Exercise Book to "Probability Theory with Simulations"

Vetier, András
Exercise Book to "Probability Theory with Simulations"
irta Vetier, András

Publication date 2014
Szerzői jog © 2014 Vetier András
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.11. PROBLEMS</td>
<td>41</td>
</tr>
<tr>
<td>4.12. 3.4 Geometrical distribution (pessimistic)</td>
<td>44</td>
</tr>
<tr>
<td>4.13. EXCEL</td>
<td>44</td>
</tr>
<tr>
<td>4.14. PROBLEMS</td>
<td>44</td>
</tr>
<tr>
<td>4.15. 3.5 Geometrical distribution (optimistic)</td>
<td>44</td>
</tr>
<tr>
<td>4.16. EXCEL</td>
<td>44</td>
</tr>
<tr>
<td>4.17. PROBLEMS</td>
<td>45</td>
</tr>
<tr>
<td>4.18. 3.6 Negative binomial distribution (pessimistic)</td>
<td>46</td>
</tr>
<tr>
<td>4.19. EXCEL</td>
<td>46</td>
</tr>
<tr>
<td>4.20. PROBLEMS</td>
<td>46</td>
</tr>
<tr>
<td>4.21. 3.7 Negative binomial distribution (optimistic)</td>
<td>47</td>
</tr>
<tr>
<td>4.22. EXCEL</td>
<td>47</td>
</tr>
<tr>
<td>4.23. PROBLEMS</td>
<td>47</td>
</tr>
<tr>
<td>4.24. 3.8 Poisson-distribution</td>
<td>48</td>
</tr>
<tr>
<td>4.25. EXCEL</td>
<td>48</td>
</tr>
<tr>
<td>4.26. PROBLEMS</td>
<td>48</td>
</tr>
<tr>
<td>4.27. 3.9 Higher dimensional discrete random variables and distributions</td>
<td>49</td>
</tr>
<tr>
<td>4.28. PROBLEMS</td>
<td>50</td>
</tr>
<tr>
<td>4.29. 3.10 Poly-hyper-geometrical distribution</td>
<td>51</td>
</tr>
<tr>
<td>4.30. EXCEL</td>
<td>51</td>
</tr>
<tr>
<td>4.31. PROBLEM</td>
<td>51</td>
</tr>
<tr>
<td>4.32. 3.11 Polynomial distribution</td>
<td>51</td>
</tr>
<tr>
<td>4.33. EXCEL</td>
<td>51</td>
</tr>
<tr>
<td>4.34. PROBLEM</td>
<td>52</td>
</tr>
<tr>
<td>4.35. 3.12 Generating a random variable with a given discrete distribution</td>
<td>52</td>
</tr>
<tr>
<td>4.36. EXCEL</td>
<td>52</td>
</tr>
<tr>
<td>4.37. PROBLEMS</td>
<td>52</td>
</tr>
<tr>
<td>4.38. 3.13 Mode of a distribution</td>
<td>53</td>
</tr>
<tr>
<td>4.39. EXCEL</td>
<td>53</td>
</tr>
<tr>
<td>4.40. PROBLEMS</td>
<td>53</td>
</tr>
<tr>
<td>4.41. 3.14 Expected value of discrete distributions</td>
<td>54</td>
</tr>
<tr>
<td>4.42. EXCEL</td>
<td>55</td>
</tr>
<tr>
<td>4.43. PROBLEMS</td>
<td>55</td>
</tr>
<tr>
<td>4.44. 3.15 Expected values of the most important discrete distributions</td>
<td>57</td>
</tr>
<tr>
<td>4.45. EXCEL</td>
<td>57</td>
</tr>
<tr>
<td>4.46. PROBLEMS</td>
<td>57</td>
</tr>
<tr>
<td>4.47. 3.16 Expected value of a function of a discrete random variable</td>
<td>58</td>
</tr>
<tr>
<td>4.48. EXCEL</td>
<td>58</td>
</tr>
<tr>
<td>4.49. PROBLEMS</td>
<td>58</td>
</tr>
<tr>
<td>4.50. 3.17 Moments of a discrete random variable</td>
<td>59</td>
</tr>
<tr>
<td>4.51. EXCEL</td>
<td>59</td>
</tr>
<tr>
<td>4.52. PROBLEMS</td>
<td>59</td>
</tr>
<tr>
<td>4.53. 3.18 Projections and conditional distributions for discrete distributions</td>
<td>60</td>
</tr>
<tr>
<td>4.54. EXCEL</td>
<td>60</td>
</tr>
<tr>
<td>4.55. PROBLEMS</td>
<td>60</td>
</tr>
<tr>
<td>4.56. 3.19 Transformation of discrete distributions</td>
<td>61</td>
</tr>
<tr>
<td>4.57. EXCEL</td>
<td>61</td>
</tr>
<tr>
<td>4.58. PROBLEMS</td>
<td>61</td>
</tr>
<tr>
<td>5. 4 Continuous random variables</td>
<td>62</td>
</tr>
<tr>
<td>5.1. PROBLEMS</td>
<td>62</td>
</tr>
<tr>
<td>5.2. 4.1 Distribution function</td>
<td>63</td>
</tr>
<tr>
<td>5.3. EXCEL</td>
<td>63</td>
</tr>
<tr>
<td>5.4. PROBLEMS</td>
<td>63</td>
</tr>
<tr>
<td>5.5. 4.2 Empirical distribution function</td>
<td>64</td>
</tr>
<tr>
<td>5.6. EXCEL</td>
<td>65</td>
</tr>
<tr>
<td>5.7. PROBLEMS</td>
<td>65</td>
</tr>
<tr>
<td>5.8. 4.3 Density function</td>
<td>65</td>
</tr>
<tr>
<td>5.9. EXCEL</td>
<td>65</td>
</tr>
<tr>
<td>5.10. PROBLEMS</td>
<td>65</td>
</tr>
<tr>
<td>5.11. 4.4 Histogram</td>
<td>67</td>
</tr>
</tbody>
</table>
Exercise Book to "Probability Theory with Simulations"

1. Preface

This exercise-book follows the structure of my text-book entitled http://www.math.bme.hu/~vetier/df/index.html Probability Theory with Simulations, containing the usual material taught at probability theory and statistics courses at most universities of the world. The same way, as in the text-book, there are five parts:

1. Probability of events
2. Discrete distributions
3. Continuous distributions in one-dimension
4. Two-dimensional continuous distributions
5. Statistics

Each part decomposes to several chapters. The majority of the problems in this exercise-book are quite ordinary, but some of them are quite unusual, because they focus on simulating randomness by Excel. At the beginning of the exercise-book, a summary of using Excel is included. A part of the problems are solved. The solution is always given in an Excel-file available by a link after the problem. Before trying to solve problems in a chapter of this exercise-book, it might be helpful to read and study the corresponding chapter in the text-book, which contains not only the theoretical basis of the topics, but a lot of solved problems, as well.

I am sure you will find mistakes in this exercise-book. I ask you to let me know them so that I could then correct them. Anyway, I am continuously working on this material, so new, corrected versions (with less or even more mistakes) will occur again and again in the future. Thanks for your cooperation.

Keep in mind that, in the simulation files, whenever you press the F9-key, the computer recalculates all the formulas, among others it gives new values to random numbers, consequently, it generates a new experiment.

2. 1 Introduction

2.1. EXCEL

2.2. The table structure of Excel

An Excel sheet consists of cells. The cells are arranged in rows and columns. The rows are numbered, the columns are identified by letters. Thus, for example, A1 is the name of the top-left cell, A2 is the cell under A1, B1 is the cell to the right of A1, and so on. In each cell, the user may write a number, or a text, or a formula. (We do not bother other possibilities.) A formula always starts with an equality sign, and in most cases, uses references to other cells. Formulas are also called functions. In the following file, there is a text in cell B2, there are numbers in cells B5, C5, D5, and and there is a formula in cell F5: =B5+C5+D5. The formula is easily verbalized: "take the sum of the contents of the cells B5, C5, D5. If the contents of the cells are, for example, the numbers 2, 3, 0.5, then in cell F5, we see the sum of these numbers, which is 5.5. If you want to see the formula itself, then you may open up the cell by double clicking on it.

http://www.math.bme.hu/~vetier/df/eg-010-01-01_Text_number_command.xls Demonstration file: Text, number, command eg-010-01-01
Exercise Book to "Probability Theory with Simulations"

Remark. When there are more formulas on a sheet, it may be useful to see not only one but all the formulas instead of the resulting values. This can be achieved by using the "Formulas", then the "Show Formulas" option, which switches from "resulting values" (as shown on Figure 1) to "formulas", and back from "formulas" to "resulting values" (as shown on Figure 2). A short-cut key stroke for this switch is Alt-. (Alt-point).

http://www.math.bme.hu/~vetier/dl/eg-010-01-02_Five_commands.xls Demonstration file: Five commands eg-010-01-02

Figure 1: Resulting values shown

Figure 2: Formulas shown

Remark. When someone makes a more composite work it is advantageous to work not only on one Excel sheet, but to use more sheets. Moreover, it is sometimes useful to connect more Excel files so that in some files we refer to data which are in another files. If we do so, then, when the files are copied or moved, it is important to pay attention whether the reference addresses remain correct.

2.3. COPY and PASTE commands

When we construct Excel files, we often need to use the same formula or similar formulas in several cells. Obviously, we do not like to type a formula many times. We prefer to write it down only once into one cell, and then copy this cell with the COPY command (Key stroke: Ctrl-C), and then paste it with the PASTE command (Key stroke: Ctrl-V) to the other cells.

Example. Squaring each element. Let a 3x3 matrix is given in region B5:D7. Assume that, in region F5:H7, we need the matrix whose elements are the squares of the elements of this matrix. In order to get the square of the top-left element of the given matrix, we will clearly write the formula =B5^2' into cell F5. If you copy now cell F5 and paste it, for example, into cell H6, then Excel first determines the relative position of cell H6 compared to cell F5. This relative position is 2 steps to the right and 1 step down. Then Excel determines the cell whose relative position compared to cell B5 is 2 steps to the right and 1 step down. This cell is obviously D6. This is why, when you copy cell F5 and paste it into cell H6, Excel, in cell H6, will write D6 instead of B5, that is, it will write the formula =D6^2 into cell H6. In the following file, we have already written the formula =B5^2 into cell F5. The reader is asked to copy cell F5 and paste it first into cell H6, and then into each cell of region F5:H7.

http://www.math.bme.hu/~vetier/dl/eg-010-03-01_Squaring_each_element_of_a_matrix.xls Demonstration file: Squaring each element eg-010-03-01

Example. Adding matrices. Let two 3x3 matrices are given in regions B5:D7 and F5:H7, respectively. Assume that, in region J5:L7, we need the sum of these matrices. In order to get the top-left element of the sum, we may write the formula =B5+F5' into cell J5. If you copy now cell J5 and paste it, for example, into cell L6, then Excel will write the formula 'V=D6+H6' into cell L6. If we copy cell J5 and paste it into each cell of region J5:H7, then we get the needed formulas in each cell of region J5:H7. In the following file, we have already written the
form ula '=D6+H6' into cell J6. The reader is asked to copy cell J5 and paste it first into cell L6, and then into each cell of region J5:L7. Thus we get the sum of the given matrices.

http://www.math.bme.hu/~vetier/df/eg-010-02-00_Adding_matrices.xls  Demonstration file: Adding matrices eg-010-02-00

2.4. Fixing references by dollar symbols

Example. Multiplying a matrix by a constant. Let a 3x3 matrix is given in region B5:D7. Assume that we need to multiply this matrix by a constant, which is in cell B9, and we want to put the product into region F5:H7. If we write the formula '=B5*B9' into cell F5, we get to top-left element correctly. However, if we copied now cell F5 and pasted it into cell H6, then Excel would clearly write not only D6 instead of B3, but D10 instead of B9. This would yield, in cell H6, obviously, the wrong formula '=D6*D10'. However, if, in cell F5, we put two $ (dollar) symbols in the reference to B9, that is, we write $B$9, and we write the formula like this: '=B5*$B$9', then in cell H6, we get the correct formula '=$D6*$B$9'. Be convinced that if we copy cell F5 and paste it into each cell of region F5:H7, we get the correct result.

http://www.math.bme.hu/~vetier/df/eg-010-03-00_Multiplying_a_matrix_by_a_constant.xls  Demonstration file: Multiplying a matrix by a constant eg-010-03-00

Example. Construction of a multiplication table. In most elementary schools, when children learn multiplication, in order to memorize the rules like "six times eight is forty-eight", they use a so called multiplication table. Such a table is given in the following file.

http://www.math.bme.hu/~vetier/df/eg-010-04-00_Multiplication_table.xls  Demonstration file: Multiplication table eg-010-04-00

The construction of this table is very simple. In cell C5 of the following file, we put the formula '=$B5*C$4'. Realize that, in this formula, there is a dollar symbol in front of the letter B and another symbol in front of the number 4. These dollar symbols, when we copy cell C5 and paste it, for example, into cell J10, keep the letter B and the number 4 fixed, yielding that the the first factor of the product will be taken from column B and row 10, and the the second factor of the product will be taken from row 4 and column J, which is the correct way to show that "six times eight is forty-eight". Copying cell C5 and pasting into each cell of region C5:N16, we get the correct multiplication table. In the following file, we have already written the correct formula into cell C5. The reader is asked to copy it and paste it first into cell J10, then into each cell of region C5:N16.

http://www.math.bme.hu/~vetier/df/eg-010-05-00_Construction_of_the_multiplication_table.xls  Demonstration file: Construction of the multiplication table eg-010-05-00

Remark. The dollar symbols, if needed, may be naturally inserted or deleted so that we simply type or delete them in the editing mode of the cell. However, there is more convenient way to do this: pressing the F4-key. When we open up the contents of a cell, and the cursor is above a reference, and we press the F4-key again and again, then the dollar symbols appear or disappear cyclically so that each of the possible four variations occur in a cyclical order. Be convinced that the formula '=$B5^2' in cell F5 in the following file changes into '=$B5$5^2', when you press the F4-key. When you press the F4-key again, you will get '=$B5$5*2', pressing the F4-key again you will get '=$B5^2'. Pressing the F4-key again, you get back '=$B5^2'. In each of the four stages, you may also copy cell F5 and paste it into each cell of region F5:H7, and observe what you get in region F5:H7.

http://www.math.bme.hu/~vetier/df/eg-010-06-00_Changes_of_dollar_symbols_when_the_F4-key_is_pressed.xls  Demonstration file: Changes of dollar symbols when the F4-key is pressed eg-010-06-00

2.5. PASTE-SPECIAL-VALUES command

When we copy and paste a cell containing a formula, we get again a formula. Excel shows the numerical result of the formula. We can see the formula itself, if we open up the cell by double-clicking on it. Sometimes it may be useful for us that after pasting the numerical result of the formula, the formula itself disappears, and the numerical value will stay in the cell. This can be achieved if we use the PASTE-SPECIAL-VALUES command, which can be performed by putting the cursor above the cell where we want to paste it, then making a right-click with the mouse, choosing PASTE-SPECIAL, and then VALUES, and then OK. In the following file, first we
constructed column C by the proper formula for the sine-function, then we put the cursor on top of this column, where the letter C stands, copied whole column, and pasted it onto column D using the PASTE-SPECIAL-VALUES techniques.

http://www.math.bme.hu/~vetier/df/eg-010-11-01_sine-table_1.xls Demonstration file: Constructing a table for the sine-function eg-010-11-01

If you now delete column C so that you put the cursor on top of this column, where the letter C stands, then you make a right-click with the mouse, and choose DELETE, then you get a table for the sine-function without formulas. This is what you see in the next file.

http://www.math.bme.hu/~vetier/df/eg-010-11-02_sine-table_2.xls Demonstration file: Constructing a table for the sine-function eg-010-11-02

Exercise. Constructing a table for the sine-function with degrees minutes. Construct the following table for the sine-function yourself. This table consists of more columns in order to involve not only integer degrees but minutes as well.

http://www.math.bme.hu/~vetier/df/eg-010-11-03_sine-table_3.xls Demonstration file: Table for the sine-function with degrees minutes eg-010-11-03

2.6. CUT and PASTE command

The CUT command (Key stroke: Ctrl-X) essentially differs from the COPY command (Key stroke: Ctrl-C). For example, open the file

http://www.math.bme.hu/~vetier/df/eg-010-03-00_Multiplying_a_matrix_by_a_constant.xls Demonstration file: Multiplying a matrix by a constant eg-010-03-00

and mark the region B5:D7, and press "Ctrl-X", then put the cursor onto B10 and press "Ctrl-V", and notice how the references change in the cells of the region F5:H7. Now open the file

http://www.math.bme.hu/~vetier/df/eg-010-03-00_Multiplying_a_matrix_by_a_constant.xls Demonstration file: Multiplying a matrix by a constant eg-010-03-00

again and mark the region F5:H7, and press "Ctrl-X", then put the cursor onto F10 and press "Ctrl-V", and notice what references you get in the cells of the region F10:H12.

2.7. Formatting

When we write, for example, a number into a cell, Excel will show it. We may like the way how Excel shows it, or, if we want some other format, we may change the format. In the following file, the number \( \pi \) is written into all the hundred cells, but the formats of the cells are different. We do not explain here, how the different formats can be achieved. We encourage the reader the be brave and make trials to change the formats.

http://www.math.bme.hu/~vetier/df/eg-010-21-01_Formatting.xls Demonstration file: Formatting eg-010-21-01

2.8. Figures

Excel has excellent abilities to construct figures. We shall mainly use the so called columns charts, line charts and scatter charts.

In order to construct a column chart, we need a series of data arranged in a row or in a column. In the following file, nine numbers are given in nine adjacent cells of row 4. If, with the mouse, you mark them, and then you choose the options: Insert, Column, 2-D column, then you get the figure we have already put there.

http://www.math.bme.hu/~vetier/df/eg-010-22-01_Constructing_a_column_chart.xls Demonstration file: Constructing a column chart eg-010-22-01
In order to construct a line chart, you should choose the options: Insert, Line, 2-D line, and you get the figure we have in the following file.

http://www.math.bme.hu/~vetier/df/eg-010-22-02_Constructing_a_line_chart.xls Demonstration file: Constructing a line chart eg-010-22-02

In order to construct a scatter chart, we need two series of data of the same size. The first of them will define the horizontal coordinates, the second will define the vertical coordinates. In the following file, the six numbers in row 4 are the horizontal coordinates, the six numbers in row 5 are the vertical coordinates. If, with the mouse, you mark the range consisting of these twelve cells, and then you choose the options: Insert, Scatter, Scatter with Straight Lines and Markers options, then you get the figure we have already put there.

http://www.math.bme.hu/~vetier/df/eg-010-22-03_Constructing_a_scatter_chart.xls Demonstration file: Constructing a scatter chart eg-010-22-03

When we have constructed a figure, then its format can be modified. We do not explain here, how the different formats can be achieved. We encourage the reader the be brave and make trials to change the formats.

2.9. Built in functions

Lots of functions are available if you enter the Formulas option, and make you choice. However, the names of the functions can be simply typed, like SIN for sine, LOG for logarithm, AVERAGE, for the average, and so on. Obviously, when you type the formulas, the arguments of the functions must be written according to strict rules.

2.10. Special usages of the mouse

"Marking a range". Move the mouse somewhere and then make a left click. A cell will be marked. Be careful to keep the cursor inside the cell so that the cursor will be a white, fat plus sign. If you keep pressing the left button of the mouse and you drag the mouse, not only the cell but a range will marked.

"Cut and paste". Now mark a range. If you move the mouse so that the cursor becomes a plus sign with arrow-heads, and you keep pressing the left button of the mouse and you drag the mouse, then the range will move the same way as if you made a CUT and PASTE command.

"Copy and paste". Now mark a cell. If you move the mouse to the button-right corner of the cell so that the cursor becomes a black plus sign, and you keep pressing the left button of the mouse and you drag the mouse along a range, then you get the same result as if you copied the cell and pasted it on the range.

"Extending a region". Now mark a range. If you move the mouse to the button-right corner of the range so that the cursor becomes a black plus sign, and you keep pressing the left button of the mouse and you drag the mouse, then - depending on the contents of the range - Excel will type "an extension" of the given range. In the following file, if you mark a green range, and - with the black plus sign - drag the mouse down, then you will get exactly what you see in the adjacent yellow range.

http://www.math.bme.hu/~vetier/df/eg-010-23-03_Extending_a_region.xls Demonstration file: Extending a region eg-010-23-03

2.11. PROBLEMS

1. Files to study Study the files related to Example 1 entitled "Coming home from Salzburg to Vac" in Section 1 of Part I of the textbook:

a. http://www.math.bme.hu/~vetier/df/ef-020-01-00_Previous_train.xls Demonstration file: The amount of time after the departure of the previous train ef-020-01-00
b. http://www.math.bme.hu/~vetier/df/ef-020-02-00_Previous_and_next_train.xls Demonstration file: Both the amount of time after the previous train and the waiting time until the next train are shown ef-020-02-00

c. http://www.math.bme.hu/~vetier/df/ef-020-03-00_Previous_train_10_experiments.xls Demonstration file: 10 experiments for the amount of time after the previous train ef-020-03-00

d. http://www.math.bme.hu/~vetier/df/ef-020-04-00_Previous_train_1000_experiments_on_a_line.xls Demonstration file: 1000 experiments on a line ef-020-04-00

e. http://www.math.bme.hu/~vetier/df/ef-020-05-00_Previous_train_10_experiments_on_a_stip.xls Demonstration file: 10 experiments on a narrow horizontal strip ef-020-05-00

f. http://www.math.bme.hu/~vetier/df/ef-020-06-00_Previous_train_1000_experiments_on_a_stip.xls Demonstration file: 1000 experiments on a narrow horizontal strip ef-020-06-00

g. http://www.math.bme.hu/~vetier/df/ef-020-07-00_Freq_relfreq_of_Unpleasant_event.xls Demonstration file: Frequency and relative frequency of the unpleasant event ef-020-07-00

h. http://www.math.bme.hu/~vetier/df/ef-020-08-00_Prob_of_Unpleasant_event.xls Demonstration file: Probability of the unpleasant event ef-020-08-00

2. Files to study Study the files related to Example 2 entitled "Random numbers" in Section 1 of Part I of the textbook:


3. Files to study Study the files related to Example 3 entitled "Pairs of random numbers" in Section 1 of Part I of the textbook:


b. http://www.math.bme.hu/~vetier/df/ef-020-12-00_Spec_triangle_with_diamond.xls Demonstration file: Special triangle combined with a diamond-shaped region - unconditional ... ef-020-12-00

c. http://www.math.bme.hu/~vetier/df/ef-020-13-00_Spec_triangle_with_diamond_Dec_RelFreq.xls Demonstration file: Special triangle combined with a diamond-shaped region - conditional ... ef-020-13-00

4. Files to study Study the files related to Example 4 entitled "Non-uniform distributions" in Section 1 of Part I of the textbook:

a. http://www.math.bme.hu/~vetier/df/ef-020-14-00_Square_of_RND.xls Demonstration file: Non-uniformly distributed points using the square of a random number ef-020-14-00


c. http://www.math.bme.hu/~vetier/df/ef-020-16-00_Rel_freq_for_non-uniform_distr.xls Demonstration file: Relative frequency for non-uniform distribution ef-020-16-00

d. http://www.math.bme.hu/~vetier/df/ef-020-17-00_Dec_Rel_freq_for_non-uniform_distr.xls Demonstration file: Conditional relative frequency for non-uniform distribution ef-020-17-00

5. Files to study Study the file related to Example 5 entitled "Waiting time for the bus" in Section 1 of Part I of the textbook: http://www.math.bme.hu/~vetier/df/ef-020-18-00_Waiting_time_for_bus_uniform_distr.xls Demonstration file: Waiting time for the bus ef-020-18-00
6. Files to study

Study the files related to Example 6 entitled “Traveling by bus and metro” in Section 1 of Part I of the textbook:


f. http://www.math.bme.hu/~vetier/df/ef-020-24-00_Bus_and_metro__Event5.xls Demonstration file: Total waiting time is more than 4, using uniform distribution ef-020-24-00


7. Files to study

Study the files related to Example 7 entitled "Dice" in Section 1 of Part I of the textbook:

a. http://www.math.bme.hu/~vetier/df/ef-020-36-00_Fair_die_1000_tosses.xls Demonstration file: Fair die, 1000 tosses ef-020-36-00

b. http://www.math.bme.hu/~vetier/df/ef-020-38-00_Fair_die_1000_tosses_REQUENCY_COMMAND.xls Demonstration file: 1000 tosses with a fair die, relative frequencies and probabilities ef-020-38-00
3.2 Outcomes and events

3.1. EXCEL

3.2. The RANDBETWEEN function

The simplest Excel function which includes randomness is the RANDBETWEEN function. If you type, for example

=RANDBETWEEN(1;6)

then Excel will return an integer number greater than or equal to 1 and smaller than equal to 6. This means that the outcomes are the numbers 1, 2, 3, 4, 5, 6. If you type

=RANDBETWEEN(-10;10)

then Excel will return an integer number greater than or equal to -10 and smaller than equal to 10. This means that the outcomes are the numbers -10, -9, -8, ..., 8, 9, 10.

3.3. The RAND function

The other Excel function which includes randomness is the RAND function. If you type

=RAND()

then Excel will return a number with 15 decimals between 0 and 1. The outcomes for this experiment are actually the numbers

0.000000000000001
0.000000000000002
0.000000000000003
...
0.999999999999997
0.999999999999998
0.999999999999999

Since these numbers constitute a "dense" subset of the (0; 1) interval, we may think as if the outcomes were all the numbers between 0 and 1, that is, the sample space were the (0; 1) interval.

Remark. When we write the RAND function correctly as
=RAND()

It may seem strange to write the pair of empty parentheses after RAND. Nevertheless, this is the correct form of this function.

Remark. Obviously, the

=6*RAND()

formula returns a number between 0 and 6. Thus, rounding up this number to an integer, that is, using the formula

=ROUNDUP( 6*RAND() ; 0 )

we get an integer number greater than or equal to 1 and smaller than or equal to 6, the same way as we get with the function

=RANDBETWEEN(1;6)

If you use and earlier versions of Excel, which does not offer the RANDBETWEEN function, then integer valued random numbers can be generated only by this "multiply-then-roundup" method.

3.4. Simulating an event with the IF function

The meaning of an event in everyday usage is rather wide. In probability theory, the meaning is rather restricted: an event means a statement related to the phenomenon so that when we make an experiment for the phenomenon, then the statement is either true or false. For example, when we work with a random number we may be interested in the event that the random number is smaller than 0.75. This statement can be simulated with Excel like this:

\[
\text{IF( RAND() < 0.75 ; "TRUE" ; "FALSE" )}
\]

In most cases it is more advantageous to write the number 1 instead of the word "TRUE" and the number 0 instead of the word "FALSE" like this:

\[
\text{IF( RAND() < 0.75 ; 1 ; 0 )}
\]

We may write the random number in a separate cell and, in the IF function, we refer to it as we do in the following file:

http://www.math.bme.hu/~vetier/df/eg-010-24-01_Simulating_an_event.xls Demonstration file: Simulating an event eg-010-24-01

3.5. PROBLEMS

1. Two tickets, one draw There are two tickets in a box: a red and a green. We choose a ticket from the box at random and observe its color.

a. What are the possible outcomes?

b. How many outcomes are there?

c. What is the sample space?

Five tickets, one draw There are five tickets in a box: a red, a white, a green, a blue and a yellow. We choose a ticket from the box at random and observe its color.

a. What are the possible outcomes?

b. How many outcomes are there?
c. What is the sample space?

2. Two tickets, two draws with replacement There are two tickets in a box: a red and a green. We choose a ticket from the box at random, replace it, then choose again, and observe the colors of both draws.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?

3. Two tickets, two draws without replacement There are two tickets in a box: a red and a green. We choose a ticket from the box at random (do not replace it), then choose again, and observe the colors of both draws.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?

4. Three tickets, two draws with replacement There are three tickets in a box: a red, a white and a green. We choose a ticket from the box at random, replace it, then choose again, and observe the colors of both draws.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?

5. Three tickets, two draws without replacement There are three tickets in a box: a red, a white and a green. We choose a ticket from the box at random (do not replace it), then choose again, and observe the colors of both draws.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?

6. Three tickets, three draws with replacement There are three tickets in a box: a red, a white and a green. We choose three times a ticket from the box at random with replacement, and observe the color of each draw.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?

7. Three tickets, three draws without replacement There are three tickets in a box: a red, a white and a green. We choose three times a ticket from the box at random without replacement, and observe the color of each draw.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?

8. Coin tossed two times A fair coin is tossed two times.
   a. What are the possible outcomes?
   b. How many outcomes are there?
9. Coin tossed three times A fair coin is tossed three times.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?

10. Coin tossed four times A fair coin is tossed four times.
    a. What are the possible outcomes?
    b. How many outcomes are there?
    c. What is the sample space?

11. Coin tossed \( 7 \) times A fair coin is tossed \( 7 \) times.
    a. What are the possible outcomes?
    b. How many outcomes are there?
    c. What is the sample space?

12. Coin tossed until a the first head A fair coin is tossed until a the first head occurs.
    a. What are the possible outcomes?
    b. How many outcomes are there?
    c. What is the sample space?

13. Letter from an English text Suppose that we choose a letter from an English text, so the possible outcomes are the 26 letters of the alphabet: 
    \( a, b, c, \ldots, x, y, z \) 
    that is, the sample space is 
    \( \{a, b, c, \ldots, x, y, z\} \) 
    Verbalize the events corresponding to the following subsets of the sample space:
    a. \( A = \{a, b, c, d, e\} \) 
    b. \( B = \{a, e, i, o, u\} \) 
    c. \( C = \{x, y, z\} \) 
    d. \( A \cap B \) 
    e. \( A \cup B \) 
    f. \( A\setminus B \) 

14. Five people with red hats Five people, call them \( 1, 2, 3, 4, 5 \), independently of each other put a red hat on their heads at random. Clearly, there are \( 2^5 = 32 \) possible variations. List in an Excel file the 32 variations in 32 rows. http://www.math.bme.hu/~vetier/df/Sol-01-01-10_Five_people_with_red_hats_Possible_outcomes.xls Solution Sol-01-01-10

15. Ten people with red hats Ten people, call them \( 1, 2, \ldots, 10 \), independently of each other put a red hat on their heads at random. Clearly, there are \( 2^{10} = 1024 \) possible variations. Construct an Excel file to list the 1024 variations in 1024 rows. http://www.math.bme.hu/~vetier/df/Sol-01-01-11_Ten_people_with_red_hats_Possible_variations.xls Solution Sol-01-01-11
3.6. 2.1 Relative frequency and probability

3.7. EXCEL

Calculating the frequency using the IF and SUM functions

The frequency (and then the relative frequency) of an event can be easily calculated using the IF and SUM functions as shown in the following file:

http://www.math.bme.hu/~vetier/df/ef-020-07-00_Freq_relfreq_of_Unpleasant_event.xls

Demonstration file: Frequency and relative frequency of the unpleasant event ef-020-07-00

Calculating the frequency using the FREQUENCY function

When we want to calculate the frequency (and the relative frequency) of more events, it is advantageous to use the FREQUENCY function as in the following file:

http://www.math.bme.hu/~vetier/df/ef-020-36-00_Fair_die_1000_tosses.xls

Demonstration file: Fair die, 1000 tosses ef-020-36-00

When you use this function you must pay attention to the special way of entering this function. Assume that the data are given in in region A1:A10 of your Excel sheet (this region is called the "Data array"), and the possible values are listed in in region C1:C6 (this region is called the "Bins array") as shown on Figure 3 entitled "Data array and Bins array"

![Figure 3: Data array and Bins array](image)

In this case, in order to use the FREQUENCY function,

1. first you have to type the

   =FREQUENCY( A1:A10 ; C1:C6 )

   formula into the cell on the right side of the first cell of the bins array, which is now the cell D1 (see Figure 4 entitled "Writing the FREQUENCY function into the first cell of the bins array")

![Figure 4: Writing the FREQUENCY function into the first cell of the bins array](image)

2. then you have to mark the whole range adjacent to the bins array, which is now the range D1:D6 (see Figure 5 entitled "Marking the whole range adjacent to the bins array")

![Figure 5: Marking the whole range adjacent to the bins array](image)

3. then to press the F2-key,

4. and finally to press the CTRL-SHIFT-ENTER key combination.

Try to do these steps correctly in the following file, where the data array is marked with the yellow color, and the bins array is marked with the blue color:
3.8. PROBLEMS

1. Math examination results
   Somebody observed the math examination results at a university from the point of view whether the students passes or fails. The sequence he got from the first 10 results is: pass, fail, fail, pass, pass, pass, fail, pass, pass, fail.
   a. Write down the sequence of relative frequencies of passing.
   b. Somebody states that the probability of passing the course is only around 1/3. Does the above sequence really contradict to this statement?

2. Two tickets, simulating two draws with replacement
   There are two tickets in a box: a red and a green. We choose a ticket from the box at random, replace it, then choose again, and observe the colors of both draws.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?
   d. Simulate in Excel as if you chose the two tickets and observed the colors of both draws.
   e. Make 1000 experiments, and calculate the relative frequency of each outcome.
   f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

3. Two tickets, simulating two draws without replacement
   There are two tickets in a box: a red and a green. We choose a ticket from the box at random (do not replace it), then choose again, and observe the colors of both draws.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?
   d. Simulate in Excel as if you chose the two tickets and observed the colors of both draws.
   e. Make 1000 experiments, and calculate the relative frequency of each outcome.
   f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

4. Three tickets, simulating two draws with replacement
   There are three tickets in a box: a red, a white and a green. We choose a ticket from the box at random, replace it, then choose again, and observe the colors of both draws.
   a. What are the possible outcomes?
   b. How many outcomes are there?
c. What is the sample space?

d. Simulate in Excel as if you chose the three tickets and observed the colors of both draws.

e. Make 1000 experiments, and calculate the relative frequency of each outcome.

f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

5. Three tickets, simulating two draws without replacement

There are three tickets in a box: a red, a white and a green. We choose a ticket from the box at random (do not replace it), then choose again, and observe the colors of both draws.

a. What are the possible outcomes?

b. How many outcomes are there?

c. What is the sample space?

d. Simulate in Excel as if you chose the three tickets and observed the colors of both draws.

e. Make 1000 experiments, and calculate the relative frequency of each outcome.

f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

6. Simulating as if a coin were tossed twice

A fair coin is tossed two times.

a. What are the possible outcomes?

b. How many outcomes are there?

c. What is the sample space?

d. Simulate in Excel as if you tossed a coin two times.

e. Make 1000 experiments, and calculate the relative frequency of each outcome.

f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

7. Simulating as if a coin were tossed three times

A fair coin is tossed three times.

a. What are the possible outcomes?

b. How many outcomes are there?

c. What is the sample space?

d. Simulate in Excel as if you tossed a coin three times.

e. Make 1000 experiments, and calculate the relative frequency of each outcome.

f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

8. Simulating as if a coin were tossed four times

A fair coin is tossed four times.

a. What are the possible outcomes?

b. How many outcomes are there?

c. What is the sample space?

d. Simulate in Excel as if you tossed a coin four times.

e. Make 1000 experiments, and calculate the relative frequency of each outcome.
f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

9. Simulating as if a coin were tossed $n$ times A fair coin is tossed $n$ times.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?
   d. Simulate in Excel as if you tossed a coin $n$ times.
   e. Make 1000 experiments, and calculate the relative frequency of each outcome.
   f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

Simulating as if a coin were tossed until the first head A fair coin is tossed until the first head occurs.
   a. What are the possible outcomes?
   b. How many outcomes are there?
   c. What is the sample space?
   d. Simulate in Excel as if you tossed a coin until the first head.
   e. Make 1000 experiments, and calculate the relative frequency of each outcome.
   f. Studying the relative frequencies make guesses how much the probabilities of the outcomes are?

10. Bus and metro When my friend comes to the university he takes a bus and then a metro. The waiting
time for the bus is uniformly distributed between 0 and 15, the waiting time for the metro is uniformly
distributed between 0 and 5. (The two waiting times are independent of each other.) The two waiting times
put together constitute a random point in a rectangle.
   a. Make a simulation with Excel for the phenomenon with 1000 experiments.
   b. Determine the relative frequency and calculate the probability of the event that the total waiting time
(waiting time for the bus plus the waiting time for the metro) is less than 7.
   c. Determine the relative frequency and calculate the probability of the event that the waiting time for bus is
less than the waiting time for metro.
   d. Replace the numbers 15, 5, 7 in your simulation by parameters.

11. Files to study Study the files in Section 3 of Part I of the textbook:
      Demonstration file: Waiting time for the bus ef-020-18-00
      Demonstration file: Traveling by bus and metro: uniformly distributed waiting times ef-020-19-00
   c. http://www.math.bme.hu/~vetier/dl/ef-020-20-00_Bus_and_metro__Event1.xls
      Demonstration file: Waiting time for bus < 4 ef-020-20-00
   d. http://www.math.bme.hu/~vetier/dl/ef-020-21-00_Bus_and_metro__Event2.xls
      Demonstration file: Waiting time for metro > 3 ef-020-21-00
      Demonstration file: Waiting time for bus < 4 AND waiting time for metro > 3 ef-020-22-00
   f. Demonstration file: Waiting time for bus < waiting time for metro ef-020-23-00
g. http://www.math.bme.hu/~vetier/df/ef-020-24-00_Bus_and_metro__Event5.xls Demonstration file: Total waiting time > 4 ef-020-24-00

h. http://www.math.bme.hu/~vetier/df/ef-020-25-00_Bus_and_metro__Event6.xls Demonstration file: Waiting time for bus is less than waiting time for metro AND total waiting time > 4 ef-020-25-00

i. http://www.math.bme.hu/~vetier/df/ef-020-26-00_Bus_and_metro__Event7.xls Demonstration file: Waiting time for bus < waiting time for metro OR total waiting time > 4 ef-020-26-00

j. http://www.math.bme.hu/~vetier/df/ef-030-01-00_Event_RelFreq_Prob__RND.xls Demonstration file: Event and relative frequency ef-030-01-00

k. http://www.math.bme.hu/~vetier/df/ef-030-02-00_Event_RelFreq_Prob__Die.xls Demonstration file: Tossing a die - probability ef-030-02-00

l. http://www.math.bme.hu/~vetier/df/ef-030-03-00_Balls_Drawn_from_Box.xls Demonstration file: Relative frequency with balls ef-030-03-00


3.9. 2.2 Random numbers

3.10. EXCEL

The two important Excel functions RANDBETWEEN and RAND were introduced earlier in the Section entitled “Outcomes and events”. The problems in this section offer a theoretical practice related to these functions.

3.11. PROBLEMS

1. Random number generator of a calculator Play with the random number generator of your calculator and/or with the RAND function of your Excel.

   a. Make many experiments, and calculate the relative frequency of getting a number between $a$ and $b$, if $0 < a < b < 1$.

   b. Be convinced that the probability of getting a number between $a$ and $b$ is equal to $b - a$ if $0 < a < b < 1$.

   c. Make a figure to show that the random numbers are distributed uniformly between 0 and 1.

   d. Make many experiments to see that the average of the random numbers is close to 0.5.

   e. Make many experiments to see that the average of the squares of the random numbers is close to 1/3.

   f. Make many experiments to see that the average of the square-roots of the random numbers is close to 2/3.

   g. Multiply each random number by 6 and be convinced that the result is a random number uniformly distributed between 0 and 6.

   h. Multiply each random number by 6, and then round up to simulate a fair die.

   i. Make many experiments to see that the square of a random numbers is not uniformly distributed.

   j. Make many experiments to see that the square-root of a random numbers is not uniformly distributed.
2. Calculating probabilities 5 independent random numbers are generated between 0 and 1 according to uniform distribution.
   a. What is the probability that the first is less then 0.95?
   b. What is the probability that the first and the second are less then 0.95?
   c. What is the probability that all the numbers are less then 0.95?
   d. What is the probability that at least one of them is less then 0.95?


3. Probabilities related to random numbers A random number \texttt{RND} is generated by a calculator or computer. Find the probabilities:
   a. \( P(\texttt{RND} < 0.7) \);
   b. \( P(\texttt{RND} < 1.7) \);
   c. \( P(\texttt{RND} < -0.7) \);
   d. \( P(0.3 < \texttt{RND}) \);
   e. \( P(-0.3 < \texttt{RND}) \);
   f. \( P(1.3 < \texttt{RND}) \);
   g. \( P(0.3 < \texttt{RND} < 0.7) \);
   h. \( P(-0.3 < \texttt{RND} < 0.7) \);
   i. \( P(-0.3 < \texttt{RND} < -0.7) \).

4. Probabilities related to random numbers A random number \texttt{RND} is generated by a calculator or computer. Find formulas for the probabilities:
   a. \( P(\texttt{RND} < z) \), where \( z \) is a number between 0 and 1;
   b. \( P(\texttt{RND} < z) \), where \( z \) is a negative number;
   c. \( P(\texttt{RND} < z) \), where \( z \) is a positive number;
   d. \( P(\texttt{RND} < z) \), where \( z \) is a real number.

5. Probabilities related to random numbers A random number \texttt{RND} is generated by a calculator or computer. Find formulas for the probabilities:
   a. \( P(a < \texttt{RND} < b) \), where \( 0 \leq a \leq b \leq 1 \);
   b. \( P(a < \texttt{RND} < b) \), where \( a \) and \( b \) are numbers between 0 and 1;
   c. \( P(a < \texttt{RND} < b) \), where \( a \) and \( b \) are arbitrary real numbers.

6. Probabilities related to the square of a random number A random number \texttt{RND} is generated by a calculator or computer. Find the probabilities
   a. \( P(\texttt{RND}^2 < 0.7) \);
b. \( P(\text{RND}^2 < 1.7); \)

c. \( P(\text{RND}^2 < z), \) where \( z \) is a positive parameter.

7. Probabilities related to the square-root of a random number A random number \( \text{RND} \) is generated by a calculator or computer. Let \( X \) be its square-root: \( X = \sqrt{\text{RND}}. \) Find the probabilities

a. \( P(X < 0.7); \)

b. \( P(X < 1.7); \)

c. \( P(X < z), \) where \( z \) is a positive parameter.

8. Probabilities related to the reciprocal of a random number A random number \( \text{RND} \) is generated by a calculator or computer. Let \( X \) be its reciprocal: \( X = 1/\text{RND}. \) Find the probabilities

a. \( P(X < 0.7); \)

b. \( P(X < 1.7); \)

c. \( P(X < z), \) where \( z \) is a positive parameter.

9. Calculating relative frequencies Calculate the relative frequency of the event \( \text{RND}_2^{} < (\text{RND}_1^{})^2 \) with Excel for

a. 10;

b. 100;

c. 1000.

experiments.

10. Studying relative frequencies A random number \( \text{RND} \) is multiplied by 6, and then the product is rounded up. The integer number we get is denoted by \( X. \) The possible values of \( X \) are clearly 1, 2, 3, 4, 5, 6. If the random number (given to 4 decimal places) is, for example, 0.2346, then \( 0.2346 \times 6 = 1.4076, \) so \( X = 2. \) Make 1000 experiments for \( X \) with Excel, and study the relative frequencies of the possible values, check that \( X \) takes the 6 possible values with equal probabilities.

11. Studying relative frequencies A random number \( \text{RND} \) is multiplied by 7, and then the product is rounded up. The integer number we get is denoted by \( X. \) The possible values of \( X \) are clearly 1, 2, 3, 4, 5, 6, 7. Make 1000 experiments for \( X \) with Excel, and study the relative frequencies of the possible values, check that \( X \) takes the 7 possible values with equal probabilities.

12. Random points in the unit square Play again with the random number generator of your calculator and/or of your computer.

a. Let both coordinates of a point defined by random numbers generated by the calculator or the computer.
   Make many experiments. Be convinced that the points are uniformly distributed on the unit square.

b. Let both coordinates of a point defined by the squares of random numbers generated by the calculator or
   the computer. Make many experiments. Be convinced that the points are not uniformly distributed on the
   unit square.

c. Let both coordinates of a point defined by the square roots of random numbers generated by the calculator
   or the computer. Make many experiments. Be convinced that the points are not uniformly distributed on the
   unit square.
d. Let the horizontal coordinate of a point defined by the square of a random number, and the vertical coordinate of that point defined by the square root of a random number. Make many experiments. Be convinced that the points are not uniformly distributed on the unit square.

13. Random points in the unit square
Two random numbers are generated by a calculator or computer: \( R_{\text{ND}_1} \) and \( R_{\text{ND}_2} \). Calculate the following probabilities. Check your results by simulation. (Remember: the relative frequency of an event should be close to the probability of the event if the number of experiments is large.)

a. \( P( R_{\text{ND}_1} < 2 R_{\text{ND}_2} ) \);

b. \( P( R_{\text{ND}_1} R_{\text{ND}_2} > 0.6 ) \);

c. \( P( ( R_{\text{ND}_1} )^2 < R_{\text{ND}_2} ) \);

d. \( P( R_{\text{ND}_1} R_{\text{ND}_2} < 0.6 ) \);

e. \( P( 0.3 < R_{\text{ND}_1} R_{\text{ND}_2} < 0.6 ) \);

f. \( P( R_{\text{ND}_2} / R_{\text{ND}_2} < 1/3 ) \);

g. \( P( R_{\text{ND}_2} / R_{\text{ND}_2} > 2 ) \).

14. Random points in the unit square
Two random numbers are generated by a calculator or computer: \( R_{\text{ND}_1} \) and \( R_{\text{ND}_2} \). Find the following probabilities:

a. \( P( R_{\text{ND}_2} < R_{\text{ND}_1} ) \);

b. \( P( R_{\text{ND}_2} < 2 R_{\text{ND}_1} ) \);

c. \( P( R_{\text{ND}_2} < 3 R_{\text{ND}_1} ) \);

d. \( P( R_{\text{ND}_2} < z R_{\text{ND}_1} ) \), where \( z \) is a positive parameter;

e. \( P( R_{\text{ND}_2} + R_{\text{ND}_1} < 1 ) \);

f. \( P( R_{\text{ND}_2} + R_{\text{ND}_1} < 0.5 ) \);

g. \( P( R_{\text{ND}_2} + R_{\text{ND}_1} < 1.5 ) \);

h. \( P( R_{\text{ND}_2} + R_{\text{ND}_1} < z ) \), where \( z \) is a positive parameter.

i. \( R_{\text{ND}_2} < ( R_{\text{ND}_1} )^2 \)

j. \( R_{\text{ND}_2} ( R_{\text{ND}_1} )^2 < 0.09 \)

15. First decimal after the decimal point
A random number \( R_{\text{ND}} \) is generated by a computer. Let us denote the first decimal after the decimal point by \( X \). If the random number (given to 10 decimal places) is, for example, 0.2346600914, then \( X \) is equal to 2.

a. What is the probability that \( X = 5 \)?

b. What is the probability that \( 3 < X < 7 \)?

c. What is the probability that \( X > 6 \)?

16. First and the second decimals after the decimal point
A random number is generated by a computer. Let us denote the first and the second decimals after the decimal point by \( X \) and \( Y \). If the random number is given to 10 decimal places, for example, is 0.5346600914, then \( X \) is 5 and \( Y \) is 3.
a. What is the probability that \( X \leq 7 \)?

b. What is the probability that \( Y \leq 4 \)?

c. What is the probability that \( X \leq 7 \) and \( Y \leq 4 \)?

### 3.12. 2.3 Classical problems

### 3.13. EXCEL

The RANDBETWEEN function was introduced above in Section 2. For example, the function

\[ =\text{RANDBETWEEN}(1;6) \]

returns an integer number greater than or equal to 1 and smaller than or equal to 6 so that each of the 6 possible outcomes have the same probability. Thus, if you want to simulate a classical problem which has, for example, 720 equally probable outcomes, then you may use the function

\[ =\text{RANDBETWEEN}(1;720) \]

Clearly, the formula

\[ =\text{ROUNDUP}(720*\text{RAND}();0) \]

has the same effect.

### 3.14. PROBLEMS

### 3.15. Combinatorial exercises

1. Permutations without repetition Using the file given below be convinced that the number of permutations without repetition of \( n \) different elements is

\[ n! \]

http://www.math.bme.hu/~vetier/dl/eg-010-05-10_Permutations_without_repetition.xls Demonstration file: Permutations without repetition eg-010-05-10

2. Permutations with repetitions Assume that we have a collection of letters so that there are \( r \) types of letters. Consider \( k_1 \) elements of the 1\textsuperscript{st} type, \( k_2 \) elements of the 2\textsuperscript{nd} type, \( k_3 \) elements of the 3\textsuperscript{rd} type, \ldots, \( k_r \) elements of the \( r \)\textsuperscript{th} type. The number of all elements is denoted by \( n \). Clearly, \( n = k_1 + k_2 + k_3 + \ldots + k_r \). Using the file given below be convinced that the number of permutations with repetitions is

\[ \frac{n!}{k_1!k_2!\ldots k_r!} \]

http://www.math.bme.hu/~vetier/dl/eg-010-05-20_Permutations_with_repetition.xls Demonstration file: Permutations with repetition eg-010-05-20

3. Variations with repetition Using the file given below be convinced that the number of variations with repetition of \( n \) different elements when \( k \) elements are taken is

\[ \frac{n!}{(n-k)!} \]
4. Variations with repetition Using the file given below be convinced that the number of variations with repetition of \( n \) different elements when \( k \) elements are taken is

\[ n^k \]


5. Combinations without repetition Using the file given below be convinced that the number of combinations without repetition of \( n \) different elements when \( k \) elements are taken is \( \binom{n}{k} \). Remember that the definition of the binomial coefficient is:

\[ \binom{n}{k} = \frac{n!}{k!(n-k)!} \]


6. Combinations with repetition Using the file given below be convinced that the number of combinations with repetition of \( n \) different elements when \( k \) elements are taken is the same as the number of combinations without repetition of \( n + (k-1) \) different elements when \( k \) elements are taken, that is

\[ \binom{n+1}{k-1} \]


7. Tickets with the numbers 1, 2, 3, 4, 5 You have 5 tickets with a number on each: 1, 2, 3, 4, 5. How many possibilities are there
   a. to arrange the 5 tickets into different orders (permutations)?
   b. to choose 3 of them if order is not important and repetition is not permitted: each digit can be used at most once?
   c. to choose 3 of them if order is important and repetition is not permitted: each digit can be used at most once?
   d. to choose 3 of them if order is important and repetition is permitted: each digit may be used more than once?

8. Tickets with the numbers 1, 2, \ldots, n You have \( n \) tickets with a number on each: 1, 2, \ldots, \( n \). How many possibilities are there
   a. to arrange the \( n \) tickets into different orders (permutations)
   b. to choose \( k \) of them if order is not important and repetition is not permitted: each digit can be used at most once?
   c. to choose \( k \) of them if order is important and repetition is not permitted: each digit can be used at most once?
   d. to choose \( k \) of them if order is important and repetition is permitted: each digit can be used more than once?

9. Three-decimal integers How many 3-decimal integers are there which consist only of the digits 1, 2, 3, 4, 5, 6, 7,
a. if repetition is not permitted: each digit can be used at most once?

b. if repetition is permitted: the digits may be used more than once?

### 3.16. Calculating probabilities

1. Fair die

   A fair die is rolled.

   a. What is the probability that we get 6?
   b. What is the probability that we get a number smaller than 6?
   c. What is the probability that we get a number greater than 2?
   d. What is the probability that we get a number greater than 2 and smaller than 6?
   e. Construct a formula in Excel to simulate a fair die.

2. Fair dice

   Two fair dice are tossed, a red and a blue.

   a. What is the probability that we get the pair red 2 and blue 6?
   b. What is the probability that we get the pair red 6 and blue 2?
   c. What is the probability that we get a 2 and a 6?
   d. What is the probability that we get a red number less than 3 and a blue number less than 5?
   e. What is the probability that we get at least one 6?
   f. Simulate in Excel as if two fair dice were tossed.

3. Fair dice

   Two fair dice are tossed, a red and a blue.

   a. What is the probability that the maximum of the two numbers is 2?
   b. What is the probability that the maximum of the two numbers is 5?
   c. What is the probability that the maximum of the two numbers is \( \frac{5}{6} \)? (Give a formula)
   d. Simulate in Excel as if two fair dice were tossed and then take their maximum.

4. Fair dice

   Two fair dice are tossed, a red and a blue.

   a. What is the probability that the sum of the two numbers is 5?
   b. What is the probability that the sum of the two numbers is 7?
   c. What is the probability that the sum of the two numbers is 9?
   d. What is the probability that the sum of the two numbers is \( \frac{5}{6} \)? (Give a formula)
   e. Simulate in Excel as if two fair dice were tossed and then take their sum.

5. Fair dice

   Two fair dice are tossed, a red and a blue. Let the "difference" mean "larger minus smaller".

   a. What is the probability that the difference between the two numbers is 0?
   b. What is the probability that the difference between the two numbers is 1?
   c. What is the probability that the difference between the two numbers is 2?
   d. What is the probability that the difference between the two numbers is 3?
e. What is the probability that the difference between the two numbers is 4?
f. What is the probability that the difference between the two numbers is 5?
g. What is the probability that the difference between the two numbers is $k$? (Give a formula)
h. Simulate in Excel as if two fair dice were tossed and then take their difference "larger minus smaller".

6. Fair dice Two fair dice are tossed, a red and a blue. Let the "difference" mean "red minus blue".
   a. What is the probability that the difference between the two numbers is 0?
   b. What is the probability that the difference between the two numbers is 1?
   c. What is the probability that the difference between the two numbers is 2?
   d. What is the probability that the difference between the two numbers is 3?
   e. What is the probability that the difference between the two numbers is 4?
   f. What is the probability that the difference between the two numbers is 5?
   g. What is the probability that the difference between the two numbers is -1?
   h. What is the probability that the difference between the two numbers is -2?
   i. What is the probability that the difference between the two numbers is -3?
   j. What is the probability that the difference between the two numbers is -4?
   k. What is the probability that the difference between the two numbers is -5?
   l. What is the probability that the difference between the two numbers is $k$? (Give a formula)
m. Simulate in Excel as if two fair dice were tossed and then take their difference "red minus blue".

7. Fair coin tossed three times A fair coin is tossed 3 times.
   a. What is the probability that we get 0 heads?
   b. What is the probability that we get 1 head?
   c. What is the probability that we get 2 heads?
   d. What is the probability that we get 3 heads?
   e. What is the probability that we get $k$ heads? (Give a formula)
   f. Simulate in Excel as if a fair coin were tossed 3 times and the number of heads were observed.

8. Fair coin tossed five times A fair coin is tossed 5 times.
   a. What is the probability that we get 0 heads?
   b. What is the probability that we get 1 head?
   c. What is the probability that we get 2 heads?
   d. What is the probability that we get 3 heads?
   e. What is the probability that we get 4 heads?
   f. What is the probability that we get 5 heads?
g. What is the probability that we get \( k \) heads? (Give a formula)

h. Simulate in Excel as if a fair coin were tossed 5 times and the number of heads were observed.

9. Fair coin tossed \( T \) times
   A fair coin is tossed \( T \) times.
   a. What is the probability that we get 0 heads?
   b. What is the probability that we get 1 head?
   c. What is the probability that we get 2 heads?
   d. What is the probability that we get \( k \) heads? (Give a formula)
   e. Simulate in Excel as if a fair coin were tossed \( T \) times and the number of heads were observed.

10. Fair die tossed three times
    A fair die is tossed 3 times. (Ace means tossing six.)
    a. What is the probability that we get 0 aces?
    b. What is the probability that we get 1 ace?
    c. What is the probability that we get 2 aces?
    d. What is the probability that we get 3 aces?
    e. What is the probability that we get \( k \) aces? (Give a formula)
    f. Simulate in Excel as if a fair die were tossed 3 times and the number of aces were observed.

11. Fair die tossed \( T \) times
    A fair die is tossed \( T \) times
    a. What is the probability that we get 0 aces?
    b. What is the probability that we get 1 ace?
    c. What is the probability that we get \( k \) aces? (Give a formula)
    d. Simulate in Excel as if a fair die were tossed \( T \) times and the number of aces were observed.

12. Two children
    Assume that each newborn baby is a boy or a girl with probabilities 0.5 - 0.5. Assume that you know that there are two children in a family, but you do not know whether the children are boys or girls. What is the probability that
    a. both children are boys;
    b. the children are of the same sex;
    c. the eldest child is a boy?
    d. Simulate in Excel two newborn babies.

13. "Five out of ninety" -lottery
    There are 90 tickets in a box: 1, 2, 3, ... 89, 90. We choose 5 of them without replacement.
    a. What is the probability that the number 55 is among the chosen tickets?
    b. What is the probability that the numbers 55 and are among the chosen tickets?

14. "Five out of ninety" -lottery
    There are 90 tickets in a box: 1, 2, 3, ... 89, 90. We choose 5 of them without replacement.
    a. What is the probability that the biggest is 55?
b. What is the probability that the biggest is 56?

c. Let $X$ denote the biggest number of the chosen ones. Find the probability that $X$ is equal to $k$. (Give a formula)

15. "Five out of ninety" -lottery There are 90 tickets in a box: 1, 2, 3, ..., 89, 90. We choose 5 of them without replacement.

a. What is the probability that the second biggest is 55?

b. What is the probability that the second biggest is 56?

c. Let $X$ denote the second biggest number of the chosen ones. Find the probability that $X$ is equal to $k$. (Give a formula)

16. "Five out of ninety" -lottery There are 90 tickets in a box: 1, 2, 3, ..., 89, 90. We choose 5 of them without replacement.

a. What is the probability that the third biggest is 55?

b. What is the probability that the third biggest is 56?

c. Let $X$ denote the third biggest number of the chosen ones. Find the probability that $X$ is equal to $k$. (Give a formula)

17. "Five out of ninety" -lottery In a box, there are 90 tickets with the numbers 1, 2, ..., 90 on them. We draw 5 tickets without replacement. Let $X$ mean the smallest number we draw, and let $Y$ mean the second largest. Find the probabilities:

a. $P(X = 27$ and $Y = 67)$;

b. $P(X = 17$ and $Y = 77)$;

c. $P(X = x$ and $Y = y$). (Give a formula)

18. Red and blue balls in a box, drawing with replacement There are 24 balls in a box. 15 of them are red, 9 of them are blue. We draw 6 of them with replacement. What is the probability that among the chosen balls the number of read is

a. exactly 2;

b. less than 2;

c. less than or equal to 2;

d. more than 2?

19. Two teams of five - five people Suppose that 10 people are divided in a random manner into two teams of 5 - 5 people. What is the probability that two particular people A and B will be in the same team?

3.17. 2.4 Geometrical problems, uniform distributions

3.18. EXCEL

Simulating a one-dimensional geometrical problem. Since the RAND function returns a random number which is uniformly distributed between 0 and 1, it can be effectively used to simulate geometrical problems. For example, the

=3*RAND()
formula returns a uniformly distributed number between 0 and 3. See the following file:

http://www.math.bme.hu/~vetier/df/eg-010-30-01_Simulating_a_one-dimensional_geometrical_problem.xls
Demonstration file: Simulating a one-dimensional geometrical problem eg-010-30-01

Simulating a two-dimensional geometrical problem. Since different occurrences of the RAND function yield independent random numbers, the point with coordinates defined by the formulas

\[ \begin{align*}
&= 3 \times \text{RAND}() \\
&= 2 \times \text{RAND}()
\end{align*} \]

will be uniformly distributed on the rectangle with vertices \((0; 0), (3; 0), (3; 2), (0; 2)\). See the following file:

http://www.math.bme.hu/~vetier/df/eg-010-30-02_Simulating_a_two-dimensional_geometrical_problem.xls
Demonstration file: Simulating a two-dimensional geometrical problem eg-010-30-02

### 3.19. PROBLEMS

1. Uniformly distributed random number We choose a uniformly distributed random number between 0 and 1.
   a. What is the probability that the number is smaller than 0.25?
   b. What is the probability that the number is smaller than \(c\)?
   c. What is the probability that the number is greater than 0.25?
   d. What is the probability that the number is greater than \(c\)?

2. Uniformly distributed random number We choose a uniformly distributed random number between 0 and 5.
   a. What is the probability that the number is smaller than 1.5?
   b. What is the probability that the number is smaller than \(c\)?
   c. What is the probability that the number is greater than 1.5?
   d. What is the probability that the number is greater than \(c\)?

3. Uniformly distributed random numbers We choose two independent uniformly distributed random numbers between 0 and 1.
   a. What is the probability that the larger is greater than 0.25?
   b. What is the probability that the larger is greater than \(c\)?
   Hint: Use the unit square as a sample space. Find the set of favorable outcomes. Calculate the probability as a ratio of certain areas.

4. Right rectangle, length of legs are random numbers We choose two independent uniformly distributed random numbers between 0 and 1, and construct a right rectangle whose legs are equal to the chosen random numbers.
   a. What is the probability that the perimeter of this rectangle is larger than 2?
   b. What is the probability that the area of this rectangle is smaller than 0.25?
   c. What is the probability that the perimeter of this rectangle is larger than 2 and the area of this rectangle is smaller than 0.25?
   d. What is the probability that the perimeter of this rectangle is larger than \(x\)? What is the probability that the area of this rectangle is smaller than \(y\)?
e. What is the probability that the perimeter of this rectangle is larger than $x$ and the area of this rectangle is smaller than $y$?

5. Line segment divided into three parts A line segment is divided into three parts by two independently chosen, uniformly distributed random points. What is the probability that center part is the shortest of the 3 parts? Hint: Use a square as a sample space. Find the set of favorable outcomes. Calculate the probability as a ratio of certain areas.

6. Uniformly distributed random numbers We choose two independent uniformly distributed random numbers between 0 and 5.
   a. What is the probability that the larger is greater than 2.5?
   b. What is the probability that the larger is greater than $c$?
   Hint: Use a square as a sample space. Find the set of favorable outcomes. Calculate the probability as a ratio of certain areas.

7. Shooting at a circular target Somebody is shooting at a circular target which has a radius of 1 meter. The target is divided into 5 parts by 4 circles of radii 0.2, 0.4, 0.6, 0.8 meters. Assuming that the probability of any subset is equal to the area of the subset divided by the area of the whole target, find the probability
   a. of the circle with radius 0.2;
   b. of each ring-like region.

8. Random point in a rectangle A random point $(X, Y)$, uniformly distributed in the rectangle with vertices $(0, 0), (2, 0), (2, 3), (0, 3)$, is considered. Calculate the probabilities:
   a. $P(Y < X)$;
   b. $P(Y < X^2)$;
   c. $P(Y/X < 1.5)$;
   d. $P(XY < 1.5)$.

9. Random point in a rectangle A random point $(X, Y)$, uniformly distributed in the rectangle with vertices $(0, 0), (2, 0), (2, 3), (0, 3)$, is considered. Give a formulas for the following probabilities:
   a. $P(Y/X < z)$;
   b. $P(XY < z)$.
   (z is a parameter)

10. Buffon's needle problem Find the probability of an intersection when the distance between the parallel lines is equal to the length of the needle.

11. Buffon's needle problem Find the probability of an intersection when the distance between the parallel lines is equal to the half of the length of the needle.

12. Bertrand's problem Choose a point on a radius of a circle according to uniform distribution on that radius. Then consider the chord passing that point and perpendicular to the radius. What is the probability that the chord is "long" (where "long" means longer than the length of a side of a regular triangle drawn into the circle). Hint: Use the circle as a sample space. Find the set of favorable outcomes. Calculate the probability as a ratio of some areas.

13. Diameters of tomatoes The diameter of a first class tomato sold in a certain shop is uniformly distributed between 6 and 9 cm, the diameter of a second class tomato is uniformly distributed between 4 and 7 cm. (The diameters of the tomatoes are independent.) What is the probability that the diameter of a second class tomato is larger than the diameter of a first class tomato?
3.20. 2.5 Basic properties of probability

3.21. PROBLEMS

1. Rain in towns A, B The probability that it will be raining in town A during a week is 3/4. The probability that it will be raining neither in town A nor in town B during a week is 1/16. The probability that it will be raining both in town A and town B during a week is 11/16. Find the probabilities:
   a. What is the probability that it will be raining in town A but not in town B during a week?
   b. What is the probability that it will not be raining in town B during a week?
   c. What is the probability that it will be raining in town B during a week?

2. Three dice
   a. What is the probability that all the three numbers are less than or equal to 5?
   b. What is the probability that all the three numbers are less than or equal to 4?
   c. What is the probability that the largest value on the three dice is 5?

3. Ten dice
   a. What is the probability that all the ten numbers are less than or equal to 5?
   b. What is the probability that all the ten numbers are less than or equal to 4?
   c. What is the probability that the largest value on the ten dice is 5?

4. Upper and lower estimates Assume that \( P(A) = 0.6 \) and \( P(B) = 0.7 \). Give upper and lower estimates for the probabilities:
   a. \( P( A \cap B ) \);
   b. \( P( A \cup B ) \);
   c. \( P( A \backslash B ) \).

3.22. 2.6 Conditional relative frequency and conditional probability

3.23. PROBLEMS

1. Numbers between 1 and 6 Choose a number between 1 and 6 at random (1 and 6 are included) so that each has the same probability. Then choose a second number also at random, but now between 1 and the first number so that each of these values have the same probability.
   a. What is the probability that the second number is 3?
   b. Make a simulation with 1000 experiments and check that the relative frequency is close to the probability.
   c. Assume that that the second number is 3.
      i. List the possible values of the first number under this condition?
ii. What is the probability that the first number is 5?

iii. Make a simulation with 1000 experiments and check that the conditional relative frequency is close to the conditional probability.

iv. What is the probability that the first number is k?

2. Election In a certain city, 30 percent of the people are Conservatives, 50 percent are Liberals, and 20 percent are Independents. Records show that in a particular election 65 percent of the Conservatives voted, 82 percent of the Liberals voted, and 50 percent of the Independents voted. If a person in the city is selected at random and it is learnt that he did not vote in the last election, what is the probability that he is a Liberal?

3. Rain and wind On a certain day, the probability of rain is 0.3, the probability of wind is 0.4. If it rains that day, then the probability of wind is 0.1.

   a. What is the probability of rain, if there is no wind that day?

   b. Assume that the above facts are true both for October 10 and November 10, and the whether conditions are independent for these two days. What is the probability that both on October 10 and November 10 there will be both rain and wind?

4. Rain in towns A, B The probability that it will be raining in town A during a week is 3/4. The probability that it will be raining neither in town A nor in town B during a week is 1/16. The probability that it will be raining both in town A and town B during a week is 11/16. Find the probabilities:

   a. On condition that it will be raining in town A what is the probability that it will be raining in town B during a week?

   b. On condition that it will not be raining in town A what is the probability that it will be raining in town B during a week?

   c. On condition that it will be raining in town B what is the probability that it will be raining in town A during a week?

   d. On condition that it will not be raining in town B what is the probability that it will be raining in town A during a week?

5. Two dice Two fair dice are tossed, a red and a blue.

   a. What is the probability that the red number is less than the blue number?

   b. On condition that the red number is less than the blue number what is the probability that the red number is 4?

   c. Make 50 or 100 or more experiments, write down the results, and calculate the conditional relative frequency. Be convinced that the conditional relative frequency is close to the conditional probability.

   d. Make 1000 simulations to check that the conditional relative frequency is close to the conditional probability.

   e. On condition that the red number is less than the blue number what is the probability that the red number is k?

   f. Replace the expression "less than" by the expression "less than or equal to" in the above sentences, and answer the questions.

6. Two dice Two fair dice are tossed, a red and a blue.

   a. On condition that the sum is greater than 5 what is the probability that the difference (in absolute value) is less than 2?

   b. On condition that the difference (in absolute value) is less than 2 what is the probability that the sum is greater than 5?
c. On condition that the red number is less than the blue number what is the probability that the red number is 1?

d. On condition that the red number is less than the blue number what is the probability that the red number is 2?

e. On condition that the red number is less than the blue number what is the probability that the red number is 3?

f. On condition that the red number is less than the blue number what is the probability that the red number is 4?

g. On condition that the red number is less than the blue number what is the probability that the red number is 5?

7. Two coins Tossing 2 coins assume that at least one head occurs.
   a. Under this condition what is the probability that tail occurs, too?
   b. Make 50 or 100 or more experiments, write down the results, and calculate the conditional relative frequency. Be convinced that the conditional relative frequency is close to the conditional probability.
   c. Make 1000 simulations to check that the conditional relative frequency is close to the conditional probability.

8. Three coins Tossing 3 coins assume that at least one head occurs.
   a. Under this condition what is the probability that tail occurs, too?
   b. Under this condition what is the probability that exactly 1 tail occurs?
   c. Under this condition what is the probability that exactly 2 tails occur?
   d. Make 50 or 100 or more experiments, write down the results, and calculate the conditional relative frequency. Be convinced that the conditional relative frequency is close to the conditional probability.
   e. Make 1000 simulations to check that the conditional relative frequency is close to the conditional probability.

9. Three coins Tossing 3 coins assume that at least one head occurs.
   a. Under this condition what is the probability that tail occurs, too?
   b. Under this condition what is the probability that exactly 1 tail occurs?
   c. Under this condition what is the probability that exactly 2 tails occur?

10. Two boxes There are two boxes: Box I contains 3 red and 7 white balls, Box II contains 8 red and 2 white balls. You get one of the boxes, you do not know which of them, but you know that the probability of Box I is 0.4, the probability of Box II is 0.6. When the box is given to you, you pick from that box 15 times with replacement.
   a. On condition that you pick red 9 times, what is the probability that Box I has been given to you?
   b. On condition that you pick red \( k \) times, what is the probability that Box I has been given to you? Find a formula and construct a table for the conditional probability with \( k = 1, 2, \ldots, 15 \). (To construct the table try to use Excel, and print out the table.)

11. Colored boxes with colored balls There are two boxes: a red and a blue. In the red box, there are 3 red and 2 blue balls. In the blue box, there are 3 red and 7 blue balls. First, we pick a ball from the red box, and put it into the blue box. Then we pick a second ball from the blue box, and put it into the red box. Finally, we pick a third ball from the red box again.
a. What is the probability that the first ball is red and the second is blue?

b. What is the probability that the second ball is blue?

c. What is the probability that the first ball is red, the second is blue and the third ball is red again?

d. What is the probability that the second is blue and the third is red?

e. What is the probability that both the first and the second balls are red?

f. What is the probability that both the second ball is red?

g. What is the probability that all the three balls are red?

h. What is the probability that both the second and the third balls are red?

i. What is the probability that the third ball is red?

j. On condition that the third ball is red, what is the probability that the second ball is red, too?

k. On condition that the third ball is red, what is the probability that the second ball is red, but the first was blue?

l. On condition that the third and second balls were red, what is the probability that the first is blue?

12. Right-handed and left-handed people In a certain country, 60 percent of the population is right-handed, 40 percent is left-handed. A right-handed person is able to hit a certain target with his left hand with a probability 0.2. For a left-handed person this probability is definitely larger, it is 0.7. A person is chosen at random.

a. What is the probability that a person hits the target with his/her left hand?

b. Assume that you get the information that a person hit the target with the left hand. What is the probability that that person is left-handed?

13. Two children Assume that each newborn baby is a boy or a girl with probabilities 0.5 - 0.5. Assume that you know that there are two children in a family, but you do not know whether the children are boys or girls. What is the probability that

a. the eldest child is a boy on condition that there is a boy (that is, one or two boys) in the family?

b. Simulate in Excel two newborn babies. Make 1000 experiments and select those results where there is a boy in the family. Check what is the proportion of those results where the eldest child is a boy among those results where there is a boy (that is, one or two boys) in the family.

14. Two children Assume that each newborn baby is a boy or a girl with probabilities 0.5 - 0.5. Assume that you know that there are two children in a family, but you do not know whether the children are boys or girls. What is the probability that

a. there is a girl in the family on condition that there is a boy (that is, one or two boys) in the family?

b. Simulate in Excel two newborn babies. Make 1000 experiments and select those results where there is a boy in the family. Check what is the proportion of those results where there is a girl in the family among those results where there is a boy (that is, one or two boys) in the family.

15. Birthdays of two people 2 randomly chosen people are asked which month and day they have their birthdays. What is the probability that they have their birthdays on different days? (There are 365 day in a year.)

16. Birthdays of three people 3 randomly chosen people are asked which month and day they have their birthdays. What is the probability that they all have their birthdays on different days? (There are 365 day in a year.)
17. Birthdays of five people 5 randomly chosen people are asked which month and day they have their birthdays. What is the probability that they all have their birthdays on different days? (There are 365 days in a year.)

18. Birthday paradox 23 randomly chosen people are asked which month and day they have their birthdays. (There are 365 days in a year.) What is the probability that the birthdays are all different? Using Excel, construct a table and a figure to show how this probability depends on $n$, and be convinced that when $n = 23$, then it is less than 0.5. Remark. This result means: if 23 randomly chosen people get together, then the probability that there will be at least two of them who have their birthdays on the same day of the year is more probable than 0.5. This fact may seem surprising because 23 compared to 365 is a rather small number!

19. When does the first birthday coincidence occur? People chosen one after the other at random are asked which month and day they have their birthdays. We stop asking as soon as we get a birthday which has already been occurred before.

   a. What is the probability that we stop at the 2nd person?
   b. What is the probability that we stop at the 3rd person?
   c. What is the probability that we stop at the 4th person?
   d. What is the probability that we stop at the 5th person?
   e. What is the probability that we stop at the $k$th person?
   f. Find a formula, and - using Excel - construct a figure to show how this probability depends on $k$.
   g. For which $k$ value is this probability maximal?

20. Red and white balls There are 2 red and 3 white balls in a box. We draw 5 times so that each time not only the chosen ball is replaced, but 2 balls of the same color is also put into the box. What is the probability that the colors of the chosen balls alternate? On condition that the colors of the chosen balls alternate, what is the probability that the first is red?

21. Red and white balls There are 2 red and 3 white balls in a box. We draw 3 times so that each time not only the chosen ball is replaced, but 2 balls of the same color is also put into the box. What is the probability that the first is a red, the second is a white, and the third is a red?

   a. On condition that the 3rd ball is red, what is the probability that the 1st is white?
   b. On condition that the 3rd ball is red, what is the probability that the 2nd is white?
   c. On condition that the 3rd ball is red, what is the probability that both the 1st and the 2nd are white?
   d. On condition that the 2nd and the 3rd ball is red, what is the probability that the 1st is white?

22. Two dice Two fair dice are tossed, a red and a blue.

   a. On condition that the sum is greater than 5 what is the probability that the difference (in absolute value) is less than 2?
   b. On condition that the difference (in absolute value) is less than 2 what is the probability that the sum is greater than 5?
   c. On condition that the red number is less than the blue number what is the probability that the red number is 1?
   d. On condition that the red number is less than the blue number what is the probability that the red number is 2?
   e. On condition that the red number is less than the blue number what is the probability that the red number is 3?
f. On condition that the red number is less than the blue number what is the probability that the red number is 4?

g. On condition that the red number is less than the blue number what is the probability that the red number is 5?

23. Hundred fair coins and one false coin There are 101 coins in my pocket. 100 coins are fair, 1 is false. False means that on both sides of this coin there are heads. We pick one out of the 101 coins, and without looking at the chosen coin we toss it

a. 5 times;

b. 10 times;

c. 20 times;

d. \( k \) times.

On condition that the we always get a head what is the probability that the false coin was picked?

24. \( n \) fair coins and one false coin There are \( n + 1 \) coins in my pocket. \( n \) are fair, 1 is false. False means that on both sides of this coin there are heads. We pick one out of the \( n + 1 \) coins, and without looking at the chosen coin we toss it \( k \) times. On condition that the we always get a head what is the probability that the false coin was picked?

25. My friend's pen My friend is a little bit scattered. When he takes a pen from his right pocket, then after using it, he places it back into his right pocket only with a probability 0.6 , and he puts it into his left pocket with a probability 0.4. When he takes a pen from his left pocket, then after using it, he places it back into his left pocket only with a probability 0.9 , and he puts it into his right pocket with a probability 0.1. When I gave him a new pen the last week, he put it into his right pocket. I know that since then he has used it exactly 3 times.

a. What is the probability that, having used it exactly 3 times, the pen is in his right pocket?

b. On condition that he has used it exactly 3 times what is the probability that the pen has never been in his left pocket?

26. Three boxes There are three boxes. The first contains 1 red ball and 1 white ball, the second contains 1 red and 2 white balls, the third contains 1 red and 3 white balls. First we chose a box so that each has the same chance, then from that box we pick one ball.

a. What is the probability that the ball is red?

b. On condition that the chosen ball is red, what is the probability that the first box was chosen?

27. Three boxes There are three boxes. The first contains 10 red and 10 white balls, the second contains 10 red and 20 white balls, the third contains 10 red and 30 white balls. First we chose a box so that each has the same chance, then from that box we pick 5 balls without replacement.

a. What is the probability that 1 of the balls is red and the other 4 are white?

b. On condition that 1 of the balls is red and the other 4 are white, what is the probability that the first box was chosen?

28. Is the number of draws even? There are four tickets in a box numbered from 1 to 4. We draw without replacement as many times as needed to get the ticket with the number 4. What is the probability that the number of draws is an even number?

29. Electrical components Suppose there are two electrical components in a machine. The probability that the first component fails during a day is 0.1. If the first component fails, the probability that the second component fails during the same day is 0.2. But if the first component works all the day, then the probability that the second component fails during that day is only 0.05. Calculate the probabilities of the events:

a. at least one of the components works all the day;
b. exactly one of the components works all the day;

c. the second components works all the day.

30. Fair and false coins in a hat A hat contains 10 coins, 7 of which are fair, and 3 of which are biased to land heads with probability 2/3. A coin is drawn from the hat and tossed twice. Assume that it lands both times. Given this information, what is the probability that it is a false coin?

31. Peter's exam Peter goes to an exam. Because of his unpreparedness, he may get a grade 3, 2, 1 with the same probability that is 1/3 for each. Grade 5 and 4, unfortunately, are impossible for him. If Peter gets a grade 3, then the probability that he arrives home with - a smiling face is 0.5; - a neutral face is 0.3; - an angry face is 0.2. If Peter gets a grade 2, then the probability that he arrives home with - a smiling face is 0.4; - a neutral face is 0.3; - an angry face is 0.3. If Peter gets a grade 1, then the probability that he arrives home with - a smiling face is 0.1; - a neutral face is 0.1; - an angry face is 0.8. What is the probability that the probability that he gets home with - a smiling face; - an angry face? If he gets home with a smiling face, then what is the probability that he got - a grade 3; - a grade 2; - a grade 1? If he gets home with a neutral face, then what is the probability that he got - a grade 3; - a grade 2; - a grade 1? If he gets home with an angry face, then what is the probability that he got - a grade 3; - a grade 2; - a grade 1? Make a simulation file, too, to make 1000 experiments (you may use the HLOOKUP function instead of many nested IF commands), and count how many times each of the following 9 cases hold: "he gets a grade 3, 2, 1" combined with "he arrives home with a smiling face, with a neutral face, with an angry face". Arrange the 9 cases into a 3 by 3 table. Construct another table about the associated 9 relative frequencies. Then construct two 3 by 3 tables about the conditional relative frequencies: - first when the grade values serve as conditions, - second when the states of his face serve as conditions.

32. Peter's exam Peter goes to an exam. He may get a grade 5, 4, 3, 2, 1 with the same probability that is 0.2 for each. If Peter gets a grade 5, then - since other influences exist, too - the probability that he arrives home with - a smiling face is 0.7; - a neutral face is 0.2; - an angry face is 0.1. If Peter gets a grade 4, then the probability that he arrives home with - a smiling face is 0.6; - a neutral face is 0.3; - an angry face is 0.1. If Peter gets a grade 3, then the probability that he arrives home with - a smiling face is 0.5; - a neutral face is 0.3; - an angry face is 0.2. If Peter gets a grade 2, then the probability that he arrives home with - a smiling face is 0.4; - a neutral face is 0.3; - an angry face is 0.3. If Peter gets a grade 1, then the probability that he arrives home with - a smiling face is 0.1; - a neutral face is 0.1; - an angry face is 0.8. What is the probability that he gets home with - a smiling face; - a neutral face; - an angry face? If he gets home with a smiling face, then what is the probability that he got - a grade 5; - a grade 4; - a grade 3; - a grade 2; - a grade 1? If he gets home with a neutral face, then what is the probability that he got - a grade 5; - a grade 4; - a grade 3; - a grade 2; - a grade 1? If he gets home with an angry face, then what is the probability that he got - a grade 5; - a grade 4; - a grade 3; - a grade 2; - a grade 1? Make a simulation file, too, to make 1000 experiments (you may use the HLOOKUP function instead of many nested IF commands), and count how many times each of the following 15 cases hold: "he gets a grade 5, 4, 3, 2, 1" combined with "he arrives home with a smiling face, with a neutral face, with an angry face". Arrange the 15 cases into a 5 times 3 table. Construct another table about the associated 15 relative frequencies. Then construct two 5 times 3 tables about the conditional relative frequencies: - first when the grade values serve as conditions, - second when the states of his face serve as conditions.

3.24. 2.7 Independence of events

3.25. EXCEL

It is an important property of the RANDBETWEEN and RAND functions that different applications of them are independent of each other. Thus, if you want to simulate two independent events so that one of them has a probability 0.3, and the other has a probability 0.4, then the first event may be defined in a cell as

=IF( RAND() < 0.3 ; 1 ; 0 )

and the other in an other cells

=IF( RAND() < 0.4 ; 1 ; 0 )
3.26. PROBLEMS

1. Three independent events $A$, $B$, $C$ represent independent events so that $P(A) = 0.6$, $P(B) = 0.7$, $P(C) = 0.8$. Find the following probabilities:
   
   a. $P(A \cap B)$;
   b. $P(A \cup B)$;
   c. $P(A \cap \overline{B})$;
   d. $P(A \cup \overline{B})$;
   e. $P(A \setminus B)$;
   f. $P(A \cap B \cap C)$;
   g. $P(A \cap B \cap \overline{C})$;
   h. $P(\overline{A} \cap \overline{B} \cap \overline{C})$;
   i. $P(A \cup B \cup C)$.

2. Three independent events $A$, $B$, $C$ represent independent events so that $P(A) = 0.6$, $P(B) = 0.7$, $P(C) = 0.8$. Find the following conditional probabilities:
   
   a. $P(A \cap B | A \cup B)$;
   b. $P(A \cup B | A \cup B \cup C)$;
   c. $P(A \cap B \cap C | A \cup B \cup C)$;
   d. $P(\overline{A} \cap \overline{B} \cap \overline{C} | A \cup B \cup C)$;
   e. $P(A \cup B | A \cup B \cup C)$.

3. Bears and or monkeys The bakery where we buy our bread advertises that from April 1, children will get a small box containing a bear with probability 0.6 or a monkey with probability 0.4. The presents on different days are declared to be independent of each other. My little daughter has decided to go there with me every day to obtain the present. She likes bears, but she does not like monkeys.
   
   a. What is the probability that on the first 3 days of April she will get bears only?
   b. What is the probability that on the first 5 days of April she will get exactly 3 bears?
   c. What is the probability that on the first 5 days of April she will get at least 3 bears?
   d. What is the probability that on the first 5 days of April she will get at least 3 bears and on the following 5 days she will get less than 3 bears?
   e. What is the probability that on the first 5 days of April she will get more bears than on the following 5 days?
   f. What is the probability that she will get her first bear on April 3?
   g. What is the probability that she will get her third bear on April 3?
h. What is the probability that she will get her third bear on April 7?

i. What is the probability that she will get her third bear on April 7 and she will get her fifth bear on April 13?

j. What is the probability that on the first 5 days of April she will get exactly 3 bears and she will get her fifth bear on April 13?

4. Five people with red hats

Five people, call them 1, 2, 3, 4, 5, independently of each other put a red hat on their heads with certain (possibly different) probabilities. Clearly, there are $2^5 = 32$ possible variations. Construct an Excel file to list the 32 variations in 32 rows, and then figure out an Excel formula to find the probability of each variation. http://www.math.bme.hu/~vetier/df/Sol-01-09-12_Five_people_with_red_hats_Probabilities.xls Solution Sol-01-09-12

5. Ten people with red hats

Ten people, call them 1, 2, ..., 10, independently of each other put a red hat on their heads with certain (possibly different) probabilities. Clearly, there are $2^{10} = 1024$ possible variations. Construct an Excel file to list the 1024 variations in 1024 rows, and then figure out an Excel formula to find the probability of each variation. http://www.math.bme.hu/~vetier/df/Sol-01-09-13_Ten_people_with_red_hats_Probabilities.xls Solution Sol-01-09-13

6. Ice-cream

Adam, Barbara and Carl, independently of each other buy ice-cream each day in summer. The probability that Adam buys ice-cream a day is 0.5. The probability that Barbara buys ice-cream a day is 0.6. The probability that Carl buys ice-cream a day is 0.7.

a. Simulate a day to show that each of them buys ice-cream or does not.

b. What is the probability that all buy ice-cream a day?

c. What is the probability that at least one of them buys ice-cream a day?

d. What is the probability that Adam buys ice-cream on condition that at least one of them buys ice-cream a day?

http://www.math.bme.hu/~vetier/df/Sol-01-09-14_Ice-cream.xls Solution Sol-01-09-14

7. All colors occur

There are 10 red, 20 white and 30 green balls in a box. We draw 7 times with replacement, and observe the colors.

a. Simulate an experiment of the 7 draws.

b. What is the probability that all colors occur among the 7 draws?


3.27. 2.8 *** Infinite sequences of events

3.28. PROBLEMS

1. Is the number of tosses divisible by 5? My friend and I play with a fair coin. We toss it until the first time a head occurs. We agree that I win if the number of tosses is divisible by 5, that is, 5 or 10 or 15, ..., and my friend wins otherwise. What is the probability that I win? What is the probability that my friend wins?

2. Three persons playing with a coin

Three persons, say A and B and C play with a fair coin. They toss the coin until the first time a head occurs. The number of tosses to get the first head is denoted by $X$. They agree that A wins if $X$ is divisible by 3, that is, if $X = 3$ or $X = 6$ or $X = 9$, ..., B wins if $X + 1$ is divisible by 3, that is, if $X = 2$ or $X = 5$ or $X = 8$, ..., C wins if $X + 2$ is divisible by 3, that is, if $X = 1$ or $X = 4$ or $X = 7$, ...

a. What is the probability that A wins?

b. What is the probability that B wins?
c. What is the probability that C wins?

3. Three persons playing with a die

Three persons, say A and B and C play with a fair die. They toss the die until the first time a six occurs. The number of tosses to get the first head is denoted by $X$. They agree that A wins if $X$ is divisible by 3, that is, if $X = 3$ or $X = 6$ or $X = 9$, ...; B wins if $X + 1$ is divisible by 3, that is, if $X = 2$ or $X = 5$ or $X = 8$, ...; C wins if $X + 2$ is divisible by 3, that is, if $X = 1$ or $X = 4$ or $X = 7$, ...

a. What is the probability that A wins?
b. What is the probability that B wins?
c. What is the probability that C wins?

3.29. 2.9 *** Drawing with or without replacement. Permutations

3.30. PROBLEMS

1. Permutations of letters

How many orders (permutations) are there for the letters

a. A, B, C;
b. A, B, C, D, E, F, G;
c. A, B, C, D, D;
d. A, B, C, D, D, D;
e. A, A, A, B, B, C, D, D, D?

2. Drawing without replacement

Assume that $A$ red and $B$ green balls are in a box. We draw $n$ times without replacement. Prove that the probability that

$$x$$

times $red$
is equal to

$$\frac{\binom{A}{x} \binom{B}{n-x}}{\binom{A+B}{n}}$$

if $\max(0, n - B) \leq x \leq \min(n, A)$

3. Drawing with replacement

Assume that $A$ red and $B$ green balls are in a box. We draw $n$ times with replacement. Prove that the probability that

$$x$$

times $red$
is equal to
where $p = \frac{A}{A+B}$.

4. A limit statement
Prove the following limit theorem: For any fixed $\gamma$, $p$ and $x = 0, 1, 2, \ldots$, if

$$A \to \infty \quad \text{and} \quad B \to \infty \text{ so that } \frac{A}{A+B} \to p$$

then

$$\left( \frac{A}{(n-x)} \right) \frac{B}{(A+B)} \to \left( \frac{n}{x} \right) p^x (1-p)^{n-x}$$

5. Heuristic explanation
Try to give a heuristic explanation for the previous limit theorem. Hint: If the number of balls in the box is large compared to the number of times we draw from the box, then does it really count whether we draw without replacement or with replacement?

**4. 3 Discrete random variables and distributions**

**4.1. EXCEL**

The following file shows several ways to visualize a discrete distribution, study them. If you are a beginner either in Excel or probability theory, then you should first study and learn to construct the figures called "Vertical bars", "Dots with connected lines" and "Connected lines only" (pages 3, 7 and 8 in the file).

*Demonstration file: Methods of visualization of discrete distributions* ef-120-01-00

Curious readers, in order to look at the most important discrete distributions, may study the following file, as well.

*Demonstration file: Most important discrete distributions* ef-120-10-00

Demonstration file: Most important discrete distributions ef-120-10-00

**4.2. PROBLEMS**

1. Calculating probabilities by summation, using Excel
The distribution of a random variable $X$ is given by:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x)$</td>
<td>0.01</td>
<td>0.04</td>
<td>0.06</td>
<td>0.09</td>
<td>0.10</td>
<td>0.15</td>
<td>0.25</td>
<td>0.20</td>
<td>0.07</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Calculate the probability that the value of the random variable is 4 or 8. *Solution* Sol-02-01-08

2. Continuation of the previous problem
Invent other events related to the random variable $X$, and calculate their probabilities by summation.

3. Calculating probabilities from a distribution
The distribution of a random variable $X$ is given by:

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(k)$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

a. Calculate the probability that $3 \leq X \leq 6$.  

---

Created by XMLmind XSL-FO Converter.
b. Calculate the probability that $1 \leq X \leq 4$.

c. Calculate the probability that $3 \leq X \leq 6$ on condition that $1 \leq X \leq 4$.

d. Calculate the probability that $1 \leq X \leq 4$ on condition that $3 \leq X \leq 6$.

\textit{Solution} \cite{Sol-02-01-01}

4. Calculating probabilities from a distribution The distribution of a random variable $X$ is given by:

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>99</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(k)$</td>
<td>1/5050</td>
<td>2/5050</td>
<td>3/5050</td>
<td>...</td>
<td>99/5050</td>
<td>100/5050</td>
</tr>
</tbody>
</table>

a. Calculate the probability that $30 \leq X \leq 60$;

b. Calculate the probability that $10 \leq X \leq 40$;

c. Calculate the conditional probability $P(30 \leq X \leq 60 \mid 10 \leq X \leq 40)$;

d. Calculate the conditional probability $P(10 \leq X \leq 40 \mid 30 \leq X \leq 60)$.

\textit{Solution} \cite{Sol-02-01-02}

5. Calculating probabilities from a distribution Study the following file and observe how the outcomes are arranged: on page 1, they are listed in a column. On page 2, they constitute a rectangular table. The values of the probabilities are taken just from the air. The event $A$, whose probability is calculated, means that $X + Y$ is greater than 6 and less than 9. \textit{Solution} \cite{Sol-02-01-20}

6. Calculating probabilities related to lottery There are 90 tickets in a box: 1, 2, 3, ..., 89, 90. We choose 5 of them without replacement as in the so called "90 choose 5 lottery".

a. What is the probability that the biggest is 55?

b. What is the probability that the biggest is 56?

c. Let $X$ denote the biggest number of the chosen ones. Find the distribution of $X$.

7. Calculating probabilities related to lottery There are 90 tickets in a box: 1, 2, 3, ..., 89, 90. We choose 5 of them without replacement.

a. What is the probability that the second biggest is 55?

b. What is the probability that the second biggest is 56?

c. Let $X$ denote the second biggest number of the chosen ones. Find the distribution of $X$.

\textit{Solution} \cite{Sol-02-01-03}

8. Calculating probabilities related to lottery There are 90 tickets in a box: 1, 2, 3, ..., 89, 90. We choose 5 of them without replacement.

a. What is the probability that the third biggest is 55?

b. What is the probability that the third biggest is 56?

c. Let $X$ denote the third biggest number of the chosen ones. Find the distribution of $X$.

9. Largest and smallest of two dice Toss 2 dice, and let $X$ be the largest, and $Y$ be the smallest of the two numbers. Find out (make a numerical table for)

a. the distribution of $X$,

b. the distribution of $Y$.

\textit{Solution} \cite{Sol-02-01-04, Sol-02-01-05}
10. Absolute value of the difference Toss 2 dice, and observe the absolute value of the difference between the two numbers on the dice. Determine the distribution of this random variable. \textbf{Solution 1} \textbf{Solution 2}

11. Number of heads with 3 dice Tossing with 3 coins observe the number of heads. Find the distribution of this random variable.

12. Number of heads with 5 dice Tossing with 5 coins observe the number of heads. Find the distribution of this random variable.

13. Number of sixes with 4 dice Tossing with 4 dice observe the numbers sixes. Find the distribution of this random variable.

14. Number of draws There are 2 red and 5 blue balls in a box. We draw without replacement until the first red is drawn. Let $X$ denote the number of draws. Determine the distribution of $X$.

15. Five people with red hats Five people, call them $1, 2, 3, 4, 5$, independently of each other put a red hat on their heads with certain (possibly different) probabilities. Clearly, there are $2^5 = 32$ possible variations.
   a. Construct an Excel file to list the 32 variations in 32 rows, and then figure out an Excel formula to find the probability of each variation.
   \textbf{Solution}

16. Ten people with red hats Ten people, call them $1, 2, \ldots, 10$, around a circular table, independently of each other put a red hat on their heads with certain (possibly different) probabilities. Clearly, there are $2^{10} = 1024$ possible variations.
   a. Construct an Excel file to list the 1024 variations in 1024 rows, and then figure out an Excel formula to find the probability of each variation. \textbf{Solution}
   \textbf{Solution}

\section{4.3. 3.1 Uniform distribution (discrete)}

\section{4.4. EXCEL}

The following file shows how a discrete uniform distribution looks like.

\textbf{Demonstration file: Discrete uniform distribution}

\section{4.5. PROBLEMS}

1. Is it uniformly distributed? If $X$ is uniformly distributed on the set $\{-5, -4, \ldots, 4, 5\}$, and $Y = X^2$, then is $Y$ also uniformly distributed?

2. Is it uniformly distributed? If $X$ is uniformly distributed on the set $\{0, 1, \ldots, 9, 10\}$, and $Y = X^2$, then is $Y$ also uniformly distributed?
3. Make a correct conclusion What conclusion can you draw for $A$ and $B$ if it is true that $X$ is uniformly distributed on the set 
$$\{A, A + 1, \ldots, B - 1, B\}$$
and $Y = X^2$ is also uniformly distributed on some set?

### 4.6. 3.2 Hyper-geometrical distribution

### 4.7. EXCEL

The following file shows how a hyper-geometrical distribution looks like.

*emph*[Demonstration file: Hyper-geometrical distribution \eg-020-60-21]*

### 4.8. PROBLEMS

1. Number of red There are 24 balls in a box, 15 of them are red, 9 of them are blue. We draw 6 of them without replacement. What is the probability that among the chosen balls the number of red is
   a. exactly 2;
   b. less than 2;
   c. less than or equal to 2;
   d. more than 2?
*emph*[Solution \sol-02-03-14]*

2. Tickets numbered and colored There are 45 tickets in a box. The tickets are numbered from 1 to 45, and they are also colored: tickets with numbers 1 through 6 are red, the other 39 of them are blue. We draw 6 tickets without replacement. What is the probability that
   a. among these 6 tickets, the number of red is an odd number, that is 1 or 3 or 5?
   b. on these 6 tickets, the smallest number is the 10 and the largest is the 33?

### 4.9. 3.3 Binomial distribution

### 4.10. EXCEL

The following files show how a binomial distribution looks like. The first file should be studied for $n = 1, 2, \ldots, 20$
, the second for $n = 20, 21, \ldots, 100$
.

*emph*[Demonstration file: Binomial distribution, when $n$ is small \eg-020-60-01]*

*emph*[Demonstration file: Binomial distribution, when $n$ is large \eg-020-60-02]*

### 4.11. PROBLEMS

1. Drawing with or without replacement There are 25 balls in a box, 10 of them are red, 15 of them are blue. We draw 7 of them
a. with replacement;

b. without replacement.

What is the probability that among the chosen balls the number of read is less than 2?

2. Tossing a die Toss a die 60 times. What is the probability that the number of sixes is

a. greater than or equal to 12;

b. less than or equal to 8;

c. strictly between 8 and 12?

3. Blue eyed girls Assume that 3/4 of girls in a country have blue eyes. If you choose 20 girls at random in that country, then what is the probability that

a. exactly 15 of them have blue eyes;

b. exactly 16 of them have blue eyes;

c. exactly 17 of them have blue eyes;

d. more than 17 of them have blue eyes;

e. less than 17 of them have blue eyes;

f. the number of blue eyed girls is between 15 and 17 (equality permitted)?

4. Teachers becoming sick There are 70 teachers in our institute. Each teacher, independently of the others, may become sick during a day with probability 0.04. What is the probability that \( k \) of them become sick during a day? Make a table and a figure - using binomial distribution - so that \( k \) runs from 0 to 20. {emph}[Solution Sol-02-04-19]

5. Tickets in a box There are \( N \) tickets in a box, numbered from 1 to \( N \). Assume that \( N = A + B \). The first \( A \) tickets are red, the other \( B \) are blue, that is, the tickets 1, 2, \ldots, \( A \) are red, and \( A + 1, A + 2, \ldots, A + B \) are blue. We draw \( n \) times with replacement. What is the probability that we draw a red ticket exactly \( k \) times? For what \( k \) values is this probability greater than 0? For which \( k \) value is the probability maximal, that is, where is the mode of the distribution?

6. Students attending or not Assume that each of 20 students attend a lesson independently of each other so that each attends with a probability 0.8. What is the probability that:

a. all attend;

b. nobody attends;

c. exactly 15 attend;

d. more than 15 attend;

e. less than 15 attend?

7. How many chairs? Assume that each of 400 students attend a lesson independently of each other so that each attends with a probability 0.6. If the number of attending students is less than or equal to the number of chairs, then all students will have a chair to sit on. If the number of attending students is greater than the number of chairs, then some students will not have a chair to sit on, which is clearly a problem. For a given number of chairs the event we may ask: how much is the probability that no problem arises. It is obvious that in order hat this probability be equal to 1, the number of chairs must be 400 (or greater). Now the question is: how many chairs are needed so that this probability be (approximately) equal to 0.99? {emph}[Solution Sol-02-04-22]

8. Air plane tickets Assume that there are 200 seats on an airplane. Since passengers may miss the flight, in order to get some extra profit the air line sells 5 extra tickets. Now assume that each of the 205 passengers
miss the flight independently of each other with a probability 0.05. If the number of missing passenger is less than or equal to 5, then the number of passengers being at the flight is greater than 200 so some of them will not have a chair seat on the plane, and thus a problem arises.

a. What is the probability that problem will arise?

b. Replace the probability 0.05 and the number of extra tickets by other values, and analyze how the probability that "problem arises" depends on the probability of missing the plane and on the number of extra seats.

c. Make an analysis how the probability that "problem arises" depends on the size of the plane.

\emph{Solution}\ Sol-02-04-23

9. Chess players Assume that each of 15 members of a group of chess players attend the club independently of each other so that each attends with a probability 0.8. If the number of attending members is odd, then one of them will not have a partner, and will be bored. What is the probability the number of attending members will be even, and so nobody will be bored?

10. Opinion survey From the respondents who participated in an opinion survey, 50 percents declared to favor a unicameral parliament, 40 percents declared to favor a bicameral parliament, and 10 percents did not answer. 400 respondents from the interviewed sample are randomly chosen (For simplicity, you may assume: with replacement). What is the probability that exactly 200 of them were in favor of a unicameral parliament?

11. Drawing with replacement If there are 10 balls in a box so that 8 are red and 2 are blue, and you draw 20 times with replacement, than how much is the probability that

a. you draw 14 times a red;

b. you draw less than 14 times a red;

c. you draw more than 14 times a red;

d. the number of red you draw is more than 12 and less than 18 times a red?

12. Drawing with replacement If there are 40 balls in a box so that 8 are red and 32 are blue, and you draw 20 times with replacement, than how much is the probability that

a. you draw 6 times a red;

b. you draw less than 6 times a red,

c. you draw more than 6 times a red;

d. the number of red you draw is more than 2 and less than 8 times a red?

13. Computer system braking down Assume that at a university, on each day, independently of other days, the central computer system brakes down with a probability 0.01.

a. What is the probability that during a year (365 days) the Central Computer System system never brakes down?

b. What is the probability that during a year there are more than 2 days when the Central Computer System system brakes down?

14. Computers infected by viruses There are 50 computers in an office. Each of them, independently of the others, may become infected by viruses during a day with probability 0.05. What is the probability that \( k \) of them become infected by viruses during a day? Make a table and a figure - using binomial distribution - so that \( k \) runs from 0 to 20.

15. Computers infected by viruses but still working There are 12 computers in an office. Each of them, independently of the others, has a virus with a probability 0.6. Each computer which has a virus still works, independently of the others, with a probability 0.7. The number of computers having a virus, but still working is a random variable, which we denote by \( W \). Calculate the following probabilities:
Exercise Book to “Probability Theory with Simulations”

a. $P(W = 5)$;

b. $P(W = k)$ for $j = 0, 1, \ldots, 12$.

c. Be convinced that the distribution you get is a binomial distribution with parameters 12, and 0.42. (Use Excel to make these calculations.)

d. Simulate $W$.

16. Which box was used? There are two boxes: a red and a blue. In the red box there are 3 balls: 1 black and 2 white. In the blue box there are 5 balls: 3 black and 2 white. One of the boxes is chosen at random so that

a. each has the 50 percent chance;

b. the red box is 5 times probable than the blue.

Then a ball is picked 20 times (with replacement) from that box. Assume that you get the information: how many times a black ball was picked, but you do not know which box was used. How would you guess the color of the box from the given information? *Solution*

### 4.12. 3.4 Geometrical distribution (pessimistic)

#### 4.13. EXCEL

The following file shows how a pessimistic geometrical distribution looks like.

*Demonstration file: Pessimistic geometrical distribution eg-020-60-12*

#### 4.14. PROBLEMS

1. Finding the first blue eyed girl Assume that 3/4 of girls in a country have blue eyes. If you choose girls at random until the first blue eyed one, then what is the probability that

   a. there are 2 not blue eyed girls before the first blue eyed girl;

   b. there are 3 not blue eyed girls before the first blue eyed girl;

   c. there are 2 or 3 not blue eyed girls before the first blue eyed girl.

   d. Let $X$ be equal to the number of not blue eyed girls before the first blue eyed girl. Find the distribution of $X$.

2. Asking for help on a highway When your car breaks down on a highway and you ask for help. Assume that each driver, independently of the other stops and helps you with a probability 0.2. What is the probability that

   a. exactly,

   b. at most,

   c. at least

   5 cars pass without giving you help before somebody will help you?

#### 4.15. 3.5 Geometrical distribution (optimistic)

#### 4.16. EXCEL
The following file shows how an optimistic geometrical distribution looks like.

\textit{Demonstration file: Optimistic geometrical distribution \texttt{neg-020-60-11}}

\section*{4.17. PROBLEMS}

1. Finding the first blue eyed girl Assume that 3/4 of girls in a country have blue eyes. If you choose girls at random until the first blue eyed one, then what is the probability that
   a. the first blue eyed girl will be the 3rd girl;
   b. the first blue eyed girl will be the 4th girl?
   c. the first blue eyed girl will be the 3rd or the 4th girl.
   d. Let $X$ be equal to the number of choices to find the first first blue eyed girl. Find the distribution of $X$.

2. Rainy days in Budapest and in New York I live in Budapest, my friend in New York. We both calculate the number of days until the first rainy day. The probability of rain in Budapest is 0.3, in New York 0.4, each day, independently of each other and other days. My observation yields $X$, the observation of my friend yields $Y$. Both $X$ and $Y$ are random variables with possible values $1, 2, 3, 4, \ldots$.
   a. Calculate the following probabilities:
      a. $P(X = 5)$;
      b. $P(X < 5)$;
      c. $P(Y = 4)$;
      d. $P(X < 15 \text{ and } X > 4)$;
      e. $P(X = 5 \text{ and } Y = 4)$,
      f. $P(X < 15 \text{ and } X > 4\text{ is } Y < 4)$,
      g. $P(Y = 2X)$;
      h. $P(X < Y \text{ and } Y < 15)$;
      i. $P(X < Y)$;
      j. Simulate $(X, Y)$.

3. Apple trees blooming Let us assume that when the apple trees are blooming in Spring, the number of flowers on a tree follows geometrical distribution with an average of 50. Each flower, independently of the others, by the end of the summer will turn into an apple with a probability $2/3$. Assuming that there are 30 apples on a tree at the end of the summer, what is the probability that there were 40 flowers on it in Spring? (You may leave the answer in the form of a sum.)

4. Tossing a coin until the first head A coin is tossed as many times as needed to get the first head.
   a. What is the probability that we have to make only 1 toss?
   b. What is the probability that we have to make 2 tosses?
   c. What is the probability that we have to make 3 tosses?
d. What is the probability that we have to make 4 tosses?

e. What is the probability that we have to make $k$ tosses? (Give a formula)

5. Tossing a die until the first six A die is tossed as many times as needed to get the first six. What is the probability that we have to make only 1 toss? What is the probability that we have to make 10 tosses? What is the probability that we have to make $k$ tosses ($k$ is a positive integer)?

4.18. 3.6 *** Negative binomial distribution (pessimistic)

4.19. EXCEL

The following file shows how a pessimistic negative binomial distribution looks like.

\emph{Demonstration file: Pessimistic negative binomial distribution \texttt{eg-020-60-32}}

4.20. PROBLEMS

1. Tossing a coin until the second head A coin is tossed as many times as needed to get the 2nd head.
   a. What is the probability that the number of tails before the 2nd head is 0?
   b. What is the probability that the number of tails before the 2nd head is 1?
   c. What is the probability that the number of tails before the 2nd head is 2?
   d. What is the probability that the number of tails before the 2nd head is $k$? (Give a formula)
   What is the probability that the number of not-aces before the 2nd ace is 0?

2. Tossing a coin until the second ace A die is tossed as many times as needed to get the 2nd ace. (Ace generally means the largest possible value, which is the six, here.)
   a. What is the probability that the number of not-aces before the 2nd ace is 0?
   b. What is the probability that the number of not-aces before the 2nd ace is 1?
   c. What is the probability that the number of not-aces before the 2nd ace is 2?
   d. What is the probability that the number of not-aces before the 2nd ace is $k$? (Give a formula)

3. Tossing a coin until the third head A coin is tossed as many times as needed to get the 3rd head.
   a. What is the probability that the number of tails before the rd head is 0?
   b. What is the probability that the number of tails before the rd head is 1?
   c. What is the probability that the number of tails before the rd head is 7?
   d. What is the probability that the number of tails before the rd head is $k$? (Give a formula)

4. Tossing a die until the third ace A die is tossed as many times as needed to get the 3rd ace.
   a. What is the probability that the number of not-aces before the 3rd ace is 0?
   b. What is the probability that the number of not-aces before the 3rd ace is 1?
   c. What is the probability that the number of not-aces before the 3rd ace is $k$? (Give a formula)

5. Tossing a coin until the $r$ th head A coin is tossed as many times as needed to get the $r$ th head.
a. What is the probability that the number of tails before the rd head is 0?
b. What is the probability that the number of tails before the rd head is 1?
c. What is the probability that the number of tails before the rd head is 2?
d. What is the probability that the number of tails before the rd head is $k$? (Give a formula)

6. Tossing a die until the $r$ th ace
A die is tossed as many times as needed to get the $r$ th ace.

a. What is the probability that the number of not-aces before the 3rd ace is 0?
b. What is the probability that the number of not-aces before the 3rd ace is 1?
c. What is the probability that the number of not-aces before the 3rd ace is 2?
d. What is the probability that the number of not-aces before the 3rd ace is $k$? (Give a formula)

### 4.21. 3.7 *** Negative binomial distribution (optimistic)***

### 4.22. EXCEL

The following file shows how a pessimistic negative binomial distribution looks like.

*Demonstration file: Optimistic negative binomial distribution*\texttt{eg-020-60-31}*

### 4.23. PROBLEMS

1. Tossing a coin until the second head
A coin is tossed as many times as needed to get the 2nd head.

a. What is the probability that we have to make 2 tosses?
b. What is the probability that we have to make 3 tosses?
c. What is the probability that we have to make 4 tosses?
d. What is the probability that we have to make $k$ tosses? (Give a formula)

2. Tossing a die until the second ace
A die is tossed as many times as needed to get the 2nd ace.

a. What is the probability that we have to make 2 tosses?
b. What is the probability that we have to make 3 tosses?
c. What is the probability that we have to make 4 tosses?
d. What is the probability that we have to make $k$ tosses? (Give a formula)

3. Tossing a coin until the third head
A coin is tossed as many times as needed to get the 3rd head.

a. What is the probability that we have to make 3 tosses?
b. What is the probability that we have to make 4 tosses?
c. What is the probability that we have to make 10 tosses?
d. What is the probability that we have to make $k$ tosses? (Give a formula)

4. Tossing a die until the third ace
A die is tossed as many times as needed to get the 3rd ace.
a. What is the probability that we have to make 3 tosses?

b. What is the probability that we have to make 4 tosses?

c. What is the probability that we have to make \( k \) tosses? (Give a formula)

5. Tossing a coin until the \( r \) th head
A coin is tossed as many times as needed to get the \( r \) th head.

a. What is the probability that we have to make \( r \) tosses?

b. What is the probability that we have to make \( r + 1 \) tosses?

c. What is the probability that we have to make \( r + 2 \) tosses?

d. What is the probability that we have to make \( k \) tosses? (Give a formula)

6. Tossing a die until the \( r \) th ace
A die is tossed as many times as needed to get the \( r \) th ace.

a. What is the probability that we have to make \( r \) tosses?

b. What is the probability that we have to make \( r + 1 \) tosses?

c. What is the probability that we have to make \( r + 2 \) tosses?

d. What is the probability that we have to make \( k \) tosses? (Give a formula)

7. Colored tickets in a box
There are \( N \) tickets in a box, numbered from 1 to \( N \). Assume that \( N = A + B \).
The first \( A \) tickets are red, the other \( B \) are blue, that is, the tickets 1, 2, \ldots, \( A \) are red, and \( A + 1 \), \( A + 2 \), \ldots, \( A + B \) are blue. We draw with replacement until the \( r \) th time we get a red ticket.

a. What is the probability that we draw exactly \( k \) times?

b. For what \( k \) values is this probability different from 0?

c. For which \( k \) value is the probability maximal, that is, where is the mode of the distribution?

4.24. 3.8 Poisson-distribution

4.25. EXCEL

The following file shows how a Poisson-distribution looks like.

\textit{Demonstration file: Poisson-distribution \texttt{\_\_020-60-41}}

4.26. PROBLEMS

1. Computers infected by viruses
There are 50 computers in an office. Each of them, independently of the others, may become infected by viruses during a day with probability 0.05.

a. What is the probability that \( k \) of them become infected by viruses during a day?

b. Make a table and a figure - using binomial distribution - so that \( k \) runs from 0 to 20.

c. Make a table and a figure - using Poisson-distribution, too. Compare them.

2. Teachers becoming sick
There are 70 teachers in our institute. Each teacher, independently of the others, may become sick during a day with probability 0.04.
a. What is the probability that \( k \) of them become sick during a day?

b. Make a table and a figure - using binomial distribution - so that \( k \) runs from 0 to 20.

c. Make a table and a figure - using Poisson-distribution, too. Compare them.

3. Earthquakes Assume that in a certain country the average number of earthquakes during a year is 1.6. What is the probability that during a year the number of earthquakes is at most 2? What is the probability that during 3 years the number of earthquakes is at most 2? How many earthquakes are the most probable during a year, and during 3 years?

4. Earthquakes in another country Assume that in an other country the probability that during a year at least one earthquake happens is 0.3. What is the probability that during 5 years the number of earthquakes is at least 3?

5. Shooting stars Assume that from a certain hill top during an August night one can see shooting stars so that the average amount of time between them is 10 minutes. What is the probability that during 15 minutes we see exactly 2 shooting stars?

6. Accidents If the average number of accidents in a city during a day is 7.2, then what is the probability that the number of accidents in that city is
   a. less than 5;
   b. more than 10;
   c. is more 5 but less than 10?
   d. How many accidents are the most likely?

7. Accidents in Buda and in Pest Assume that the average number of serious accidents in Buda during a day is 1.6, while in Pest it is 2.5. What is the probability that during a day there is at least 1 serious accident in Pest, but no serious accident happens in Buda? How many serious accidents are the most probable in the whole territory of Budapest?

8. Magnetic tapes Suppose that a certain type of magnetic tape contains, on the average, 3 defects per 1000 feet. What is the probability that a roll of tape 1200 feet long contains no defects?

9. Twins Assume that the average number of twin births in a city during a day is 0.7. What is the probability that the number of twin births in that city during 3 days is 2? What is the most probable number of twins birth in that city during 3 days?

10. Triple twins Assume that the average number of triple twins during a year in a country is 1.5. What is the probability that there are
   a. no triple twins;
   b. more than 2 triple twins during two and a half years? (Explain why the distribution you use is justified.)

11. How many fish? Some years ago I met an old fisherman. He was fishing in a big lake, in which many small fish were swimming regardless of each other. He raised his from time to time, and picked the fish if there were any. He told me that out of 100 cases the net is empty only 15 times or so, and then he added: "If you could guess the number of fish in the net when I raise the net out of the water the next time, I would give you a big sack of money." I am sure he would not have said such a promise if he knew that his visitor was a well educated person in probability theory! I was thinking a little bit, then I made some calculation, and then I said a number. Which number did I say?

4.27. 3.9 Higher dimensional discrete random variables and distributions
The following files shows some two-dimensional distributions.

1. Drawing twice from a box without replacement There are 6 tickets in a box numbered from 1 through 6. You choose two tickets, one after the other, without replacement. The first number is denoted by $X$, the second denoted is denoted by $Y$. Find the distribution of the two-dimensional random variable $(X, Y)$.

2. Drawing twice from a box without replacement There are 6 tickets in a box numbered from 1 through 6. You choose two tickets without replacement. The first number is denoted by $X$, the second is denoted denoted by $Y$. Find the distribution of the two-dimensional random variable $(X, Y)$.

3. Drawing twice from a box with replacement There are 6 tickets in a box numbered from 1 through 6. You choose two tickets with replacement. The first number is denoted by $X$, the second is denoted denoted by $Y$. Find the distribution of the two-dimensional random variable $(X, Y)$.

4. Drawing twice from a box with replacement There are 6 tickets in a box numbered from 1 through 6. You choose two tickets with replacement. The first number is denoted by $X$, the second is denoted denoted by $Y$. Find the distribution of the two-dimensional random variable $(X, Y)$.

5. Drawing without replacement, when there are repeated tickets in the box There are 6 tickets in a box with the numbers 1, 2, 2, 3, 3, 3. You choose two tickets without replacement. The first number is denoted by $X$, the second is denoted denoted by $Y$. Find the distribution of the two-dimensional random variable $(X, Y)$.

6. Drawing with replacement, when there are repeated tickets in the box There are 6 tickets in a box with the numbers 1, 2, 2, 3, 3, 3. You choose two tickets with replacement. The first number is denoted by $X$, the second is denoted denoted by $Y$. Find the distribution of the two-dimensional random variable $(X, Y)$.

7. Calculating a conditional distribution There are 6 tickets in a box numbered from 1 through 6. First you choose a ticket, and you denote the number on it by $X$. Then you remove from the box all the tickets which have a number greater than $X$. Then you choose a second number from the box, and you denote the number on it by $Y$. Obviously $Y \leq X$.

   a. Find the distribution of the two-dimensional random variable $(X, Y)$.

   b. What is the probability that $Y = k$?

   c. Assume that $Y = k$. Find the distribution of $X$ under this condition.

4.28. PROBLEMS
8. Calculating a conditional distribution There are \( n \) tickets in a box numbered from 1 through \( n \). First you choose a ticket, and you denote the number on it by \( X \). Then you remove from the box all the tickets which have a number greater than \( X \). Then you choose a second number from the box, and you denote the number on it by \( Y \). Obviously \( Y \leq X \).

a. Find the distribution of the two-dimensional random variable \( (X, Y) \).

b. What is the probability that \( Y = k \) ?

c. Assume that \( Y = k \). Find the distribution of \( X \) under this condition.

9. Calculating a conditional distribution There are \( n \) tickets in a box numbered from 1 through \( n \). First you choose a ticket, and you denote the number on it by \( X \). Then you remove from the box all the tickets which have a number greater than \( X \). Then you choose a second number from the box, and you denote the number on it by \( Y \). Obviously \( Y \leq X \).

a. What is the probability that \( X + Y = 5 \)?

b. Assume that \( Y = 5 \). Find the distribution of \( X + Y \) under this condition.

10. Calculating a conditional distribution There are \( n \) tickets in a box numbered from 1 through \( n \). First you choose a ticket, and you denote the number on it by \( X \). Then you remove from the box all the tickets which have a number greater than \( X \). Then you choose a second number from the box, and you denote the number on it by \( Y \). Obviously \( Y \leq X \).

a. What is the probability that \( X + Y \leq 5 \)?

b. Assume that \( X + Y \leq 5 \). Find the distribution of \( X + Y \) under this condition.

### 4.29. 3.10 *** Poly-hyper-geometrical distribution

#### 4.30. EXCEL

The following file shows a Poly-hyper-geometrical distribution.

\[ \text{Demonstration file: Poly-hyper-geometrical distribution} \]

#### 4.31. PROBLEM

1. Determining a poly-hyper-geometrical distribution There are 30 tickets in a box. 5 of them are red, 10 of them are white, 15 of them are green. You choose 7 tickets without replacement. \( X \) denotes how many times a red ticket is chosen, \( Y \) denotes how many times a white ticket is chosen. Recall that the distribution of \( X \) is the hyper-geometrical distribution with parameters \( (5; 25; 7) \), and the distribution of \( Y \) is the hyper-geometrical distribution with parameters \( (10; 20; 7) \). Find now the distribution of the two-dimensional random variable \( (X, Y) \).

### 4.32. 3.11 *** Polynomial distribution

#### 4.33. EXCEL

The following file shows a Polynomial distribution.

\[ \text{Demonstration file: Polynomial distribution} \]
4.34. PROBLEM

1. Determining a polynomial distribution There are 30 tickets in a box. 5 of them are red, 10 of them are white, 15 of them are green. You choose 7 tickets with replacement. $X$ denotes how many times a red ticket is chosen, $Y$ denotes how many times a white ticket is chosen. Recall that the distribution of $X$ is the binomial distribution with parameters $n = 7$, $\mu = 5/30$, and the distribution of $Y$ is the binomial distribution with parameters $n = 7$, $\mu = 10/30$. Find now the distribution of the two-dimensional random variable $(X, Y)$.

4.35. 3.12 Generating a random variable with a given discrete distribution

4.36. EXCEL

The following file shows how a random variable with a given distribution can be simulated.

\textit{Demonstration file: Generating a random variable with a given discrete distribution}

4.37. PROBLEMS

1. Simulating random variables with discrete uniform distribution Simulate with Excel
   a. a random variable $X$ uniformly distributed on the set $1, 2, 3, 4, 5, 6$;
   b. a random variable $X$ uniformly distributed on the set $11, 12, 13, 14, 15, 16$;
   c. a random variable $X$ uniformly distributed on the set $1, 4, 9, 16, 25, 36$;
   d. a random variable $X$ uniformly distributed on the set $1, 2, l...n$
      , where $n$ is a parameter;
   e. a random variable $X$ uniformly distributed on the set $A, A + 1, l...B$
      , where $A$ and $B$ are parameters.

2. Simulating a random variable with non-uniform discrete distribution Simulate with Excel a random variable which has the following distribution:

   \begin{align*}
   \begin{array}{|c|cccccccc|}
   \hline
   k & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
   \hline
   p(k) & 0.1 & 0.1 & 0.1 & 0.1 & 0.3 & 0.2 & 0.1 \\
   \hline
   \end{array}
   \end{align*}

3. Simulating a random variable with binomial distribution Simulate with Excel a random variable which has the binomial distribution with parameters
   a. 10 and 0.5;
   b. 10 and 0.7;
   c. $n$ and $P$, where $n$ and $P$ are parameters.

4. Simulating a random variable with Poisson-distribution Simulate with Excel a random variable which has the Poisson-distribution with parameter
4.38. 3.13 Mode of a distribution

4.39. EXCEL

When a discrete distribution is given by a table in Excel, its mode can be easily identified. This is shown in the next file.

\emph{Demonstration file: Calculating - with Excel - the mode of a discrete distribution \ref{ef-140-01-00}}

The next file shows the modes of some important distributions:

\emph{Demonstration file: Modes of binomial, Poisson and negative binomial distributions \ref{ef-130-50-00}}

4.40. PROBLEMS

1. Finding the mode Find the mode(s) of the following distribution:

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(k)$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

2. Finding the mode Find the mode(s) of the following distribution:
3. Finding the mode of a binomial distribution
   Find the mode(s) of the binomial distribution with parameters
   a. 10 and 0.5;
   b. 10 and 0.7;
   c. $\eta$ and $p$, where $\eta$ and $p$ are parameters.

4. Finding the mode of a Poisson distribution
   Find the mode(s) of the Poisson-distribution with parameter
   a. 2.8;
   b. $\lambda$, where $\lambda$ is a parameter.

5. Finding the mode of a pessimistic geometrical distribution
   Find the mode(s) of the pessimistic geometrical distribution with parameter
   a. 1/6;
   b. $p$, where $p$ is a parameter.

6. Finding the mode of an optimistic geometrical distribution
   Find the mode(s) of the optimistic geometrical distribution with parameter
   a. 1/6;
   b. $p$, where $p$ is a parameter.

7. Finding the mode of a pessimistic negative binomial distribution
   Find the mode(s) of the pessimistic negative binomial distribution with parameters
   a. 3 and 0.5;
   b. 3 and 0.7;
   c. 3 and $p$;
   d. $\tau$ and $p$, where $\tau$ and $p$ are parameters.

8. Finding the mode of an optimistic negative binomial distribution
   Find the mode(s) of the optimistic negative binomial distribution with parameters
   a. 3 and 0.5;
   b. 3 and 0.7;
   c. 3 and $p$;
   d. $\tau$ and $p$, where $\tau$ and $p$ are parameters.

9. Finding the mode of the birthday problem
   People chosen at random are asked which month and day they have their birthdays. We stop asking as soon as we get a birthday which has already been occurred before. Let $X$ denote the number of people asked.
   a. What is the distribution of $X$?
   b. Find the mode of $X$.

### 4.41. 3.14 Expected value of discrete distributions
4.42. EXCEL

The following file shows how the expected value of a discrete distribution can be calculated if the distribution is given by a table in Excel.

\textit{Demonstration file: Calculating the expected value of a discrete distribution ef-150-01-00}

4.43. PROBLEMS

1. Absolute value of the difference with two dice Toss 2 dice, and observe the absolute value of the difference between the two numbers on the dice.
   a. Calculate the expected value of this random variable.
   b. Make 1000 simulations and be convinced that the average of the experimental results is close to the expected value.

2. Number of heads with three coins Tossing with 3 coins observe the number of heads. Find the expected value of this random variable.

3. Number of tosses Toss a coin until you get the first time a head. How much is the expected value of the random variable $X$ defined as the number of tosses.

4. Number of sixes with four dice Tossing with 4 dice observe the numbers sixes. Find the expected value of this random variable.

5. Maximum with two dice Tossing with 2 dice observe the maximum of the 2 numbers we toss. Find the expected value of this random variable.

6. Number of draws There are 2 red and 5 blue balls in a box. We draw without replacement until the first red is drawn. Let $X$ denote the number of draws. Calculate the expected value of $X$.

7. Number of tosses Toss a pair of coins until you get the first time that both coins are heads. How much is the expected value of the random variable $X$ defined as the number of tosses.

8. Number of tosses Toss a die until you get the first time an ace. How much is the expected value of the random variable $X$ defined as the number of tosses.

9. Number of tosses Toss a pair of dice until you get the first time that both dice are aces. How much is the expected value of the random variable $X$ defined as the the number of tosses.

10. Number of injured people Assume that when a 5 passenger car has an accident, then the number $X$ of injured people, independently of any other factors, has the following distribution: $P(X = 0) = 0.4$, $P(X = 1) = 0.2$, $P(X = 2) = 0.1$, $P(X = 3) = 0.1$, $P(X = 4) = 0.1$, $P(X = 5) = 0.1$, and when an 8 passenger bus has an accident, then the number $Y$ of injured people, independently of any other factors, has the following distribution: $P(Y = 0) = 0.50$, $P(Y = 1) = 0.10$, $P(Y = 2) = 0.10$, $P(Y = 3) = 0.05$, $P(Y = 4) = 0.05$, $P(Y = 5) = 0.05$, $P(Y = 6) = 0.05$, $P(Y = 7) = 0.05$, $P(Y = 8) = 0.05$.
   a. How much is the expected value of the number of injured people when a 5 passenger car has an accident?
   b. How much is the expected value of the number of injured people when an 8 passenger bus has an accident?
   c. How much is the expected value of the number of injured people when a 5 passenger car hits an 8 passenger bus?
11. Mobile phone calls during an hour
Assume that the average number of mobile phone calls a man gets during an hour is 2.5. What is the probability that he gets

a. exactly 0;

b. exactly 1;

c. exactly 2;

d. exactly 3;

e. less than 2;

f. more than 2 calls during an hour?

12. Mobile phone calls during two hours (Continuation of the previous problem.) Consider now the random variable: "the number of calls during 2 hours". Based on your common sense, figure out how much the its expected value is during 2 hours. What is the probability that he gets

a. exactly 0;

b. exactly 1;

c. exactly 2;

d. exactly 3;

e. less than 2;

f. more than 2 calls during 2 hours?

13. Expected value of the number of draws
There are four tickets in a box numbered from 1 to 4. We draw without replacement as many times as needed to get the ticket with the number 4.

a. What is the probability that the number of draws is an even number?

b. How much is the expected value of the number of draws?

14. Comparison of the expected values
A discrete distribution is defined by the formula: \( p(x) = \frac{x^2}{30} \) \( (x = 1, 2, 3, 4) \). Sketch the graph of the distribution. How much is its expected value? An other discrete distribution is defined by the formula: \( p(x) = \frac{x^2}{7.5} (x = \frac{1}{2}, 1, \frac{3}{2}, 2) \). Sketch the graph of the distribution. How much is its expected value? Compare them.

15. Comparison of the expected values
Calculate the numerical value of the expected value of the following distributions:

<table>
<thead>
<tr>
<th>k</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(k) )</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

a.

<table>
<thead>
<tr>
<th>k</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(k) )</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

b.

<table>
<thead>
<tr>
<th>k</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(k) )</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

c.

<table>
<thead>
<tr>
<th>k</th>
<th>33</th>
<th>36</th>
<th>39</th>
<th>42</th>
<th>45</th>
<th>48</th>
<th>51</th>
<th>54</th>
<th>57</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(k) )</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

d.
### 4.44. 3.15 Expected values of the most important discrete distributions

**4.45. EXCEL**

The following file gives the expected value of geometrical, binomial and Poisson-distributions.

\textit{Demonstration file: Expected values of the most important discrete distributions nef-150-07-01}

### 4.46. PROBLEMS

1. Lottery players Assume that during a year (52 weeks), the expected value of the number of lottery players winning 5 hits on the "5-lottery" is 3.4, the expected value of the number of lottery players winning 6 hits on the "6-lottery" is 7.6. What is the probability that
   
a. during a year, on the "5-lottery" nobody has 5 hits;
   
b. during a month (4 weeks), on the "5-lottery" nobody has 5 hits? (Assume that the "intensity" of lottery players is uniform, during the year.)
   
c. during a year, on the "5-lottery" 5 players have 5 hits and on the "6-lottery" 6 players have 6 hits;
   
d. during a year, the number of players having 5 hits on the "5-lottery" plus the number of players having 6 hits on the "6-lottery" is exactly 7?

2. Number of tosses Toss a coin until you get the first time a head. How much is the expected value of the random variable \( X \) defined as the number of tosses.

3. Number of tosses Toss a die until you get the first time an ace. How much is the expected value of the random variable \( X \) defined as the number of tosses.

4. Number of tosses Toss a pair of dice until you get the first time that both dice are aces. How much is the expected value of the random variable \( X \) defined as the number of tosses.

5. Discovering a formula for an expected value Tossing a die until the first six, let \( X \) be the number of tosses. Make actually 10 experiments for this random variable, and - analyzing the experimental results - set up a simple relation between the average of the observed X-values and the relative frequency of six. Then
imagining a large number of experiments, discover how the expected value of \( X \) can be expressed in terms of the probability of tossing a six.

6. Discovering a formula for an expected value Tossing a die until the third six, let \( X \) be the number of tosses. Make actually or just imagine 10 experiments for this random variable, and - analyzing the experimental results - set up a simple relation between the average of the observed \( X \)-values and the relative frequency of six. Then imagining a large number of experiments, discover how the expected value of \( X \) can be expressed in terms of the probability of tossing a six.

7. Discovering a formula for an expected value Finally imagine that we toss a false die until the \( r \)th six, (false means that the probability of six is not necessarily 1/6, but it is, say, equal to \( p \)) and let \( X \) be the number of tosses. Imagine a large number of experiments for this random variable, and figure out a simple relation between the average of the observed \( X \)-values and the relative frequency of six in order to conclude a formula for the expected value of \( X \).

8. Deriving the expected value of the optimistic negative binomial distribution Here is a method to find the expected value of the optimistic negative binomial distribution: Imagine that we toss a false die until the 3rd six. Let \( X \) be the number of tosses to get the 3rd six. Introduce the following random variables, as well: \( X_1 \) = the number of tosses to get the 1st six. \( X_2 \) = the number of tosses after the 1st six to get the 2nd six. \( X_3 \) = the number of tosses after the 2nd six to get the 3rd six. On one hand, \( X_1, X_2, X_3 \) obviously follow the optimistic geometrical distribution with parameter \( p \), so their expected value is \( \frac{1}{p} \). On the other hand, a simple relation between \( X \) and \( X_1, X_2, X_3 \) can be noticed. From this relation it is easy to derive the formula for the expected value of \( X \).

4.47. 3.16 Expected value of a function of a discrete random variable

4.48. EXCEL

The following file shows how the expected value of a function of a discrete random variable can be calculated if the distribution of the random variable is given by a table in Excel.

\textbf{Demonstration file: Calculating - with Excel - the expected value of a function for a discrete distribution [ef-160-01-00]}

4.49. PROBLEMS

1. Expected values of some functions of a discrete random variable The distribution of a random variable \( X \) is given by:

\[
\begin{array}{cccccccccc}
  & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
 k & 0.1 & 0.1 & 0.1 & 0.1 & 0.3 & 0.2 & 0.1 \\
\end{array}
\]

How much is its expected value of
a. \( X \);
b. \( 1/X \);
c. \( X^2 \)?

2. Expected values of some functions of a discrete random variable Assume that the random variable \( X \) follows the discrete distribution: \( p(x) = \frac{x^2}{30} (x = 1, 2, 3, 4) \). How much is its expected value of
a. \( X \)?
b. $1/X$?

c. $X^2$?

4.50. 3.17 Moments of a discrete random variable

4.51. EXCEL

The following files show how the moments of a discrete distribution can be calculated if the distribution is given by a table in Excel.

Demonstration file: Calculating the second moment of a discrete distribution

4.52. PROBLEMS

1. Second moments of some functions of a discrete random variable The distribution of a random variable $X$ is given by:

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(k)$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

How much is the second moment of

a. $X$;

b. $1/X$;

c. $X^2$?

2. Second moments of some functions of a discrete random variable We toss two fair dice. How much is the second moment of

a. the difference ("red die minus blue die");

b. the minimum?

3. Second moments of some functions of a discrete random variable Assume that the random variable $X$ follows the discrete distribution: $p(x) = \frac{x^3}{30} (x = 1, 2, 3, 4)$. How much is the second moment of

a. $X$;

b. $1/X$;

c. $X^2$?

4. Third moments of some function of a discrete random variable The distribution of a random variable $X$ is given by:

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(k)$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

How much is the third moment of

a. $X$;
b. \( 1/X \);  

c. \( X^2 \);  

5. Third moments of some function of a discrete random variable Assume that the random variable \( X \) follows the discrete distribution: \( p(x) = \frac{x^2}{50} (x = 1, 2, 3, 4) \). How much is the third moment of  

a. \( X \);  
b. \( 1/X \);  
c. \( X^2 \);  

4.53. 3.18 Projections and conditional distributions for discrete distributions

4.54. EXCEL

Here are some files to study the relations between projections and conditional distributions for discrete distributions.  

*emphasis* Demonstration file: Construction from conditional distributions, discrete case (version A) \( \text{ef}-200-75-00 \)  

*emphasis* Demonstration file: Construction from conditional distributions, discrete case (version B) \( \text{ef}-200-76-00 \)  

*emphasis* Demonstration file: Projections and conditional distributions, discrete case (version A) \( \text{ef}-200-77-00 \)  

*emphasis* Demonstration file: Projections and conditional distributions, discrete case (version B) \( \text{ef}-200-78-00 \)  

4.55. PROBLEMS

1. Unconditional and conditional probabilities and distributions Assume that the distribution of the random variable \((X,Y)\) is  

<table>
<thead>
<tr>
<th>( y/x )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>3</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
</tbody>
</table>

a. Check that the sum of all probabilities is equal to 1.  
b. Calculate the probability \( P(X + Y = 6) \).  
c. Calculate the probability \( P(X + Y >= 6) \).  
d. Calculate the probability \( P(X - Y >= 3) \).  
e. Calculate the conditional probability \( P(X - Y >= 3 | X + Y >= 6) \).  
f. Determine the distribution of \( X \).
2. Unconditional and conditional probabilities and distributions Assume that the distribution of the random variable \((X, Y)\) is

\[
\begin{array}{cccccc}
5 & 0.003 & 0.001 & 0.005 & 0.005 & 0.001 & 0.002 \\
4 & 0.001 & 0.004 & 0.005 & 0.005 & 0.001 & 0.007 \\
3 & 0.005 & 0.004 & 0.000 & 0.005 & 0.004 & 0.000 \\
2 & 0.003 & 0.007 & 0.000 & 0.010 & 0.000 & 0.000 \\
1 & 0.000 & 0.007 & 0.015 & 0.000 & 0.000 & 0.000 \\
\end{array}
\]

| y / x | 1 | 2 | 3 | 4 | 5 | 6 |

\begin{align*}
a. & \text{ Check that the sum of all probabilities is equal to } 1. \\
b. & \text{ Calculate the probability } P(X + Y = 6). \\
c. & \text{ Calculate the probability } P(X + Y \geq 6). \\
d. & \text{ Calculate the probability } P(X - Y \geq 3). \\
e. & \text{ Calculate the conditional probability } P(X - Y \geq 3 | X + Y \geq 15). \\
f. & \text{ Determine the distribution of } X. \\
g. & \text{ Determine the distribution of } Y. \\
h. & \text{ Determine the conditional distributions of } Y \text{ on condition that } X = 3. \\
i. & \text{ Determine the conditional distributions of } Y \text{ on condition that } X = 4. \\
j. & \text{ Determine the conditional distributions of } Y \text{ on condition that } X = x \text{ for all } x. \\
k. & \text{ Determine the conditional distributions of } X \text{ on condition that } Y = y \text{ for all } y. \\
l. & \text{ Calculate the conditional probability } P(2 \leq Y \leq 5 | X = x) \text{ for all } x. \\
m. & \text{ Are } X \text{ and } Y \text{ independent of each other?}
\end{align*}

4.56. 3.19 Transformation of discrete distributions

4.57. EXCEL

The following files give simple numerical examples for transformations of a discrete distributions.

\textit{Demonstration file: Transformation of a discrete distribution \textasciitilde 020-03-01}
4.58. PROBLEMS

1. Transformations of a discrete distribution The distribution of a random variable $X$ is given by:

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(k)$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Find the distribution of
a. $2X$;
b. $X^2$;
c. $|X - 3|$.

2. Transformations of a discrete distribution Assume that the distribution of the random variable $(X,Y)$ is

<table>
<thead>
<tr>
<th>$y/x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>3</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Find the distribution of
a. $X + Y$;
b. $X - Y$;
c. $2X + Y$.

3. Transformations of a discrete distribution Assume that the distribution of the random variable $(X,Y)$ is

<table>
<thead>
<tr>
<th>$y/x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>4</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>3</td>
<td>0.005</td>
<td>0.004</td>
<td>0.000</td>
<td>0.005</td>
<td>0.005</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.003</td>
<td>0.007</td>
<td>0.000</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.007</td>
<td>0.015</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Find the distribution of
a. $X + Y$;
b. $X - Y$;
c. $2X + Y$.

5. 4 Continuous random variables

5.1. PROBLEMS
1. Are they continuous? Which of the following random variables are continuous?

   a. the temperature measured in Celsius in the center of a city at noon time in summer;
   b. the number of days when the temperature is higher than 30 Celsius in the center of a city at noon time during a summer;
   c. the amount of time when, the temperature is higher than 30 Celsius in the center of a city during a day in summer,
   d. the number of students who attend a lecture at a university;
   e. the weight (in kg) of the highest student among all students who attend a lecture at a university;
   f. the total weight of all students who attend a lecture at a university;

5.2. 4.1 Distribution function

5.3. EXCEL

The following file shows the graphs of the distribution functions of the most important continuous distributions.

\textbf{Demonstration file: Distribution functions of the most important continuous distributions \ref{200-57-50-distr}}

5.4. PROBLEMS

1. Calculating probabilities from the distribution function Assume that the distribution function of a random variable \( X \) is \( F(x) = \frac{x}{2} \) for \( 0 \leq x \leq 2 \). How much is
   
   a. \( P(X < 1) \);
   b. \( P(X > 3) \);
   c. \( P(X > 0.7) \);
   d. \( P(0.3 < X < 0.7) \);
   e. \( P(0.3 < X < 0.7 | X > 1) \)?

2. Calculating probabilities from the distribution function Assume that the distribution function of a random variable \( X \) is \( F(x) = \frac{1}{2} + \frac{1}{x} \arctan(x) \), or equivalently, \( F(x) = \frac{1}{2} + \frac{1}{x} \tan^{-1}(x) \), for all \( x \). How much is
   
   a. \( P(X < 1) \);
   b. \( P(X > 3) \);
   c. \( P(X < -1) \);
   d. \( P(X > -3) \);
   e. \( P(1 < X < 5) \);
   f. \( P(1 < X < 5 | X > 3) \)?

\textbf{Solution} \ref{Sol-03-02-01}
3. Calculating probabilities from the distribution function. Assume that the distribution function of a random variable is 
\[ F(x) = \begin{cases} 1 & \text{if } x \leq 0, \\ 0 & \text{otherwise}. \end{cases} \]
How much is
a. \( P(X < 1); \)
b. \( P(X > 3); \)
c. \( P(X < -1); \)
d. \( P(X > -3); \)
e. \( P(1 < X < 5); \)
f. \( P(1 < X < 5 | X > 3)? \)

4. Finding the distribution function. A random number is generated by a calculator or computer. Let \( X \) be its
a. cube;
b. logarithm;
c. square;
d. square-root;
e. reciprocal.
Find the distribution function of \( X \) in each case.

5. Finding the distribution function. Calculate \( F(x) \) if
a. \( X = \text{RND}^2; \)
b. \( X = \text{RND}^3; \)
c. \( X = \sqrt{\text{RND}}; \)
d. \( X = 1/\text{RND}; \)
e. \( X = 1/\text{RND}^2; \)
f. \( X = -\ln(\text{RND}); \)

6. Finding the distribution function. Four random numbers are generated: \( \text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4. \)
a. Let \( X_1 \) denote the largest value of the values \( \text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4. \) Find the distribution function \( X_1. \)
b. Let \( X_2 \) denote the second largest value of the values \( \text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4. \) Find the distribution function of \( X_2. \)

7. Finding the distribution function. Ten random numbers are generated by a calculator or computer. Let \( \tilde{X} \) be the 8th smallest of them. Find the distribution function of \( \tilde{X}. \)

8. Constructing the graphs of distribution functions. Make an Excel file to visualize the graphs of the distribution functions of the random variables given in the previous problems.

5.5. 4.2 *** Empirical distribution function
5.6. EXCEL

The following file shows the construction of an empirical distribution function

\verb|Demonstration file: Empirical distribution function|  

5.7. PROBLEMS

1. Empirical distribution function of the maximum of two random numbers The following file is a small modification of the previous one. Some cells on page "1" are colored light yellow, some cells on page "s3" are colored dark yellow. Modify the light yellow cells: replace the power of the random number by the maximum of two random numbers. Write the Excel formula of the distribution function of the maximum of two random numbers into the dark yellow cells, which is, \(x^2\). You will see that the empirical distribution function will be oscillating around the distribution function. \verb|Demonstration file: Empirical distribution function|

2. Empirical distribution function of your own random variable Choose a distribution so that you can easily simulate it and its distribution function can be easily given by a formula like in the previous problem. Now use this distribution to do the steps of the previous problem.

5.8. 4.3 Density function

5.9. EXCEL

The following file shows the graphs both of the distribution functions and of the density functions of the most important continuous distributions.

\verb|Demonstration file: Distribution functions of the most important continuous distributions|

5.10. PROBLEMS

1. Finding the distribution function from the density function Assume that the density function of the random variable \(X\) is \(f(x) = 1/8 \cdot x\) if \(0 < x < 2\). Find the distribution of \(X\).

2. Finding the distribution function from the density function Assume that the density function of the random variable \(X\) is \(f(x) = 1/8 \cdot x\) if \(0 < x < 2\). Find the distribution function of \(1/X\).

3. Finding the distribution function from the density function Assume that the density function of \(X\) is \(f(x) = 3 \cdot x^2\) if \(0 < x < 1\). Find the distribution function of \(X^6\).

4. Calculating probabilities from the density function Assume that the density function of a random variable \(X\) is \(f(x) = e^{-x}\), if \(x > 0\), and \(f(x) = 0\) otherwise. How much is
   a. \(P(X < 1)\);
   b. \(P(X > 3)\);
   c. \(P(1 < X < 5)\);
   d. \(P(1 < X < 5|X > 3)\)?
5. Calculating probabilities from the density function Assume that the density function of a random variable $X$ is 
$$f(x) = \frac{1}{\sqrt{1-x^2}} \quad -1 < x < 1$$
How much is
a. $P(X < 0.5)$;
b. $P(X > 0.3)$;
c. $P(X > 0.7)$;
d. $P(0.3 < X < 0.7)$;
e. $P(0.3 < X < 0.7 | X > 0.5)$?
6. Finding the value of the constant $c$ Assume that the density function of the random variable $X$ is 
$$f(x) = cx^2 \quad (0 < x < 3),$$
where $c$ is a constant. Find the value of $c$, and determine the distribution of $X$.
7. Finding the density function A random number is generated by a calculator or computer. Let $X$ be its
a. cube;
b. logarithm;
c. square;
d. square-root;
e. reciprocal.
Find the density function of $X$ in each case.
8. Finding the density function Calculate $f(x)$ if
a. $X = \text{RND}^2$;
b. $X = \text{RND}^3$;
c. $X = \sqrt{\text{RND}}$;
d. $X = 1/\text{RND}$;
e. $X = 1/\text{RND}^2$;
f. $X = -\ln(\text{RND})$.
9. Finding the density function Four random numbers are generated: $\text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4$.
   a. Let $X_1$ denote the largest value of the values $\text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4$. Find the density function of $X_1$.
   b. Let $X_2$ denote the second largest value of the values $\text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4$. Find the density function of $X_2$.
10. Finding the density function Ten random numbers are generated by a calculator or computer. Let $X$ be the 8th smallest of them. Find the density function of $X$.
11. Constructing the graphs of density functions Make an Excel file to visualize the graphs of the density functions of the random variables given in the previous problems.
12. "Linear" density function: Make an Excel file to visualize the graph of the density function which is linear on the interval $[0, 1]$, that is, $f(x) = 3x$ for $0 < x < 1$. Use $a$ as a parameter. Pay attention to the fact that the total area under the density function must be equal to 1.

13. "Linear" density function: Make an Excel file to visualize the graph of the density function which is linear on the interval $[A; B]$, that is, $f(x) = ax + b$ if $A < x < B$. Use $b$ as a parameter. Pay attention to the fact that the total area under the density function must be equal to 1.

5.11. 4.4 *** Histogram

5.12. EXCEL

The following files show a histogram construction.

\textbf{Demonstration file: Histogram construction} \textit{ref-200-04-00}

\textbf{Demonstration file: Histogram construction for standard normal distribution} \textit{ref-200-05-00}

5.13. PROBLEM

1. Histogram construction: The following file is a small modification of the previous one. Some cells on page "1" are colored light yellow, some cells on page "s3" are colored dark yellow. Modify the light yellow cells so that you put there a random variable whose values are between $-3$ and $3$. Write the Excel formula of the density function of this random variable into the dark yellow cells. You will see that the histogram will be oscillating around the density function.

\textbf{Demonstration file: Histogram construction} \textit{eg-030-03-02}

5.14. 4.5 Uniform distributions

5.15. EXCEL

The following file shows the distribution functions and the density functions of uniform distributions, and gives also a simulation for this distribution.

\textbf{Demonstration file: Uniform distribution: distribution function, density function, simulation} \textit{ref-200-07-00}

5.16. PROBLEMS

1. Calculating probabilities: Assume that $X$ follows uniform distribution on the interval $[0, 5]$. Calculate the probabilities and conditional probabilities:

   a. $P(X < 3)$;
   b. $P(1 < X)$;
   c. $P(1 < X \text{ and } X < 3)$;
   d. $P(1 < X \mid X < 3)$;
   e. $P(X < 3 \mid 1 < X)$. 
2. Calculating probabilities Assume that $X$ follows uniform distribution on the interval $[0.8, 12.3]$. Calculate the probabilities and conditional probabilities:
   
   a. $P(X < 3.3)$;
   
   b. $P(1.8 < X)$;
   
   c. $P(1.8 < X$ and $X < 3.3)$;
   
   d. $P(1.8 < X | X < 3.3)$;
   
   e. $P(X < 3.3 | 1.8 < X)$.

3. Waiting time for the bus Assume that the waiting time for the bus follows uniform distribution between 0 and $c$ minutes, but the constant $c$ is not known. How much is $c$, if
   
   a. $P($waiting time $< 5) = 0.25$;
   
   b. $P($waiting time $< 5) = 0.5$;
   
   c. $P($waiting time $< 5 |$ waiting time $< c/3) = 0.5$.

5.17. 4.6 Distributions of some functions of random numbers

5.18. EXCEL

The following files give simulations for random variables derived from uniformly distributed random numbers generated by a computer.

\textbf{Demonstration file: Square of a random number, simulation and density function}

\textbf{Demonstration file: Square root of a random number: simulation and density function}

\textbf{Demonstration file: Random number to a positive power, simulation}

\textbf{Demonstration file: Random number to a positive power times a constant, simulation}

\textbf{Demonstration file: Reciprocal of a random number, simulation}

\textbf{Demonstration file: A random number raised to a negative exponent, simulation}

\textbf{Demonstration file: A random number raised to a negative exponent times a constant, simulation}

\textbf{Demonstration file: Product of two random numbers, simulation}

\textbf{Demonstration file: Ratio of two random numbers, simulation}

5.19. PROBLEMS

1. Finding the distribution function and the density function Find the distribution function and the density function of the following random variables:
   
   a. $X = 5 \ast \text{RND}$;
   
   b. $X = 2 + 5 \ast \text{RND}$.
c. $X = -2 + 5 \times \text{RND}$;
d. $X = \text{RND}^2$;
e. $X = \text{RND}^{-2}$;
f. $X = \text{RND}^3$;
g. $X = c \times \text{RND}$, where $c$ is a positive constant;
h. $X = c \times \text{RND}$, where $c$ is a negative constant;
i. $X = 1/\text{RND}^2$;
j. $X = \ln(\text{RND})$;
k. $X = -\ln(\text{RND})$;
l. $X = 25\sqrt{\text{RND}}$;
m. $X = t(\text{RND})$, where $y = t(x)$ is a strictly increasing continuously differentiable function;
n. $X = t(\text{RND})$, where $y = t(x)$ is a strictly decreasing continuously differentiable function.

2. Finding the distribution function and the density function Two random numbers are generated: $\text{RND}_1$, $\text{RND}_2$. Find the distribution function and the density function of
   a. their maximum;
b. $\text{RND}_1 \times \text{RND}_2$;
c. $\text{RND}_1 / \text{RND}_2$;
d. $\text{RND}_2 / \text{RND}_1$;
e. $\text{RND}_2 / \text{RND}_2$;
f. $\text{RND}_2 / \text{RND}_2$.

3. Making simulations Make simulations of the random variables of the previous problems.

5.20. 4.7 *** Arc-sine distribution

5.21. EXCEL

The following files show the arc-sine distribution.

*Demonstration file: Arc-sine distribution

*Demonstration file: Arc-sine distribution

5.22. PROBLEMS

1. Constructing the graph of the density function Construct the graph of the density function of the arc-sine distribution.
2. Arc-sine distribution on the interval \((-\frac{\pi}{2}, 0)\). Generalize the notion of the arc-sine distribution so that the interval is replaced by the interval \((-\frac{\pi}{2}, \frac{\pi}{2})\). Set up the formula of the
a. distribution
b. density
function of this generalized distribution. This distribution may be called arc-sine distribution with parameter \(A\).

3. Arc-sine distribution on the sky. The moon of Jupiter called Io goes around Jupiter on a (practically) circular path. The Earth is in plane of this circle. If, using a telescope, you look at Io, then you may see it between the two extreme positions. If the time instant of your experiment is chosen at random, then the position of Io is a random point between the two extreme positions. Check that the position of IO follows the arc-sine distribution.

5.23. 4.8 *** Cauchy distribution

5.24. EXCEL

The following files show the arc-sine distribution.

\[\text{Demonstration file: Cauchy distribution} \] 

\[\text{Demonstration file: Cauchy distribution} \]

5.25. PROBLEMS

1. Constructing the graph of the density function. Construct the graph of the density function of the Cauchy distribution.

2. Cauchy distribution with parameters \(A\) and \(B\). Assume that the random variable \(X\) follows the Cauchy distribution. Choose two constants, say \(A\) and \(B\), and consider the random variable \(Y = AX + B\). Find the formula of the distribution function and density function of \(Y\). We may call the distribution of \(Y\) "the Cauchy distribution with parameters \(A\) and \(B\)."

5.26. 4.9 *** Beta distributions

5.27. EXCEL

The following files show beta distributions.

\[\text{Demonstration file: Beta distributions} \]

\[\text{Demonstration file: Beta distributions} \]

\[\text{Demonstration file: Beta distributions} \]

5.28. PROBLEMS

1. Constructing the graph of the density function. Construct the graphs of the density functions of the beta distributions related to
Exercise Book to "Probability Theory with Simulations"

a. size 2 and rank 1;
b. size 2 and rank 2.

2. Constructing the graph of the density function Construct the graphs of the density functions of the beta distributions related to
a. size 3 and rank 1;
b. size 3 and rank 2.

3. Constructing the graph of the density function Construct the graphs of the density functions of the beta distributions related to
a. size 5 and rank 1;
b. size 5 and rank 2.

4. Making simulations Make simulations of the random variables with the distributions of the previous problems.

5. Calculating probabilities Three people, independently of each other, arrive to a party according to uniform distribution between 7pm and 7:30pm. What is the probability that
a. the first of them arrives
   i. between 7:10pm and 7:15pm?
   ii. between 7:15pm and 7:20pm?
b. the second of them arrives
   i. between 7:10pm and 7:15pm?
   ii. between 7:15pm and 7:20pm?
c. the third of them arrives
   i. between 7:10pm and 7:15pm?
   ii. between 7:15pm and 7:20pm?

6. Making simulations Make simulations for the previous problems, and check that the relative frequency is always close to the probability.

5.29. 4.10 Exponential distribution

5.30. EXCEL

The following files show beta distributions.

*emph*[Demonstration file: Exponential distribution \ef-200-16-00]

*emph*[Demonstration file: Exponential distribution \ef-200-17-00]

5.31. PROBLEMS

1. Calculating probabilities $X$ has an exponential distribution with parameter $1/3$. Find the probabilities:
Exercise Book to "Probability Theory with Simulations"

a. \( P(X < 3.5) \);

b. \( P(X > 2.5) \);

c. \( P(2.5 < X < 3.5) \);

d. \( P(X < 3.5 \mid X > 2.5) \);

e. \( P(X > 2.5 \mid X < 3.5) \).

2. Making simulations Make simulations for the previous problem, and check that the (conditional) relative frequency is always close to the (conditional) probability.

3. Finding a constant \( X \) has an exponential distribution with parameter \( 1/3 \). Find the constant \( c \) so that \( P(X < c) = 0.5 \).

4. Finding a constant \( X \) has an exponential distribution with parameter \( 1/3 \). Find the constant \( c \) so that \( P(X > c) = 0.7 \).

5. Life-time of an object with the memoryless property Assume that the life-time \( X \) of an object has the memoryless property, and thus it follows exponential distribution. Assume that \( P(X > 4.5) = 0.3 \). Find the probabilities:

a. \( P(X < 1) \);

b. \( P(X > 3) \);

c. \( P(1 < X < 5) \);

d. \( P(1 < X < 5 \mid X > 3) \).

6. Life-time of a cup Let \( X \) be the life-time of a cup in a self-service restaurant (measured in months). Assume that the life-time follows an exponential distribution, and the average life-time is 3.5 months. Calculate the probabilities

a. \( P(3 < X) \);

b. \( P(13 < X \mid 10 < X) \).

7. Earth-quakes The amount of time between two earth-quakes on an island follows exponential distribution with an average of 3 years.

a. Assuming that 3 years have passed since the last earth-quake, what is the probability that there will be an earth-quake during the next year?

b. Assuming that there will be an earth-quake during the next year, what is the probability that it will happen in the second half of the year?

8. Electric bulbs The life-time of a certain type of electric bulbs has an exponential distribution with an average life time of 150 hours.

a. What is the probability that a bulb of this type will work at least 200 hours?

b. What is the probability that a bulb of this type having been used successfully for 100 hours will work at least 200 more hours in addition to the 100 hours?

9. Life-time of a transistor Assume that the life time of a transistor has an exponential distribution with an average life time of 5 months.
a. What is the probability that the transistor has a life time between 2 and 3 months?

b. On condition that the transistor has a life time between 1 and 4 months, what is the probability that the transistor has a life time between 2 and 3 months?

10. Two electrical components Suppose that two independent electrical components are exponentially distributed with a common average of 2.5 years. What is the probability that at least one of them has a lifetime greater than 3 years?

11. Calculating probabilities and expected values Assume that $X$ is exponentially distributed with parameter 2.

   a. What is the probability that $X$ is between 1 and 4?

   b. On condition that $X > 1$, what is the probability that $X$ is between 1 and 4?

   c. On condition that $X < 4$, what is the probability that $X$ is between 1 and 4?

   d. On condition that $X < 5$, what is the probability that $X$ is between 1 and 4?

   e. Find the expected value of the distance between $X$ and 3.

   f. Find the expected value of the distance between $X$ and its expected value.

   g. Find $c$ so that $P(X < c) = 0.25$.

   h. Find $c$ so that $P(X > c) = 0.25$.

12. Calculating probabilities and expected values Assume that $X$ is exponentially distributed with parameter $\lambda$.

   a. What is the probability that $X$ is between $a$ and $b$?

   b. On condition that $X > a$, what is the probability that $X$ is between $a$ and $b$?

   c. On condition that $X < b$, what is the probability that $X$ is between $a$ and $b$?

   d. On condition that $X < d$, what is the probability that $X$ is between $a$ and $b$?

   e. Find the expected value of the distance between $X$ and its expected value.

   f. Find $c$ so that $P(X < c) = p$.

   g. Find $c$ so that $P(X > c) = q$.

13. Calculating probabilities and expected values The life-time of an electrical component of a computer is assumed to satisfy the memoryless property. If its average life-time is 2.5 years, then what is the probability that such a component survives

   a. more than 1 year?

   b. more than 2 years?

   c. more than 3 years?

   d. more than $x$ years?

14. Average life-time of an electrical component The life-time of an electrical component of a computer is assumed to satisfy the memoryless property. Assume that the probability that such a component survives more than 2.5 years is 0.7. How much is the average life-time of this electrical component?
15. Reciprocal of the arrival time The random arrival time, denoted by \( X \), of the goods I am waiting for has an exponential distribution with an expected value of 3 hours. My profit, denoted by \( Y \), from the goods is proportional to the reciprocal of the arrival time: \( Y = \frac{1}{X} \). Determine the distribution function and the median of \( Y \).

5.32. 4.11 *** Gamma distribution

5.33. EXCEL

The following files show beta distributions.

- \textit{Demonstration file: Gamma distribution \( \text{ef}-200-25-00 \)}
- \textit{Demonstration file: Gamma distribution \( \text{ef}-200-26-00 \)}
- \textit{Demonstration file: Gamma distribution \( \text{ef}-200-19-00 \)}
- \textit{Demonstration file: Gamma distribution \( \text{ef}-200-28-00 \)}
- \textit{Demonstration file: Gamma distribution \( \text{ef}-200-29-00 \)}
- \textit{Demonstration file: Gamma distribution \( \text{ef}-200-30-00 \)}
- \textit{Demonstration file: Gamma distribution \( \text{ef}-200-31-00 \)}
- \textit{Demonstration file: Gamma distribution \( \text{ef}-200-32-00 \)}

5.34. PROBLEMS

1. Calculating a probabilities \( X \) has a gamma distribution with parameter \( n = 2 \) and \( \lambda = \frac{1}{3} \). Find the probabilities:
   - a. \( P(X < 6.5) \);
   - b. \( P(X > 5.5) \);
   - c. \( P(5.5 < X < 6.5) \);
   - d. \( P(X < 6.5 \mid X > 5.5) \);
   - e. \( P(X > 5.5 \mid X < 6.5) \).

2. Calculating a probabilities \( X \) has a gamma distribution with parameter \( n = 3 \) and \( \lambda = \frac{1}{3} \). Find the probabilities:
   - a. \( P(X < 9.5) \);
   - b. \( P(X > 8.5) \);
   - c. \( P(8.5 < X < 9.5) \);
   - d. \( P(X < 9.5 \mid X > 8.5) \);
   - e. \( P(X > 8.5 \mid X < 9.5) \).
Exercise Book to “Probability Theory with Simulations”

3. Making simulations

Make simulations for the previous problem, and check that the (conditional) relative frequency is always close to the (conditional) probability.

5.35. 4.12 Normal distributions

5.36. EXCEL

The following files show normal distributions.

\textit{Demonstration file: Normal distribution}\ef{200-35-05}
\textit{Demonstration file: Normal distribution}\ef{200-33-00}
\textit{Demonstration file: Normal distribution}\ef{200-64-00}
\textit{Demonstration file: Normal distribution}\ef{200-35-00}
\textit{Demonstration file: Normal distribution}\ef{200-36-00}

5.37. PROBLEMS

1. Calculating probabilities

Assume that $X$ follows standard normal distribution. Calculate the following probabilities:

a. $P(X > 1.5)$;

b. $P(-1.3 < X < 2.5)$;

c. $P(-1.3 < X < 2.5|X > 1.5)$;

d. $P(X > 1.5) - 1.3 < X < 2.5)$.

\textit{Solution}\ef{Sol-03-13-01}

2. Calculating probabilities

Assume that $X$ follows normal distribution with parameters $\mu = 220, \sigma = 10$. Calculate the following probabilities:

a. $P(X > 225)$;

b. $P(215 < X < 229)$;

c. $P(215 < X < 229|X > 225)$;

d. $P(X > 225|215 < X < 229)$.

3. Height of a man

Assume that the height of a randomly chosen man in Hungary measured in cm is normally distributed with parameters $\mu = 180$ and $\sigma = 10$. What is the probability that the height of a randomly chosen man in Hungary is

a. less than 170?

b. greater than 185?

c. between 170 and 185?

d. smaller than 165 on condition that he is smaller than 185?
4. Height of a man Assume that the height of a randomly chosen man in a country has a normal distribution. The expected value is 175 cm, the standard deviation is 15 cm.

a. What is the probability that a randomly chosen man is
   i. shorter than 150 cm?
   ii. taller than 200 cm?
   iii. has a height between 150 and 200 cm?

b. Which is the height \( c \) so that
   i. the probability that a randomly chosen man is shorter than \( c \) is equal to 0.9?
   ii. the probability that a randomly chosen man is taller than \( c \) is equal to 0.8?

5. Weight of a sack of potato Let us assume that the weight \( X \) (measured in kg) of a sack of potato follows normal distribution with parameters \( \mu = 10 \) and \( \sigma = 1.1 \). Sketch a nice graph of the density function of \( X \). Calculate \( P(9.9 < X < 10.1) \).

6. Simulation of a standard normal random variable There are several ways to simulate a standard normal random variable. Check that each of the following simulations yields a standard normal random variable.

a. \( X = \text{RND} \)

b. \( X = \text{RND} \)

c. \( X = (\text{RND}_1 + \text{RND}_2 + \ldots + \text{RND}_{12}) - 6 \)
   where \( \text{RND}_1, \text{RND}_2, \ldots, \text{RND}_{12} \)
   are independent random variables, uniformly distributed between 0 and 1;

d. \( X = \sqrt{2 \ln(\text{RND}_1)} \cos(\text{RND}_2) \), where \( \text{RND}_1, \text{RND}_2 \)
   are independent random variables, uniformly distributed between 0 and 1.

Making several thousand simulations, you may be convinced that the last two simulations are much more efficient than the first two in the sense that they use much less time.

7. Height of a woman Assume that the height of a randomly chosen woman in a country has a normal distribution. The expected value is 175 cm, the standard deviation is 15 cm.

a. What is the probability that a randomly chosen man is
   i. shorter than 150 cm?
   ii. taller than 200 cm?
   iii. has a height between 150 and 200 cm?

b. Which is the height \( c \) so that
   i. the probability that a randomly chosen man is shorter than \( c \) is equal to 0.9?
   ii. the probability that a randomly chosen man is taller than \( c \) is equal to 0.8?

8. Height of teenager boys Let us assume that the height \( X \) of teenage boys (measured in cm-s) in a certain country follows normal distribution with parameters \( \mu = 172, 2 \) and \( \sigma = 12, 6 \). Using Excel, make the
graphs of the density function and of the distribution function. Calculate with Excel, for 4 decimal precision, the following unconditional and conditional probabilities:

(a) \( P(X < 200) \);
(b) \( P(X > 160) \);
(c) \( P(165 < X < 185) \);
(d) \( P(165 < X | X < 185) \);
(e) \( P(165 < X < 185 | X < 190) \).

9. Height of teenager boys in other countries The height \( X \) of teenager boys in other countries also follows normal distribution, but with other parameters. Let us assume that in European countries \( \mu = 12.6 \), but \( \mu \) may be different in different countries. For some reason, we want to know how the probabilities

(a) \( P(\text{height is more than } 160 \text{ cm}) \);
(b) \( P(\text{height is more than } 160 \text{ cm, but less than } 180 \text{ cm}) \),
depend on \( \mu \). Construct the graphs of these functions with Excel. \textit{Solution \Sol-03-13-03}

10. Height of teenager boys in Asian countries Let us assume that in Asian countries \( \sigma \) is less, say it is \( \sigma = 6.3 \). How the probabilities

(a) \( P(\text{height is more than } 160 \text{ cm}) \);
(b) \( P(\text{height is more than } 160 \text{ cm, but less than } 180 \text{ cm}) \),
depend on \( \mu \) now? Construct the graphs of these functions, and compare them to the previous ones.

11. Heights of teenager boys and girls Let us assume that the height of teenager boys (measured in cm-s) in a certain country follows normal distribution with parameters \( \mu = 182.2 \) and \( \sigma = 12.6 \), and the height of teenager girls follows normal distribution with parameters \( \mu = 162.2 \) and \( \sigma = 9.1 \). Let us assume that 75 percent of students in a high school are girls, 25 percent are boys.

(a) Using Excel, sketch the graph of the density function and of the distribution function of the height of a randomly chosen girl.
(b) Using Excel, sketch the graph of the density function and of the distribution function of the height of a randomly chosen boy.
(c) \( P(\text{height of a randomly chosen girl} < 175) =? \)
(d) \( P(\text{height of a randomly chosen boy} < 175) =? \)
(e) A student is chosen at random. The height of this student is \( X \). \( P(X < 175) =? \)
(f) \( P(165 < X < 185) =? \)
(g) \( P(165 < X < 185 | X < 190) =? \)
(h) On condition that \( X < 160 \), what is the probability that the student is a boy?
(i) On condition that \( X > 180 \), what is the probability that the student is a boy?
(j) On condition that \( 165 < X < 175 \), what is the probability that the student is a boy?
(k) Find a formula for the distribution function of \( X \).
1. Find a formula for the density function of $X$.

m. Using Excel, sketch the graph of the distribution function and of the density function.

12. Temperature measured in Celsius degrees The temperature $X$ (on May 1, at a certain place) measured in Celsius has a normal distribution with expected value 15.5 and standard deviation 4.5. Calculate the probabilities:

a. $P(X < 10 \text{ or } X > 20)$;

b. $P(X < 5 \text{ or } X > 25)$;

c. $P(X < 10 \text{ or } X > 20\mid X < 5 \text{ or } X > 25)$;

d. Find $a$ so that $P(X < a) = 0.25$;

e. Find $b$ so that $P(X < b) = 0.75$.

13. Temperature measured also in Kelvin degrees

a. The temperature $X$ (on May 1, at a certain place) measured in Celsius degrees has a normal distribution with expected value 15.5 and standard deviation 4.5. Calculate the probabilities:

i. $P(X < 10 \text{ or } X > 20)$;

ii. $P(X < 5 \text{ or } X > 25)$;

iii. $P(X < 10 \text{ or } X > 20\mid X < 5 \text{ or } X > 25)$.

b. The same temperature measured in Kelvin degrees is denoted by $Z$. The conversion formula between $X$ and $Z$ is $Z = X + 273.15$. The same temperature measured in Réaumur degrees is denoted by $W$. The conversion formula between $X$ and $W$ is $W = 8/10X$. The same temperature measured in Fahrenheit is denoted by $Y$. The conversion formula between $X$ and $Y$ is $Y = 9/5X + 32$. Figure out

i. the expected value of $Z$, $W$ and $Y$;

ii. the standard deviation of $Z$, $W$ and $Y$;

iii. make nice figures for the density functions of $X$, $Z$, $W$ and $Y$, as well. Figures made by a computer are appreciated.

14. Comparing Excel functions Check that, in Excel, NORMSDIST$(x)$ and NORMDIST$(x; 0; 1; \text{TRUE})$ mean the same thing.

15. Comparing Excel functions Check that, in Excel, NORMSDIST$(\frac{x - \mu}{\sigma})$ and NORMDIST$(x; \mu; \sigma; \text{TRUE})$ mean the same thing.

16. Amount of pollution Assume that the amount of pollution (measured in milligramms) in a liter of water of a lake follows normal distribution with expected value 6.5 and standard deviation 0.5. What is the probability that the average of

a. 4 experimental results is in the interval $[6.4; 6.6]$?

b. 25 experimental results is in the interval $[6.4; 6.6]$?

17. Finding a critical value $c$ Assume that a measurement result $X$ has a normal distribution with expected value 220 and standard deviation 10.

a. What is the probability that $X$ is larger than 225?
b. If we know that the measurement result $X$ is larger than 210, then what is the probability that $X$ is larger than 225?

c. How much is the critical value $c$ for which $P(X > c) = 0.9$?

d. How much is the critical value $c$ for which $P(X > c|X > 210) = 0.9$?

18. How does the probability depend on $\mu$? $X$ has a normal distribution with an expected value $\mu$ and standard deviation 5. Using Excel, make a graph of the following probabilities for $150 < \mu < 200$:

a. $P(X < 180)$;

b. $P(X > 170)$;

c. $P(170 < X < 180)$;

d. $P(165 < X < 185)$.

Take then the value of the standard deviation as a parameter, and study how the graphs change when you change the parameter.

**5.38. 4.13 *** Distributions derived from normal**

**5.39. EXCEL**

The following files show distributions derived from normal distributions.

- emph{Demonstration file: Log-normal distribution \ref{200-38-00}}
- emph{Demonstration file: Chi-square distribution, n=3 \ref{200-39-00}}
- emph{Demonstration file: Chi-square distribution \ref{200-40-00}}
- emph{Demonstration file: Chi-distribution, n=3 \ref{200-41-00}}
- emph{Demonstration file: Chi-distribution \ref{200-42-00}}
- emph{Demonstration file: Student-distribution (T-distribution) and random variable, n=3 \ref{200-43-00}}
- emph{Demonstration file: Student-distribution (T-distribution) and random variable \ref{200-44-00}}
- emph{Demonstration file: F-distribution, m=3, n=4 \ref{200-45-00}}
- emph{Demonstration file: F-distribution \ref{200-46-00}}

**5.40. PROBLEMS**

1. Simulation Simulate random variables following the distributions derived from normal.

2. Densities Construct the graphs of the density functions of the distributions derived from normal.

3. Distribution functions Construct the graphs of the distribution functions of the distributions derived from normal.

**5.41. 4.14 ***Generating a random variable with a given continuous distribution**
5.42. EXCEL

Study the following file, and notice that if we define the random variable as $X = \text{RND}^c$, that is, we plug RND into $x = y^c$, then the empirical distribution function approaches the function $y = x^{1/c}$, which is the inverse of the function $x = y^c$. This fact means that the theoretical distribution function of $X = \text{RND}^c$ is really the inverse of the function into which we plug RND.

*Demonstration file: Empirical distribution function ref:200-01-00*

5.43. PROBLEMS

1. Generating a random variable which follows a given continuous distribution Choose a continuous distribution so that the formula of its distribution function and the formula of the inverse of the distribution function is known for you. Define a random variable by plugging a random number into the inverse of the distribution function. Then, making more experiments, construct the empirical distribution function of the random variable, and be convinced that the empirical distribution function approaches the given distribution function.

2. Uniformly distributed random variable on the interval $[0; 1]$ Simulate a uniformly distributed random variable on the interval $[0; 1]$.

3. Uniformly distributed random variable on $[0; B]$ Simulate a uniformly distributed random variable on the interval $[0; B]$.

4. Uniformly distributed random variable on $[A; B]$ Simulate a uniformly distributed random variable on the interval $[A; B]$.

5. "Linear" density function Assume that the density function of a random variable is linear on the interval $[0; 1]$, that is, $f(x) = ax + b$ if $0 < x < 1$. Use $b$ as a parameter.
   - a. Calculate the distribution function.
   - b. Determine the inverse of the distribution function.
   - c. Make a point-cloud for a random variable which follows the distribution defined by such a density function.

*Solution Sol-03-15-01*

6. Density linearly increasing on $[0; 1]$ Assume that the density function of a continuous distribution is linearly increasing on the interval $[0; 1]$, that is, $f(x) = cx$, if $0 < x < 1$, and 0 otherwise, where $c$ is a constant. How much is $c$? Determine the distribution function $F(x)$. Find the inverse of the distribution function. Use the inverse of the distribution function to simulate a random variable $X$ which follows this distribution.

7. Density linearly increasing on the interval $[0; B]$ Assume that the density function of a continuous distribution is linearly increasing on the interval $[0; 1]$, that is, $f(x) = cx$, if $0 < x < B$, and 0 otherwise, where $c$ is a constant. How much is $c$? Determine the distribution function $F(x)$. Find the inverse of the distribution function. Use the inverse of the distribution function to simulate a random variable $X$ which follows this distribution.

8. Density linearly decreasing on $[0; 1]$ Assume that the density function of a continuous distribution is linearly decreasing on the interval $[0; 1]$, that is, $f(x) = c(1 - x)$, if $0 < x < 1$, and 0 otherwise, where $c$ is a constant. How much is $c$? Determine the distribution function $F(x)$. Find the inverse of the distribution function. Use the inverse of the distribution function to simulate a random variable $X$ which follows this distribution.
9. Density linearly decreasing on \([0; 1]\) Assume that the density function of a continuous distribution is linearly increasing on the interval \([a; b]\), that is, 
   \[ f(x) = \begin{cases} \frac{1}{b-a}x + a, & \text{if } a < x < b, \\ 0, & \text{otherwise} \end{cases} \]
   where \(a\) and \(b\) are constants. How much is \(E(X)\)? Determine the distribution function \(F(x)\). Find the inverse of the distribution function. Use the inverse of the distribution function to simulate a random variable \(X\) which follows this distribution.

10. Density quadratically increasing on \([0; 1]\) Assume that the density function of a continuous distribution is linearly increasing on the interval \([0; 1]\), that is, 
    \[ f(x) = cx^2, \quad \text{if } 0 < x < 1, \quad \text{and } 0 \text{ otherwise}, \]
    where \(c\) is a constant. How much is \(E(X)\)? Determine the distribution function \(F(x)\). Find the inverse of the distribution function. Use the inverse of the distribution function to simulate a random variable \(X\) which follows this distribution.

11. Density quadratically increasing on \([0; B]\) Assume that the density function of a continuous distribution is linearly increasing on the interval \([0; B]\), that is, 
    \[ f(x) = cx^2, \quad \text{if } 0 < x < B, \quad \text{and } 0 \text{ otherwise}, \]
    where \(c\) is a constant. How much is \(E(X)\)? Determine the distribution function \(F(x)\). Find the inverse of the distribution function. Use the inverse of the distribution function to simulate a random variable \(X\) which follows this distribution.

12. Generating a random variable which follows a normal distribution Generate a random variable which follows
    a. the normal distribution with \(\mu = 75\) and \(\sigma = 5\);
    b. the normal distribution with \(\mu = 175\) and \(\sigma = 5\);
    c. the normal distribution with \(\mu = 175\) and \(\sigma = 25\);
    d. the normal distribution with \(\mu\) and \(\sigma\), where \(\mu\) and \(\sigma\) are parameters.

13. Generating a random variable which follows an exponential distribution Generate a random variable which follows
    a. the exponential distribution with \(\lambda = 2.5\);
    b. the exponential distribution with \(\lambda = 5\);
    c. the exponential distribution with an expected value 2.
    d. the exponential distribution with \(\lambda\), where \(\lambda\) is parameter.

5.44. 4.15 Expected value of continuous distributions

5.45. EXCEL

The following file shows the expected values of some continuous distributions.

\texttt{Demonstration file: Expected value of continuous distributions Neff-200-57-61}

5.46. PROBLEMS

1. Calculating the expected value from the density function Assume that the density function of a random variable \(X\) is
   a. \(f(x) = 2x\), if \(0 \leq x \leq 1\);
b. $f(x) = x/2$, if $0 \leq x \leq 2$;

c. $f(x) = \frac{1}{\sqrt{x}}$, if $0 \leq x \leq 1$.
How much is the Expected value of $X$.

2. Calculating the expected value from the distribution function Assume that the distribution function of a random variable $X$ is

a. $F(x) = \frac{x}{2}$, if $0 \leq x \leq 2$;

b. $F(x) = 1 - e^{-x}$, if $x > 0$, and $f(x) = 0$ otherwise.
How much is the Expected value of $X$.

3. Calculating the expected value of random variables A random number is generated by a calculator or computer. Let $X$ be its

a. cube;

b. logarithm;

c. square;

d. square-root;

e. reciprocal.
Find the expected value of $X$ in each case.

4. Calculating the expected value of random variables Calculate the expected value of $X$, if

a. $X = \text{RND}^2$;

b. $X = \text{RND}^3$;

c. $X = \sqrt{\text{RND}}$;

d. $X = 1/\text{RND}$;

e. $X = 1/\text{RND}^2$;

f. $X = -\ln(\text{RND})$.

5. Calculating the expected value of random variables Four random numbers are generated: $\text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4$.

a. Let $X_1$ denote the largest value of the values $\text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4$. Find the expected value of $X_1$.

b. Let $X_2$ denote the second largest value of the values $\text{RND}_1, \text{RND}_2, \text{RND}_3, \text{RND}_4$. Find the expected value of $X_2$.

6. Calculating the expected value of random variables Ten random numbers are generated by a calculator or computer. Let $X$ be the 8th smallest of them. Find the expected value of $X$.

5.47. 4.16 Expected value of a function of a continuous random variable

5.48. EXCEL
The following file shows the expected values of some functions of a random number.

\emph{Demonstration file: Expected value of some functions of a random number \ref{200-58-00}}

5.49. PROBLEMS

1. Calculating the expected value of a function of a random variable The density function of $X$ is $f(x) = 5x^4 \ (0 < x < 1)$. Let $Y = X^2$.
   a. Calculate the expected value of $Y$ without using the density function of $Y$, but using the density function of $X$.
   b. Calculate the distribution function and the density function of $Y$, and then the expected value of $Y$ using the density function of $Y$.

2. Calculating the expected value of a function of a random variable Calculate the expected value of
   a. $X = \text{RND}$;
   b. $X = \text{RND}^2$;
   c. $X = \text{RND}^3$;
   d. $X = \sqrt{\text{RND}}$.
   Make simulation (1000 experiments) for each of the random variables above, and check that the average of the experimental result is close to the expected value.

3. Calculating the expected value of a distance $X$ has an exponential distribution with parameter $1/3$. Let us consider the distance
   a. between the random value $X$ and the value 1, which is $|X - 1|$.
   b. between the random value $X$ and the value 2, which is $|X - 2|$.
   c. between the random value $X$ and a constant $c$, which is $|X - c|$.
   Calculate the expected value of this distance.

4. Calculating the expected value of a squared distance $X$ has an exponential distribution with parameter $1/3$.
   a. Let us consider the squared distance between the random value $X$ and the value 1, which is $(X - 1)^2$.
      Calculate the expected value of $(X - 1)^2$.
   b. Let us consider the squared distance between the random value $X$ and the value 2, which is $(X - 2)^2$.
      Calculate the expected value of $(X - 2)^2$.
   c. Let us consider the squared distance between the random value $X$ and a constant $c$, which is $(X - c)^2$.
      Calculate the expected value of $(X - c)^2$.

5. Calculating the expected value of a distance Calculate the expected value of the distance from the expected value for the random variables:
   a. $X = 5 \times \text{RND}$;
   b. $X = 10 \times \text{RND}$;
c. \( X = \text{RND}^3 \).

6. Calculating the expected value of a squared distance Calculate the expected value of the squared distance from the expected value for the random variables:

a. \( X = 5 \times \text{RND} \);

b. \( X = 10 \times \text{RND} \);

c. \( X = \text{RND}^3 \).

7. Approximating the average by a constant 100 independent random numbers are generated between 0 and 1 according to uniform distribution. Determine a constant number which can be considered as a good approximation for the average of their cubes.

5.50. 4.17 *** Median

5.51. EXCEL

The following file shows the median of the exponential distribution.

\( \text{Demonstration file: Median of the exponential distribution} \)

5.52. PROBLEMS

1. Calculating the median Calculate the median of the distribution with the distribution function: \( F(x) = 2x(0 < x < 1) \). Compare it to the expected value.

2. Calculating the median Calculate the median of the distribution with the distribution function: \( F(x) = 2 - 2x(0 < x < 1) \). Compare it to the expected value.

3. Calculating the median Calculate the median of the exponential distribution with parameter \( \lambda \). Compare it to the expected value.

5.53. 4.18 Standard deviation, etc.

5.54. EXCEL

The following files show the meaning of the standard deviation.

\( \text{Demonstration file: Standard deviation for a data-set} \)

\( \text{Demonstration file: Standard deviation for a random variable} \)

The following file shows the standard deviation of some distributions.

\( \text{Demonstration file: Standard deviation for a data-set} \)

5.55. PROBLEMS

1. Comparing the empirical variance to the theoretical We have learnt that the variance of the uniform distribution on the \([0;1]\) interval is 1/12. Generate 1000 experimental results for the random variable
RND

and using the VARP command, be convinced that the variance of the generated data set is approximately 1/12.

2. Comparing the empirical variance to the theoretical Generate 1000 experimental results for the random variable

\[ RND_1 + RND_2 \]

and using the VARP command, be convinced that the variance of the generated data set is approximately 2/12.

3. Comparing the empirical variance to the theoretical Generate 1000 experimental results for the random variable

\[ RND_1 + RND_2 + RND_3 \]

and using the VARP command, be convinced that the variance of the generated data set is approximately 3/12.

4. Comparing the empirical variance to the theoretical Generate 1000 experimental results for the random variable

\[ RND_1 + RND_2 + \ldots + RND_{12} \]

and using the VARP command, be convinced that the variance of the generated data set is approximately 1.

5. Comparing the empirical variance to the theoretical We have learnt that the variance of the standard normal distribution is 1. Generate 1000 random numbers by the NORMSINV(RAND()) or by the NORMINV(RAND();0;1) command, and using the VARP command, be convinced that the variance of the generated data set is approximately 1.

6. Comparing the empirical variance to the theoretical Generate 1000 random numbers by the NORMINV(RAND();10;3) command, and using the VARP command. Be convinced that the variance of the generated data set is approximately 9.

7. Comparing the empirical variance to the theoretical Generate 1000 random numbers by the NORMINV(RAND();10;4) command, and using the VARP command. Be convinced that the variance of the generated data set is approximately 16.

8. Summation rule for the variance Let us take now the sum of the independent random variables generated in the previous 2 problems, that is, let us generate 1000 random numbers by the NORMINV(RAND();10;3) + NORMINV(RAND();20;4) command. Be convinced that the variance of the generated data set is approximately 25.

9. Average rule for the variance \( X \) has a normal distribution with parameters \( \mu = 300 \) and \( \sigma = 50 \).

   a. Calculate the probability that \( X \) is between 285 and 315.

   b. We make 25 experiments for \( X \). Calculate the probability that the average of the experimental results is between 285 and 315.

10. 25 sacks of potato Let us assume that the weight \( X \) (measured in kg-s) of a sack of potato follows normal distribution with parameters \( \mu = 10 \) and \( \sigma = 0.1 \). Assume that 25 sacks are examined. Let \( \overline{X}_{25} \) be their average weight. Sketch a nice graph of the density function of \( \overline{X}_{25} \). Calculate \( P(9.9 < \overline{X}_{25} < 10.1) \).

5.56. 4.19 *** Poisson-processes
5.57. EXCEL

The following files simulate Poisson processes with different intensity functions. First, second and third occurrences are observed in them.

\emph{Demonstration file: First (second and third) occurrence, homogeneous Poisson-process [ef-200-21-00]}

\emph{Demonstration file: First (second and third) occurrence, "trapezoid shaped" intensity function [ef-200-22-00]}

\emph{Demonstration file: First (second and third) occurrence, linearly increasing intensity function [ef-200-23-00]}

\emph{Demonstration file: First (second and third) occurrence, decreasing intensity function [ef-200-24-00]}

5.58. PROBLEMS

1. Number of telephone calls during an hour Assume that during the working hours the average number of telephone calls arriving to our department in an hour is 4.5. What is the probability that, between 11 and 12am,
   a. no calls arrive?
   b. 1 call arrives?
   c. 2 calls arrives?
   d. less than 5 calls arrive?
2. Number of telephone calls during 20 minutes Assume that during the working hours the average number of telephone calls arriving to our department in an hour is 4.5. What is the probability that, between 11 and 11:20am,
   a. no calls arrive?
   b. 1 call arrives?
   c. 2 calls arrives?
   d. less than 5 calls arrive?
3. Number of telephone calls on condition that ... Assume that during the working hours the average number of telephone calls arriving to our department in an hour is 4.5. On condition that less than 5 calls arrive between 11 and 12am, what is the probability that 2 calls arrive
   a. between 11 and 12am?
   b. between 11 and 11:20am?

5.59. 4.20 *** Transformation from line to line

5.60. EXCEL

Here are some files to study transformations from line to line.

\emph{Demonstration file: Uniform distribution transformed by a power function with a positive exponent [ef-200-95-00]}
5.61. PROBLEMS

1. Transformation from line to line The density function of $X$ is $f(x) = 5x^4$ $(0 < x < 1)$. Let $Y = X^2$. Calculate and visualize the distribution function and the density function of $Y$. Calculate the expected value, the variance and standard deviation of $Y$ without using the density function of $Y$, but using the the density function of $X$. Calculate the expected value, the variance and standard deviation of $Y$ using the density function of $Y$.

2. Salt crystals The salt crystals have a cubic shape so that their size is random. Assume that the length $X$ of a side of a salt crystal cube (measured in mm) has an exponential distribution with an expected value 0.5.
   a. Calculate the expected value of the surface $Y$ of a salt crystal ($Y = 6 * X^2$).
   b. Determine the formula of the distribution function $G(y) = P(Y < y) = ...$ and the formula of the density function $g(y)$ of the surface of a randomly chosen salt crystal.
   c. Determine the formula of the distribution function $H(v) = P(V < v) = ...$ and the formula of the density function $h(v)$ of the volume $V$ of a randomly chosen salt crystal ($V = X^3$).
   d. Which is the critical volume-value for which the half of the salt crystals have a smaller volume than the critical value, and half of the salt crystals have a larger volume than the critical value?

3. Log-normal distribution Assume that a measurement result $X$ has a normal distribution. Determine the density function of the random variable $Y = \exp(X)$. The distribution of $Y$ is called: log-normal distribution, since the logarithm of $Y$ has a normal distribution. Try to figure out: for what kind of real life problems we can use log-normal distribution.

6.5 Two-dimensional random variables and distributions

6.1. PROBLEMS

1. Calculating probabilities by summation, using Excel The distribution of a two-dimensional random variable $(X, Y)$ is given by:
Calculate the probability that $6 < X + Y < 9$. Solution

2. Continuation of the previous problem Invent other events related to the two-dimensional random variable $(X, Y)$, and calculate their probabilities by summation.

3. A two dimensional discrete random variable Simulate with Excel
   a. a random variable $Y$ uniformly distributed on the set $1, 2, \ldots, n$, where $n$ is a parameter;
   b. a random variable $X$ uniformly distributed on the set $1, 2, 3, 4, 5, 6$, and then a random variable $Y$ uniformly distributed on the set $1, 2, \ldots, X$
      , and then consider the two-dimensional random variable $(X, Y)$. What is the distribution of $(X, Y)$?

4. Working with two-dimensional density functions For each of the cases below, the density function $f(x, y)$ of a two-dimensional random point $(X, Y)$ is defined. In each case, determine the density function $f_1(x)$ of $X$ and the density function $f_2(y)$ of $Y$. Try to visualize the densities by mass distributions. Try to draw the graphs of the densities (a surface for $f(x, y)$, a curve for $f_1(x)$ and for $f_2(y)$). Imagine that, using a computer, we generate a point-cloud for $(X, Y)$ and project it onto the $x$- and $y$-axes to get the point-clouds for $X$ and $Y$. Try to draw (with your pencil or pen) the point-clouds which you think would resemble to the point-clouds made by the computer.
   a. $f(x, y) = 1 \quad (0 < x < 1, \quad 0 < y < 1);$  
   b. $f(x, y) = 1/6 \quad (0 < x < 3, \quad 0 < y < 2);$  
   c. $f(x, y) = 4xy \quad (0 < x < 1, \quad 0 < y < 1);$  
   d. $f(x, y) = 6xy^2 \quad (0 < x < 1, \quad 0 < y < 1);$  
   e. $f(x, y) = e^{-x-y} \quad (x > 0, \quad y > 0);$  
   f. $f(x, y) = 10e^{-5x-2y} \quad (x > 0, \quad y > 0);$  
   g. $f(x, y) = 2 \quad (x > 0, \quad y > 0, \quad 0 < x + y < 1);$  
   h. $f(x, y) = 2 \quad (0 < x < y < 1);$  
   i. $f(x, y) = 6(y-x) \quad (0 < x < y < 1);$  
   j. $f(x, y) = 3x \quad (0 < x < y < 1);$  
   k. $f(x, y) = 3(1-y) \quad (0 < x < y < 1).$
   Moreover, calculate the probabilities:
a. $P(X < 0.5)$;

b. $P(X + Y < 1)$;

c. $P(Y > X + 0.5)$;

d. $P(Y > \sqrt{X})$.

6.2. 5.1 Uniform distribution on a two-dimensional set

6.3. PROBLEMS

1. Uniform distribution on a square Assume that $(X, Y)$ follows uniform distribution on the unit square $\{(x, y): 0 < x < 1, 0 < y < 1\}$ . Calculate the probabilities and conditional probabilities:

a. $P(X < 0.8)$;

b. $P(Y > 0.3)$;

c. $P(X < 0.8 \text{ and } Y > 0.3)$;

d. $P(X < 0.8 \mid Y > 0.3)$;

e. $P(Y > 0.3 \mid X < 0.8)$;

f. $P(X + Y < 0.8)$;

g. $P(Y < X/2)$;

h. $P(Y < X^2)$.

2. Uniform distribution on a triangle A random point $(X, Y)$, uniformly distributed in the triangle with vertices $(0, 0), (2, 0), (0, 3)$, is considered. Calculate the probabilities:

a. $P(Y < X)$;

b. $P(Y < X^2)$.

c. Determine both the distribution and the density function of $Y/X$.

6.4. 5.2 *** Beta distributions in two-dimensions

6.5. EXCEL

The following files show two-dimensional beta point-clouds. Study them.

\textbf{Demonstration file: Two-dimensional beta point-cloud related to size 2 and ranks 1 and 2} \ref{200-69-00}

\textbf{Demonstration file: Two-dimensional beta point-cloud related to size 3 and ranks 1 and 2} \ref{200-70-00}
The following file serves to study two-dimensional point-clouds for arrival times.

6.6. PROBLEMS

1. Smallest and biggest of 10 random numbers Three independent random numbers (uniformly distributed between 0 and 1) are generated. Let $X$ be the smallest, $Y$ be the biggest of them. Determine the density function $f(x, y)$ of $(X, Y)$, visualize the density by a mass distribution, try to draw the graph (a surface) for $f(x, y)$, and if you have access to a computer, then generate a point-cloud for $(X, Y)$. Moreover, calculate the probabilities:
   a. $P(X < 0.5)$;
   b. $P(X + Y < 1)$;
   c. $P(Y > X + 0.5)$;
   d. $P\left(Y > \sqrt{X}\right)$.

2. Smallest and second smallest of 3 random numbers Three independent random numbers (uniformly distributed between 0 and 1) are generated. Let $X$ be the smallest, $Y$ be the second smallest of them. Determine the density function $f(x, y)$ of $(X, Y)$, visualize the density by a mass distribution, try to draw the graph (a surface) for $f(x, y)$, and if you have access to a computer, then generate a point-cloud for $(X, Y)$. Moreover, calculate the probabilities:
   a. $P(X < 0.5)$;
   b. $P(X + Y < 1)$;
   c. $P(Y > X + 0.5)$;
   d. $P\left(Y > \sqrt{X}\right)$.

3. Second smallest and biggest of 3 random numbers Three independent random numbers (uniformly distributed between 0 and 1) are generated. Let $X$ the second smallest, $Y$ the biggest of them. Determine the density function $f(x, y)$ of $(X, Y)$, visualize the density by a mass distribution, try to draw the graph (a surface) for $f(x, y)$, and if you have access to a computer, then generate a point-cloud for $(X, Y)$. Moreover, calculate the probabilities:
   a. $P(X < 0.5)$;
   b. $P(X + Y < 1)$;
c. $P(Y > X + 0.5)$;

d. $P\left( Y > \sqrt{X} \right)$.

4. Smallest and second smallest of 4 random numbers

Four independent random numbers (uniformly distributed between 0 and 1) are generated. Let $X$ be the smallest, $Y$ be the biggest of them. Determine the density function $f(x, y)$ of $(X, Y)$, visualize the density by a mass distribution, try to draw the graph (a surface) for $f(x, y)$, and if you have access to a computer, then generate a point-cloud for $(X, Y)$.

Moreover, calculate the probabilities:

a. $P(X < 0.5)$;

b. $P(X + Y < 1)$;

c. $P(Y > X + 0.5)$;

d. $P\left( Y > \sqrt{X} \right)$.

5. Smallest and second smallest of 4 random numbers

Four independent random numbers (uniformly distributed between 0 and 1) are generated. Let $X$ be the smallest, $Y$ be the second smallest of them. Determine the density function $f(x, y)$ of $(X, Y)$, visualize the density by a mass distribution, try to draw the graph (a surface) for $f(x, y)$, and if you have access to a computer, then generate a point-cloud for $(X, Y)$.

Moreover, calculate the probabilities:

a. $P(X < 0.5)$;

b. $P(X + Y < 1)$;

c. $P(Y > X + 0.5)$;

d. $P\left( Y > \sqrt{X} \right)$.

6. Second smallest and third smallest of 4 random numbers

Four independent random numbers (uniformly distributed between 0 and 1) are generated. Let $X$ be the second smallest, $Y$ be the third smallest of them. Determine the density function $f(x, y)$ of $(X, Y)$, visualize the density by a mass distribution, try to draw the graph (a surface) for $f(x, y)$, and if you have access to a computer, then generate a point-cloud for $(X, Y)$.

Moreover, calculate the probabilities:

a. $P(X < 0.5)$;

b. $P(X + Y < 1)$;

c. $P(Y > X + 0.5)$;

d. $P\left( Y > \sqrt{X} \right)$.

6.7. 5.3 Projections and conditional distributions

6.8. EXCEL
Here are some files to visualize projections and conditional distributions. Study them.

- **Demonstration file: $X = RND_1$, $Y = X RND_2$, projections and conditional distributions**
- **Demonstration file: Two-dim beta distributions, $n = 10$, projections and conditional distributions**
- **Demonstration file: Two-dim beta distributions, $n \leq 10$, projections and conditional distributions**
- **Demonstration file: Conditional distributions, uniform on parallelogram**
- **Demonstration file: Conditional distributions, (RND1;RND1RND2)**
- **Demonstration file: Conditional distributions, uniform on triangle**
- **Demonstration file: Conditional distributions, Bergengoc bulbs**
- **Demonstration file: Conditional distributions, standard normal**
- **Demonstration file: Conditional distributions, normal**

### 6.9. PROBLEMS

1. Two dice, largest and smallest Toss 2 dice and let $X$ be the largest and $Y$ be the smallest of the two numbers. Find out (make a numerical table for)
   a. the distribution of $(X,Y)$;
   b. the conditional distributions of $Y$ on condition that $X$ is given, and
   c. the conditional distributions of $X$ on condition that $Y$ is given.

2. Working with conditional distributions, discrete case Assume that $X$ follows discrete uniform distribution between 1 and 10 (1 and 10 included), and if $X = x$, then $Y$ follows discrete uniform distribution between 1 and $x$ (1 and $x$ included). Find the conditional distribution of $Y$ on condition that
   a. $X = 4$;
   b. $X = 5$;
   c. $X = x$.
   d. What are the possible values of $(X,Y)$? Find the distribution of $(X,Y)$.

3. Working with conditional distributions, discrete case Assume that $X$ follows discrete uniform distribution between 1 and 10 (1 and 10 included), and if $X = x$, then $Y$ follows discrete uniform distribution between 1 and $x$ (1 and $x$ included). What are the possible values of $Y$? Find the distribution of $Y$. Find the conditional distributions of $X$ on condition that
   a. $Y = 6$;
   b. $Y = 7$;
   c. $Y = y$. 
4. A particle landing in semi-circle Assume that a particle lands on the semi-circle $x^2 + y^2 \leq 1$, $y \geq 0$ according to uniform distribution, but we are able to observe only the coordinate of the landing point.

a. What is the conditional density function of the $Y$ coordinate, if $X = 1/3$?

b. What is the conditional density function of the $Y$ coordinate, if $X = x$? (Pay attention to give the domain of the conditional density function correctly.)

5. Conditional density and distribution functions The density function of $(X, Y)$ is $f(x, y) = 15xy^2$ ($0 < y < x < 1$). Calculate and visualize by graphs of the density functions and distribution functions.

a. $f_1(x);

b. f_{2 | 1}(y|x);

c. f_2(y);

d. f_{1 | 2}(x|y);

e. F_1(x);

f. $F_{2 | 1}(y|x);

g. f_2(y);

h. $F_{1 | 2}(x|y)$.

6. Calculating the conditional density functions Assume that the density function of the two-dimensional random variable $(X, Y)$ is $f(x, y) = 6(y - x)$ ($0 < x < y < 1$). Find a. the density function of $X$, and b. the conditional density function of $Y$ on condition that $X = x$. (Pay attention to the correct definition of the domain of the formulas.)

7. Working with conditional distributions, continuous case Let $X \sim \text{RND}_1$ and $Y = X \ast \text{RND}_2$.

a. Find the formula of the density function of $X$. (Denote the variable of the density function of $X$ by $x$. Pay attention to set up the domain of the formula correctly.) Calculate the probabilities:

i. $P(X < 0.1)$;

ii. $P(X < 0.2)$;

iii. $P(0.1 < X < 0.2)$.

b. Find the formula of the (conditional) density function of $Y$ on condition that $X = 0.25$. Calculate the conditional probabilities:

i. $P(Y < 0.1 \mid X = 0.25)$,

ii. $P(Y < 0.2 \mid X = 0.25)$;

iii. $P(0.1 < Y < 0.2 \mid X = 0.25)$.
c. Find the formula of the (conditional) density function of $Y$ on condition that $X = 0.5$. Calculate the conditional probabilities:
   
   i. $P(Y < 0.1 \mid X = 0.5)$;
   
   ii. $P(Y < 0.2 \mid X = 0.5)$;
   
   iii. $P(0.1 < Y < 0.2 \mid X = 0.5)$.

d. Find the formula of the (conditional) density function of $Y$ on condition that $X = 0.75$. Calculate the conditional probabilities:
   
   i. $P(Y < 0.1 \mid X = 0.75)$;
   
   ii. $P(Y < 0.2 \mid X = 0.75)$;
   
   iii. $P(0.1 < Y < 0.2 \mid X = 0.75)$.

e. Find the formula of the (conditional) density function of $Y$ on condition that $X = x$. Calculate the conditional probabilities:
   
   i. $P(Y < 0.1 \mid X = x)$;
   
   ii. $P(Y < 0.2 \mid X = x)$,
   
   iii. $P(0.1 < Y < 0.2 \mid X = x)$.
   
   (In each case, denote the variable of the (conditional) density function of $Y$ by $y$. Pay attention to set up the domain of the formula correctly.)

f. Find the formula of the density function of $(X, Y)$. (Denote the variables of the density function of $(X, Y)$ by $x$ and $y$. Pay attention to set up the domain of the formula correctly. Make a figure on the plane about the domain.) Calculate the (at least four of) following probabilities and conditional probabilities:
   
   i. $P(X + Y < 1)$;
   
   ii. $P(X + Y < 0.5)$;
   
   iii. $P(Y < 0.25)$;
   
   iv. $P(Y < 0.5)$;
   
   v. $P(Y < 0.75)$;
   
   vi. $P(Y < 0.5 \ast X)$;
   
   vii. $P(X \ast Y < 0.5)$;
   
   viii. $P(Y < 0.25 \ast X \cap X \ast Y < 0.5)$;
   
   ix. $P(Y < 0.25 \ast X \mid X \ast Y < 0.5)$;
   
   x. $P(X \ast Y < 0.5 \mid Y < 0.25 \ast X)$.
   
   (In each case, make a figure on the plane about the event(s) involved in that problem, and then set up the associated double integral(s), and then calculate the double integral(s).)
g. You may make computer simulations, and then probabilities can be approximated by relative frequencies. Try to do so.

h. Find the formula of the (unconditional) density function of $Y$ so that you determine the distribution function of $Y$ by calculating the probability $P(Y < y)$, and then you differentiate with respect to $y$.

8. Determining density function $f(x, y)$ Let us define a two-dimensional random variable as follows.

a. Let the horizontal coordinate $X$ be the square-root of a random number, and let the vertical coordinate $Y$ be uniformly distributed between 0 and $X$.

b. Let the horizontal coordinate $X$ be the square-root of a random number, and let the vertical coordinate $Y$ be uniformly distributed between 0 and $X^2$.

c. Let the horizontal coordinate $X$ be the square-root of a random number, and let the vertical coordinate $Y$ be $X$ multiplied by the square-root of another random number.

d. Let the horizontal coordinate $X$ be the cube-root of a random number, and let the vertical coordinate $Y$ be uniformly distributed between 0 and $X$.

e. Let the horizontal coordinate $X$ follow exponential distribution with parameter 1, and let the vertical coordinate $Y$ follow exponential distribution with parameter $X$.

f. Let the horizontal coordinate $X$ follow exponential distribution with parameter 1, and let the vertical coordinate $Y$ follow exponential distribution with parameter $1/X$.

g. Let the horizontal coordinate $X$ follow standard normal distribution, and let the vertical coordinate $Y$ follow normal distribution with parameters $X/2$ and 1.

h. Let the horizontal coordinate $X$ follow standard normal distribution, and let the vertical coordinate $Y$ follow normal distribution with parameters $2X$ and 1.

Make 1000 experiments and visualize the point-cloud. What is the two-dimensional density function $f(x, y)$?

6.10. 5.4 Normal distributions in two-dimensions

6.11. EXCEL

The following files show two-dimensional normal random variables:

\textbf{Demonstration file: Height and weight}\hspace{1em}\texttt{def-200-65-00}

\textbf{Demonstration file: Height and weight, ellipse, eigen-vectors (projections and conditional distributions are also studied)}\hspace{1em}\texttt{def-200-82-00}

\textbf{Demonstration file: Two-dim normal distributions normal distributions, projections and conditional distributions}\hspace{1em}\texttt{def-200-83-00}

\textbf{Demonstration file: Measuring voltage}\hspace{1em}\texttt{def-200-66-00}

6.12. PROBLEMS

1. Expected values, standard deviations, probabilities Let us consider the two-dimensional random variable $(X, Y)$, which follows normal distribution with parameters $\mu_1 = 26$, $\sigma_1 = 4$, $\mu_2 = 14$, $\sigma_2 = 2$, $r = 0.6$.

a. How much is the expected value and the standard deviation of $X$?
b. How much is the probability that $22 < X < 26$?

c. How much is the probability that $24 < X < 28$?

d. How much is the expected value and the standard deviation of $Y$?

e. How much is the probability that $12 < Y < 14$?

f. How much is the probability that $13 < Y < 15$?

2. Conditional expected values, standard deviations, probabilities Let us consider the two-dimensional random variable $(X, Y)$, which follows normal distribution with parameters $\mu_1 = 26, \sigma_1 = 4, \mu_2 = 14, \sigma_2 = 2, r = 0.6$. How much is the conditional expected value and the conditional standard deviation of $Y$ on condition that

a. $X = 25$?

b. $X = 26$?

c. $X = 27$?

d. $X = x$?

3. Conditional expected values, standard deviations, probabilities Let us consider the two-dimensional random variable $(X, Y)$, which follows normal distribution with parameters $\mu_1 = 26, \sigma_1 = 4, \mu_2 = 14, \sigma_2 = 2, r = 0.6$. How much is the probability that $12 < Y < 14$ on condition that

a. $X = 25$?

b. $X = 26$?

c. $X = 27$?

4. Conditional expected values, standard deviations, probabilities Let us consider the two-dimensional random variable $(X, Y)$, which follows normal distribution with parameters $\mu_1 = 26, \sigma_1 = 4, \mu_2 = 14, \sigma_2 = 2, r = 0.6$. How much is the probability that $13 < Y < 15$ on condition that

a. $X = 25$?

b. $X = 26$?

c. $X = 27$?

6.13. 5.5 Independence of random variables

6.14. EXCEL

The following file serves to understand the notion of dependence and independence:

*Demonstration file: Lengths dependent or independent* 

6.15. PROBLEMS

1. Independent binomially distributed random variables Assume that $X$ and $Y$ are independent of each other, and $X$ follows binomial distribution with parameters 7 and 0.5, and $Y$ follows binomial distribution with parameters 5 and 0.3. Determine the distribution of $(X,Y)$. 
2. Passenger car and bus Assume that when a 5 passenger car has an accident, then the number \( X \) of injured people has the following distribution: \( P(X = 5) = 0.1 \), \( P(X = 6) = 0.05 \), \( P(X = 7) = 0.05 \), \( P(X = 8) = 0.05 \), and when an 8 passenger bus has an accident, then the number \( Y \) of injured people has the following distribution: \( P(Y = 5) = 0.05 \), \( P(Y = 6) = 0.05 \), \( P(Y = 7) = 0.05 \), \( P(Y = 8) = 0.05 \). When a 5 passenger car hits an 8 passenger bus, then

a. what is the probability that nobody is injured?

b. what is the probability that exactly 1 person is injured?

c. what is the probability that exactly 2 persons are injured?

d. how much is the expected value of the total number of injured people?

e. what is the probability that exactly 2 persons are injured in the car on condition that the total number of injured people is 5?

3. Working with two-dimensional density functions For each of the cases below, the density function \( f(x, y) \) of a two-dimensional random point \((X, Y)\) is defined. In each case, determine the density function \( f_1(x) \) of \( X \) and the density function \( f_2(y) \) of \( Y \), and then decide whether \( X \) and \( Y \) are independent or not.

a. \( f(x, y) = 1 \) \((0 < x < 1, \ 0 < y < 1)\);

b. \( f(x, y) = 1/6 \) \((0 < x < 3, \ 0 < y < 2)\);

c. \( f(x, y) = 4xy \) \((0 < x < 1, \ 0 < y < 1)\);

d. \( f(x, y) = 6xy^2 \) \((0 < x < 1, \ 0 < y < 1)\);

e. \( f(x, y) = e^{-x-y} \) \((x > 0, \ y > 0)\);

f. \( f(x, y) = 10e^{-5x-2y} \) \((x > 0, \ y > 0)\);

g. \( f(x, y) = 2 \) \((x > 0, \ y > 0, \ 0 < x + y < 1)\);

h. \( f(x, y) = 2 \) \((0 < x < y < 1)\);

i. \( f(x, y) = 6(y - x) \) \((0 < x < y < 1)\);

j. \( f(x, y) = 3x \) \((0 < x < y < 1)\);

k. \( f(x, y) = 3(1 - y) \) \((0 < x < y < 1)\).

6.16. 5.6 Generating a two-dimensional random variable

6.17. PROBLEMS

1. Generating first \( X \), then \( Y \) Consider the distribution on the plane with the density function

\[
f(x, y) = \frac{2y}{x^2} \quad \text{if} \quad 0 < y < x < 1.
\]
Simulate a two-dimensional random variable \((X, Y)\), which follows this distribution, as follows:

a. Determine \(f_1(x), F_1(x), F_1^{-1}(u)\),

b. then simulate \(X\) as \(F_1^{-1}(\text{RND}_1)\).

c. Determine \(f_{2|1}(y|x), F_{2|1}(y|x), F_{2|1}^{-1}(v|x)\),

d. then simulate \(Y\) as \(F_{2|1}^{-1}(\text{RND}_2|X)\).

e. Making 1000 experiments for \((X, Y)\), construct a point-cloud of the 1000 experimental results, and check that the density suggested by the point-cloud in the triangle defined by the inequalities \(0 < y < x < 1\) is in accordance with the density \(f(x, y)\).

2. Generating first \(Y\), then \(X\) Consider the distribution on the plane with the density function

\[
f(x, y) = \frac{2y}{x^2} , \text{ if } 0 < y < x < 1.
\]

Simulate a two-dimensional random variable \((X, Y)\), which follows this distribution, as follows:

a. Determine \(f_2(y), F_2(y), F_2^{-1}(v)\);

b. then simulate \(Y\) as \(F_2^{-1}(\text{RND}_2)\).

c. Determine \(f_{1|2}(x|y), F_{1|2}(x|y), F_{1|2}^{-1}(u|y)\);

d. then simulate \(X\) as \(F_{1|2}^{-1}(\text{RND}_1|X)\).

e. Making 1000 experiments for \((X, Y)\), construct a point-cloud of the 1000 experimental results, and check that the density in the region defined by the inequalities \(0 < x < y\) is in accordance with the density \(f(x, y)\).

3. Generating first \(X\), then \(Y\) Consider the distribution on the plane with the density function

\[
f(x, y) = \frac{4y}{x} , \text{ if } 0 < y < x < 1.
\]

Simulate a two-dimensional random variable \((X, Y)\), which follows this distribution, as follows:

a. Determine \(f_1(x), F_1(x), F_1^{-1}(u)\);

b. then simulate \(X\) as \(F_1^{-1}(\text{RND}_1)\).

c. Determine \(f_{2|1}(y|x), F_{2|1}(y|x), F_{2|1}^{-1}(v|x)\);

d. then simulate \(Y\) as \(F_{2|1}^{-1}(\text{RND}_2|X)\).

e. Making 1000 experiments for \((X, Y)\), construct a point-cloud of the 1000 experimental results, and check that the density suggested by the point-cloud in the triangle defined by the inequalities \(0 < y < x < 1\) is in accordance with the density \(f(x, y)\).

4. Generating first \(X\), then \(Y\) Consider the distribution on the plane with the density function
Simulate a two-dimensional random variable \((X, Y)\), which follows this distribution, as follows:

a. Determine \(f_1(x), F_1(x), F_1^{-1}(u)\);

b. then simulate \(X\) as \(F_1^{-1}(\text{RND}_1)\).

c. Determine \(f_{2|1}(y|x), F_{2|1}(y|x), F_{2|1}^{-1}(v|x)\);

d. then simulate \(Y\) as \(F_{2|1}^{-1}(\text{RND}_2|X)\).

e. Making 1000 experiments for \((X, Y)\), construct a point-cloud of the 1000 experimental results, and check that the density in the region defined by the inequalities \(0 < x < y\) is in accordance with the density \(f(x, y)\).

5. Water level in a water reservoir Let us assume that the water level of a water reservoir is measured on such a scale that the minimum level corresponds to 0, the maximum level corresponds to 1. Let us denote the water level on a day by \(X\), and the water level two days later by \(Y\). Let us assume that the two-dimensional random variable \((X, Y)\) has the following density function:

\[
f(x, y) = \frac{4}{5}(1 + xy) \quad \text{if} \quad 0 < x < 1, \quad 0 < y < 1
\]

Simulate \((X, Y)\). Solution: The density function of \(X\) is:

\[
f_1(x) = \int_{-\infty}^{\infty} f(x, y) \, dy = \int_{0}^{1} \frac{4}{5}(1 + xy) \, dy = \frac{4}{5} \left(1 + \frac{x}{2}\right) = \frac{4}{5} + \frac{2}{5}x \quad \text{if} \quad 0 < x < 1
\]

The distribution function of \(X\) is:

\[
F_1(x) = \int_{-\infty}^{x} f_1(x) \, dx = \int_{0}^{x} \frac{4}{5} \left(1 + \frac{x}{2}\right) \, dx
\]

\[
= \frac{4}{5} \left(x + \frac{x^2}{4}\right) = \frac{4}{5}x + \frac{1}{5}x^2 \quad \text{if} \quad 0 < x < 1
\]

The inverse of the distribution function is the solution for \(x\) of the equation \(F_1(x) = u\), which is now the following quadratic equation:

\[
\frac{4}{5}x + \frac{1}{5}x^2 = u
\]

The solution is

\[
x = -2 + \sqrt{4 + 5u}
\]

thus \(X\) can be simulated as

\[
X = -2 + \sqrt{4 + 5 \text{RND}_1}
\]

The conditional density function of \(Y\) on condition that \(X = x\) is:
The conditional distribution function of \( Y \) on condition that \( X = x \) is:

\[
F_{2\mid1}(y|x) = \int_{-\infty}^{y} f_{2\mid1}(y|x) \, dy = \int_{0}^{y} \frac{1 + xy}{1 + \frac{x}{2}} \, dy = \frac{y + x\frac{y^2}{2}}{1 + \frac{x}{2}}
\]

The inverse of the distribution function of the conditional distribution is the solution for \( y \) of the equation \( F_{2\mid1}(y|x) = \nu \), which is now the following quadratic equation:

\[
\left( \frac{2}{2 + x} \right) y + \left( \frac{x}{2 + x} \right) y^2 = \nu
\]

or equivalently

\[
x y^2 + 2y - (2 + x)\nu = 0
\]

The solution is

\[
y = \frac{-2 + \sqrt{4 + 4x(2 + x)\nu}}{2x}
\]

thus \( Y \) can be simulated as

\[
Y = \frac{-2 + \sqrt{4 + 4X(2 + X)\text{RND}_2}}{2X}
\]

6.18. 5.7 Properties of the expected value, variance and standard deviation

6.19. EXCEL

The following file shows the "average"-property of the standard deviation.

6.20. PROBLEMS

1. Square root law for the standard deviation of the sum Let us assume that \( X \) and \( Y \) are independent random variables, and \( X \) follows normal distribution with parameters \( \mu_1 = 10 \) and \( \sigma_1 = 3 \), and \( Y \) follows normal distribution with parameters \( \mu_2 = 15 \) and \( \sigma_2 = 4 \). Calculate the probabilities:

a. \( P(X + Y < 20) \);

b. \( P(X + 2Y < 50) \);
2. Married couples Assume that the weight of a randomly chosen Hungarian woman is normally distributed with an expected value of 60 kg and standard deviation 4 kg, and the weight of a randomly chosen Hungarian man is normally distributed with an expected value of 75 kg and standard deviation 7 kg. Let us also assume that the weights of the wife and of the husband are independent of each other.

a. How much is the expected value and the standard deviation of
   i. the total weight of a randomly chosen married couple?
   ii. the total weight of 10 randomly chosen married couples?
   iii. the average weight of 10 randomly chosen men?
   iv. the average weight of 100 randomly chosen men?
   v. the average weight of 10 randomly chosen women?
   vi. the average weight of 100 randomly chosen women?

b. What is the probability that
   i. the total weight of a randomly chosen married couple is between 130 and 140 kg?
   ii. the total weight of 10 randomly chosen married couples is between 1300 and 1400 kg?
   iii. the average weight of 10 randomly chosen men is between 74 and 76 kg?
   iv. the average weight of 100 randomly chosen men is between 74 and 76 kg?
   v. the average weight of 10 randomly chosen women is between 59 and 61 kg?
   vi. the average weight of 100 randomly chosen women is between 59 and 61 kg?

c. How much must \( n \) be in order that
   i. the average weight of \( n \) randomly chosen men is between 74 and 76 kg with a probability 0.9?
   ii. the average weight of \( n \) randomly chosen men is between 74 and 76 kg with a probability 0.99?
   iii. the average weight of \( n \) randomly chosen women is between 59 and 61 kg with a probability 0.9?
   iv. the average weight of \( n \) randomly chosen women is between 59 and 61 kg with a probability 0.99?

3. Expected value, variance and standard deviation of sums of random numbers Determine the expected value, the variance and the standard deviation of the following random variables:

a. \( RND \);

b. \( RND_1 + RND_2 \);

c. \( RND_1 + RND_2 + RND_3 + RND_4 + RND_5 + RND_6 + RND_7 + RND_8 + RND_9 + RND_{10} + RND_{11} + RND_{12} \)

4. Continuation of the previous problem How many measurements are needed to guarantee that the average of the experimental results is in the interval \([6.4; 6.6]\) with a probability

a. 0.99;

b. 0.999?
5. Non-independent random variables Try to define (by simulation) non-independent random variables so that the variance of the sum is not necessarily equal to the sum of the variance of the terms.

6. Measuring voltages Let us assume that the voltage $X$ (measured in Volt-s) in a certain electrical circuit follows normal distribution with parameters $\mu = 2210$ and $\sigma = 2.5$. Using Excel, sketch the graph of the density function and of the distribution function of $X$.

   a. Using Excel, make 1000 experiments for $X$, and visualize them on a nice figure.
   
   b. $P(119 < X < 221) = ?$
   
   c. $P(118 < X < 222) = ?$

   Assume that 5 independent measurements are performed for the voltage. Let $\overline{X}_5$ mean their average. Accept the fact that $\overline{X}_5$ follows a normal distribution, too.

   a. Using Excel, sketch the graph of the density function and of the distribution function of $\overline{X}_5$.
   
   b. Using Excel, make 1000 experiments for $\overline{X}_5$, and visualize them on a nice figure.
   
   c. $P(119 < \overline{X}_5 < 221) = ?$
   
   d. $P(118 < \overline{X}_5 < 222) = ?$

   Assume that 50 independent measurements are performed for the voltage. Let $\overline{X}_{50}$ mean their average.

   a. How much is the expected value of $\overline{X}_{50}$?
   
   b. How much is the standard variation of $\overline{X}_{50}$?

   Accept the fact that $\overline{X}_{50}$ follows a normal distribution, too.

   a. Using Excel, sketch the graph of the density function and of the distribution function of $\overline{X}_{50}$.
   
   b. Using Excel, make 1000 experiments for $\overline{X}_{50}$, and visualize them on a nice figure.
   
   c. $P(119 < \overline{X}_{50} < 221) = ?$
   
   d. $P(118 < \overline{X}_{50} < 222) = ?$

7. CO$_2$ pollution The CO$_2$ pollution is measured on the main square of a town through 25 days. The 25 measurement results are considered independent of each other. At each measurement the probability that the CO$_2$ pollution is above the acceptance level is 1/5.

   a. Using normal distribution find the approximate value of the probability that during 25 days at least 6 times the CO$_2$ pollution is above the acceptance level.

   b. Assume that the amount of CO$_2$ pollution has a uniform distribution on the interval $[0, 2]$. Using normal distribution find the approximate value of the probability that during 25 days the average of the measurement results is greater than $az \ 1 + \frac{\sqrt{a}}{10} \approx 1.17$.

### 6.21. 5.8 Transformation from plane to line

### 6.22. EXCEL

The following files show transformations from plane to line.
The following files show projections from plane to axes.

1. Exponential random variables with different parameters $X$ and $Y$ are independent, exponential random variables with parameters $\lambda_1$ and $\lambda_2$ (life-times of electrical bulbs of different types). Let $Z = X + Y$. Calculate and visualize the distribution function and the density function of $Z$.

2. Exponential random variables with equal parameters $X$ and $Y$ are independent, exponential random variables with the same parameter $\lambda$ (life-times of electrical bulbs of the same type). Let $Z = X + Y$. Calculate and visualize by graphs the distribution function and the density function of $Z$.

3. Transformation from plane to line The density function of $(X, Y)$ is $f(x, y) = 15xy^2$ $(0 < y < x < 1)$. Let $Z = X + Y$. Calculate and visualize by graphs the distribution function and the density function of $Z$. Calculate the expected value, the variance and standard deviation of $Z$ without using the density function of $Z$. Calculate the expected value, the variance and standard deviation of $Z$ using the density function of $Z$.

4. Transforming a uniform distribution from a rectangle to a line Let $X$ be uniformly distributed between 0 and 2, $Y$ uniformly distributed between 0 and 3, $X$ and $Y$ independent and $Z = X + Y$.
   a. Calculate the probability that $Z > 2$.
   b. Find the expected value of $Z$.
   c. Find the standard deviation of $Z$.

5. Transformation from plane to line The density function of $(X, Y)$ is $f(x, y) = 15xy^2$ $(0 < y < x < 1)$. Let $Z = X + Y$. Calculate and visualize by graphs the distribution function and the density function of $Z$. Calculate the expected value, the variance and standard deviation of $Z$ without using the density function of $Z$. Calculate the expected value, the variance and standard deviation of $Z$ using the density function of $Z$.

6. Transformation from plane to plane

6.4. 5.9 *** Transformation from plane to plane

6.5. EXCEL

The following files give some transformation from plane to plane.
6.26. PROBLEMS

1. Transformation from square to triangle Consider the uniform distribution on the unit square,

   \[ \{(x, y) : 0 < x < 1, \ 0 < y < 1\} \]

   which has the constant density function

   \[ f(x, y) = 1 \]

   Transform it by the transformation \( u = x y, \ v = y \). Find the new density function.

2. Transformation from square to triangle Consider the continuous distribution on the unit square

   \[ \{(x, y) : 0 < x < 1, \ 0 < y < 1\} \]

   with the density function

   \[ f(x, y) = 4xy \]

   Transform it by the transformation \( u = x y, \ v = y \). Find the new density function.

6.27. 5.10 *** Sums of random variables. Convolution

6.28. EXCEL

The following files show how the distribution of the sum can be calculated:

\[ \text{Demonstration file: Summation of independent random variables, fair dice} \]

\[ \text{Demonstration file: Summation of independent random variables, unfair dice} \]

6.29. PROBLEMS

1. Getting a "triangle-shaped" distribution Consider the uniform distribution on the sets \( 0, \ldots, 5 \). Convolve it by itself. You should get a "triangle-shaped" distribution on the set \( 0, \ldots, 10 \). Making the numerical calculations, you may use Excel.

2. Getting a "triangle-shaped" distribution Consider the uniform distribution on the sets \( 0, \ldots, n \). Convolve it by itself. You should get a "triangle-shaped" distribution on the set \( 0, \ldots, 2n \).

3. Getting a "trapezoid-shaped" distribution Consider the uniform distributions on the sets \( 0, \ldots, n_1 \) and \( 0, \ldots, n_2 \), where \( n_1 \neq n_2 \). Convolve them. You should get a "trapezoid-shaped" distribution on the set \( 0, \ldots, n_1 + n_2 \). Making the numerical calculations, you may use Excel.

4. Getting a "trapezoid-shaped" distribution Consider the uniform distributions on the sets \( 0, \ldots, 5 \) and \( 0, \ldots, 8 \). Convolve them. You should get a "trapezoid-shaped" distribution on the set \( 0, \ldots, 13 \).
5. Convolving binomial distributions Convol\(e\) the binomial distribution with parameters 2 and 0.5 by the binomial distribution with parameters 3 and \(p\). You should get the binomial distribution with parameters 5 and \(p\). Making the numerical calculations, you may use Excel.

6. Convolving binomial distributions Convol\(e\) the binomial distribution with parameters 2 and \(p\) by the binomial distribution with parameters 3 and \(p\). You should get the binomial distribution with parameters 5 and \(p\).

7. Convolving Poisson-distributions Convol\(e\) the Poisson-distribution with parameter 2 by the Poisson-distribution with parameter 3. You should get the Poisson-distribution with parameter 3. Making the numerical calculations, you may use Excel.

8. Convolving Poisson-distributions Convol\(e\) the Poisson-distribution with parameter \(\lambda_1\) by the Poisson-distribution with parameter \(\lambda_2\). You should get the Poisson-distribution with parameter \(\lambda_1 + \lambda_2\).

6.30. 5.11 Limit theorems to normal distributions

6.31. EXCEL

Here is a file to study the binomial approximation of normal distribution:

\emph{Demonstration file: Binomial approximation of normal distribution \(\text{ef-300-14-00}\)}

Here is a file to study how convolutions approximate normal distributions:

\emph{Demonstration file: Convolution with uniform distribution \(\text{ef-300-15-00}\)}

\emph{Demonstration file: Convolution with asymmetrical distribution \(\text{ef-300-16-00}\)}

\emph{Demonstration file: Convolution with U-shaped distribution \(\text{ef-300-17-00}\)}

\emph{Demonstration file: Convolution with randomly chosen distribution \(\text{ef-300-18-00}\)}

This is a file to study how gamma distributions approximate normal distributions:

\emph{Demonstration file: Gamma distribution approximates normal distribution \(\text{ef-300-19-00}\)}

Here are files to study the two-dimensional central limit theorem:

\emph{Demonstration file: Two-dimensional central-limit theorem, rectangle \(\text{ef-300-20-00}\)}

\emph{Demonstration file: Two-dimensional central-limit theorem, parallelogram \(\text{ef-300-21-00}\)}

\emph{Demonstration file: Two-dimensional central-limit theorem, curve \(\text{ef-300-22-00}\)}

6.32. PROBLEMS

1. Draw with replacement 25 times from the box containing the tickets with the numbers 3, 5, 7, 9 so that each ticket has the same chance.
   a. What is the expected value of the sum of the 25 draws?
   b. What is the standard deviation of the sum of the 25 draws?
   c. What is the probability that the sum of the 25 draws is more than 140? (Use normal approximation.)
2. 400 independent random numbers which are uniformly distributed between 0 and 1 are generated. What is the (approximate) probability that their average is between 0.45 and 0.55? Give your answer using normal approximation.

3. Toss a die 600 times. What is the probability that the number of sixes is
   a. greater than or equal to 120?
   b. less than or equal to 80?
   c. strictly between 80 and 120?

4. Assume that in a country 2/3 of the people are for party "A", and 1/3 of them are for party "B". Choosing 500 people at random, what is the probability that the relative frequency of the people being for party "A" among the chosen ones is between $\frac{2}{3} - 0.1$ and $\frac{2}{3} + 0.1$?

5. The diameter of a first class tomato sold in a certain shop is uniformly distributed between 6 and 9 cm, the diameter of a second class tomato is uniformly distributed between 4 and 7 cm. (The diameters of the tomatoes are independent.) If we put 25 first class tomatoes and 25 second class tomatoes in a line, then how much is the expected value and the standard deviation of the total length of the 50 tomatoes being in the line?

6. From the respondents who participated in an opinion survey, 50 percents declared to favor a unicameral parliament, 40 percents declared to favor a bicameral parliament, and 10 percents did not answer. 400 respondents from the interviewed sample are randomly chosen (For simplicity, you may assume: with replacement).
   a. What is the probability that exactly 200 of them were in favor of a unicameral parliament?
   b. What is the approximate probability that the number of respondents who favor a unicameral parliament is at least 190 and at most 210?

7. 6 Regression in one-dimension

7.1. PROBLEMS

1. Minimizing the the expected value of $|X - c|$ A certain type of electric bulbs has exponential life-time $X$ with parameter 2.5.
   a. What is the number $c$ for which the expected value of $|X - c|$ is minimal?
   b. Check your result with simulation. Take the optimal $c$ value, call it $c_{opt}$, and another $c$ value, call it $c_{other}$. Simulate 1000 experiments, $X_1, \ldots, X_{1000}$, and approximate the expected values of $|X - c_{opt}|$ and $|X - c_{other}|$ by the averages

   $$\frac{|X_1 - c_{opt}| + \ldots + |X_{1000} - c_{opt}|}{1000}$$

   and

   $$\frac{|X_1 - c_{other}| + \ldots + |X_{1000} - c_{other}|}{1000}$$

   respectively. You may be convinced that the first average, in most cases, is smaller than the second one.

2. Minimizing the the expected value of $(X - c)^2$ A certain type of electric bulbs has exponential life-time $X$ with parameter 2.5.
a. What is the number \( c \) for which the expected value of \( (X - c)^2 \) is minimal?

b. Check your result with simulation. Take the optimal \( c \) value, call it \( c_{\text{opt}} \), and another \( c \) value, call it \( c_{\text{other}} \). Simulate 1000 experiments, \( X_1, \ldots, X_{1000} \), and approximate the expected values of \( (X - c_{\text{opt}})^2 \) and \( (X - c_{\text{other}})^2 \) by the averages

\[
\frac{(X_1 - c_{\text{opt}})^2 + \ldots + (X_{1000} - c_{\text{opt}})^2}{1000}
\]

and

\[
\frac{(X_1 - c_{\text{other}})^2 + \ldots + (X_{1000} - c_{\text{other}})^2}{1000}
\]

respectively. You may be convinced that the first average, in most cases, is smaller than the second one.

7.2. 6.1 Regression in two-dimensions

7.3. EXCEL

Here are files to study regression problems:

- Demonstration file: Regression
- Conditional distributions, expected value, median, \( (RND_1 RND_2;RND_1) \)
- Conditional distributions, expected value, median, \( (RND_1 RND_2;RND_1/RND_2) \)

7.4. PROBLEMS

1. Water level in a reservoir - regression curves

Let us assume that the water level of a water reservoir is measured in such a scale that the minimum level corresponds to 0, the maximum level corresponds to 1. Let us denote the water level on a day by \( X \), and the water level two days later by \( Y \). Let us assume that the two-dimensional random variable \( (X, Y) \) has the following density function:

\[
f(x, y) = \frac{4}{5} (1 + xy) \quad \text{if} \quad 0 < x < 1, \quad 0 < y < 1
\]

a. Find the conditional expected value of \( Y \) on condition that \( X = x \).

b. Find the conditional median of \( Y \) on condition that \( X = x \).

Solution:

a. The density function of \( X \) is:

\[
f_1(x) = \int_{-\infty}^{\infty} f(x, y) \, dy = \int_{0}^{1} \frac{4}{5} (1 + xy) \, dy = \frac{4}{5} \left( 1 + \frac{x}{2} \right) = \frac{4}{5} + \frac{2}{5} x \quad \text{if} \quad 0 < x < 1
\]

The conditional density function of \( Y \) on condition that \( X = x \) is:
The conditional expected value of $Y$ on condition that $X = x$ is:

$$E(Y|X = x) = \int_{-\infty}^{\infty} y \ f_{2|1}(y|x) \ dy = \int_{0}^{1} y \ \frac{1 + xy}{1 + \frac{x}{2}} \ dy = \frac{2 + \frac{x}{2}}{1 + \frac{x}{2}}$$

**\texttt{Solution Sol-05-02-01}**

b. The conditional distribution function of $Y$ on condition that $X = x$ is:

$$F_{2|1}(y|x) = \int_{-\infty}^{y} f_{2|1}(y|x) \ dy = \int_{0}^{y} \frac{1 + xy}{1 + \frac{x}{2}} \ dy = \frac{y + \frac{y^2}{2}}{1 + \frac{x}{2}}$$

$$= \left( \frac{2}{2 + x} \right) y + \left( \frac{x}{2 + x} \right) y^2 \quad \text{if} \quad 0 < x < 1, \ 0 < y < 1$$

The conditional median is the solution to the equation

$$F_{2|1}(y|x) = \frac{1}{2}$$

that is,

$$\left( \frac{2}{2 + x} \right) y + \left( \frac{x}{2 + x} \right) y^2 = \frac{1}{2}$$

The solution, that is, the conditional median of $Y$ on condition that $X = x$ is:

$$y = \frac{-2 + \sqrt{4 + 2x(2 + x)}}{2x}$$

**\texttt{Solution Sol-05-02-02}**

2. Checking the results of the previous problem by simulation

3. Take the regression curve defined by the conditional expected values of $Y$, and denote this function by $k_{\text{opt}}(x)$:

$$k_{\text{opt}}(x) = \frac{\frac{1}{2} + \frac{x}{2}}{1 + \frac{x}{2}}$$

Take an arbitrary other function $k_{\text{other}}(x)$, too. Simulate 1000 experiments, $X_1, \ldots, X_{1000}$, and approximate the expected values of $(Y - k_{\text{opt}}(X))^2$ and $(Y - k_{\text{other}}(X))^2$ by the averages

$$\frac{(Y_1 - k_{\text{opt}}(X_1))^2 + \ldots + (Y_{1000} - k_{\text{opt}}(X_{1000}))^2}{1000}$$

and

$$\frac{(Y_1 - k_{\text{other}}(X_1))^2 + \ldots + (Y_{1000} - k_{\text{other}}(X_{1000}))^2}{1000}$$

respectively You may be convinced that the first average, in most cases, is smaller than the second one.
4. Take the regression curve defined by the conditional medians of $Y$, and denote this function by $k_{opt}(x)$:

$$k_{opt}(x) = \frac{-2 + \sqrt{4 + 2x(2 + x)}}{2x}$$

Take an arbitrary other function $k_{other}(x)$, too. Simulate 1000 experiments, $X_1, \ldots, X_{1000}$, and approximate the expected values of $(Y - k_{opt}(X))^2$ and $|Y - k_{other}(X)|$ by the averages

$$\frac{|Y_1 - k_{opt}(X_1)| + \ldots + |Y_{1000} - k_{opt}(X_{1000})|}{1000}$$

and

$$\frac{|Y_1 - k_{other}(X_1)| + \ldots + |Y_{1000} - k_{other}(X_{1000})|}{1000}$$

respectively. You may be convinced that the first average, in most cases, is smaller than the second one.

### 7.5. 6.2 Linear regression

### 7.6. PROBLEMS

1. Water level in a reservoir - linear regression

Let us assume that the water level of a water reservoir is measured in such a scale that the minimum level corresponds to 0, the maximum level corresponds to 1. Let us denote the water level on a day by $X$, and the water level two days later by $Y$. Let us assume that the two-dimensional random variable $(X, Y)$ has the following density function:

$$f(x, y) = \frac{4}{5} (1 + xy) \quad \text{if} \quad 0 < x < 1, \quad 0 < y < 1$$

a. Calculate the expected value of $X$ and $Y$.

b. Calculate the second moments, the variances and the standard deviations of $X$ and $Y$.

c. Calculate the covariance between $X$ and $Y$.

d. Calculate the correlation coefficient between $X$ and $Y$.

e. Set up the equation of the regression line when $Y$ is estimated from $X$.

2. Checking the results of the previous problem by simulation

Check by simulation that the regression line gives the best linear estimation. Take the regression line, and denote the linear function by $k_{opt}(x)$. Take an arbitrary other line $k_{other}(x)$, too. Simulate 1000 experiments, $X_1, \ldots, X_{1000}$, and approximate the expected values of $(Y - k_{opt}(X))^2$ and $(Y - k_{other}(X))^2$ by the averages

$$\frac{(Y_1 - k_{opt}(X_1))^2 + \ldots + (Y_{1000} - k_{opt}(X_{1000}))^2}{1000}$$

and

$$\frac{(Y_1 - k_{other}(X_1))^2 + \ldots + (Y_{1000} - k_{other}(X_{1000}))^2}{1000}$$

respectively. You may be convinced that the first average, in most cases, is smaller than the second one.
7.7. 6.3 Confidence intervals

7.8. EXCEL

Here are files to study construction of confidence intervals.

\textit{Demonstration file: Finite confidence interval for the expected value, using standard normal distribution}\footnote{ef-300-23-00}

\textit{Demonstration file: Infinitely long confidence interval for the expected value, using standard normal distribution}\footnote{ef-300-23-50}

\textit{Demonstration file: Finite confidence interval for the expected value, using t-distribution}\footnote{ef-300-24-00}

7.9. PROBLEMS

1. Constructing a confidence interval Assume that the weight of a randomly chosen child in a certain school is normally distributed with an unknown expected value and a standard deviation 4 kg. What is the confidence interval for the expected value if we have the following experimental results: 15.4 kg, 26.4 kg, 24.1 kg, 20.5 kg, 15.8 kg, 26.4 kg, 25.0 kg, 21.6 kg, 86.4 kg, 27.7 kg, and we want that the confidence interval contains the expected value with a probability
   a. 0.95 ?
   b. 0.68

2. How long is the confidence interval Assume that the height of men in a certain country is normally distributed with an unknown expected value $\mu$ and a standard deviation 20 cm. Assume that you have measured 5 men, and their heights are: 173.2 cm; 176.4 cm; 185.3 cm; 177.7 cm; 173.0 cm. Using our method associated to the probability value
   a. $\bar{p} = 0.90$,
   b. $\bar{p} = 0.95$,
   what confidence interval do you declare for $\mu$? How long is the interval?

3. How long is the confidence interval Assume that you have measured 20 men, and their heights are: 173.2 cm; 176.9 cm; 185.3 cm; 177.7 cm; 173.0 cm; 183.5 cm; 166.4 cm; 169.2 cm; 187.7 cm; 192.7 cm; 163.4 cm; 186.3 cm; 185.9 cm; 177.7 cm; 173.2 cm; 170.0 cm; 175.4 cm; 175.3 cm; 197.1 cm; 174.3 cm. Using our method associated to the probability value
   a. $\bar{p} = 0.90$,
   b. $\bar{p} = 0.95$,
   what confidence interval do you declare for $\mu$? How long is the interval?

4. U-test Assume that the weight of a randomly chosen woman in a certain country is normally distributed with an unknown expected value and a standard deviation 4 kg. What is the confidence interval for the expected value, if we want that it contains the expected value with probability 0.95, and we have the following experimental results:
   a. 65.4 kg, 76.4 kg, 74.1 kg, 70.5 kg;
   b. 65.8 kg, 76.4 kg, 75.0 kg, 76.6 kg, 86.4 kg, 77.7 kg;
   c. 65.4 kg, 76.4 kg, 74.1 kg, 70.5 kg, 65.8 kg, 76.4 kg, 75.0 kg, 76.6 kg, 86.4 kg, 77.7 kg?
7.10. 6.4 U-tests

7.11. EXCEL

In the following simulation files, the average of the experimental results is directly simulated. The experimental results themselves are not simulated.

*Demonstration file: U-test 1: Case of "less than", when $n$ is given
  $\bar{x} = 300 - 25 - 20$

*Demonstration file: U-test 1: Case of "less than", when $n$ is given, version B
  $\bar{x} = 300 - 25 - 30$

*Demonstration file: U-test 1: Case of "less than", when $n$ is given, version C
  $\bar{x} = 300 - 25 - 40$

*Demonstration file: U-test 2: Case of "less than", when $n$ is calculated
  $\bar{x} = 300 - 26 - 00$

*Demonstration file: U-test 3: Case of "equality", when $n$ is given
  $\bar{x} = 300 - 27 - 20$

*Demonstration file: U-test 3: Case of "equality", when $n$ is given, version B
  $\bar{x} = 300 - 27 - 30$

*Demonstration file: U-test 3: Case of "equality", when $n$ is given, version C
  $\bar{x} = 300 - 27 - 40$

*Demonstration file: U-test 4: Case of "equality", when $n$ is calculated
  $\bar{x} = 300 - 28 - 00$

*Demonstration file: U-test 5: Case of "equality", when an interval is considered instead of the point $\mu_0$
  $\bar{x} = 300 - 29 - 00$

Here are files to study the U-test for "two populations".

*Demonstration file: U-test, two populations (Version B)
  $\bar{x} = 300 - 29 - 40$

*Demonstration file: U-test, two populations (Version C)
  $\bar{x} = 300 - 29 - 50$

7.12. PROBLEMS

Use the above files to answer the following problems.

1. U-test Assume that the weight of a randomly chosen man in a certain country is normally distributed with an unknown expected value $\mu$ and a standard deviation 10 kg. You are testing the hypothesis that $\mu = 82$ with a method for which it is true that when the hypothesis holds, then the probability of accepting the hypothesis is $p = 0.95$. Find the critical values (symmetrical about 82), if you use a sample of size 5. Find the critical values (symmetrical about 82), if you use a sample of size 10. Would you accept the hypothesis $\mu = 82$, if you have the following experimental results:
   a. 85.4 kg, 76.4 kg, 84.1 kg, 80.5 kg, 87.7 kg;
   b. 85.4 kg, 76.4 kg, 84.1 kg, 80.5 kg, 85.8 kg, 76.4 kg, 75.0 kg, 76.6 kg, 86.4 kg, 87.7 kg?

2. U-test Assume that the weight of a randomly chosen man in a certain country is normally distributed with an unknown expected value $\mu$ and a standard deviation 7 kg. If you use the significance level $p = 0.95$, then would you accept the hypothesis $\mu = 82$, if we have the following experimental results:
   a. 85.4 kg, 76.4 kg, 84.1 kg, 80.5 kg;
   b. 85.8 kg, 76.4 kg, 75.0 kg, 76.6 kg, 86.4 kg, 87.7 kg;
   c. 85.4 kg, 76.4 kg, 84.1 kg, 80.5 kg, 85.8 kg, 76.4 kg, 75.0 kg, 76.6 kg, 86.4 kg, 87.7 kg?
7.13. 6.5 *** T-tests

7.14. EXCEL

Here are files to study the T-test for "equality".

- Demonstration file: T-test, equality, sample average, 1 experiment (version A) \ef-300-31-00
- Demonstration file: T-test, equality, sample average, 1000 experiments (version A) \ef-300-32-00
- Demonstration file: T-test, equality, sample average, 1 experiment (version B, no figures) \ef-300-33-00
- Demonstration file: T-test, equality, sample average, 1 experiment (version B with figures) \ef-300-34-00

Here are files to study the T-test for "less-than".

- Demonstration file: T-test, less-than, sample average, 1 experiment (version A) \ef-300-37-00
- Demonstration file: T-test, less-than, sample average, 1000 experiments (version A) \ef-300-38-00
- Demonstration file: T-test, less-than, sample average, 1 experiment (version B, no figures) \ef-300-39-00
- Demonstration file: T-test, less-than, sample average, 1 experiment (version B with figures) \ef-300-40-00

7.15. PROBLEMS

Use the above files to answer the following problems.

1. t-test Assume that the weight of a randomly chosen man in a certain country is normally distributed with an unknown expected value $\mu$ and a standard deviation. You are testing the hypothesis that $\mu = \bar{X}$ with a method for which it is true that when the hypothesis holds, then the probability of accepting the hypothesis is $\alpha = 0.95$. Find the critical values (symmetrical about 82), if you use a sample of size 5. Find the critical values (symmetrical about 82), if you use a sample of size 10. Would you accept the hypothesis $\mu = \bar{X}$, if you have the following experimental results:
   a. 85.4 kg, 76.4 kg, 84.1 kg, 80.5 kg, 87.7 kg;
   b. 85.4 kg, 76.4 kg, 84.1 kg, 80.5 kg, 85.8 kg, 76.4 kg, 75.0 kg, 76.6 kg, 86.4 kg, 87.7 kg?

2. U-test Assume that the weight of a randomly chosen man in a certain country is normally distributed with an unknown expected value $\mu$ and standard deviation. If you use the significance level $\alpha = 0.95$, then would you accept the hypothesis $\mu = \bar{X}$, if we have the following experimental results:
   a. 85.4 kg, 76.4 kg, 84.1 kg, 80.5 kg;
   b. 85.8 kg, 76.4 kg, 75.0 kg, 76.6 kg, 86.4 kg, 87.7 kg;
   c. 85.4 kg, 76.4 kg, 84.1 kg, 80.5 kg, 85.8 kg, 76.4 kg, 75.0 kg, 76.6 kg, 86.4 kg, 87.7 kg?

7.16. 6.6 *** Chi-square-test for fitness

7.17. EXCEL
Here are files to study chi-square-test for fitness.

\textit{Demonstration file: Chi-square-test for fitness (version B, no figures)} \cite{ef-300-50-00}

\textit{Demonstration file: Chi-square-test for fitness (version C, with figures)} \cite{ef-300-51-00}

Chi-square-test for normality. Details will be written here later.

\textit{Demonstration file: Chi-square-test for normality} \cite{ef-300-50-10}

\textit{Demonstration file: Chi-square-test for independence} \cite{ef-300-50-50}

\section*{7.18. PROBLEMS}

Use the above files to answer the following problems.

1. Is the die fair? Somebody made 80 tosses with a die. Here are the results:

\begin{verbatim}
2 1 6 4 1 3 3 5 2 5
3 4 6 1 4 4 5 2 3 5
5 6 1 6 1 6 4 4 1 1
2 1 5 6 1 1 5 3 2 3
6 2 2 4 4 1 5 5 2 3
5 3 6 5 4 6 3 6 6 4
1 6 5 1 5 4 3 3 2 1
6 3 6 6 2 4 1 4 5 6
\end{verbatim}

Would you accept that the die is fair?

2. Is the die fair? Somebody made 100 tosses with a die. Here are the results:

\begin{verbatim}
4 2 6 4 4 3 4 5 5 4
6 1 4 4 4 6 4 4 4 6
5 1 3 6 1 4 1 5 5 5
6 5 6 5 6 2 3 2 4 4
2 5 4 3 3 5 5 6 5 6
6 6 1 4 5 1 4 1 6 6
2 2 5 2 6 6 5 4 4 1
4 4 4 1 6 4 4 1 6 5
6 5 3 4 5 6 6 4 2 6
6 6 3 5 6 4 6 5 5 6
\end{verbatim}

Would you accept that the die is fair?

3. Are the coins fair? Somebody states that he tossed 2 fair coins 50 times, and checked how many heads he got with the 2 coins. Here are the results:

\begin{verbatim}
0 0 1 0 1 1 0 2 0 1
0 2 1 1 1 0 2 1 0 1
2 0 1 0 1 1 0 2 2 2
2 1 1 2 1 1 0 1 1 2
0 2 1 1 2 0 1 2 2 0
\end{verbatim}

Would you accept that the coins are fair?
4. Are the coins fair? Somebody states that he tossed 2 fair coins 60 times, and checked how many heads he got with the 2 coins. Here are the results:

```
2 2 0 1 1 1 1 1 1 0
1 2 1 2 1 2 1 1 1 0
1 1 1 1 1 0 2 2 1 0
0 1 1 0 2 1 1 2 1 1
2 2 0 2 1 1 0 1 1
1 2 1 0 2 0 0 1 2 1
```

Would you accept that the coins are fair?

7.19. 6.7 *** Chi-test for standard deviation (Chi-square-test for variance)

7.20. EXCEL

Here are files to study Chi-test for standard deviation (Chi-square-test for variance).

\textit{Demonstration file: Chi-test for standard deviation (Chi-square-test for variance), 1 experiment (version A)} \ref{300-52-00}

\textit{Demonstration file: Chi-test for standard deviation (Chi-square-test for variance), 1000 experiments (version A)} \ref{300-53-00}

\textit{Demonstration file: Chi-test for standard deviation (Chi-square-test for variance) (version B, no figures)} \ref{300-54-00}

\textit{Demonstration file: Chi-test for standard deviation (Chi-square-test for variance) (version C, with figures)} \ref{300-55-00}

7.21. PROBLEMS

Use the above files to answer the following problems.

1. Is the standard deviation 10? Here are 10 experimental results for a normally distributed random variable: 85.30, 90.81, 106.55, 99.45, 93.75, 96.84, 118.09, 99.29, 92.71, 90.73. Would you accept that the standard deviation of the random variable is 10?

2. Is the standard deviation 25? Here are 12 experimental results for a normally distributed random variable: 123.67, 83.08, 118.91, 119.85, 88.77, 58.75, 124.30, 117.37, 87.14, 148.49, 47.97, 57.26. Would you accept that the standard deviation of the random variable is 20?

7.22. 6.8 *** F-test for equality of variances (of standard deviations)

7.23. EXCEL

Here are files to study F-tests for variances (standard deviation).

\textit{Demonstration file: F-test for equality of variances (of standard deviations) (version B, no figures)} \ref{300-56-00}
Use the above file to answer the following problems.

1. Are the standard deviations equal? Here are 11 experimental results for a normally distributed random variable \( \chi \): 123.67, 83.08, 118.91, 119.85, 88.77, 85.75, 124.30, 117.37, 87.14, 148.49, 47.97. Here are 9 experimental results for a normally distributed random variable \( \gamma \): 136.66, 112.20, 44.14, 58.42, 143.21, 155.20, 121.47, 50.81, 114.76. Would you accept that the standard deviation of \( \chi \) is equal to the standard deviation of \( \gamma \)?

2. Are the standard deviations equal? Here are 13 experimental results for a normally distributed random variable \( \chi \): 57.55, 56.33, 44.28, 63.74, 52.30, 58.25, 51.55, 41.41, 45.14, 37.53, 54.12, 54.73, 56.41. Here are 14 experimental results for a normally distributed random variable \( \gamma \): 15.33, 19.64, 21.70, 24.96, 20.67, 21.58, 30.04, 17.45, 14.35, 14.21, 17.43, 23.26, 18.94, 20.77. Would you accept that the standard deviation of \( \chi \) is equal to the standard deviation of \( \gamma \)?

7.25. 6.9 *** Test with ANOVA (Analysis of variance)

7.26. EXCEL

This is a file to study ANOVA (Analysis of variance).

Use the above file to answer the following problems.

1. Are all the expected values equal? Four normally distributed random variables are considered. We have 12 measurement results for the first: 14.74, 10.6, 13.38, 20.63, 9.34, 9.44, 16.52, 9.04, 19.48, 16.21, 21.86, 12.69. Measurement results for the second: 16.04, 9.18, 17.81, 8.94, 16.37, 14.77, 8.60, 13.76, 11.44. Measurement results for the third: 21.76, 17.88, 13.38, 8.06, 9.98, 15.85, 10.03, 13.55, 12. Measurements results for the fourth: 13.12, 12.95, 13.60, 10.79, 0.26, 10.72, 7.17, 4.51, 4.98, 5.83, 11.82, 15.31. Would you accept that all the expected values equal?

2. Are all the expected values equal? Four normally distributed random variables are considered. We have 12 measurement results for the first: 18.98, 18.73, 15.31, 6.70, 18.73, 17.00, 11.21, 18.48, 14.99, 19.09, 16.00, 14.96. Measurement results for the second: 8.05, 8.35, 5.61, 8.28, 10.22, 6.62, 10.93, 1.13, 13.65. Measurement results for the third: 27.17, 21.73, 13.04, 14.66, 17.13, 19.94, 15.84, 15.85, 12 measurement results for the fourth: 12.55, 3.34, 6.48, 10.6, 6.13, 14.08, 10.57, 10.53, 0.54, 5.12, 9.17, 22.58. Would you accept that all the expected values equal?