to be observed=number of women in the sample.

a) Find the chance that there will be 220 or less women in the sample.

b) Find the chance that there will be 230 or less women in the sample.

c) Find the chance that there will be 180 or less women in the sample.

d) Find the chance that there will be 170 or less women in the sample.

e) Find the chance that the number of women in the sample is more than 180 but less than 220.

f) Find the chance that the number of women in the sample is more than 170 but less than 230.

g) It is almost sure (has a 95% chance) that there will be no more than ____ women in the sample.

h) It is almost sure (has a 99% chance) that there will be no more than ____ women in the sample.

i) It is almost sure (has a 99.9% chance) that there will be no more than ____ women in the sample.

j) It is quite unprobable (has a 5% chance) that there will be less than ____ women in the sample.

k) It is quite unprobable (has a 1% chance) that there will be less than ____ women in the sample.

l) It is quite unprobable (has a 0.1% chance) that there will be less than ____ women in the sample.

m) It is almost sure (has a 95% chance) that the number of women will be between 200±____ in the sample.
Normal approximation (exercises)

o) It is almost sure (has a 99% chance) that the number of women will be between 200±_____ in the sample.

3) There are exactly 50% men and 50% women in a big city. A random sample of size 1600 is selected from among them. Variable to be observed=proportion of women in the sample.
   a) Find the chance that there will be 51% or more women in the sample.
   b) Find the chance that there will be 51% or less women in the sample.
   c) Find the chance that there will be 52% or more women in the sample.
   d) Find the chance that there will be 53% or more women in the sample.
   e) Find the chance that there will be 54% or more women in the sample.
   f) There is a 95% chance that the proportion of women in the sample will be inside 50%±____%.
   g) There is a 99% chance that the proportion of women in the sample will be inside 50%±____%.

4) There are exactly 50% men and 50% women in a big city. A random sample of size 400 is selected from among them. Variable to be observed=proportion of women in the sample.
   a) Find the chance that there will be 51% or more women in the sample.
   b) Find the chance that there will be 51% or less women in the sample.
   c) Find the chance that there will be 52% or more women in the sample.
   d) Find the chance that there will be 53% or more women in the sample.
   e) Find the chance that there will be 54% or more women in the sample.
   f) There is a 95% chance that the proportion of women in the sample will be inside 50%±____%.
   g) There is a 99% chance that the proportion of women in the sample will be inside 50%±____%.

5) In Trolland 20 percent of the voting population favours the Dwarf Party. (This fact is unknown to the Trolls yet.) A random sample of size 1000 is selected by a polling firm wanting to know the proportion of those in favor of the Dwarf Party in the population.
   a) There is a 95% chance that their error will not exceed ____%.
   b) There is a 99% chance that their error will not exceed ____%.
   c) There is a ____% chance that their estimate will be inside 20%±1.3%.
   d) There is a ____% chance that their estimate will be inside 20%±2.5%. (That is, that the number of those in favor of the Dwarf Party will be between 187 and 213 in the sample.)

6) The average height of men liable to conscription is 180 cms in the Burgundy population. (The SD of the heights is 15 cms.) A random sample of 100 is selected from among these men.
   a) Find the chance that the mean height will be above 185 cms in the sample.
   b) Find the chance that the mean height will be above 183 cms in the sample.
   c) Find the chance that the mean height will be above 181 cms in the sample.

7) Lucifer light bulbs have a mean lifetime m=1000 hours (the SD of lifetimes d=250 hours), according to specifications. A random sample of 100 bulbs is taken, the lifetime of each bulb is being determined, the average of lifetimes is then calculated.
   a) Find the chance that the average lifetime falls under 900 hours in the sample (on condition the specifications are correct).
8) Numbered cards in a box; the player gets in dollars the amount shown on the card he has drawn. The average of the box (that is, of the cards in the box) is 10$, the SD of the box is 20$. One hundred draws are made (draws are with replacement) then the average of the draws computed.

a) Find the chance that the average of the draws is under 8$.

b) Find the chance that the average of the draws is under 9$.

c) There is only a 5% chance that the average is under _______$.

d) There is only a 1% chance that the average is under _______$ (Fill in the blanks.)

9) "Every third wins!" in a lottery game, the slogan says. It means that one third of all lottery tickets in the game is a winning ticket, distributed randomly among the non-winning ones. 300 lottery tickets are bought.

a) The chance is ____% that there will be at least 86 winning tickets among the 300.

b) The chance is ____% that the number of winning tickets among the 300 is 120 or more.

c) There is a 99% chance that the number of winning tickets among the 300 is ___ or more.

d) There is a 99% chance that the number of winning tickets among the 300 is inside 100±______.

e) There is a 95% chance that the number of winning tickets among the 300 is ___ or more.

f) There is a 95% chance that the number of winning tickets among the 300 is inside 100±______.

(Fill in the blanks.)

10) 40% of the voting population is in favour of a proposal. A random sample of n=1000 is taken by a polling firm.

a) Find the chance that their error exceeds 1%. (That is, that they get a percentage higher than 41% or lower than 39%).

b) Find the chance that their error exceeds 2%.

c) Find the chance that their error exceeds 3%.

d) Find the chance that their error exceeds 4%.

e) Find the chance that their error exceeds 5%.

11) 20% of the voting population is in favour of a proposal. A random sample of n=1000 is taken by a polling firm.

a) Find the chance that their error exceeds 1%.

b) Find the chance that their error exceeds 2%.

c) Find the chance that their error exceeds 3%.

d) Find the chance that their error exceeds 4%.

e) Find the chance that their error exceeds 5%.

12) 10% of the voting population is in favour of a proposal. A random sample of n=1000 is taken by a polling firm.

a) Find the chance that their error exceeds 1%.

b) Find the chance that their error exceeds 2%.
c) Find the chance that their error exceeds 3%.
d) Find the chance that their error exceeds 4%.
e) Find the chance that their error exceeds 5%.

13–15) Same as 10–12 but with n=500.
16–18) Same as 10–12 but with n=2500.

19) Tosses are made with a fair coin.
a) With 100 tosses there is a 95% chance that the number of heads is inside 50%±____%.
b) With 400 tosses there is a 95% chance that the number of heads is inside 50%±____%.
c) With 900 tosses there is a 95% chance that the number of heads is inside 50%±____%.
d) With 1,600 tosses there is a 95% chance that the number of heads is inside 50%±____%.
e) With 10,000 tosses there is a 95% chance that the number of heads is inside 50%±____%.
f) With 40,000 tosses there is a 95% chance that the number of heads is inside 50%±____%.

(Fill in the blanks.)

20) Tosses are made with a coin loaded so that the chance of tossing a head is p=0.40.
a) With 100 tosses there is a 95% chance that the number of heads is inside 40%±____%.
b) With 400 tosses there is a 95% chance that the number of heads is inside 40%±____%.
c) With 900 tosses there is a 95% chance that the number of heads is inside 40%±____%.
d) With 1,600 tosses there is a 95% chance that the number of heads is inside 40%±____%.
e) With 10,000 tosses there is a 95% chance that the number of heads is inside 40%±____%.
f) With 40,000 tosses there is a 95% chance that the number of heads is inside 40%±____%.

(Fill in the blanks.)

21) Tosses are made with a coin loaded so that the chance of tossing a head is p=0.49.
a) With 100 tosses there is a 95% chance that the number of heads is inside 49%±____%.
b) With 400 tosses there is a 95% chance that the number of heads is inside 49%±____%.
c) With 900 tosses there is a 95% chance that the number of heads is inside 49%±____%.
d) With 1,600 tosses there is a 95% chance that the number of heads is inside 49%±____%.
e) With 10,000 tosses there is a 95% chance that the number of heads is inside 49%±____%.
f) With 40,000 tosses there is a 95% chance that the number of heads is inside 49%±____%.

(Fill in the blanks.)

22) We are to decide with confidence if a coin is fair or loaded so as in exercise 20. Find the number of tosses necessary, approximately.

23) We are to decide with confidence if a coin is fair or loaded so as in exercise 21. Find the number of tosses necessary, approximately.
24) The scores of the Crow IQ test have a population SD of $d=10$ points. The unknown population mean is to be estimated from a sample. (Simple random sampling is used.)

a) Taking a sample of size 100 it is almost certain (exactly, has a chance of 99%) that the deviation of the sample mean from the population mean is not more than ____ points. (Then the measurement can be said to be almost sure to have an error less than this value.)

b) Taking a sample of size 400 the chance is 99% that the deviation of the sample mean from the population mean is not more than ____ points.

c) Taking a sample of size 900 the chance is 99% that the deviation of the sample mean from the population mean is not more than ____ points.

d) Taking a sample of size 1000 the chance is 99% that the deviation of the sample mean from the population mean is not more than ____ points.

e) The researchers would like to estimate the present population average of the IQ scores with a precision of 0.25 points (that is, they want the S.E. of the measurement to be not more than 0.25 points). Find the necessary sample size.

f) The researchers would like to estimate the present population average of the IQ scores with a precision of 0.25 points (that is, they want the chance that the error of the measurement exceeds 0.25 points to be not more than 1%). Find the necessary sample size.

Readings

Chapter 14. Roulette/2: normal approximation (exercises)

For the roulette, see Freedman–Pisani–Purves: Statistics, pp. 281-286.

exercise set A (100 games)

1) Jean plays European roulette (with 37 slots) a hundred times, putting 1 dollar on Red each time. If he wins he gets his dollar back with winnings of 1 dollar. If he loses he gets nothing. (He wins with 18 of the 37 slots.)

a) Find the chance of his 'being ahead' after the 100 games. (That is, the chance that the sum of what he gets exceeds the sum of what he pays.)

b) Find the chance of his net gain from the 100 games being over 30 percent of his bets. (That is, the chance that the sum of what he gets minus the sum of what he pays exceeds 30 dollars.)

2) Jean plays European roulette a hundred times, putting 1 dollar on First dozen each time. If he wins he gets his dollar back with winnings of 2 dollars. If he loses he gets nothing. (He wins with 12 of the 37 slots.)

a) Find the chance of his 'being ahead' after the 100 games. (That is, the chance that the sum of what he gets exceeds the sum of what he pays.)

b) Find the chance of his net gain from the 100 games being over 30 percent of his bets.

3) Jean plays European roulette a hundred times, putting 1 dollar on First row each time. If he wins he gets his dollar back with winnings of 11 dollars. If he loses he gets nothing. (He wins with 3 of the 37 slots.)

a) Find the chance of his 'being ahead' after the 100 games. (That is, the chance that the sum of what he gets exceeds the sum of what he pays.)

b) Find the chance of his net gain from the 100 games being over 30 percent of his bets.

exercise set B (1000 games)

1) Jean plays European roulette a thousand times, putting 1 dollar on Red each time.

a) Find the chance of his 'being ahead' after the 1,000 games.

b) Find the chance of his net gain from the 1,000 games being over 30 percent of his bets that is, over 300 dollars.

c) Find the chance of his net gain from the 1,000 games being over 30 dollars.

d) Find the chance of his net gain from the 1,000 games being over 95 dollars.

2) Jean plays European roulette a thousand times, putting 1 dollar on First dozen each time.

a) Find the chance of his 'being ahead' after the 1,000 games.

b) Find the chance of his net gain from the 1,000 games being over 30 percent of his bets that is, over 300 dollars.

c) Find the chance of his net gain from the 1,000 games being over 30 dollars.

d) Find the chance of his net gain from the 1,000 games being over 95 dollars.

3) Jean plays European roulette a thousand times, putting 1 dollar on First row each time.

a) Find the chance of his 'being ahead' after the 1,000 games.
b) Find the chance of his net gain from the 1,000 games being over 30 percent of his bets that is, over 300 dollars.

c) Find the chance of his net gain from the 1,000 games being over 30 dollars.

d) Find the chance of his net gain from the 1,000 games being over 95 dollars.

exercise set C (10,000 games)

1) Jean plays European roulette ten thousand times, putting 1 dollar on Red each time.
   a) Find the chance of his 'being ahead' after the 10,000 games.
   b) Find the chance of his net gain from the 10,000 games being over 30 percent of his bets that is, over 3,000 dollars.
   c) Find the chance of his net gain from the 10,000 games being over 30 dollars.
   d) Find the chance of his net gain from the 10,000 games being over 300 dollars.

2) Jean plays European roulette ten thousand times, putting 1 dollar on First dozen each time.
   a) Find the chance of his 'being ahead' after the 10,000 games.
   b) Find the chance of his net gain from the 10,000 games being over 30 percent of his bets that is, over 3,000 dollars.
   c) Find the chance of his net gain from the 10,000 games being over 30 dollars.
   d) Find the chance of his net gain from the 10,000 games being over 300 dollars.

3) Jean plays European roulette ten thousand times, putting 1 dollar on First row each time.
   a) Find the chance of his 'being ahead' after the 10,000 games.
   b) Find the chance of his net gain from the 10,000 games being over 30 percent of his bets that is, over 3,000 dollars.
   c) Find the chance of his net gain from the 10,000 games being over 30 dollars.
   d) Find the chance of his net gain from the 10,000 games being over 300 dollars.

exercise set D

1) Find the expected value of the net gain from 100 games, putting 1 dollar
   (a) on Red   (b) on First dozen   (c) on First row each time.

2) Find the standard deviation of the net gain from 100 games, putting 1 dollar
   (a) on Red   (b) on First dozen   (c) on First row each time.

1’) Find the expected value of the net gain from 1,000 games, putting 1 dollar
   (a) on Red   (b) on First dozen   (c) on First row each time.

2’) Find the standard deviation of the net gain from 1,000 games, putting 1 dollar
1”) Find the expected value of the net gain from 10,000 games, putting 1 dollar
(a) on Red  (b) on First dozen  (c) on First row
each time.

2”) Find the standard deviation of the net gain from 10,000 games, putting 1 dollar
(a) on Red  (b) on First dozen  (c) on First row
each time.

exercise set E

1) We want a net gain of at least 50%. (That is, spending 100 dollars, we want a net gain of at least 50 dollars.)
Which gives you a better chance,
a) putting 1 dollar on Red a hundred times?
b) putting 1 dollar on First dozen a hundred times?
c) putting 1 dollar on First row a hundred times?

2) We want a net gain of at least 50%. (That is, spending 10,000 dollars, we want a net gain of at least 5,000 dollars.)
Which gives you a better chance,
a) 100 games, putting 100 dollars on Red each time?
b) 1,000 games, putting 10 dollars on Red each time?
c) 10,000 games, putting 1 dollar on Red each time?
d) (1 game) putting all 10,000 dollars on Red at once?
Chapter 15. Some kinds of distributions (theory)

1. Discrete distributions

*Geometric distribution (with parameter p):*

**Example 15.1.**

Rolling a fair die until the first ace shows. Denote $X$ the number of rolls needed. (E.g. $X=1$ if an ace shows at once, $X=2$ if the first toss is not an ace but the second is an ace, etc.)

- More generally, an experiment is repeated until some event $A$ occurs. The probability of $A$ happening is $p$ in each experiment ($p$ is also called the probability of success). The experiments are independent. The variable $X$ shows the number of trials until the first success, that is, until the first occurrence of $A$.

$$P(X = k) = (1 - p)^{k-1} \cdot p$$  \hspace{1cm} (15.1)

Its expected value is $\frac{1}{p}$, its S.D. $\sqrt{\frac{1-p}{p^3}}$.

*Negative binomial distribution (with parameters $p$ and $r$):*

**Example 15.2.**

Rolling a fair die until the third ace shows (the aces need not be consecutive). Denote $X$ the number of rolls needed for this. (E.g. $X=3$ if the first 3 rolls are all aces; $X=4$ if only 2 of the first 3 rolls are aces and the fourth roll is an ace, etc.)

- More generally, an experiment is repeated until some event $A$ occurs the $r$th time. The probability of $A$ happening is $p$ in each experiment; the experiments are independent. The variable $X$ shows the number of trials until the $r$th occurrence of $A$. Then for $k \geq r$

$$P(X = k) = \binom{k-1}{r-1} p^r \cdot (1-p)^{k-r}$$  \hspace{1cm} (15.2)

(The probability of a sequence with $r$ successes and $k-r$ failures is $p^r (1-p)^{k-r}$; the number of sequences with a success on the $k$th position and $r-1$ successes on the positions 1..($k-1)$ is $\binom{k-1}{r-1}$ that is, the number of combinations the positions of the $r-1$ successes can be selected from the first $k-1$ positions.)

The sum of $r$ random variables with geometric distributions of parameter $p$ gives a negative binomial distribution with parameters $p$ and $r$.

True or false: a negative binomial distribution with parameters $p$ and $r$ is the sum of $r$ independent random variables with geometric distributions of parameter $p$.

$$E(X) = \frac{1}{p}, \text{ D}(X) = \sqrt{\frac{r(1-p)}{p^3}}$$

*Hypergeometric distribution (with parameters $N$, $M$ and $n$):*

**Example 15.3.**

10 balls in a box, 6 blue, 4 yellow; five draws are made, without replacement. What is the chance that exactly 3 blue balls are among the 5 drawn?
Some kinds of distributions (theory)

\[ P(3 \text{ blue}) = \binom{3}{1} \binom{1}{1} \binom{1}{1} \]

– more generally: in a box (in a population) there are altogether N elements, of those M of this kind, N–M of some other kind. n draws are made, without replacement, from the box. Denote X the number of elements of this kind in the sample. Then, after some combinatorial considerations,

\[ P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \]

(15.3)

\[ \text{E}(X) = \frac{M}{N}, \text{D}(X) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \]

Binomial distribution (with parameters p and n):

Example 15.4.

balls in a box, one third of the balls are blue the remaining two thirds are yellow. Ten draws are made with replacement. What is the chance that, of the ten balls drawn, exactly 4 are blue?

\[ P(4 \text{ blue}) = \binom{10}{4} \left( \frac{1}{3} \right)^4 \left( \frac{2}{3} \right)^6 \]

Example 15.5.

ten rolls with a fair dice. What is the chance of getting exactly 2 aces?

\[ P(2 \text{ aces}) = \binom{10}{2} \left( \frac{1}{6} \right)^2 \left( \frac{5}{6} \right)^8 \]

- more generally, an experiment is repeated n times. The variable X shows the number of times an event A with probability p occurs. The experiments are independent.

\[ P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \]

(The probability of a good sequence, that is, of k successes with (n-k) failures, equals \( p^k (1-p)^{n-k} \); the number of good sequences equals the number of combinations the positions of the k successes can be selected from the n positions, that is, \( \binom{n}{k} \).)

Remark: both the hypergeometric and the binomial distributions can be obtained by summing a number of very simple variables.

a. Hypergeometric distribution with parameters N, M and n:

Denote \( X_i \) (i=1,...,n) a variable indicating if the ith draw happens to be of this kind, its value being 1 if the draw is of this kind and 0 if it is of the other kind. (\( X_i \) is also called an indicator variable.) Then \( X = \sum_{i=1}^{n} X_i \).

(The experiment could be modelled putting N numbered cards in a box, writing an 1 on M of the cards and a 0 on (N-M) of the cards, then making n draws, without replacement. Denote X the sum of the draws.)

b. Binomial distribution (with parameters p and n):

Denote \( X_i \) (i=1,...,n) a variable indicating if event A occurred at the ith repetition of the experiment. Let \( X_i = 1 \) if A happens and \( X_i = 0 \) if A does not happen. Then \( X = \sum_{i=1}^{n} X_i \).

(The experiment could be modelled putting cards numbered 0 and 1 in a box, the proportion of the 1s being p and the proportion of the zeroes (1–p), then making n draws from the box, with replacement. Denote X the sum of the draws.)
Some kinds of distributions (theory)

The $x$-s in examples a) and b) are called indicator variables, indicating with 1 if some event happens and 0 if it does not happen. Their distribution is called Bernoulli distribution /with parameter $M/N$ in (a); with parameter $p$ in (b)/.

Poisson distribution (with parameter $\lambda$)

– if you take a series of binomial distributions with parameters $(n,p)$, increasing $n$ and decreasing $p$ such that the expected value, $np$, remains constant¹ you get a Poisson distribution with parameter $np=:\lambda$. Then

$$P(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

(15.4)

E.g., the number of raindrops falling into a given square on the street in a minute has approximately this kind of distribution, $\lambda$ being the average number of raindrops in a minute here. (An interesting example can be found in Lady Luck² by Warren Weaver: the distribution of the number of cavalry men in the British army kicked to death by horses each year had been found to be very close to a Poisson distribution.)

$$E(X)= \lambda \quad , \quad D(X)= \sqrt{\lambda}$$

2. Continuous distributions

Uniform distribution on interval $[a;b]$ :

- denoted $U[a;b]$
- its density function $f_x(x) = \frac{1}{b-a}$ on interval $[a;b]$; elsewhere $f_x(x)=0$

exponential distribution (with parameter $\lambda$):

- its density function $f(t) = \lambda \cdot e^{-\lambda t}$ if $t > 0$, and $f(t) = 0$, if $t < 0$.

Example 15.6.

a. an atom liable to fission is observed, denote $t$ the time from now until the fission actually occurs. Then the distribution of $t$ is exponential.

b. The distribution of the time interval between two subsequent breakings of threads on a power-loom is usually considered exponential.

$$E(X)= \frac{1}{\lambda} \quad , \quad D(X)= \frac{1}{\lambda^2}$$

gamma distribution with parameters $\lambda$ and $n$

– is the sum of $n$ independent exponential variables with parameter $\lambda$, so

$$E(X)= \frac{n}{\lambda} \quad , \quad D(X)= \frac{\sqrt{n}}{\lambda}$$

The normal distributions have a central role in probability theory and mathematical statistics.

Normal distribution (with E.V. = $m$ and S.D. = $d$):

– its density function $f(x) = \frac{1}{d \sqrt{2\pi}} \cdot e^{-\frac{(x-m)^2}{2d^2}}$

Check its expected value really being $m$ and its S.D. really being $d$.

A special case of normal distributions is the

1¹ $n \rightarrow \infty, \quad p \rightarrow 0$ . The Poisson distribution is then the limiting distribution of the binomials, that is, $\lim_{n \rightarrow \infty, p \rightarrow 0} \{P_{\text{Bernoulli}}(X = k)\} = P_{\text{Poisson}}(X = k)$

**Standard Normal Distribution** (a normal distribution with E.V.=0 and S.D.=1)

- with density function \( f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \)

Approximately normally distributed are some population variables, e.g.: body sizes (heights; weights; numbers of hairs on persons’ scalps); size errors in production (inside diameters of bearings; actual weights of one-kilogram breads); or measurement errors in repeated measures of the same quantity (see Freedman, Chapter 24).

Besides, approximately normally distributed are some random variables. For example, given an experiment of selecting a sample of size \( n \) from a numeric population and then calculating (a) the sample sum or (b) the sample mean, both are going to be distributed close to normal. (The sample sum and the sample mean both are random variables as their actual values depend on the sample randomly selected, that is, drawn from a box. Both disperse around their respective expected values.)

Accordingly, in real studies, dealing with sample means coming from big samples, these means can be characterized with normal distributions, and on these grounds it can be decided whether the sample observed conforms to our expectations or deviates significantly from what has been expected.

**Some Important Properties.** The sum of two or more normally distributed, independent random variables is also normally distributed, as are the linear combinations of some independent, normally distributed variables:

- let \( X \) be a normal variable \( \in \) \( \mathbb{R} \) a nonzero constant; then \( c.X \) is also normally distributed;
- let \( X \) and \( Y \) be independent, normally distributed variables; then \( X+Y \) is also normally distributed;
- let \( X \) and \( Y \) be independent, normally distributed variables; then \( X−Y \) is also normally distributed;
- let \( X_1, X_2, \ldots, X_n \) be independent, normally distributed variables and \( a_1, a_2, \ldots, a_n \in \mathbb{R} \) constants; then \( Y:= \frac{1}{c} \sum_{i=1}^{n} a_i X_i \) is also normally distributed (if only \( \sum_{i=1}^{n} a_i^2 \neq 0 \))

The following distributions derived from the normal distribution will have important roles in hypothesis testing.

**Chi-squared distribution:** summing the squares of \( n \) independent standard normal random variables we get a random variable with a chi-square distribution of \( n \) degrees of freedom. That is, let \( X_1, X_2, \ldots, X_n \) be independent random variables with standard normal distributions; let \( Y := \sum_{i=1}^{n} X_i^2 \); then the distribution of \( Y \) is a chi-squared distribution (denoted by \( \chi^2 \)) with \( n \) degrees of freedom.

E.V. = \( n \); variance = \( 2n \)

**T Distribution, alias Student Distribution:**

Let \( X \) and \( Y \) be independent random variables, the distribution of \( X \) being standard normal, that of \( Y \) being a chi-squared distribution with \( n \) degrees of freedom; let \( z := \frac{1}{\sqrt{n}} \frac{X}{\sqrt{Y/n}} \). Then the distribution of \( Z \) is a \( t \) distribution with \( n \) degrees of freedom (also called Student distribution with \( n \) degrees of freedom).

Its E.V. = 0 if \( n>1 \); its variance = \( \frac{n}{n-2} \) if \( n>2 \).

**F Distribution:** let \( X \) and \( Y \) be independent random variables, with chi-squared distributions of \( n \) and \( m \) degrees of freedom, respectively; let \( Z := \frac{X}{m} \). Then the distribution of \( Z \) is an \( F \) distribution with \( (n,m) \) degrees of freedom.

**Readings**


\( n \) has to be sufficiently large.
Chapter 16. Some kinds of distributions – expected values, standard deviations, probabilities (questions, exercises)

(in part calculus exercises)

1) True or false: a binomial variable with parameters \((n, p)\) is the sum of \(n\) independent, identically distributed indicator variables.

2) True or false: a hypergeometric variable with parameters \((N, M, n)\) is the sum of \(n\) independent, identically distributed indicator variables.

3) True or false: a binomial variable with parameters \((n, p)\) is the sum of \(n\) independent indicator variables.

4) True or false: a hypergeometric variable with parameters \((N, M, n)\) is the sum of \(n\) identically distributed indicator variables.

5) True or false: a negative binomial variable with parameters \((p, r)\) is the sum of \(r\) independent, identically distributed geometric variables of parameter \(p\).

6) Prove the two formulas giving the E.V. and the S.D. of binomial distributions with parameters \((n, p)\).

7a) Prove the formula giving the E.V. of hypergeometric distributions with parameters \((N, M, n)\).

b*) Prove the formula giving the S.D. of hypergeometric distributions with parameters \((N, M, n)\).

8*) Prove the formula giving the E.V. of the geometric distribution with parameter \(p\).

9) Prove the formulas giving the E.V. and the S.D. of the exponential distribution with parameter \(\lambda\).

10) Prove the formulas giving the E.V. and the S.D. of the gamma distribution with parameters \((\lambda, n)\).

11) The lifetime$^1$ of single atoms of a radioactive isotope is characterized with an exponential distribution of parameter \(\lambda\). Find the half-life of this isotope. (That is, the time during which one half of the atoms in a given volume fissions.)

12) Let \(X\) be exponentially distributed with parameter \(\lambda\). True or false:

\[ P(X > t + t_0 \mid X > t) = P(X > t) \]

(If it is true that means that, if the lifetime of an atomic particle is characterized with an exponential distribution, then the chance that this single particle survives the first \(t\) hours equals the chance that it survives another \(t\) hours given it has already survived the first \(t_0\) hours. This property is called the memoryless property of the exponential distribution. Accordingly, radioactive particles may be thought of as not growing old.)

13) Let the distribution of \(X\) be a geometric distribution with parameter \(p\). Show that

\[ P(X > k + k_0 \mid X > k_0) = P(X > k) \]

(E.g., is it true, that, rolling a fair dice, the chance that the first ace shows on the second roll equals the chance that the first ace shows on the twelfth roll given that no aces have been shown during the first ten rolls. If it were true, the dice could be thought of as not having a memory.)

---

$^1$lifetime of a single atom := the time from now on until the fission of this atom actually occurs.
14) Dwarves in Trolland's forests are dwarfish, their heights being distributed normally with a population mean of 140 centimetres and a population standard deviation of 10 centimetres.

   a) Find the percentage of the population exceeding 160 centimetres in height.
   b) Exceeding 165 centimetres.
   c) Find the percentage of the population with heights between 130 cms and 150 cms.
   d) Find the percentage of the population with heights between 130 cms and 160 cms.
   e) Find the height in cms the first decile of the population is at.

15) Let X, Y and Z be normally distributed, independent random variables with E.V.s of 5, 10 and 20, and with S.D.s of 10, 20 and 40, respectively.

   Let W:=X+Y+Z.
   a) Find the distribution of W.
   b) E(W)=?
   c) D(W)=?
   d) P(W > 10)=?

15') Let X, Y and Z be normally distributed, independent random variables with E.V.s of 5, 10 and 20, and with S.D.s of 10, 20 and 40, respectively.

   Let W:=X+Y−Z.
   a) Find the distribution of W.
   b) E(W)=?
   c) D(W)=?
   d) P(W > 10)=?
Chapter 17. Law of large numbers (LLN) and central limit theorem (CLT) (theory)

1. The law of large numbers (LLN)

a. LLN for probabilities and relative frequencies:

The same experiment is repeated a lot of times and it is observed every time if an event A with probability p occurs. The experiments are independent. Denote $X_n$ the relative frequency of event A in the first n experiments. Then, for every real $\varepsilon > 0$,

$$\lim_{n \to \infty} \left| P(|X_n - p| > \varepsilon) \right| = 0$$  (17.1)

That is: the chance that the sample proportion (the relative frequency) deviates from the population proportion (the probability) with more than a given positive number (\varepsilon) converges to zero, as the sample size increases.

b. LLN for variables on the interval measurement level and their expected values:

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed random variables with a common expected value of $M$ and a common standard deviation of $D$. Denote $Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ the mean of the first n of the X variables. Then, for every real $\varepsilon > 0$,

$$\lim_{n \to \infty} \left( P(|Y_n - M| > \varepsilon) \right) = 0$$  (17.2)

That is: the chance that the sample mean deviates from the population mean with more than a given positive number (\varepsilon) converges to zero, as the sample size increases.

(The proof in both cases is, essentially, an application of Chebyshev's inequality. An expression of the type "$X - E(X)$" stands inside the modulus, in both formulas. Writing $kD(X)$ of the Chebyshev's $P(|X-E(X)| < kD(X))$ in the place of \varepsilon (and applying the square root law and a replacement $D(X) = \sqrt{ Var(X) }$) we get

$$\varepsilon = k \cdot \frac{SD}{\sqrt{n}}.\text{ Hence } k = k_0 = \sqrt{n} \cdot \frac{\varepsilon}{SD} , \text{ so } \lim_{n \to \infty} k_n = \infty .$$

Therefore the $\frac{1}{k_n}$ specifying the upper bound for the probability, on the Chebyshev's right side, converges to zero.)

2. Central limit theorem (CLT)

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed random variables (with expected value $M$ and standard deviation $D$). Denote

$$Y_n = \frac{\left( \sum_{i=1}^{n} X_i \right) - nM}{\sqrt{n} \cdot D}$$

the standardized of the sum of the first n of the X variables. Then the distribution of $Y_n$ will be close to a standard normal distribution, if only n is big enough. More exactly, for every real $x \in \mathbb{R}$

$$\lim_{n \to \infty} P(Y_n < x) = \Phi(x)$$

where $\Phi(x)$ is the standard normal cumulative distribution function.
Law of large numbers (LLN) and central limit theorem (CLT) (theory)

Less exactly, it is known that $Y_n$ has an E.V.=$M$ and an S.D.=$\sigma/\sqrt{n}$. The point is then that the sum is distributed normally, its distribution is close to that of a normal variable with E.V.=$M$ and S.D.=$\sigma/\sqrt{n}$.

That is, the distribution of the sum (and therefore of the mean) of a sample from an essentially arbitrary population is close to normal, if the sample is big enough.

Readings


Chapter 18. Hypothesis testing/1 – z test, introductory exercises (exercises)

exercise set A

1) You have a dice and want to know if it is fair, considering the average of the rolls. A hundred rolls are made and the sum of the rolls is 308. Is the dice loaded? or is it only chance variation?

2) You have a dice and want to know if it is fair; you have a suspicion that the aces may be a bit too frequent. The dice has been rolled 180 times; of the 180 rolls 42 have been aces. Is the dice loaded? or is it only chance variation?

3) Gamblers draw, with replacement, from a closed box in which the average of the numbers is said to be 1000, the SD of the numbers is 1500. (The gambler wins, in forints, the number he has drawn. The draws are to be paid for.) Two hundred draws have been observed and the mean of the draws have been only 820 (quite a bit less then the 1000 it should have been).

Could it be chance variation? Or is something wrong with the box?

4) A European roulette is tested: are red numbers too frequent? (There are 37 slots /numbers/ on the roulette wheel, each having equal chances. Of these 18 are red.) 370 games have been observed; 195 were red. Can it be chance variation? Is something wrong with the roulette?

exercise set B

1) You have a dice and want to know if it is fair – considering the average of the rolls. A hundred rolls are made – the sum of the rolls happens to be 308. (Is the dice loaded? or is it chance variation?)

– What does the null hypothesis say? What does the alternative hypothesis say?

– What test statistic would you use? (The test statistic is the number you observe to make your decision: the random variable you look at deciding if the dice is fair.)

– Find the expected value of the test statistic, given the null hypothesis is true. (The null hypothesis is the one stating that the difference is only chance variation, that there is nothing wrong with the dice; the assumption used in computing the P-value.)

– Find the S.E. of the test statistic, given the null hypothesis is true.

– Deviations of what kind make an evidence for the alternative? deviations upwards? deviations downwards? both? (What makes an evidence for the alternative hypothesis? Only if the test statistic is higher than the expected value? Or only if the test statistic is lower than the expected value? Or both? The alternative hypothesis is the one stating that the difference is real and can not be explained with chance variation; that something is wrong. The "so large a difference" the P-value is the probability of, always means a difference that is an evidence for the alternative.)

– Find the distribution of the test statistic, given the null hypothesis is true.

– What size is the difference between the observed value and the expected value (of the test statistic)? How many S.E.s?

2–4) Exercises 2–4 of set A with these questions added:

– what does the null say? what does the alternative say?

– what test statistic would you use?
– find the expected value of this test statistic, given the null is true.

– find the S.E. of the test statistic, given the null is true.

– deviations of what kind make an evidence for the alternative? deviations upwards? deviations downwards? both?

– find the distribution of the test statistic, given the null is true.

– what size is the difference between the observed value and the expected value? how many S.E.s?

e**xercise set C**

1a) A dice is tested if it is fair, considering the average of the numbers rolled. One hundred rolls are made, the sum of the rolls is 325. Can it be chance variation? Or is the dice loaded?

1b) A dice is tested if it is fair, considering the average of the numbers rolled. One hundred rolls are made, the sum of the rolls is 320. Can it be chance variation? Or is the dice loaded?

1c) A dice is tested if it is fair, considering the average of the numbers rolled. One hundred rolls are made, the sum of the rolls is 308. Can it be chance variation? Or is the dice loaded?

1d) A dice is tested if it is fair, considering the average of the numbers rolled. One hundred rolls are made, the sum of the rolls is 305. Can it be chance variation? Or is the dice loaded?

1e) What does the null state in the above examples?

1f) What does the alternative state in the above examples?

1g) Find the P-values for each.

1’a) A dice is tested whether it is fair, because of a suspicion that the numbers rolled are too low on the average. One hundred rolls are made, the sum of the rolls is 325. Can it be chance variation? Or is the dice loaded?

1’b) A dice is tested whether it is fair, because of a suspicion that the numbers rolled are too low on the average. One hundred rolls are made, the sum of the rolls is 320. Can it be chance variation? Or is the dice loaded?

1’c) A dice is tested whether it is fair, because of a suspicion that the numbers rolled are too low on the average. One hundred rolls are made, the sum of the rolls is 308. Can it be chance variation? Or is the dice loaded?

1’d) A dice is tested whether it is fair, because of a suspicion that the numbers rolled are too low on the average. One hundred rolls are made, the sum of the rolls is 305. Can it be chance variation? Or is the dice loaded?

1’e) What does the null state in the above examples?

1’f) What does the alternative state in the above examples?

1’g) Find the P-values for each.

2) You have a dice and want to know if it is fair or shows aces too frequently. 180 rolls are made;

a) of these 15 have been aces.

b) of these 38 have been aces.

c) of these 39 have been aces.

d) of these 42 have been aces.

e) of these 43 have been aces.

Can it be chance variation? Or is something wrong with the dice?

f) What does the null say in the above examples?
g) What does the alternative say in the above examples?

h) Find the $P$-values for a)...e) each.

2) You have a dice and want to know if it is fair, considering the chance of aces. (You take it to be wrong whether aces are too frequent or too rare.) 180 rolls are made:

a) of these 15 have been aces.
b) of these 38 have been aces.
c) of these 39 have been aces.
d) of these 42 have been aces.
e) of these 43 have been aces.

Can it be chance variation? Or is something wrong with the dice?

f) What does the null say in the above examples?
g) What does the alternative say in the above examples?
h) Find the $P$-values for a)...e) each.

3) Gamblers draw, with replacement, from a closed box in which the average of the numbers is said to be 1000, the SD of the numbers is 1500. (The gambler wins, in forints, the number he has drawn. Of course, the draws are to be paid for.)

Two hundred draws are observed and

a) the average of the draws is 900.
b) the average of the draws is 1100.
c) the average of the draws is 820.
d) the average of the draws is 790.
e) the average of the draws is 740

Can it be chance variation? Or is something wrong with the box?

f) What does the null say in the above examples?
g) What does the alternative say in the above examples?
h) Find the $P$-values for a)...e) each.

4) A European roulette is tested: are red numbers too frequent or too rare? (There are 37 slots /numbers/ on the roulette wheel, each having equal chances. Of these 18 are red.) 370 games are observed; of these

a) 170 are red
b) 190 are red
c) 198 are red
d) 157 are red

can it be chance variation? Is something wrong with the roulette?
e) What does the null say in the above examples?
f) What does the alternative say in the above examples?

g) Find the P-values for a)...d) each.

4') A European roulette is tested: are red numbers too frequent? (There are 37 slots /numbers/ on the roulette wheel, each having equal chances. Of these 18 are red.) 370 games are observed; of these

a) 170 are red
b) 190 are red
c) 198 are red
d) 157 are red

Can it be chance variation? Is something wrong with the roulette?

e) What does the null say in the above examples?

f) What does the alternative say in the above examples?

g) Find the P-values for a)...d) each.

exercise set D

1) Dice are tested whether the numbers rolled are too big on the average. 100 rolls are made with each. If a dice proves to be loaded on the (*) level (that is, with P<5%) it is to be destroyed.

a) This job being done by trained workers having no learning in statistics, instructions have to be exact and simple. It goes like this: "The dice with an average ________ is to be destroyed." (Fill in the blank. Options: (a) less than ....; (b) greater than ....; (c) less than .... or greater than ..... The than-numbers are to be specified, too.)

b) Assume the dice are all perfect. Is it possible, though, there being one or more among them proving to be loaded and therefore being destroyed?

c) If yes, what proportion of the dice is to be destroyed, approximately?

2) Dice are tested, considering the average of the numbers rolled. (Dice are taken to be faulty whether the average is too big or too low.) 100 rolls are made with each dice. If a dice proves to be loaded on the (*) level (that is, with P<5%) it is to be destroyed.

a) This job is made by trained workers, not learned in statistics, so the instructions have to be exact and simple. It says: "The dice with an average ________ is to be destroyed." (Fill in the blank. Options: (a) less than ....; (b) greater than ....; (c) less than .... or greater than ..... The than-numbers are to be specified, too.)

b) Assume the dice are all perfect. Is it possible, though, there being one or more among them proving to be loaded and therefore being destroyed?

c) If yes, what proportion of the dice is to be destroyed, approximately?

Readings

Chapter 19. Hypothesis testing/2 – one-sample z test (exercises)

1) 100 tosses are made with a coin; 62 are heads. What do you say, can it be chance variation? Or is something the matter with the coin?

2) Prepacked oranges are sold in a hypermarket. The average weight of the packs is said to be 1000 grams, the S.D. of the weights to be 50 grams. The consumer protection makes an examination, weighting 25 randomly selected packs of these oranges. The average of the 25 weights is 965 grams.

Does it prove that the packs weigh less than stated? or can it be chance variation? (The S.D in the sample was close to the stated population value.)

3) In exercise 2), if the weights of the packs were as stated, the distribution of the mean weight of 25 randomly selected packs (a random variable) would be close to a ______ distribution with E.V.______ and S.E.______. Then the chance of the mean being only 965 grams or less would be about______, that is, it would be ____ (more/less) than 5%. Therefore the difference ______(can/can not) be explained with chance variation.

4) Instead of the expected 100, only 81 of the first 600 rolls with a new dice have been aces. Does it prove that the dice is faulty? Or can the difference be explained with chance variation?

5) The proportion of black-haired people in a population was 20% five years ago. A sample of the population is now examined to see if the proportion has remained the same. The sample is a simple random sample of size n=1000.

A) Which is to be the null hypothesis, the change or the absence of change?

A1) If the sample proportion of black-haired people were much less than 20%, would it prove that their population proportion has decreased, too?

A2) If the sample proportion of black-haired people were exactly 20%, would it prove that their population proportion has remained exactly 20%, too?

B) The selection having been done, only 172 black-haired have been found in the sample, instead of the expected 200.

B1) Is this difference an evidence?

B2) ...an evidence for what?

B3) ...an evidence on what level?

6) The population proportion of illiterate people was 40% in Troland five years ago. The state television has started a campaign to popularize reading books. A sample survey is to be done to test whether the campaign has been effective.

A) Which is to be proved: (I) that the campaign has been effective or (II) that is has not been effective?

B) Which can be proved? (with the means of hypothesis testing)

B1) If the proportion of illiterate people in the sample were exactly 40%, would it prove that there has been no change in the population?

B2) If the proportion of illiterate people in the sample were much less than 40%, would it prove that there must have been some change in the population as well?

\footnote{the way the concept is used in hypothesis testing}
C) The research having been done, 360 illiterate people are found in the sample of 1000.

C1) Does it make an evidence?

C2) ...an evidence for what?

C3) ...an evidence on what level?

7) 100 tosses are made with a fair coin, the heads are counted. Find the chance that the number of heads exceeds 62.

a) The distribution of the number of heads is ______ . (Options: binomial / hypergeometric)

b) The number of tosses being large enough the distribution of the number of heads is close to a normal distribution with E.V. =______ and S.D. =_____.

c) therefore the chance that the the number of heads exceeds 62, is about ______.

8) It is is to be decided if a coin shows heads too often. 100 tosses are made; of these 62 are heads. The observed significance (that is, the P-value) of the difference is _____ therefore the coin is presumably ______ (fair/faulty).

Readings

Chapter 20. Hypothesis testing/3 – decision rules; one- and two-tailed tests; questions (exercises)

1) A and B discuss about a coin. A says it is fair, B says it is loaded. They make an experiment to decide the debate – of the 400 tosses made 225 is heads. Does it prove B’s statement? If yes – on what level? Find the P-value,

a) if B only says that the coin is loaded some way.

b) if B says that the coin shows disproportionately much heads.

c) if B says that the coin shows disproportionately few heads.

2) In testing a hypothesis the evidence proved to be on the (**) level. True or false: "this means that with 99% probability the alternative is true”.

3) In X county two years ago the proportion of those smoking among the adult population was 40%. A sample survey is conducted to discover if the two-years anti-smoking campaign has been effective.

"If the number of those smoking in the sample of 100 is less than ____ that is an evidence on the (*) level for the effectiveness of the campaign."

– Fill in the blank. Describe the steps.

– What does the null say?

– What does the alternative say?

4) Testing a hypothesis the evidence happened to be on the (*) level (more exactly, the P-value is 4%) with a one-tailed alternative hypothesis.

Had the alternative hypothesis been two-tailed, what level would the evidence be? Find the appropriate P-value.

5) In X county two years ago the proportion of those smoking among the adult population was 40%. A sample survey is conducted to discover if the two years anti-smoking campaign has been effective.

"If the number of those smoking in the sample of 1000 is less than ____ that is an evidence on the (*) level for the effectiveness of the campaign."

– (Fill in the blank.) Just describe the steps of filling in the blanks, it is not necessary to actually find the number.

– What does the null say?

– What does the alternative say?

6) In the above example, the null hypothesis is stating something about the proportion of smokers in the ____.

(options: sample / population)

7) True or false (decide for each):

a) a bigger difference means a stronger evidence.

b) a bigger difference means a smaller P-value.

c) a smaller P-value means a stronger evidence.

8/a) Difference between what is 7/a) and 7/b) about? (options: null / alternative)
Hypothesis testing/3 – decision rules; one- and two-tailed tests; questions (exercises)

b) What is the small P-value in 7/c) evidence for? (options: null / alternative)

9) Dice are tested by trained employees against the hypothesis that aces show too rarely. A die is discarded if the difference is statistically significant on the (*) level. 10,000 dice are to be tested this week. Assuming all 10,000 to be perfect, what is to be expected – will there be any that will be discarded?

10) An experiment has been made while testing a hypothesis with a one-tailed alternative. The difference proved to be statistically significant (P-value 1.75%). Find the P-value, assuming a two-tailed alternative.

11) True or false: “The P-value is the probability of the null hypothesis being true.”

B/1) A coin is tested whether it shows heads too often. The decision rule: 200 tosses are made and if the number of heads is 114 or more the coin is discarded.

a) what does the null say? what does the alternative say?

b) what is the experiment?

c) what random variable is observed? (what is the test statistic?)

d) assuming all coins tested being exactly fair, what do you say, will there be any discarded?

e) if yes, about what percentage would be discarded that way?

f) find the critical region; find the acceptance region.

g) find the probability of errors of type I.

B/2) In testing 10,000 coins happening to be exactly fair, the decision rule detailed in B/1) is applied. Find the expected value of the number of coins to be discarded.

B/3) A coin is tested whether it shows too much heads. The first 200 tosses are observed; 114 are heads. Is the coin loaded? Or is the difference chance variation? Find the P-value.

B/4) A coin is tested whether it shows too much heads. For the decision, the first 200 tosses are to be observed; if the number of heads is _____ or more that is evidence on the (*) level that the coin is loaded. (Fill in the blank.)

C/1) A coin is tested whether it is loaded. The decision rule: 200 tosses are made and if the number of heads is 114 or more, or the number of tails is 114 or more the coin is discarded.

a) what does the null say? what does the alternative say?

b) what is the experiment?

c) what random variable is observed? (what is the test statistic?)

d) assuming all coins tested being exactly fair, what do you say, will there be any discarded?

e) if yes, about what percentage would be discarded that way?

f) find the critical region; find the acceptance region.

g) find the probability of errors of type I.

C/2) In testing 10,000 coins happening to be exactly fair, the decision rule detailed in C/1) is applied. Find the expected value of the number of coins to be discarded.

C/3) A coin is tested whether it is loaded. The first 200 tosses are observed; 114 are heads. Is the coin loaded? Or is the difference chance variation? Find the P-value.

D/1) A fruit machine is tested by the consumer protection in respect of the chance of winning. Which is the more appropriate: a one-tailed test or a two-tailed test?
D/2) A fruit machine is tested by the mechanician of the producer in respect of the chance of winning (that is, whether the settings are all correct). Which is the more appropriate: a one-tailed test or a two-tailed test?

D/3) A fruit machine is tested by the owner wanting to know whether it has been tinkered with to provide the gamblers from his money with more than their due. Which is the more appropriate: a one-tailed test or a two-tailed test?

E/1) Dice are tested by trained employees against the hypothesis that aces show too rarely. A die is discarded if the difference is statistically significant on the (*) level. 10,000 dice are to be tested this week. Assuming all 10,000 to be perfect, what is to be expected – how many of the ten thousand will be discarded, approximately?

E/2) True or false: "The P-value shows the probability that the alternative hypothesis is true."

Readings

Chapter 21. Hypothesis testing/4 – e.g. probabilities of errors (exercises)

1) A coin is tested whether it is fair or shows heads too often. 100 tosses will be observed.
   a) what observed values would make evidence on the (*) level for the alternative?
   b) what observed values would make evidence on the (**) level for the alternative?
   c) what observed values would make evidence on the (***) level for the alternative?
   d) Which is the alternative: that the coin is fair or that it shows heads too often?

2) A coin is tested whether it is fair or shows heads too often. 100 tosses will be made. The decision rule is, if the number of heads is 55 or more (which means a deviation of 10% or more from the expected) then the coin is taken to be faulty, otherwise it is accepted to be fair.

   Find the probability of error of type I of this decision rule. (That is, find the chance that a fair coin is to be taken faulty, with this process.)

3) A coin is tested whether it is fair or shows heads too often. 100 tosses will be made. The decision rule is, if the number of heads is 55 or more (which means a deviation of 10% or more from the expected) then the coin is taken to be faulty, otherwise it is accepted to be fair. It is also known that the loaded coins in the shipment all have a probability of exactly 65% for tossing a head.

   Find the probability of error of type II of this decision rule. (That is, find the chance that a loaded coin is to be accepted fair, with this process.)

What does the null say? what does the alternative say now?

4) 100,000 coins are to be tested by well-trained workers applying the decision rule of exercise 2). Actually, the workers are being tested: the coins are all perfectly fair; it is the workers’ error rates the researchers are interested in.

   – what do you say, will there be any of the 100,000 coins to taken faulty?
   – if yes – about how many will there be such?

5) A coin is tested whether it is fair or loaded (either way). 100 tosses will be observed.
   a) What observed values would make evidence on the (*) level for the alternative?
   b) What observed values would make evidence on the (**) level for the alternative?
   c) What observed values would make evidence on the (***) level for the alternative?
   d) Which is the alternative: that the coin is fair or that it is loaded either way?

6) It is to be decided upon coin if it is fair. The probability of tossing a head is exactly 60% with the loaded coins. (They are the products of a factory having made wrist watches before watches went digital.)

   a) Testing to be based upon n=100 tosses:
      a1) a decision rule is needed with 5% probability of error of type I. Describe the decision process. Find the probability of error of type II.
      a2) a decision rule is needed with 1% probability of error of type I. Describe the decision process. Find the probability of error of type II.
Hypothesis testing/4 – e.g.

a3) a decision rule is needed with 0.1% probability of error of type I. Describe the decision process. Find the probability of error of type II.

a4) a decision rule is needed where the probabilities of error of type I and of type II are about equal. Describe the decision process. Find the probabilities of errors of type I and of type II.

b) Testing to be based upon n=400 tosses:

b1) a decision rule is needed with 5% probability of error of type I. Describe the decision process. Find the probability of error of type II.

b2) a decision rule is needed with 1% probability of error of type I. Describe the decision process. Find the probability of error of type II.

b3) a decision rule is needed with 0.1% probability of error of type I. Describe the decision process. Find the probability of error of type II.

b4) a decision rule is needed where the probabilities of error of type I and of type II are about equal. Describe the decision process. Find the probabilities of errors of type I and of type II.

c) Testing to be based upon n=900 tosses: a decision rule is needed where the probabilities of error of type I and of type II are about equal. Describe the decision process. Find the probabilities of errors of type I and of type II.

7) In Burgundy the threshold for a party to enter Parliament is 10%. The proportion of those in favour of the Black Party in the population must be somewhere around the threshold. The government wishes to time the general elections so that the Blacks would not get into Parliament. They are going to make a sample survey to see if it is the right time now. (The sample size is n=1600.)

a) What is it the government wants evidence for?

b) What shall the null be now? what shall the alternative be?

c) What do the hypotheses make their statements about? (options: population proportion of Black Party supporters / sample proportion of Black Party supporters)

d) Specify a decision rule. "If the proportion of Black Party supporters in the sample will be _____ (find an appropriate number) or ______ (options: more / less) that is an evidence on the (**) level that the blacks ______ (options: will / will not) get into Parliament".

8) In Trolland the threshold for a party to enter Parliament is 10%. The proportion of those in favour of the Black Party in the population must be somewhere around the threshold. The government wishes to time the general elections so that the Blacks would get into Parliament. They are going to make a sample survey to see if it is the right time now. (The sample size is n=1600.)

a) What is it the government wants evidence for?

b) What shall the null be now? what shall the alternative be?

c) What do the hypotheses make their statements about? (options: population proportion of Black Party supporters / sample proportion of Black Party supporters)

d) Specify a decision rule. "If the proportion of the Black Party supporters in the sample will be _____ (find an appropriate number) or ______ (options: more / less) that is an evidence on the (**) level that the blacks ______ (options: will / will not) get into Parliament".

10) Rolls of the Bread&Roll Ltd. weigh 70 grams on the average, according to specifications from the producer. (It is also known, from the small print, that the S.D of these weights is 10 grams.) The consumer protection is going to test on a sample of size n=25 whether the rolls are too small.

a) Find the sample mean constituting a (*) level evidence for the rolls being too light.

b) Find the sample mean constituting a (**) level evidence for the rolls being too light.
c) Find the sample mean constituting a (***) level evidence for the rolls being too light.

(The data for the S.D. has been checked and found o.k.)

11) Rolls of the Bread&Roll Ltd. weigh 70 grams on the average, according to specifications from the producer. (It is also known, from the small print, that the S.D of these weights is 10 grams.) Consumer protection is going to test on a sample of size n=25 whether the rolls are too small. The decision rule to be applied is as follows: if the average weight of the rolls in the sample is 65 grams or less, it is taken as evidence for the rolls being too light, therefore their informing the public prosecutor's office is deemed unavoidable.

Find the probability of error of type I for this decision process.

12) Assume the Bread&Roll Ltd. have been informed about the coming investigation of the consumer protection and adjusted their production line with the utmost precision to work according to specifications (that is, average=70 grams, SD=10 grams). Comes the consumer protection and executes the decision process specified in exercise 11.

a) Is it possible that the cheating (the rolls being too light) will be proven by consumer protection?

b) If so, what is the chance for this?

13) Rolls of the Bread&Roll Ltd. weigh 70 grams on the average, according to specifications from the producer. (It is also known, from the small print, that the S.D of these weights is 10 grams.) The factory control is going to test on a sample of size n=25 whether the rolls are of the right weight.

a) Find the sample mean constituting a (*) level evidence for the rolls not being of the right weight.

b) Find the sample mean constituting a (**) level evidence for the rolls not being of the right weight.

c) Find the sample mean constituting a (***) level evidence for the rolls not being of the right weight.

(The data for the S.D. has been checked and found o.k.)

14) A preliminary task from the winter exam of first year students (1,000 students) of the School for Statistical Quality Control is, as follows: the student gets a coin and, observing 50 tosses, has to decide on a (*) level whether it is fair (that is, he has to apply a decision process with a 5% probability of error of type I). If his decision is false (classifying a fair coin as loaded or a loaded one as fair) he is not allowed to take the exam this semester. The students are making the tests according to the rules; no one cheats. The coins are all perfectly fair.

a) Will there be any of the 1,000 students making a faulty decision (classifying his coin as loaded)?

b) If you say yes, could you tell how many of the students will have decided this way, approximately?

c) Find the chance that the number of students having such bad luck will be 35 or less.

15) This exercise is about weights of pre-packed ice creams. These are said to be 40 grams on the average with an S.D. of 10 grams. The consumer protection suspects that the packs are somewhat short of weight. To publish their findings they need evidence on the (**) level (otherwise they would risk a libel suit). They want to make their sample big enough to guarantee a chance of at least 95% for getting an evidence on the (**) level given their suspicion is right.

Find the necessary sample size if the suspected size of the shortage is

a) 4 grams per packs on the average.

b) 2 grams per packs on the average.

c) 1 grams per packs on the average.

15') This exercise is about weights of pre-packed ice creams. These are said to be 40 grams on the average with an S.D. of 10 grams. The consumer protection suspects that the packs are somewhat short of weight. To publish their findings they need evidence on the (**) level (otherwise they would risk a libel suit). They want to make
Find the necessary sample size if the suspected size of the shortage is

a) 4 grams per packs on the average.

b) 2 grams per packs on the average.

c) 1 grams per packs on the average.

16) This exercise is, among others, about weights of loaves of bread. The Troll Royal Chancellor is angry with the proprietor of the Crown Jewel Bakery; but Trolland is a constitutional state. So what else could he do, he makes the bakery checked very often to see if their breads are short of weight. Every day an inspector, sent to the bakery, selects a simple random sample of size 25 from the 1000-gram loaves and makes the testing. (One-sample z-test is applied assuming, rightly, that the S.D. of the weights is 45 grams.) Would an evidence on the (***) level been found on the shortage of weights one day the bakery would be closed up for a year that same day.

Daily testing is continuous for a year (meaning 365 days and 365 tests).

Assume the production line to be set exactly according to prescriptions, producing loaves of 1000 grams on the average, with an S.D. of 45 grams.

a) Is it yet possible that the bakery will be closed up some day during the year?

b) If yes – with what chance?

c) Find the chance that closing up of the bakery will be initiated during the first 30 days.

Readings


Chapter 22. Bivariate concepts (theory)

Two-dimensional vector variables

This handout is about how the interdependence of two random variables is to be dealt with.

The random variables \( X \) and \( Y \), defined with a common experiment, each take real numbers as their values. The pair of values taken by them can be described with a vector \((x;y)\), therefore the \((X;Y)\) variable pair can be thought of as a single variable taking two-dimensional vectors \((x;y)\) as values. In this context we speak of vector variables, in this case two-dimensional vector variables.

1. Joint distribution, marginal distributions, conditional distributions (discrete variables)

Joint distributions: Denote \( x_i \, \quad (i=1,...,I) \) the values taken by \( X \); denote \( y_j \, \quad (j=1,...,J) \) the values taken by \( Y \). Then the numbers \( p_{ij} (i=1,...,I; j=1,...,J) \) together, where \( p_{ij} = P(X=x_i \land Y=y_j) \), are called their joint distribution.

Obviously, \( p_{ij} \geq 0 \quad (i=1,...,I; j=1,...,J) \) and \( \sum_{i=1}^{I} p_{ij} = 1 \).

The joint distribution gives the chances for every possible combination of values for the variable pair \( X,Y \).

Marginal distributions:

Denote \( p_+ = \sum_{j=1}^{J} p_{ij} \), denote \( p_{+j} = \sum_{i=1}^{I} p_{ij} \);

then \( p_+ = P(X=x_i) \), similarly \( p_{+j} = P(Y=y_j) \),

that is, the numbers \( (p_+; i=1,...,I) \) give the (univariate) distribution of \( X \),

likewise the numbers \( (p_{+j}; j=1,...,J) \) give the distribution of \( Y \).

These are called marginal distributions, in this context.

Ignoring what values \( X \) (or \( Y \)) takes we get a unidimensional distribution: the marginal for \( Y \) (or, for \( X \)). Examining three variables, instead of the two, the joint distribution gives the probabilities of value triplets \((x,y,z)\). Therefore, ignoring the values taken by one of the three – say, by \( X \) – a two-dimensional marginal distribution would result, the joint distribution of \( Y \) and \( Z \), specifying event probabilities of the type \( P=(Y=y_j, \, Z=z_k) \). And, ignoring the values taken by two of the three – say, by \( X \) and \( Z \) – a unidimensional marginal distribution would result, the distribution of \( Y \), specifying event probabilities of the type \( P(Y=y_j) \).

Conditional distributions:

– the conditional distribution of variable \( X \) given that \( Y=y_j \) is given by \( P(X=x_i|Y=y_j) = \frac{p_{ij}}{p_{+j}} (i=1,...,I) \).

Denoted also as \( p_{i|j} \) (short, but may be ambiguous).

Obviously \( p_{i|j} \geq 0 \) and \( \sum_{j=1}^{J} p_{i|j} = 1 \); it is a distribution.

The conditional distribution of \( Y \) given that \( X=x_i \) is defined the same way.

The number of conditional distributions of \( X \), given \( Y \), equals the number of the values assumed by \( Y \): for every \( y_j \) there belongs a conditional distribution of \( X \).

Less exactly: we watch how \( X \) behaves when \( Y=y_i \); how \( X \) behaves when \( Y=y_i \); and so on, describing the "behavior" of \( X \) with the maximum possible resolution: its distribution.
Independence: the discrete random variables X and Y are independent, if

\[ p_{i,j} = p_{i,+} p_{+,j} \quad (i=1,..I; \ j=1,..J) \]

that is, if the joint distribution equals the product of the marginal distributions.

Not assuming independence, it follows from the multiplication rule for event probabilities that \( p_{i,j} = p_{i,+} p_{+,j} \). So, for independence to hold \( p_{i,j} = p_{i,+} \) is necessary for every pair \((i,j)\) \((i=1,..I;\ j=1,..J)\): the conditional distributions\(^1\) of X on Y all have to be equal with the marginal distribution of X. Therefore the variables X and Y are independent if and only if all conditional distributions of X on Y are identical, that is: if the behavior of X is exactly the same no matter what value Y assumes.

Similarly, X and Y are independent if and only if all conditional distributions of Y on X are identical: independence is a symmetric relation.

2. Joint cumulative distribution function, joint density, marginal densities, conditional densities (joint distribution; continuous variables)

Joint cumulative distribution function of the X, Y variable pair

\[ F_{X,Y}(x,y) = P(X < x \land Y < y) \]

– its properties:

a. \( 0 \leq F_{X,Y}(x,y) \leq 1 \)

b. \( \lim_{x \to +\infty} F_{X,Y}(x,y) = \lim_{y \to +\infty} F_{X,Y}(x,y) = 1 \)

c. \( \lim_{x \to -\infty} F_{X,Y}(x,y) = 0 \)

d. for every \( x \), \( F_{X,Y}(x,y) \) is nondecreasing and continuous from the left;

e. for every \( y \), \( F_{X,Y}(x,y) \) is nondecreasing and continuous from the left;

Marginal distribution functions:

As \( F_X(x) = P(X < x) = P(X < x \land Y < +\infty) \), so \( F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x,y) \)

this way the marginal cumulative distribution function of X could be obtained from the joint cumulative distribution function. (Similarly for Y.)

The marginal cumulative distribution function of X is identical to the univariate cumulative distribution function of X seen in handout (6).

Independence:

X and Y are independent \( \iff \{ F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \forall x,y \in \mathbb{R}^2 \} \),

that is, if the joint cumulative distribution function equals the product of the marginal cumulative distribution functions.

Explanation:

\( \implies \) : obviously, the product property follows from independence;

\( \impliedby \) : it will only be shown that from the product property it follows that the probability of every little \([x_1,x_2] \times [y_1,y_2]\) cell can be written as a product – which, in the case of discrete variables, would be equivalent with independence.

\(^1\)That is, the conditional distributions of X given that \( Y=y_j \), for \( j=1,..J \).
1. For an arbitrary X and Y,

\[ P(\{X \in [x_1, x_2] | Y \in [y_1, y_2]\}) = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - (F_{X,Y}(x_1, y_2) - F_{X,Y}(x_1, y_1)) \]

(Explain.)

2. If, further, \( F_{X,Y}(x, y) = F_X(x)F_Y(y) \), then the above equals

\[ \cdots = F_X(x_2)F_Y(y_2) - F_X(x_2)F_Y(y_1) - F_X(x_1)F_Y(y_2) + F_X(x_1)F_Y(y_1) = \frac{|F_{X,Y}(x_2) - F_{X,Y}(x_1)|F_Y(y_2) - F_Y(y_1)|}{P(X \in [x_1; x_2]) \cdot P(Y \in [y_1; y_2])} \]

Joint density (bivariate density): The function \( f_{X,Y}(x, y) : \mathbb{R}^2 \to \mathbb{R} \) is the joint density of the random vector variable \((X; Y)\) if, for any \( A \subset \mathbb{R}^2 - \text{re} \)

\[ P((X, Y) \in A) = \int_A f_{X,Y}(x, y) \text{d}A. \]

Vector variables having a joint density are called jointly continuous variables.

If there exists either \( x_0 \) such that \( P(X=x_0)>0 \), or \( y_0 \) such that \( P(Y=y_0)>0 \), or a pair \((x_0, y_0)\) such that \( P(X=x_0, Y=y_0)>0 \), then \((X; Y)\) cannot be jointly continuous. (Explain.)

Properties of joint densities:

a. \( f_{X,Y}(x, y) \geq 0 \)

b. \( \int f_{X,Y}(x, y) \text{d}y = 1 \)

Marginal densities: in the bivariate context, the univariate \(^{1}\) density \( f_x(x) \) of X is called the marginal density of X. It can be obtained, assuming \((X, Y)\) to be jointly continuous, by integrating the joint density with respect to y with x held constant:

\[ f_x(x) = \int f_{X,Y}(x, y) \text{d}y \]

Less exactly, taking the intersection, in 3 dimensions, of the joint density \( f_{X,Y}(x, y) \) with the plane \( x=x_0 \), the area of the intersection equals the value of the marginal (that is, univariate) density \( f_x(x) \) of X in \( x_0 \).

Like in the discrete case, summing across y-values marginals for X are obtained.

Conditional densities: as with conditional distributions in the discrete case, we want to know the behavior of Y, given \( X=x_0 \). In what way does the distribution of Y change when X assumes this, instead of that, value? We have to look at the intersections of the joint density with the planes \( x=x_0 \). More exactly, their shapes. But the areas of these intersections are not equal. (As we have seen, the area of such an intersection equals the density of X in \( x_0 \).) As a density must have an area of 1, the conditional densities are obtained transforming these intersections to an area of 1:

\[ f_{Y/X=x_0}(y) = \frac{f_{X,Y}(x_0, y)}{f_X(x_0)} \]

defined in x only if \( f_X(x) > 0 \).

Consequence: obviously

\[ f_{X,Y}(x, y) = f_X(x) \cdot f_{Y/X=x_0}(y) \quad ('multiplication\ rule') \]

Independence: assuming \((X; Y)\) to be jointly continuous

\((X \text{ and } Y \text{ are independent}) \iff (f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)) \)

\(^{1}\)indeed, only for sets not too "ugly", e.g. intervals, unions of a finite set of intervals, and so on (the so-called Borel-sets; see measure theory).

\(^{2}\)e.g. see handouts 6, 7 and 9.
Bivariate concepts

Explanation: Y being independent of X means that the conditional densities \( f_{Y|X=x} \) of Y are all identical, independent of the value x (all being therefore identical to the marginal \( f_Y(y) \) of Y). So it follows, according to the 'multiplication rule', that the joint density equals the product of this common conditional density and of the marginal density \( f_X(x) \) and backwards, if the joint density can be obtained as the product of the two marginal densities, \( f_{X,Y} = f_X f_Y \), then, obviously, the conditional densities \( f_{Y|X=x} \) all have to be identical, independent of x.

3. Conditional expected value

As has been the case with conditional distributions and conditional densities, once more the question is the behavior of Y given that X equals some value – but with a lower resolution. The 'behavior' of Y is now summarized into one single number: its expected value.

**Conditional expected value, discrete distributions:**

\[
E(Y|X = x_i) = \sum_{j=1}^{d} P_{ij} y_j = \sum_{j=1}^{d} \frac{P_{ij}}{P_X(x_i)} y_j
\]  

(22.1)

according to its name, the conditional expected value of Y shows the expected value of Y, given that X=x_i.

Also: conditional expectation.

statement: (an analogue of the total probability formula):

\[
E(Y) = \sum_{i=1}^d E(Y|X = x_i)P(X = x_i)
\]  

(22.2)

The expected value of Y can be computed calculating the conditional expectations \( E(Y|X=x_i) \) of Y on X first, and then calculating the average of these, weighted according to their respective probabilities \( P(X=x_i) \) : that is, computing the expected value of the conditional expected values. (Explain.)

**Conditional probability as a function, and as a random variable:**

The conditional expectation defined above is, in fact, a function, mapping the values \( x_i \) to the conditional expectations \( E(Y|X=x) \). Yet it is a random variable, too: it depends on chance what value X is assuming (X is a random variable), so it depends on chance, too, what the expected value of Y given X is.

Denote \( h(x) = E(Y|X=x) \) the conditional expectation of Y given X as a function; then the random variable we are talking about is \( E(Y|X) = h(X) \).

**Example 22.1.**

three cards, numbered 1, 1 and 2, in a box; two draws are made from the box, without replacement. Denote X the first draw, Y the second draw. Given that the first draw is 1, the expected value of the second equals 1.5. Given that the first draw is 2, the expected value of the second equals 1. As a function, the conditional expectation \( E(Y|X) \) maps \( x=1 \) to \( h(x)=1.5 \); and maps \( x=2 \) to \( h(x)=1 \).

As a random variable the conditional expectation \( E(Y|X) \) assumes the values 1.5 and 1 with probabilities 2/3 and 1/3, respectively. This random variable is characterized by its extremely strong \( ^4 \) dependence on X as well.

Given that X=1 its value equals 1.5, with certainty; given that X=2 its value equals 1, with certainty.

Follows:

A reformulating:

\[
E(Y) = E( E(Y|X) )
\]

that is: the expected value of the conditional expectations of Y is the expected value of Y. (The average of the group means weighted with group weights, gives the total mean.)

\( ^4 \)This dependence is deterministic, not stochastic. Knowing the value of X, the value of \( E(Y|X) \) is known with certainty.
Conditional expected value, jointly continuous distributions:

\[
E(Y|X = x_0) = \int y \cdot f_{Y|X=x_0}(y) dy = \frac{\int y \cdot f_{XY}(x_0, y) dy}{f_X(x_0)}
\]  
(22.3)

(This is the expected value of the conditional density of Y given that \(X=x_0\)).

statement (an analogue of the total probability formula)

\[
E(Y) = \int y f_Y(y) dy = \int y \left[ \int f_{X,Y}(x,y) dx \right] dy = \int \left[ \int y f_{X,Y}(x,y) dy \right] dx = E(E(Y|X))
\]  
(22.4)

explanation:

1st equality: per definition

2nd equality: marginal density expressed via joint density

3rd equality: exchanging \(dx dy\) to \(dy dx\) in the integral. Doing the integrating with respect to \(y\) on the right side, the conditional expectation \(E(Y | X=x)\) is almost obtained only the weighting has been with the "intersection" instead of the conditional density. So the expression \(\int y f_{X,Y}(x,y) dy\) equals the conditional expectation multiplied with \(f_X(x)\) (see 2.2.3). From this follows the

4th equality.

5th equality: see the equation defining \(E(g(X))\) on page 3 of handout (6), with \(g(x) = E(Y | X=x)\) here.

That is, the expected value equals the average of the conditional expectations weighted with probabilities (densities) belonging to them.

**conditional expectations and independence:**

statement: \(X\) and \(Y\) are independent \(\rightarrow E(Y | X)\) is a constant (and \(E(X | Y)\) is a constant, too) (Explain.)

It follows, that, if \(X\) and \(Y\) are independent then \(E(Y | X) = E(Y)\) and \(E(X | Y) = E(X)\).

remark: the reverse is false. Even if both \(E(Y | X)\) and \(E(X | Y)\) are constant, \(X\) and \(Y\) may be dependent.

Example: consider a random vector variable \((X;Y)\) assuming the values \((1;0), (0;1), (-1;0)\) and \((0;-1)\) with equal probabilities.

statement: from \(E(Y | X)\) it does not follow that \(E(X | Y)\) is a constant, too.

Example: consider a variable \((X;Y)\) assuming the value \((1;0)\) with probability 0.50 and the values \((2;-1)\) and \((2;1)\) with probabilities 0.25 each.

Unlike independence, the property of one variable having a constant conditional expectation on the other is not symmetric.

**4. Conditional variance:**

Once again, the question is the behavior of \(Y\) given that \(X\) equals some value – considering now how the variance of \(Y\) changes with \(X\) changing.

Let \((X;Y)\) be a bivariate random vector variable, and \(Z=(Y-E(Y | X))^2\), then

\[
\text{var}(Y | X) := E(Z | X)
\]

With a denotation more redundant but perhaps more easily understood:
Denote \( m, m_x \) the conditional expectation of \( Y \), given that \( X=x \); that is, \( m_x = E(Y \mid X=x) \). Then
\[
\text{var}(Y \mid X=x) := E( (Y-m_x)^2 \mid X=x )
\]
statement:
\[
\text{var}(Y) = E( \text{var}(Y \mid X) ) + \text{var}( E(Y \mid X) )
\]
(Explanation later, at 6.3)

The total variance of a variable \( Y \) can be decomposed into two parts, one being the average of the variances inside groups defined by \( X \) (the so-called within groups component of the total variance), the other being the variance of the group averages (the so-called between groups component). (Both the average with the within groups component and the variance with the between groups component are weighted, according to the sizes – probabilities – of the groups.)

**conditional variances and independence:** if \( X \) and \( Y \) are independent then \( \text{var}(Y \mid X) \) is a constant (and \( \text{var}(X \mid Y) \) is a constant, too) (Explain.)

It follows, that, if \( X \) and \( Y \) are independent then \( \text{var}(Y \mid X) = \text{var}(Y) \) and \( \text{var}(X \mid Y) = \text{var}(X) \).

(Explain.)

### 5. The expected value of a function of a random vector variable

Let \( g: \mathbb{R}^2 \rightarrow \mathbb{R} \) be a function mapping pairs of real numbers to single real numbers; let \( (X,Y) \) be a jointly continuous bivariate random vector variable with joint density \( f_{X,Y}(x,y) \). Then
\[
E(g(X,Y)) = \int_{\mathbb{R}^2} g(x,y) \cdot f_{X,Y}(x,y) \, dx \, dy \tag{22.5}
\]
Like in the univariate case, the density \( f_{X,Y}(x,y) \) can be thought of as a weighting function defining weights for the averaging of the function \( g(x,y) \).

Accordingly,

- the expected value of \( X \) is obtained, applying formula (5.1) with \( g(x,y) \) here being \( g(x,y)=x \),
  \[
  E(X) = \int_{\mathbb{R}^2} x \cdot f_{X,Y}(x,y) \, dx \, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_X(x) \, dx \tag{22.6}
  \]
  \[
  E(X^2) = \int_{\mathbb{R}^2} x^2 \cdot f_{X,Y}(x,y) \, dx \, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 f_X(x) \, dx \tag{22.7}
  \]

- The expected value of the product of two variables \( X, Y \) is:
  \[
  E(XY) = \int_{\mathbb{R}^2} xy \cdot f_{X,Y}(x,y) \, dx \, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f_X(x) \, dx \, dy \tag{22.8}
  \]
  \[
  f_{Y \mid X=x} \cdot f_X(x) \, dx \, dy = \int_{x} \frac{x}{x} \cdot f_X(x) \cdot E(Y \mid X=x) \, dx = \int_{-\infty}^{\infty} x \cdot E(Y \mid X=x) \, dx
  \]
  \[
  = E(X \cdot E(Y \mid X))
  \]
  **Once again (5.1) has been applied, \( g(x,y) \) being here \( g(x,y)=xy \).**

A statement more or less analogue to the total probability formula results:
Bivariate concepts (theory)

\[ E(XY) = E(X \cdot E(Y | X)) = E(Y \cdot E(X | Y)) \]  
\[ (5.4) \]

6. Covariance

1. def.: \[ \text{cov}(X,Y) = E ( (X-E(X)) \cdot (Y-E(Y)) ) \]

Covariance shows how much X and Y vary in accord: whether, when one is large compared to its expected value, the other is large, too: whether they deviate at the same times and in the same directions, compared to their respective averages. If their deviations mostly take place synchronously and in the same directions – if when X is large, Y is large, too – then the products averaged here tend to be mostly positive so their average will be positive, too. If the deviations take place more or less synchronously, but mostly in opposite directions – if when X is large, Y is small and vice versa – then the products averaged here tend to be mostly negative, so the average will be negative, too.

2. statement, an alternative formula for the covariance:

\[ \text{cov}(X,Y)=E(XY)-E(X)E(Y) \]

proof: \[ \text{cov}(X,Y) = E ( (X-E(X)) \cdot (Y-E(Y)) ) = E ( XY - X \cdot E(Y) - Y \cdot E(X) + E(X)E(Y) ) = E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) = E(XY) - E(X)E(Y) \]

3. covariance and independence:

3.1 statement: if X and Y are independent \( \implies \text{cov}(X,Y)=0 \)

proof: if X and Y are independent then \( E(Y | X) = E(Y) \). Besides, \( E(XY) = E(X \cdot E(Y | X)) \) (see 5.4). So, if X and Y are independent then \( E(XY) = E(X \cdot E(Y | X)) = E(X \cdot \cdot E(Y)) = E(Y) \cdot E(X) \)

3.2, statement: if \( E(Y | X) = E(Y) \) then \( \text{cov}(X,Y)=0 \).

(Similarly, if \( E(X | Y) = E(X) \).)

Proof goes like proof of 6.3.1 having used only that \( E(Y | X) = E(Y) \) if independence holds.

3.3, statement: from \( \text{cov}(X,Y)=0 \), it does not follow that X and Y be independent.

Example: consider a random vector variable \((X,Y)\) assuming the values \((-1;0), (0;1), (1;0)\) and \((0;-1)\) with equal probabilities of 0.25.

(To be proven: (a) that \( \text{cov}(X,Y)=0 \), and (b) that X and Y are not independent.)

3.4, statement: from \( \text{cov}(X,Y)=0 \) it does not follow that either \( E(Y | X) = E(Y) \) or that \( E(X | Y) = E(X) \)

Example: consider a vector variable \((X,Y)\) assuming the values \((-1;0), (0;1), (0;-1), (1;0)\) and \((2;0)\) with equal probabilities of 0.20.

(To be proven: (a) that \( \text{cov}(X,Y)=0 \), and (b) that neither \( E(Y|X) \) nor \( E(X|Y) \) are constant.)

So, being uncovariated or uncorrelated (that is, if the covariance of two variables equals 0), though in a sense it means something like independence, is a weaker statement than the constancy of the conditional expectations, which, in turn, is weaker than independence.

4. some properties of covariance:

(covariance as a symmetric, positive definite, bilinear function)

(0) \( \text{cov} (X,c)=0 \)

(1) \( \text{cov}(X;Y)=\text{cov}(Y;X) \)

(2a) \( \text{cov}(\alpha \cdot X;Y) = \alpha \cdot \text{cov}(X;Y) \)
(2b) \( \text{cov}(X; a \cdot Y) = a \cdot \text{cov}(X; Y) \)

(3a) \( \text{cov} (X_r + X_s; Y) = \text{cov}(X_r; Y) + \text{cov}(X_s; Y) \)

(3b) \( \text{cov} (X; Y_r + Y_s) = \text{cov}(X; Y_r) + \text{cov}(X; Y_s) \)

ált.:

(4a) \( \text{cov}(\sum \alpha_i X_i; Y) = \sum \alpha_i \text{cov}(X_i; Y) \)

and

(4b) \( \text{cov}(X; \sum \alpha_i Y_i) = \sum \alpha_i \text{cov}(X; Y_i) \)

(5) \( \text{cov}(X; X) \geq 0 \)

and \( (\text{cov}(X; X) = 0 \iff X \equiv c ) \)

(Explain.)

Remark: obviously,

(6) \( \text{cov}(X; X) = \text{var}(X) \).

5, some consequences:

(1) \( \text{cov} (\alpha X + \beta_1 Y + \gamma_1; \alpha X + \beta_2 Y + \gamma_2) = \text{cov}(X; Y) \)

(2) \( \text{var} (\alpha X + \beta Y + \gamma) = \alpha^2 \text{var}(X) + \beta^2 \text{var}(Y) + 2\alpha \beta \text{cov}(X; Y) \)

or

(3) \( \text{cov}(\sum \alpha_i X_i; \sum \beta_j Y_j) = \sum \alpha_i \beta_j \text{cov}(X_i; Y_j) \)

and

(4) \( \text{var}(\sum \alpha_i X_i + \sum \beta_j X_j) = \sum \alpha_i^2 \text{var}(X_i) + \sum \beta_j^2 \text{var}(X_j) \)

6, some more identities:

(1) \( E(h(X; Y)) = E ( E(h(X,Y) \mid X ) ) \)

– explain (double integral)

(2) \( E(Y) = E ( E(Y \mid X ) ) \)

- explain (simple; almost done at (3.2.1))

(3) \( \text{var}(Y) = E ( \text{var}(Y \mid X ) ) + \text{var} ( E(Y \mid X ) ) \)

(4) \( \text{cov}(X; Y) = \text{cov} ( X; E(Y \mid X ) ) \)

(3) proof of:

\[
\text{var}(Y) = E [ (Y - E(Y))^2 ] = \int \int (y - E(Y))^2 \cdot f_{XY}(x, y) = \int \int (y - E(Y))^2 \cdot f_{X}(x) \cdot (y - E(Y))^2 \cdot f_{Y|X=y}(y) \, dy \, dx = \int f_{x}(x) \cdot \left( \text{var}(Y \mid X = x) + (E(Y \mid X = x) - E(Y))^2 \right) \, dx = E(\text{var}(Y \mid X)) + E[(E(Y \mid X) - E(Y))^2]
\]

where

• equality before 2nd row (4th equality) is based upon (2.2.3), with the joint density expressed;

• equality before 3rd row (5th equality) is based upon Steiner's identity\(^{5}\) applied to the distribution of Y on condition X=x, with E(Y) as the A in the Steiner.

\(^{5}\)see handout (22.a)
Exercise: prove (4).

(Hint: formula $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$ might be a good starting point;

- it is to be shown that the $E(X)E(Y)$ parts are equal,
- it is to be shown that the $E(XY)$ parts are equal – see also 5.4.)

7, two more properties of the covariance:

7.1, $| \text{cov}(X;Y) | \leq (\text{var}(X) + \text{var}(Y)) / 2$

proof: $0 \leq \text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X;Y)$

7.2: $|\text{cov}(X;Y)| \leq \sqrt{\text{var}(X) \cdot \text{var}(Y)}$

/explanation:

a. subtracting a constant from $X$ (or from $Y$) does not change neither variance nor covariance, so both $X$ and $Y$

   may be thought of as having $0$ expected values.

b. consider two random variables with finite ranges; their joint distribution can be modelled with a box with cards, the cards all having two numbers (the value of $X$ written on the top, the value of $Y$ written on the bottom of the card) so that the proportions of the pairs in the box correspond to the joint distribution. Then $\text{var}(X) = \frac{1}{n} \sum_{i=1}^{n} x_i^2$, $\text{var}(y) = \frac{1}{n} \sum_{i=1}^{n} y_i^2$, $\text{cov}(X;Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$. So it is to be shown that

$$|\sum_{i,j} r_{i,j}| \leq \sqrt{\sum_i^2 \sum_j^2}$$

- squared:

$$|\sum_{i,j} r_{i,j}| \leq \sqrt{\sum_i^2 \sum_j^2}$$

- handling $i=j$ and $i<>j$ pairs separately, then carrying the left to the right,

we obtain that

$$0 \leq \sum_{i,j} (x_i^2 + y_j^2 - 2r_{i,j}x_iy_j)$$

is to be shown. The right side equals

$$- \sum_{i,j} (x_iy_j - r_{i,j})^2$$

which is a sum of squares, therefore cannot be negative. (Finished.)

The absolute value of the covariance of two variables is less than either the arithmetic or the geometric mean of

the two variances.

7.3: the correlation coefficient (Pearson’s correlation):

def: $\text{corr}(X;Y) := \frac{\text{cov}(X;Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$

Denoted also $r(X;Y)$

statement: $-1 \leq \text{corr}(X;Y) \leq 1$

proof: 7.2

One of the most widely used measures to show how closely two variables on the interval measurement level are

interrelated. It takes its values from the interval $[-1 \leq 1]$.

if $\text{corr}(X;Y)=1$ or $\text{corr}(X;Y)=-1$ it signifies a linear relationship $X = aY + b$ between the two variables, with $a<>0$

. ($a>0$ when $\text{corr}(X;Y)=1$, $a<0$ when $\text{corr}(X;Y)=-1$).

---

*that is, having $X$ substituted with $X'=X-E(X)$, $Y$ with $Y'=X-E(Y)$, variances and covariance remain unchanged, $\text{var}(X)=\text{var}(X')$, $\text{var}(Y)=\text{var}(Y')$, $\text{cov}(X,Y)=\text{cov}(X',Y')$
7. Variance in a direction; the covariance matrix

Considering the random variable \((X;Y)\) as a variable assuming its values from the plane \( \mathbb{R}^2 \), it is not evident why it is to be characterized with just the variance of its 'shadows' on the x-axis and on the y-axis (that is, the variance of its marginal distributions). There are many possibilities for coordinating a vector space even if the axes are assumed orthogonal. Let us see how the variance of the shadow (that is, the projection) on a given axis of a vector variable can be calculated.

X and Y can be assumed such that \(E(X)=E(Y)=0\), as it does not change neither the variances nor the covariance of X and Y. (See 6.7.2(a) and footnote 6).

Denote \(\alpha\) the angle between the x-axis and the axis \(e\) to be projected upon; then the projection of a given \((x;y)\) value of the variable \((X,Y)\) on axis \(e\) is \(x \cos \alpha + y \sin \alpha\), so the variance we are looking for is the variance of the variable \(Z := \cos \alpha X + \sin \alpha Y\).

By (5.2),
\[
\text{var}(Z) = \cos^2 \alpha \cdot \text{var}(X) + \sin^2 \alpha \cdot \text{var}(Y) + 2 \sin \alpha \cos \alpha \cdot \text{cov}(X;Y)
\]

Rephrased: denote \(\mathbf{u} = \begin{bmatrix} \cos \alpha \\ \sin \alpha \end{bmatrix}\) the unit-length direction vector of axis \(e\); and denote \(D\) the 2x2 matrix with \(\text{var}(X)\) and \(\text{var}(Y)\) in its main diagonal and \(\text{cov}(X;Y)\) in the remaining two positions:
\[
D = \begin{bmatrix}
\text{var}(X) & \text{cov}(X;Y) \\
\text{cov}(X;Y) & \text{var}(Y)
\end{bmatrix}
\]

Then
\[
\text{var}(Z) = \mathbf{u}^T D \mathbf{u}
\]

where \(Z = \mathbf{u}^T \begin{bmatrix} X \\ Y \end{bmatrix}\).

This way the variance of \((X;Y)\) in any direction can be obtained applying the covariance matrix.

Def.: The above \(D\) matrix is the covariance matrix of the random vector variable \((X;Y)\).

8. Supplement – variance, covariance, correlation, and geometry

Consider a finite box model for the joint distribution of \((X_1, X_2, \ldots, X_n)\): \(N\) cards in a box, \(n\) numbers on each card (the values of \((X_1, X_2, \ldots, X_n)\) for that case), proportions of cards according to proportions by the joint distribution. Assume that \(E(X_1)=E(X_2)=\ldots=E(X_n)=0\).

Let it be represented in the \(N\)-dimensional space, the so-called variable space:

the variable \(X_i\) is represented by the vector whose
1st coordinate=\(X_i\), \(j\)th coordinate equals the value of \(X_j\) in the \(j\)th case (on the \(j\)th card), and so on,

so the \(n\) variables are represented by \(n\) points in the \(N\)-dimensional space (or, by \(n\) positional vectors running from the origo to the \(n\) points, respectively.)

Then
\[
\text{var}(X_i) = \frac{1}{N} \sum_{i} X_i^2 = \frac{1}{N} \cdot |X_i|^2
\]
\[
\text{cov}(X_i,X_j) = \frac{1}{N} \sum_{i} X_{i}X_{j} = \frac{1}{N} \cdot (X_i,X_j)
\]
Bivariate concepts (theory)

and

\[\text{corr}(X, Y) = \frac{\mathbb{E}(X, Y) - \mathbb{E}(X) \mathbb{E}(Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{\langle X, Y \rangle}{\sqrt{|X| \cdot |Y|}} = \cos(X, Y)\]

that is,

- the standard deviation more or less corresponds to the length of a vector,
- the variance more or less corresponds to the squared length of a vector,
- the covariance more or less corresponds to the scalar product of the two vectors, and
- the correlation coefficient corresponds to the cosine of the angle between the two vectors; so
- being uncorrelated corresponds to the orthogonality of the two vectors.

Question: what does the equality in the proof of 7.1 correspond to?

9. Bivariate joint normal distributions

1: the bivariate standard normal distribution:

1.1: the distribution of the bivariate random variable with component variables X and Y is bivariate standard normal, if it has a joint density

\[f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\]

Let X and Y both be univariate standard normal variables and let them be independent. Then these two, taken together as components of a vector variable, make a bivariate standard normal variable. (As X and Y are independent, their two-dimensional, joint density can be obtained as the product of their respective univariate densities.) Then

- the conditional densities (or the intersections of the joint density with planes of the kind x=x_0, or with planes of the kind y=y_0) are all univariate standard normal densities;
- the joint density is, as if a univariate normal density would have been rotated around the z-axis;

(Explanation: the nominator in the exponent is the distance squared of point (x;y) from the origin, so the points with a common value of the joint distribution fulfil the equality \(x^2+y^2=c\) with some non-negative \(c\) – these points are on a circle with center \((x=0,y=0)\). Therefore the level sets of the joint distribution are circle-shaped with center \((x=0,y=0)\).)

- the level sets of the joint distribution are concentric circles;
- the covariance matrix of the joint distribution is \(D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\); the two-dimension identity matrix.

1.2: Likewise, the joint distribution the component variables \(X_1,X_2,...,X_n\) make is an n-dimensional normal distribution if the \(X_i\)s are independent, univariate standard normal variables.

(The k-dimensional conditional distributions will be k-dimensional standard normal distributions here, the level sets n-dimensional spheres.)

2: bivariate normal distributions:

\footnote{more exactly, the 3-dimensional graph of the joint density is such, as if the 2-dimensional graph of a univariate normal density with zero expected value would have been rotated around the z-axis.}
Bivariate concepts (theory)

2.1.1, their derivation:

A bivariate normal distribution is obtained, making a linear transformation $A$ on the bivariate standard normal variable $Z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$. The distribution of the resulting random vector variable $AZ = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is called bivariate normal.

Less exactly (an example):

1. 1st step: make the vector variable extend 3-folds in the $x$-direction and compress 2-folds in the $y$-direction (that is, its $x$-coordinate is multiplied by 3, its $y$-coordinate is multiplied by 0.50). The density of the resulting vector variable is no longer rotationally symmetric: the graph is elongated in the $x$ direction. The level sets will be concentric ellipses with their center in the origin, their length of their long axis is 6 times the length of the short axis; the long axes lie on the $x$-axis of the two-dimensional plane; the short axes on the $y$-axis.

2. 2nd step: rotate the plane e.g. to the left with angle $\alpha$ – the joint density rotates with it, so the elongation now is not parallel with the $x$-axis but oblique; the long axes of the level-set ellipses likewise are making an angle of $\alpha$ with the $x$-axis, the short axes are making an angle of $\alpha$ with the $y$-axis.

The first step corresponds to a multiplication with the matrix $A_1 = \begin{bmatrix} 3 & 0 \\ 0 & 0.5 \end{bmatrix}$; the second step corresponds to a multiplication with the matrix $A_2 = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$. The two, taken together, corresponds to a multiplication with the matrix $A = A_2A_1 = \begin{bmatrix} 3\cos(\alpha) & -0.5\sin(\alpha) \\ 0.5\cos(\alpha) & 0.5\sin(\alpha) \end{bmatrix}$.

Remark: two-dimensional normal distributions are obtained this way only if the linear transformation $A$ is such that does not map the two-dimensional space into one dimension. (That is, linear transformations with nonzero determinants, transformations with rank 2, regular transformations).

E.g., the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ would map every point $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ onto the $x$-axis;

$A = \begin{bmatrix} 0 & 6 \\ 2 & 1 \end{bmatrix}$ would map every point $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ onto the $y$-axis;

and a multiplication with the matrix $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ maps every point to points with coordinates $x=y$, so these points all lie on the line $x=y$, that is, in one dimension.

If these transformations were used in the derivation $AZ=X$ of the vector variable $X$, all values of $X$ would lie on one single line, that is, in one dimension, so the distribution of $X$ would not be a bivariate (two-dimensional) normal.

2.1.2: statements 1) – conditional distributions, conditional expectations and conditional variances with bivariate normal distributions:

With a bivariate normal distribution,

a. all conditional distributions are univariate normal distributions;

b. all marginal distributions are univariate normal distributions;

c. the level sets are concentric ellipses;

further:

(d) all conditional distributions $(Y|X=x)$ have the same variance, that is, the conditional variances $\text{var}(Y|X=x)$ are all equal regardless of $x$;

(d”) all conditional distributions $(X|Y=y)$ have the same variance, that is, the conditional variances $\text{var}(X|Y=y)$ are all equal regardless of $y$, azonosak;

(e’) the conditional expectation function $E(Y|X=x)$ is a linear function of $x$;

(e”) the conditional expectation function $E(X|Y=y)$ is a linear function of $y$. 

Created by XMLmind XSL-FO Converter.
The slope of the conditional expectation functions:

\[(f') \text{ the slope of } E(Y|X) \text{ is } \frac{\text{cov}(X,Y)}{\text{cov}(X)} \frac{D(Y)}{D(X)} \]

\[(f'') \text{ the slope of } E(X|Y) \text{ is } \frac{\text{cov}(X,Y)}{\text{var}(Y)} \frac{D(X)}{D(Y)} \]

Consequence.1:

From the above (a),(d),(e) and (f) it follows that if X and Y are the component variables of a bivariate normal distribution then

- X and Y are independent \(\iff\) \(E(Y|X)\) is constant \(\iff\) \(\text{cov}(X,Y)=0\),

and

- X and Y are independent \(\iff\) \(E(X|Y)\) is constant \(\iff\) \(\text{cov}(X,Y)=0\).

The conditional distributions \(Y|X=x\) are all univariate normal distributions with equal variances, they can differ in their expected values only. These expected values – the values \(E(Y|X=x)\) – all lie on a linear function of \(x\), on a line \(ax+b\). If the covariance equals zero

- \(\Rightarrow\) the slope of the line \(ax+b\) equals zero

- \(\Rightarrow\) the conditional expectations \(E(Y|X=x)\) are all equal

- \(\Rightarrow\) the conditional distributions \(Y|X=x\) are all identical.

In short, with joint normal distributions, the uncorrelatedness of the component variables is a sufficient condition for their independence.

Consequence.2: (e’) and (e”) mean that with joint normal distributions the conditional expectation function \(E(Y|X)=x\) and \(E(X|Y)=y\) can be found among the linear functions: that is, with linear regression.

Remark: the line of the conditional expectation function (the line of the linear regression) is not identical to the line of longer axis of the level sets – slants less.

The regression line estimating \(y\) from \(x\)-values lies between the line of the long axes and the \(x\)-axis; the regression line estimating \(x\) from \(y\)-values lies between the line of the long axes and the \(y\)-axis. So the line estimating \(x\) from \(y\) and the line estimating \(y\) from \(x\) are not the same.

2.1.3, statements (2): calculations with the defining transformation matrix \(A\):

Let \(Z = [z_1 \ z_2]\) be bivariate standard normal; let the bivariate normal \(X = [x_1 \ x_2]\) be derived from \(Z\) with the transformation \(X = AZ\) (multiplying \(X\) with the 2x2 matrix \(A\)). Then

a. (a) the covariance matrix \(D\) of the new variable \(X\) is \(D = A \cdot A^T\).

(This follows from the statements concerning variances and covariances of linear combinations of variables in 6.5.)

Consequence: variances and covariances of component variables of a bivariate normal \(X\) are easily obtained, if we know the defining transformation matrix \(A\).

b. the joint density of the new \(X\) derived this way is

\[f_{x_1,x_2}(x_1,x_2) = \frac{1}{d\pi} \cdot \frac{1}{2\pi} e^{-\frac{1}{2\pi} [x_1^2 + x_2^2 - \text{cov}(x_1,x_2)]} \]

with the coefficients \(b\) being the elements of corresponding indices of matrix \([b]\) is \((A^{-1})^T \cdot A^T\).
Bivariate concepts (theory)

Remark: \((A^{-1})^T A^{-1} = D^{-1}\), that is, the matrix defining coefficients in the joint density is the inverse of the covariance matrix.

2.1.4:

All the way up to now it has been assumed in (9) that the expected value of the vector variables in question is zero \(E(X_1)=0\) and \(E(X_2)=0\). In general, \(E(X_i)=m_i\) and \(E(X_j)=m_j\). Then

- the level sets are concentric ellipses with their common center in \((m_1, m_2)\)
- the lines of the conditional expectation functions are crossing in \((m_1, m_2)\).

That is, the joint distribution we have talked about by now is translated, in the geometrical sense, from the origin to \((m_1, m_2)\).

So the joint density is

\[
\begin{align*}
 f_{X, Y}(x_1, x_2) &= \frac{1}{2\pi \det(A)} e^{-\frac{1}{2}(b_1(x_1-m_1)^2 + b_2(x_2-m_2)^2 + 2b_1b_2(x_1-m_1)(x_2-m_2))} \\
&= \frac{1}{2\pi \sqrt{\det(D)}} e^{-\frac{1}{2}(x_1^2 + x_2^2 + 2r x_1 x_2)} \quad (22.9)
\end{align*}
\]

2.1.5, an alternative formula for the joint density:

the distribution of \(X\) and \(Y\) is bivariate normal if their joint density

\[
\begin{align*}
 f_{X, Y}(x_1, x_2) &= \frac{4}{2\pi d_1 d_2 \sqrt{1-r^2}} e^{-\frac{1}{2} \left( \frac{(x_1-m_1)^2}{d_1^2} + \frac{(x_2-m_2)^2}{d_2^2} - \frac{2r(x_1-m_1)(x_2-m_2)}{d_1 d_2} \right)} \\
&= \frac{4}{2\pi d_1 d_2 \sqrt{1-r^2}} e^{-\frac{1}{2} \left( \frac{(x_1-m_1)^2}{d_1^2} + \frac{(x_2-m_2)^2}{d_2^2} - \frac{2r(x_1-m_1)(x_2-m_2)}{d_1 d_2} \right)} \quad (22.10)
\end{align*}
\]

with \(m_1\) and \(m_2\) denoting the expected values of \(X_1\) és \(X_2\), \(d_1\) and \(d_2\) denoting the standard deviations of \(X_1\) és \(X_2\), and \(r\) denoting their correlation coefficient.

Readings


10. A) Steiner’s identity (theory)

A heavily used formula in statistics is Steiner’s identity (also known as ANOVA-equality).

segédtétel:

a. Steiner’s identity for numeric populations (for boxes):

Let \(x_i \in \mathbb{R}\) for \(i = 1, \ldots, m\) be real numbers, \(m\) their average \((m = \frac{\sum x_i}{n}\)\), and \(A\) an arbitrary real number. Then

\[
\sum_{i=1}^{n}(x_i - A)^2 = \sum_{i=1}^{n}(x_i - m)^2 + n(m - A)^2 \quad (22.11)
\]

That is: summing the squared deviations of numbers in a numeric population from an arbitrary (strange) number \(A\), the sum is equal to the

- sum of the squared deviations of the numbers in the population from their average

plus

- \(n\) times the squared difference between number \(A\) and the average of the population (\(n\) denoting the population size=the number of cards in the box=the number of numbers in the population).

Consequence: Assume we are looking for an optimal kind of middle value for a population, considering that the best from which the sum of the squared deviations of the numbers is smallest. This optimal middle value would then be the arithmetic mean.
explanation: \[ \sum_{i=1}^{n} (x_i - A)^2 = \sum_{i=1}^{n} (x_i - m) + (m - A)^2 = \]
\[ \sum_{i=1}^{n} [x_i - m]^2 + (m - A)^2 + 2(x_i - m)(m - A) \Rightarrow \sum_{i=1}^{n} [x_i - m]^2 + \sum_{i=1}^{n} (m - A)^2 + 2(m - A) \sum_{i=1}^{n} (x_i - m) \]

It remains to be proved that \( \sum_{i=1}^{n} (x_i - m) = 0 \) exercise: prove this.

b. Steiner’s identity for random variables:

Let \( X \) be a numeric random variable with expected value \( m \) and standard deviation \( d \). Then

\[ E( (X-A)^2 ) = d^2 + (m-A)^2 \]

proof:

\[ E( (X-A)^2 ) = E ( (X-m)+(m-A))^2 ) \Rightarrow \]
\[ = E ( (X-m)^2 + (m-A)^2 + 2(X-m)(m-A) ) \Rightarrow \]
\[ = E((X-m)^2) + E((m-A)^2) + 2(m-A)E(X-m) \Rightarrow \]
\[ = var(X) + (m-A)^2 + 2(m-A), 0 = \]
\[ = var(X) + (m-A)^2 \]
Chapter 23. Bivariate distributions (exercises)

some exercises about bivariate distributions

set A

1) There are four cards in a box, numbered 0, 0, 1 and 2. The experiment is two draws from the box, without replacement. Denote X the first draw, Y the second draw.

a) find the joint distribution (make a table);

b) find the marginal distributions;

c) find the conditional distributions \(Y \mid X=x\) for \(x=0,1,2\);

d) find the conditional expectation function \(E(Y \mid X)\);

e) find the distribution of the conditional-expectation-as-a-random-variable \(E(Y \mid X)\);

c') find the conditional distributions \(X \mid Y=y\) for \(y=0,1,2\);

d') find the conditional expectation function \(E(X \mid Y)\);

e') find the distribution of \(E(X \mid Y)\) as a random variable;

f) find the covariance of \(X\) and \(Y\);

g) find the correlation coefficient (Pearson's) of \(X\) and \(Y\).

1') Four cards in a box, numbered 0, 0, 10 and 20. Two draws from the box, without replacement. Denote X the first draw, Y the second draw.

a) find the joint distribution;

b) find the marginal distributions;

c) find the conditional distributions \(Y \mid X=x\) for \(x=0,10,20\);

d) find the conditional expectation function \(E(Y \mid X)\);

e) find the distribution of \(E(Y \mid X)\) as a random variable;

c') find the conditional distributions \(X \mid Y=y\) for \(y=0,10,20\);

d') find the conditional expectation function \(E(X \mid Y)\);

e') find the distribution of \(E(X \mid Y)\) as a random variable;

f) find the covariance of \(X\) and \(Y\);

g) find the correlation coefficient of \(X\) and \(Y\).

1") Four cards in a box, numbered 0, 0, 10 and 200. Two draws from the box, without replacement. Denote X the first draw, Y the second draw.

a) find the joint distribution;

b) find the marginal distributions;

c) find the conditional distributions \(Y \mid X=x\) for \(x=0,10,200\);
d) find the conditional expectation function \( E(Y \mid X) \);

e) find the distribution of \( E(Y \mid X) \) as a random variable;

c') find the conditional distributions \( (X \mid Y=y) \) for \( y=0,10,200 \);

d') find the conditional expectation function \( E(X \mid Y) \);

e') find the distribution of \( E(X \mid Y) \) as a random variable;

f) find the covariance of \( X \) and \( Y \);

g) find the correlation coefficient of \( X \) and \( Y \).

2) Consider a random vector variable \((X;Y)\) assuming the values \((0;0)\), \((0;1)\), \((1;0)\) and \((2;2)\) with equal probabilities of 0.25. Make a graphic representation of the cumulative distribution function \( F_{X,Y}(x,y) \). (Hint: the areas with different values of the function be denoted by different colours.)

2') The variable \((X;Y)\) assumes the values \((1;1)\), \((2;2)\) and \((3;3)\) with equal probabilities of 1/3. Make a graphic representation of the cumulative distribution function \( F_{X,Y}(x,y) \). (Areas with different colours...)

2") The variable \((X;Y)\) assumes the values \((3;1)\), \((2;2)\) and \((1;3)\) with equal probabilities of 1/3. Make a graphic representation of the cumulative distribution function \( F_{X,Y}(x,y) \). (Areas with different colours...)

3) The variable \((X;Y)\) is uniformly distributed on the rectangle \([-2;2] \times [-1;1]\). (That means that it has a joint density being constant over the area and zero outside the area.)

a) find the joint density;

b) find the marginal densities;

c) find the marginal cumulative distribution functions;

d) find the conditional densities \( f_{Y \mid X=x_0}(y) \) belonging to \( x_0=-2, x_0=0, \) and \( x_0=2 \);

e) find the conditional expectation functions \( E(Y \mid X) \) and \( E(X \mid Y) \);

f) find the covariance and the correlation coefficient for the two variables.

4) The variable \((X;Y)\) is uniformly distributed on the parallelogram with vertices \((-2;-1), (-2;0), (2;0)\) and \((2;1)\).

a) find the joint density;

b) find the marginal densities;

c) find the marginal cumulative distribution functions;

d) find the conditional densities \( f_{Y \mid X=x_0}(y) \) belonging to \( x_0=-2, 0,2 \);

d') find the conditional densities \( f_{Y \mid X=x_0}(y) \) belonging to \( x_0 \) (in general, as a function of \( x_0 \));

e) find the conditional densities \( f_{X \mid Y=y_0}(x) \) belonging to \( y_0=-0.5; 0;0.5 \);

e') find the conditional densities \( f_{X \mid Y=y_0}(x) \) belonging to \( y_0 \) (in general, as a function of \( y_0 \));

f) find the conditional expectation functions \( E(Y \mid X) \) and \( E(X \mid Y) \);

g) find the covariance and the correlation coefficient for the two variables.

4') The variable \((X;Y)\) is uniformly distributed on the parallelogram with vertices at \((-20;-1), (-20;0), (20;0)\) and \((20;1)\).

---

\[ \text{the rectangle } [-2;2] \times [-1;1] = \{ (x,y) \in \mathbb{R}^2 : (-2 \leq x \leq 2) \land (-1 \leq y \leq 1) \} \]
Bivariate distributions (exercises)

104

Created by XMLmind XSL-FO Converter.

a) find the joint density;
b) find the marginal densities;
c) find the marginal cumulative distribution functions;
d) find the conditional densities \( f_{Y|X=x_0}(y) \) belonging to \( x_0 = -20, 0.20 \);
d') find the conditional densities \( f_{Y|X=x_0}(y) \) belonging to \( x_0 \) (in general, as a function of \( x_0 \ ));
e) find the conditional densities \( f_{X|Y=y_0}(x) \) belonging to \( y_0 = -0.5; 0; 0.5 \);
e') find the conditional densities \( f_{X|Y=y_0}(x) \) belonging to \( y_0 \) (in general, as a function of \( y_0 \ ));
f) find the conditional expectation functions \( E(Y | X) \) and \( E(X | Y) \);
g) find the covariance and the correlation coefficient for the two variables.

Comment upon the changes as compared to exercise 4.

5) The variable \((X; Y)\) is uniformly distributed on the hexagon, not convex, that has its vertices at \((-1; -1)\), \((-1; 0)\), \((0; 1)\), \((1; 0)\), \((1; -1)\) and \((0; 0)\), in this order.
a) find the joint density;
b) find the marginal densities;
c) are \(X\) and \(Y\) independent?
d) find the conditional densities \( f_{Y|X=x_0}(y) \) belonging to \( x_0 = -1, -0.5, 0, 0.5, 1 \);
e) find the conditional densities \( f_{X|Y=y_0}(x) \) belonging to \( y_0 = -0.5; 0; 0.5 \);
f) find the conditional expectation functions \( E(Y | X) \) and \( E(X | Y) \); sketch the graph of each;
g) find the conditional variances \( V(X | Y=y_0) \) belonging to \( y_0 = -0.5, 0, 0.5, 1 \);
h) find the conditional variance function \( V(Y | X) \), mapping \( x_0 \) to \( V(Y | X=x_0) \);
i) find the covariance and the correlation coefficient for the two variables.

5') Now, the variable \((X; Y)\) is uniformly distributed on the, not convex, hexagon that has its vertices at \((9; 4), (9; 5), (10; 6), (11; 5), (11; 4)\) and \((10; 5)\), in this order.
a) find the joint density;
b) find the marginal densities;
c) are \(X\) and \(Y\) independent?
d) find the conditional densities \( f_{Y|X=x_0}(y) \) belonging to \( x_0 = 9, 9.5, 10, 10.5, 11 \);
e) find the conditional densities \( f_{X|Y=y_0}(x) \) belonging to \( y_0 = 4.5, 5, 5.5 \);
f) find the conditional expectation functions \( E(Y | X) \) and \( E(X | Y) \); sketch the graph of each;
g) find the conditional variances \( V(X | Y=y_0) \) belonging to \( y_0 = 4.5, 5, 5.5, 6 \);
h) find the conditional variance function \( V(Y | X) \);
i) find the covariance and the correlation coefficient for the two variables.
Comment upon the changes as compared to exercise 5.

6) The variable \((X;Y)\) is uniformly distributed on the rhombus with vertices \((-2;0)\), \((0;1)\), \((2;0)\), and \((0;-1)\).

a) find the joint density;

b) find the marginal densities;

c) are \(X\) and \(Y\) independent?

d) find the conditional densities \(f_{Y|X=x_0}(y)\) belonging to \(x_0=-1.5, -0.5, 0, 0.5, 1.5\);

e) find the conditional densities \(f_{X|Y=y_0}(x)\) belonging to \(y_0=-0.75, -0.5, 0, 0.5, 0.75\);

f) find the conditional expectation functions \(E(Y \mid X)\) and \(E(X \mid Y)\); sketch the graph of each;

g) find the conditional variance function \(V(Y \mid X)\);

h) find the conditional variance function \(V(X \mid Y)\);

i) find the covariance and the correlation coefficient for the two variables.

7) The variable \((X;Y)\) is jointly continuous; its joint density is \(f_{X,Y}(x,y) = \frac{1}{2} + \frac{xy}{2}\) on the rectangle \([-1;1] \times [0;1]\); outside the rectangle it is zero.

a) check if it is a joint density indeed;

b) find its marginal density \(f_X(x)\);

c) find its marginal density \(f_Y(y)\);

d) find the conditional densities \(f_{Y|X=x_0}(y)\) belonging to \(x_0=-1, x_0=-0.5, x_0=0, x_0=0.5\) and \(x_0=1\); sketch the graph of each. (Preferably in a single coordinate system with different colours for the different lines.)

e) find the conditional densities \(f_{X|Y=y_0}\) belonging to \(y_0=0, y_0=\frac{1}{2}, y_0=\frac{3}{4}\) and \(y_0=1\); sketch the graph of each. (Preferably in a single coordinate system with different colours for the different lines.)

f) find the conditional expectation functions \(E(Y \mid X)\) and \(E(X \mid Y)\); sketch the graph of each;

g) find the covariance and the correlation coefficient for the two variables.

8) \(f_X(x)=x+1\) if \(-1 \leq x \leq 0\), \(f_X(x)=1-x\) if \(0 \leq x \leq 1\), and \(f_X(x)=0\) elsewhere;

\(f_Y(y)=y+1\) if \(-1 \leq y \leq 0\), \(f_Y(y)=1-y\) if \(0 \leq y \leq 1\), and \(f_Y(y)=0\) elsewhere;

further, \(X\) and \(Y\) are known to be independent.

a) Find the joint density.

b) Describe of what geometrical shape the point set \(\{(x,y,z) \in \mathbb{R}^3 : z=f(x,y)\}\) is in the three-dimensional space.

set \(B\)

1) Two tosses with a fair coin. Three variables are considered, \(A\):={is the first toss a head?}; \(B\):={is the second toss a head?}, \(C\):={is the number of heads tossed an even number?}. (Values: 1=yes, 0=not.)

– find the joint distribution table (2x2x2) for the three variables

– find the 2-dimensional marginal distributions

\(^{\dagger}\)this point set is the 3-dimensional graph of the bivariate joint density function.
– are A and B independent?
– are A and C independent?
– are B and C independent?
– are the three variables "really" independent?
– give an example of an experiment with three variables of the same 2-dimensional marginals as seen above but with the three variables "really" independent.

set C

Denote here \( r(X;Y) \) the correlation coefficient (Person's) of variables \( X \) and \( Y \).

1) Let \( X \) and \( Y \) be uncorrelated random variables with standard deviations \( d=1 \) and expected values \( m=0 \).

Let \( Z:=3X+4Y \).

a) find the S.D. for \( Z \);
b) find \( \text{cov}(X,Z) \)
c) find \( r(X,Z) \)

2) Let \( X \) and \( Y \) be uncorrelated random variables with standard deviations \( d=1 \) and expected values \( m=0 \).

Let \( V:=X+5Y; \ W:=3X+10Y. \)
a) find the S.D.s for \( V \), and for \( W \)
b) find \( \text{cov}(V,W) \)
c) find \( r(V,W) \)

3) \( \text{var}(X)=4, \ \text{var}(Y)=9, \ \text{cov}(X,Y)=3. \)

\( V := X + 10Y; \ W := 100X + 2Y. \)
a) find the S.D.s for \( V \), and for \( W \)
b) find \( \text{cov}(V,W) \)
c) find \( r(V,W) \)

4) \( \text{var}(X)=25, \ \text{var}(Y)=100, \ r(X,Y)=0.5. \)

\( V := 2X + Y; \ W := Y - 2X. \)
a) find the S.D.s for \( V \), and for \( W \)
b) find \( \text{cov}(V,W) \)
c) find \( r(V,W) \)

set D

1) Variable \( X \) is independent of variable \( Y \), variable \( Y \) is independent of variable \( Z \). Does it follow that variable \( X \) is independent of variable \( Z \)?

2) Variable \( X \) is positively correlated with variable \( Y \); variable \( Y \) is positively correlated with variable \( Z \). Does it follow that variable \( X \) is positively correlated with variable \( Z \)?

3) \( r(X,Y)>0.90; \ r(Y,Z)>0.90. \) Does it follow that \( X \) is positively correlated with \( Z \)?

3') \( r(X,Y)>0.60; \ r(Y,Z)>0.60. \) Does it follow that \( X \) is positively correlated with \( Z \)?

3") Find the smallest positive number \( a \) that makes the following statement true:

'If \( r(X,Y)>a \) and \( r(Y,Z)>a \) then \( X \) and \( Z \) are positively correlated'
4.) "If $X_1, X_2, ..., X_7$ are such that $r(X_i, X_{i+1}) > a$ (i=1,...,6),
then $r(X_1, X_7) > 0$ ."

Find what numbers a make the above statement true.

5.) Let $X, Y$ and $Z$ be uncorrelated, $r(X,Y)=r(Y,Z)=r(X,Z)=0$

Is it possible that there exists a variable $W$ positively correlated with all three?

(E.g., might there be a $W$ such that $r(X,W)=r(Y,W)=r(Z,W) = \frac{1}{\sqrt{3}} \approx 0.577$ ?

6.) Let $X_1, ..., X_7$ be random variables, pairwise uncorrelated, with $d=1$ and $m=0$. Let

$Y = \frac{1}{n} \sum_{i=1}^{n} X_i$. Find $r(Y, X_i)$.

set E

1) Let $X$ and $Y$ be independent standard normal random variables.
   a) Specify their joint distribution.
   b) What shape are the level sets of the joint distribution?

2) Let $X$ and $Y$ be independent univariate normal random variables with expected values $m_x$ and $m_y$, and with standard deviations $d_x$ and $d_y$.
   a) Specify their joint distribution.
   b) What shape are the level sets of the joint distribution?

3) (a preliminary for exercise 4) The function $f(x)$ is the density of a univariate distribution. Its formula is known to be $f(x) = c \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}$, where $c$ is some constant.
   a) Does it follow that $f(x)$ is the density of some normal distribution?
      If yes...
   b) ...find the expected value of this normal distribution.
   c) ...find the S.D. of this normal distribution.

4) The random variable $(X;Y)$ is bivariate normal, with a joint density

$f_{X,Y}(x,y) = \frac{4}{2\pi d_x d_y \sqrt{1-r^2}} e^{-\frac{1}{2}(x-m_x)^2 + \frac{1}{2}(y-m_y)^2 - 2r \frac{d_x}{d_y} (x-m_x)(y-m_y)}$.
   a) Show that all conditional distributions $(Y \mid X=x_0)$are univariate normal distributions.
   b) b) Find the conditional expectation of $Y$ given that $X=x_0$ (that is, find the value of $E(Y \mid X=x_0)$).
   c) Find the value of the conditional variance $V(Y \mid X=x_0)$.
   d) Find what points in 9.2.1.2 of handout (22) these calculations constitute a proof (or at least an explanation) to.
      (Remark: a), b) and c) are to be solved by the appropriate analysis of the joint density function.)

set F:

"Some automatization, e.g. excel, seems useful here."
1) The bivariate normal \( X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \) is obtained by applying two transformations to the random vector values of the standard bivariate normal successively.

- 1st step: the second coordinate is multiplied with 0.25.
- 2nd step: the resulting vector is rotated to the left with an angle of \( 30^\circ \)

a) Find the transformation matrix \( A_1 \) of the first transformation step.

b) Find the transformation matrix \( A_2 \) of the second transformation step.

c) Find the transformation matrix \( A \) defining variable \( X \).

(Check where \( [1,0] \), 1st coordinate vector, is mapped by \( A \); check where \( [0,1] \), is mapped.)

d) What shape are the level sets of the joint distribution of \( X \)?

e) Find the covariance matrix \( D \) for the vector variable \( X \).

f) Find the variances and SDs for the component variables of \( X \). ( var\((X_1) = \? \); D\((X_1) = \? \); var\((X_2) = \? \); D\((X_2) = \? \) )

g) Find the covariance of the component variables. ( cov\((X_1,X_2) = \? \) )

h) Find the correlation coefficient for the component variables.

i) Find the direction in which the S.D. of the vector variable \( X \) is largest. How much is it?

j) Find the direction in which the S.D. of the vector variable \( X \) is smallest. How much is it?

1*) (continues the previous exercise)

a) Find the inverse matrix of the transformation matrix \( A_1 \) of the 1st step.

b) Find the inverse matrix of the transformation matrix \( A_2 \) of the 2nd step.

c) Find the inverse matrix of matrix \( A \) of the transformation defining \( X \).

(Check if the \( AA^{-1} \) product equals the identity.)

d*) Specify the joint density of variable \( X \).

2) Same as exercise 1) with the difference that the y-coordinate is multiplied with 0.50 instead of 0.25 in the 1st transformation step.

– Describe the change in the shape of the level sets, as compared to exercise 1).

– How do the marginal variances and the covariance change, as compared to exercise 1)?

– (it would be the point:) How does the correlation coefficient change, as compared to exercise 1)?

3) Same as 1)-2) but with the y-coordinates multiplied with 0.90, instead of 0.25 and 0.50, in the 1st step...

– describe the change in the shape of the level sets, as compared to exercise 1) and 2)

– how does the correlation coefficient change, as compared to exercise 1) and 2)?

4) Same as 1)..3) but with the y-coordinates multiplied with 0.10 in the 1st step...

– describe the change in the shape of the level sets, as compared to exercise 1)..3)

– how does the correlation coefficient change, as compared to exercise 1)..3)?
6) Describe the changes you see in the correlation of the component variables if the angle of rotation in the 2nd step is 45° instead of 30°. (As above, with shrinkages of 0.10, 0.25, 0.50 and 0.90.)

7) Describe the changes in the correlation of the component variables if the angle of rotation is only 5° (With shrinkages of 0.10, 0.25, 0.50 and 0.90, as above.)
Chapter 24. Estimation: concepts (theory)

1. Estimations

The situation: some parameter (numeric property) of a population is to be measured (e.g. the mean or the standard deviation of a variable, the covariance or the correlation of two variables) but there is no way to observe the whole population. Therefore a sample is selected and the population parameter is estimated from the observed values of the sample items.

This is what we have done, approximating the population average with the sample average or the population S.D. with the S.D. calculated from the sample. Or this is when, wanting to know the covariance of two variables in the population, their sample covariance is calculated and used as a substitute for the population value.

An estimation consists of first selecting a sample from the population, and then obtaining a single value from the sample values as a result of some calculations. These, taken together, define a random variable, called estimation.

(In probability theory it was (1) boxes, (2) experiments (draws from the box), and then (3) some operation resulting in a number (the sum of the draws; the mean of the draws; the number of '1's among the draws; the product of the draws, the largest number among the numbers drawn, etc). Here it is (1) a population from which (2) a sample is selected, and then (3) some calculations follow. The latter, the function that defines what operations to do with the sample values to obtain the estimate, is called the estimator.)

But the estimation as a random variable has something special to it: it is a random variable we expect something from. An estimation is expected to be close to the population parameter we wish to estimate with it. It has to be close to the target.

Now, denote X the estimation as a random variable, and \( \theta \) (theta) the population parameter X is to be close to.

2. Characterizing a single estimation: bias, standard error and root mean square error, variance and mean square error

The closer an estimation is to the parameter \( \theta \)-the better it is deemed.

So an estimation is deemed good if

a. its expectation equals \( \theta \), and, moreover,

b. does so with a small standard error, or,

c. the middle error is small.

So an estimation can be characterized with its

a. bias: \( \text{bias} := E(X) - \theta \)

b. standard error: \( \text{standard error} := \sqrt{E((X - E(X))^2)} \)

c. r.m.s. error (short for root mean square error): \( \text{r.m.s. error} := \sqrt{E((X - \theta)^2)} \)

In this handout sampling always means simple random sampling (SRS) with replacement (n draws, with replacement, from a box representing the population). In this context the elements of a sample (1st draw, 2nd draw, etc.) are independent, identically distributed random variables.
(Two more: the variance is the S.E., squared; the mean square error is the r.m.s.e., squared. Or:

- the variance is :=E((X-E(X))^2)
- the m.s.e. (short for mean square error) is :=E((X-\theta)^2).

The bias shows how much the expected value of the estimation is above or below the parameter;

- the standard error shows the middle deviations between the random values of the estimation and their expected value (in what circle the estimation disperses around its center);
- the r.m.s. error shows the middle deviations between the random values of the estimation and the parameter (in what circle the estimation disperses around the parameter)

Statement: \( \text{rmse}^2 = \text{torzítás}^2 + \text{SH}^2 \)

proof: denote X the estimation as a random variable, \( \theta \) the parameter to estimate. Taking version (b) of Steiner's identity with \( \theta \) as the 'strange' value A, we obtain:

\[
E((X - \theta)^2) = \text{var}(X) + (E(X) - \theta)^2
\]

The expression on the left is the r.m.s.e, squared;

- the first expression on the right side is the S.E. squared,
- the second expression on the right is the bias squared

Def.: an estimation X is unbiased if \( E(X) = \theta \).

Def.: Let X and Y be unbiased estimators of the same \( \theta \). X is more efficient than Y if \( \text{S.E.}(X) < \text{S.E.}(Y) \).

3. Characterizing estimation series: asymptotic unbiasedness and consistency

Chapters, questions and statements in estimation theory deal with what happens when "the same" estimation is done with larger and larger samples. (E.g., computing the sample mean from a sample of size 100, then from a sample of size 1000, and so on.) These statements deal with estimation series.

Dealing with estimations which are not unbiased, it is good if at least their bias converges to zero, selecting larger and larger samples. This is what the following definition is about.

Def.: Let \( X_n \) be a series of estimations for a parameter \( \theta \); this estimation series is asymptotically unbiased if

\[
\lim_{n \to \infty} E(X_n) = \theta.
\]

From another viewpoint, we want to know, whether, selecting larger and larger samples, the probability that the error of the estimate exceeds some fixed positive number converges to zero. (Less exactly, can we expect, that the bigger the sample, the less the probability of an error of a fixed size, this probability converging to zero?) If an estimation series has this property it is called consistent.

Def.: Let \( X_n \) be a series of estimations for a parameter \( \theta \) this estimation series is consistent if, for any \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} P (|X_n - \theta| > \varepsilon) = 0
\]

Statement: if \( X_n \) is a series of unbiased estimations for a parameter \( \theta \) and \( \lim_{n \to \infty} \text{S.H.}(X_n) = 0 \), then \( X_n \) is consistent.

Explain (hint: Chebyshev).

Statement: if \( X_n \) is an asymptotically unbiased estimation series for a parameter \( \theta \) and \( \lim_{n \to \infty} \text{S.H.}(X_n) = 0 \), then \( X_n \) is consistent. Explain.

Statement: the uncorrected sample variance \( S_n^2 = \frac{\sum (x_i - \bar{y})^2}{n} \) is a biased estimate of the population variance.
Denote $D^2$ the population variance, $m$ the population mean.

proof: according to version (a) of Steiner's identity with m as the A in the Steiner,

$$\sum(x_i - m)^2 = \sum(x_i - \bar{x})^2 + n(m - m)^2$$

- $E((x_i - m)^2) = \text{var}(X) = D^2$, therefore the expected value of the left side is $n \cdot D^2$;
- $E(x_i - m) = m$ so $E((x_i - m)^2) = \text{var}(x_i) = \frac{D^2}{n}$; hence (knowing that the variance of the sample mean $\text{var}(\bar{x}) = \frac{D^2}{n}$ the expected value of the second expression on the right side is $n \cdot \frac{D^2}{n} = D^2$.
- Now, carrying the second expression on the right side to the left side, we obtain that the expected value of the first expression on the right side is $(n-1) \cdot D^2$.

But then

$$E(S^2) = \frac{(n-1) \cdot D^2}{n} = \frac{n-1}{n} D^2 \neq D^2$$

Def.: The estimator $S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$ is called corrected sample variance.

Statement: with any population, the corrected sample variance is an unbiased estimator of the population variance (assuming simple random sampling and a population having a finite variance.)

Statement: the uncorrected sample variance is an asymptotically unbiased estimator of the population variance (assuming simple random sampling and a population having a finite variance.)

Remark: the uncorrected sample variance is also called empirical variance as opposed to sample variance denoting the corrected sample variance.

4. Confidence intervals

The situation is more or less what it has been. Wanting to know the value of some population parameter, a sample is selected — but, through some calculations from the sample, an interval is now produced instead of a single value, and this interval is expected to contain the population parameter we are looking for.

In some cases the chance that the interval contains the parameter is known. This chance is then called the confidence level and the interval a confidence interval. (The usual denotation for a 'confidence interval with a confidence level of 0.95' is CI.)

For example, we are going to find a 99%-confidence interval for the mean $\mu$ of a population with a known standard deviation, $\sigma$. A sample of 100 is selected. If the sample mean can be assumed normal (the population distribution being not very skewed and the extreme values rare enough), then

$$P(m < \bar{x} < m + 2.58S\sigma) = 0.99$$

(where $S\sigma = \sigma / \sqrt{100}$), hence

$$P(\bar{x} - 2.58S\sigma < \bar{x} < \bar{x} + 2.58S\sigma) = 0.99$$

So

$$[\bar{x} - 2.58S\sigma, \bar{x} + 2.58S\sigma]$$

is a 99%-confidence interval for the population mean.

Readings

1The kind seen earlier, the point-estimations, could be likened to operations of the artillery: the soldiers keep on shooting and the question is, how widely the shots disperse around their centre (how precise the cannon is), how much the distance between the centre of the shots and the target is (whether the aiming is good), and how widely the shots disperse around the target (that is, cannon and aiming taken together, how precise the shooting is). Interval estimations are more like trying to catch a fly with a swatter: the questions are, whether the fly is caught, and with what chance we get the fly.

2that is, with some combinations of sampling, calculations and population distribution

112

Chapter 25. Estimations (exercises)

estimations – a few exercises

1) Three cards in a box (population), numbered 1, 3 and 11. A sample of size 3 is selected from the box, with replacement. Let the estimator be

a) the sample mean.

b) the sample median.

Find the distribution of the estimation as a random variable in both cases. (This distribution is called the sampling distribution.)

The parameter to be estimated is the population mean.

– find for each of the estimations if it is unbiased;
– find the S.E. of the estimations;
– find the r.m.s.e. of the estimations.

2) Make a simulation consisting of 100 experiments described in exercise (1), computing the sample mean each time. Find the empirical distribution for the sample mean (values of sample mean observed, with their proportions, in a table). Compare this with the probability distribution of the sample mean, calculated in exercise (1).

3) (simulation.) Select a sample of 11 from a population distributed uniformly over (0,1). Repeat 1000 times. Compute (a) the sample mean (b) the sample median each time. Find the empirical distributions for both. Compare them. Describe what you observe, concerning (1) the bias (2) the S.E. of the estimations.

4) (simulation.) A quick method for estimating the S.D. of approximately normal populations uses the (max – min)/4 estimator. Let us see how it performs.

(a) Select a sample of 5, 1000 (or 10,000) times, from a population of standard normal distribution, computing each time both – the estimate '(max – min)/4', and
– the standard estimate of the standard deviation, S*

Compare their empirical distributions. Describe what you observe, concerning (1) the bias (2) the S.E. of the estimations.

(b) Repeat, for samples of 25 instead of 5. Describe the changes compared to (a).

5) (simulation) Compare two estimators for the population mean, (a) the sample mean and (b) the mean of the sample maximum and the sample minimum. Select samples of 5 e.g. 1000 times from a population distributed U[0;1]. Compute both estimates each time. Compare how they perform concerning (1) bias and (2) spread.

5') (simulation) Compare two estimators for the population mean, (a) the sample mean and (b) the mean of the sample maximum and the sample minimum. Select samples of 5 e.g. 1000 times from a population of standard normal distribution. Compute both estimates each time. Compare how they perform concerning (1) bias and (2) spread.

---

For additional notes:

1 three draws from the box, with replacement
2 take the difference of the largest and the smallest in the sample; divide by four. That is, compute (range/4).
3 \( s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \)
4 \( \frac{(\text{max} + \text{min})}{2} \)
5°) (simulation) Compare two estimators for the population mean, (a) the sample mean and (b) the mean of the sample maximum and the sample minimum. Let variable $X$ be distributed uniformly over $[1;2]$; let $Y := X^3$.

a) Find the expected value of $Y$.

b) Select samples of 5 e.g. 1000 times from a population with a distribution identical to the distribution of $Y$. Compute both estimates each time. Compare how they perform concerning (1) bias and (2) spread.

`generate independent random values of U[1;2] and raise them to the third power.`