Eötvös Loránd University

Mathematics I.

Budapest

Fall 2020

Contents

1 Functions

The first two chapters basically consist of a review of the secondary school curriculum. We aim to establish our main topic in the spring semester: the Newton-Leibniz's differential calculus (differentiation, integration), but the knowledge of functions will appear in several topics already during the autumn semester. In addition to repetition, new concepts will also appear, which we will need to describe and characterize various social and economic processes and to solve problems.

1.1 Motivating example and introduction

Functions surround our daily lives. Functions describe the relationship between two (or more) variables.

Examples:

- We often study time-varying processes. For example, the distance of a moving body from the starting point, the change of a loan or a bank deposit, the decay of a radioactive material. In these cases, one variable is time, and the function describes the change in time for the other variable.
- In economic life, we often examine the change in demand for a given product as a function of the price of the product. Or: a functional relationship can be described between the quantity to be produced from a product and the cost of making a product.

Examples of problems that can be solved using functions:

- There is a uniform tariff system for taxi companies in Budapest: the basic fee is 700 HUF, in addition 300 HUF per kilometer has to be paid. (We can also say that the amount you pay depends on the length of the trip you take.) How much HUF do we pay for a 4.5-kilometer trip? For how many kilometers did we take a taxi if we pay 2,860 HUF at the end of the trip?
- Increasing the price of a product can reduce the demand for that product. How do we determine the price of a product to get the most income out of it?

1.2 Theory

1.2.1 Functions

A function is a rule that assigns to each object in a set D exactly one object in a set R. The set D is called the *domain* of the function, and the set of assigned objects in R is called the *range*. For functions in this class, the domain and range will be collections of real numbers and the function itself will be denoted by a letter such as f. The value that the function f assigns to the number x in the domain is then denoted by $f(x)$, which is often given by a formula, such as $f(x) = x^2 + x - 1$. The notation of $x \mapsto f(x)$ can also be used. We can think of a function as a "mapping" from numbers in D to numbers in R, or as a "machine" that takes a given number from D (input) and converts it into a number in R (output) through a process indicated by the functional rule.

No matter how you choose to think of a functional relationship, it is important to remember that a function assigns one and only one number in the range to each number in the domain. The inverse is not always true: given the following function: $g(x) = x^2$, we can see that $g(5) = g(-5)$.

The set of ordered pairs $\{(x, f(x))|x \in D\}$ are the graph of the function and represents a curve in the plane giving a pictorial representation of the function.

1.2.2 Transformations of functions

One simple way to create a new function from an old one is to add a parameter. We call these modifications transformations of the given rule.

$f(x) + b$ is the graph of $f(x)$ shifted up by b units ($b > 0$) or shifted down by b units ($b < 0$) (vertical shift)

$f(x + b)$ is the graph of $f(x)$ shifted left by b units ($b > 0$) or shifted right by b units ($b < 0$) (horizontal shift)

$af(x)$ is the graph of $f(x)$ vertically squeezed ($0 < a < 1$) or stretched ($a > 1$) (vertical stretch)

$f(ax)$ is the graph of $f(x)$ horizontally stretched ($0 < a < 1$) or squeezed ($a > 1$) (horizontal stretch)

$-f(x)$ is the graph of $f(x)$ reflected about the x-axis. This means that the signs on the all the y coordinates are changed to the opposite sign. (reflection about the x-axis)

$f(-x)$ is the graph of $f(x)$ reflected about the y-axis. This means that the signs on the all the x coordinates are changed to the opposite sign. (reflection about the y-axis).

1.2.3 Operations with functions

Functions with overlapping domains can be added, subtracted, multiplied and divided. The meaning of $f + g$, $f - g$, $f \cdot g$, f/g ($g \neq 0$) is evident, for example if $f(x) = x^2$, $g(x) = 2x + 5$, then $(f \cdot g)(x) = f(x) \cdot g(x) = x^2 \cdot (2x + 5) = 2x^3 + 5x^2$.

Given functions $f(u)$ and $g(x)$, the *composition* $(f \circ g)(x) = f(g(x))$ is the function of x formed by substituting $u = g(x)$ for u in the formula for $f(u)$. Example: $f(x) = x^2, g(x) = 2x + 5$, then $f(g(x)) = (2x + 5)^2$ and $g(f(x)) = 2x^2 + 5$. It is important, that the composition is not a commutative operation!

1.2.4 Rate of change

The *change* of f on the $[a; b]$ interval: $f(b) - f(a)$. The average rate of change on this interval:

$$\frac{f(b) - f(a)}{b - a}.$$

It is a measure of how much the function changed per unit on average over that interval. The geometric meaning of this quantity is the slope of the straight line drawn through the points $(a; f(a))$ and $(b; f(b))$ of the graph of the function.

Example: What is the change and the average rate of change of $f(x) = x^2 + 4x + 5$ between $x = 1$ and $x = 3$? $f(1) = 10$, $f(3) = 26$, thus, the function increased by 16 units in 2 units, so the average rate of change was 8 over this interval.

1.2.5 Linear and quadratic functions

A linear function has the form $f(x) = mx + b$. A function is linear if its slope, or rate of change, is the same at every point.

A quadratic function has the form $f(x) = ax^2 + bx + c$, where a, b, and c are numbers with a not equal to zero. The shape of the graph is a parabola, whose main characteristics are its zeros (there are 0, 1, or 2 of these, called horizontal intercepts), and the location, value, and nature (minimum or maximum) of its extreme value.

1.2.6 Sequences

A sequence is obtained if a function is given only for discrete values. Sequences are functions interpreted on natural numbers. The linear functions correspond to the arithmetic and the exponential functions (discussed in detail next week) to the geometric series.

1.3 Exercises

1.3.1

The function $P = f(t)$ gives the population of a city (in millions) as a function of t, the years that have passed since 1970.

a) What does the expression $f(12) = 35$ mean regarding the population of the city? Formulate your answer as a complete sentence.

b) Using f, find a formula for the function j, where $j(t)$ denotes the population (in ten thousands) of the city as a function of the years passed since 1985.

c) What is the meaning of the function $g(t) = f(t - 1970) - f(t - 1970 - 1)$?

d) Let $h(t)$ denote the growth of the population in percents in year t compared to the previous year. Find the expression for h using f. Find the expression for h using j.

e) In 1990 the population of the city was 30 million, in 1991 it was 29.5 million, while in 1992 there were 31 million people living in this particular city. At which points can you evaluate f, g or h? If possible, do these evaluations.

Solution:

a) The population of the city in 1982 was 35 million.

b) $j(t) = f(15 + t) \cdot 100$

c) How much the population changed from year $t - 1$ to year t.

d) The question could be clarified as "Let $h(t)$ denote the percentage growth of the population from year $t - 1$ to year t...." (percentage of change represent the degree of change over time). Then
$h(t) = \dfrac{f(t-1970) - f(t-1970-1)}{f(t-1970-1)} = \dfrac{j(t-1985) - j(t-1985-1)}{j(t-1985-1)}$.

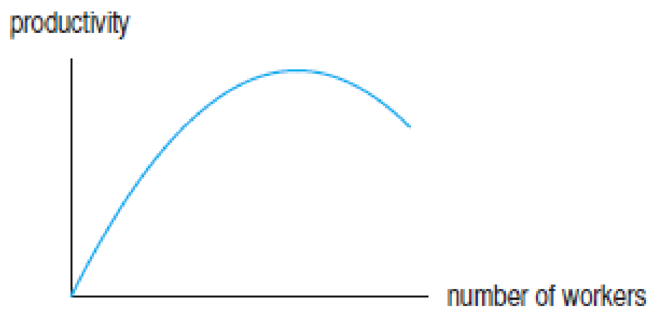
e) $f(20) = 30$ million, $f(21) = 29.5$ million, $f(22) = 31$ million

$g(1991) = f(21) - f(20) = -0.5$ million, $g(1992) = f(22) - f(21) = 1.5$ million

$h(1991) = \dfrac{f(21) - f(20)}{f(20)} \approx -0.017$ (1.7% decrease), $h(1992) = \dfrac{f(22) - f(21)}{f(21)} = 0.051$ (5.1% increase)

1.3.2

The figure below tells you about an assembly line whose productivity is represented as a function of the number of workers on the line:



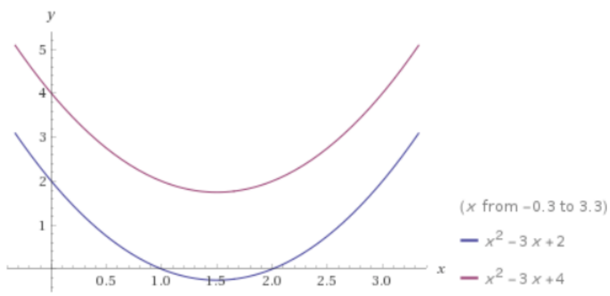
What can you tell about this curve, can you explain its shape? Which function has a graph similar to this?

1.3.3

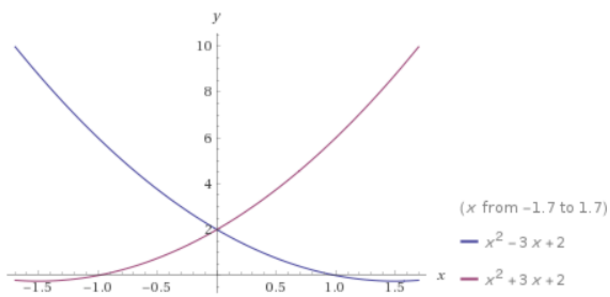
Plot the graph of f, if $f(x) = x^2 - 3x + 2$. Plot the following functions as well: $g(x) = f(x) + 2$, $h(x) = f(x+3)$, $i(x) = -2f(x)$, $j(x) = f(x/2)$. In what situations do such transformations occur for functions in practice?

Solution:

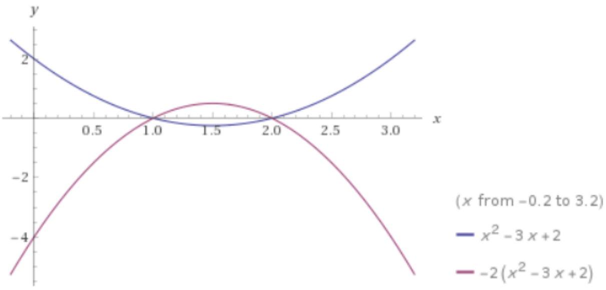
$g(x) = f(x) + 2 = x^2 - 3x + 4$ (move up by 2):



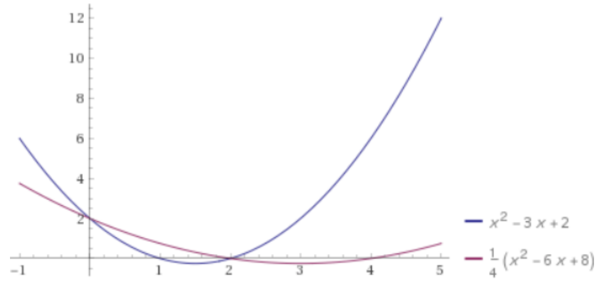
$h(x) = f(x+3) = (x+3)^2 - 3(x+3) + 2$ (move left by 3):



$i(x) = -2f(x)$ (reflect about the x-axis and squeeze vertically):



$j(x) = f(x/2)$ (stretch horizontally):



Such transformations occur: shifting or rescaling variables, changing units, changing reference points

1.3.4

The function $S = f(t)$ gives the average annual sea level S (in meters) in Aberdeen, Scotland (in meters), as a function of t, where t is the number of years before 2008. Write mathematical expressions which have the following meanings:

a) In 1983 the average annual sea level in Aberdeen was 7.019 meters.

b) The average annual sea level was the same in 1978 and 1879.

c) The average annual sea level in 2008.

d) The average annual sea level in centimeters, given as a function of the number of years passed since 1980.

e) The average annual sea level in Aberdeen increased by 1 millimeter from 2007 to 2008.

Solution:

a) $f(25) = 7.019$

b) $f(30) = f(129)$

c) $f(0)$

d) $f(28-t) \cdot 100$

e) $f(0) = f(1) + 0.001$

1.3.5

a) What is the general formula for a linear function? How many points do determine the parameters? Let $f(x) = 3 - 2x$. Plot the graph of f. Where does the graph intersect the axes?

b) Give the formula for the linear function g, which intersects the x axis at $(-1, 0)$ and the y axis at $(0, 8)$.

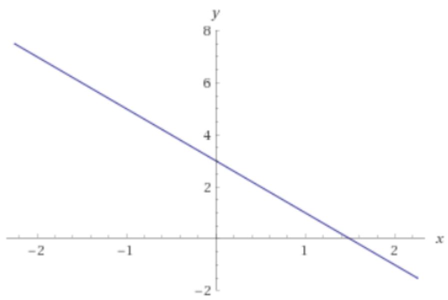
c) Provide the formula for a linear function h_1 so that the point $(6, 5)$ lies on the graph of h_1. Give the formula for a linear function h_2 so that both $(6, 5)$ and $(10, 1)$ lie on the graph of h_2. How many points do exactly determine the function?

d) Let $f(x) = 2x + 5$. Calculate the change and the (average) rate of change of f with respect to x on the interval $[-1, 2]$. Do the same for the interval $[-3, -2]$. What did you notice?

Solution:

a) linear function: $f(x) = ax + b$; 2 points determine

The graph intersects the y-axis at 3, and the x-axis at $\frac{3}{2}$.



b) $g(x) = 8x + 8$

c) $h_1(x) = ax + b = ax + 5 - 6a,\ a \in \mathbb{R}$
$h_2(x) = -x + 11$
2 points determine

d) $f(x) = 2x + 5$
$[-1, 2]$ change: $f(2) - f(-1) = 9 - 3 = 6$, rate of change: $\dfrac{f(2) - f(-1)}{2 - (-1)} = 2$
$[-3, -2]$ change: $f(-2) - f(-3) = 1 - (-1) = 2$, rate of change: $\dfrac{f(-2) - f(-3)}{-2 - (-3)} = 2$
the rate of change is the same

1.3.6

Is the relationship between x and y linear? If it is, give the appropriate expression for the function.

x	5,2	5,3	5,4	5,5	5,6
y	27,8	29,2	30,6	32,0	33,4

Solution: yes, the relationship is linear (rate of change is the same: $\frac{1.4}{0.1} = 14$, $y = 14x - 45$

1.3.7

a) What is the general formula for a quadratic function? How many points do determine the parameters? Let $f(x) = x^2 + 3x + 2$ and $g(x) = 9 - x^2$. Plot f and g as well. What is the essential difference between the two graphs?

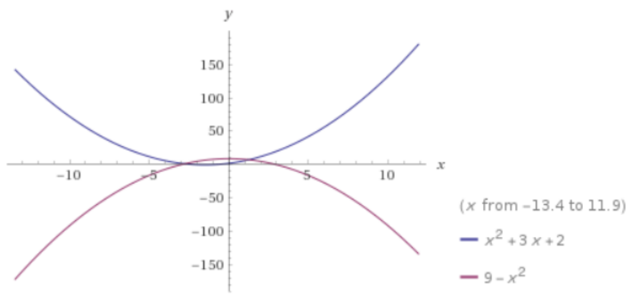
b) Where does f intersect the axes? Determine its zeros. Does it have extreme values (i.e. minimum and/or maximum)? If yes, determine the point(s) where f has an extreme value. What is its maximal/minimal value (if there is one)? How can these be determined by the original expression of f and after completing the square?

c) Give the formula for a quadratic function whose graph intersects the x axis at $(1, 0)$ and the y axis at $(0, -1)$. Provide the formula for a quadratic function which has maximum at $x = -5$. Give the formula for a quadratic function which has a minimal value of 4.

d) Find the formula for a quadratic function f so that the points $(1, -12)$ and $(4, 12)$ both lie on the graph of f. Give the formula for a quadratic function g so that the points $(1, -12)$, $(4, 12)$ and $(2, -10)$ all lie on the graph of g. How many points do uniquely determine the formula of a quadratic function?

Solution:

a) quadratic function: $f(x) = ax^2 + bx + c$; 3 points determine



b) f intersects the y-axis at 2, and the x-axis at -1 and -2; $f(0) = 2$ and $f(-1) = f(-2) = 0$
min $f(x) = -0.25$ at $x = -1.5$, no maximum
$f(x) = x^2 + 3x + 2 = (x + 1.5)^2 - 0.25$

c) $f_1(x) = ax^2 + (-a+1)x - 1$, $a \in \mathbb{R}$
$f_2(x) = -(x+5)^2 + c$, $c \in \mathbb{R}$
$f_3(x) = (x-a)^2 + 4$, $a \in \mathbb{R}$

d) $f(x) = ax^2 + (8-5a)x + 4a - 20$, $a \in \mathbb{R}$
$g(x) = 3x^2 - 7x - 8$
3 points determine

1.3.8

a) Give the change and the (average) rate of change of $f(x) = x^2$ on the interval $[2, 3]$. Do the same for the interval $[4, 5]$. What do you notice?

b) Give the change and the rate of change of $f(x) = 2x^2 + 3x + 1$ on the interval $[2, 3]$. Do the same for $[4, 5]$. Can you give the general formula for the rate of change of this f for the interval $[a, b]$?

Solution:

a) $f(x) = x^2$

$[2, 3]$ change: $f(3) - f(2) = 9 - 4 = 5$, rate of change: $\dfrac{f(3)-f(2)}{3-2} = 5$

$[4, 5]$ change: $f(5) - f(4) = 25 - 16 = 9$, rate of change: $\dfrac{f(5)-f(4)}{5-4} = 9$ The two rates are different.

b) $f(x) = 2x^2 + 3x + 1$

$[2, 3]$ change: $f(3) - f(2) = 28 - 15 = 13$, rate of change: $\dfrac{f(3)-f(2)}{3-2} = 13$

$[4, 5]$ change: $f(5) - f(4) = 66 - 45 = 21$, rate of change: $\dfrac{f(5)-f(4)}{5-4} = 21$

$[a, b]$ change: $f(b) - f(a) = 2b^2 + 3b + 1 - (2a^2 + 3a + 1) = 2(b^2 - a^2) + 3(b-a) = 2(b-a)(b+a) + 3(b-a)$,

rate of change: $\dfrac{f(b)-f(a)}{b-a} = \dfrac{2(b-a)(b+a)+3(b-a)}{b-a} = 2(a+b) + 3$

1.3.9

Write the change and the rate of change of the function $f(x) = x^3$ for the following intervals: $[1, 3]$, $[1.5, 2.5]$, $[1.75, 2.25]$. Can you give a simplified expression for the rate of change of f on an interval $[a, b]$?

Solution: $f(x) = x^3$

$[1, 3]$ change: $f(3) - f(1) = 26$, rate of change: $\dfrac{f(3)-f(1)}{3-1} = 13$

(can be calculated similarly for intervals $[1.5, 2.5]$ and $[1.75, 2.25]$)

$[a, b]$ change: $f(b) - f(a) = b^3 - a^3 = (b - a)(b^2 + ab + a^2)$, rate of change: $\dfrac{f(b) - f(a)}{b - a} = b^2 + ab + a^2$

1.3.10

* For a given function $f(x) = ax^2 + bx + c$ we know that $ac + bc + c^2$ is negative. Show that f has two different roots!

Solution: Need discriminant $D = b^2 - 4ac > 0$.

If $ac + bc + c^2 < 0$, then $-ac > bc + c^2$, so $b^2 - 4ac > b^2 + 4bc + 4c^2 = (b + 2c)^2 \geq 0$, therefore $D > 0$.

1.3.11

Suppose that the total cost of manufacturing q units of a certain product is $C(q) = 0.001q^2 + 0.9q + 2$ thousand dollars. The product can be sold for unit price of 1500 dollars.

a) Compute the cost of manufacturing 10 units.

b) Compute the cost of manufacturing the 10th unit.

c) Compute the profit if 10 units are produced and sold.

d) What is the maximal profit? How many units of product have to be manufactured (and sold) to reach this maximal profit?

Solution:

a) $C(10) = 0.001 \cdot 10^2 + 0.9 \cdot 10 + 2 = 11.1$

b) $C(9) = 0.001 \cdot 9^2 + 0.9 \cdot 9 + 2 = 10.181$

   cost of manufacturing the 10th unit: $C(10) - C(9) = 11.1 - 10.181 = 0.919$, i.e., 919 dollar

c) profit = revenue - manufacturing cost = $1.5 \cdot 10 - C(10) = 3.9$ i.e. 3900 dollar

d) profit on q units: revenue - manufacturing cost = $\$1.5 \cdot q - C(q) = -0.001q^2 + 0.6q - 2 = -0.001\left(q^2 - 600q + \dfrac{2}{0.001}\right) = -0.001\left((q - 300)^2 - 300^2 + 2000\right) = -0.001\left((q - 300)^2 - 88000\right) = -0.001(q - 300)^2 - 88$ (in thousand of dollars), therefore the maximum of this function is at $q = 300$, so 300 units have to be manufactured, and the maximum profit is 88 (in thousands of dollars) = \$88000

1.3.12

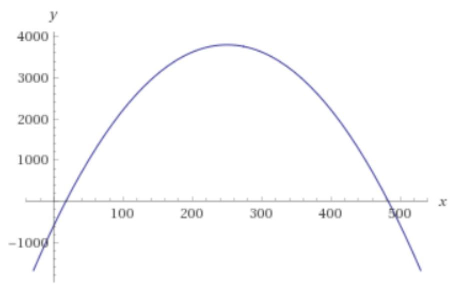
Suppose that when the price of a certain product is p hundred dollars per unit, then x hundred units will be purchased by consumers, where $p = -0.05x + 38$. The cost of producing x hundred units is $C(x) = 0.02x^2 + 3x + 574.77$ hundred dollars.

a) Express the profit P obtained from the sale of x hundred units as a function of x. Sketch the graph of the profit function.

b) Use the profit curve P to determine the level of production x that results in maximum profit. What unit price p corresponds to maximum profit?

Solution:

a) profit: $P(x) = (-0.05x+38)x - (0.02x^2+3x+574.77) = -0.07x^2+35x-574.77 = -0.07(x^2 - 500x - 8211) = -0.07\left((x-250)^2 - 250^2 - 8211\right) = -0.07\left((x-250)^2 - 54289\right) = -0.07(x - 250)^2 + 3800.23$ hundred dollars



b) $x = 250$ hundred units = 25000 units, $p = -0.05 \cdot 250 + 38 = 25.5$ dollar

1.3.13

An environmental study of a certain community suggests that the average daily level of carbon monoxide in the air will be $c(p) = 0.5p + 1$ ppm (parts per million) when the population is p thousand. It is estimated that t years from now the population of the community will be $p(t) = 10 + 0.1t^2$ thousand.

a) Express the level of carbon monoxide in the air as a function of time.

b) When will the carbon monoxide level reach 6.8 ppm?

Solution:

a) $C(p(t)) = 0.5(10 + 0.1t^2) + 1 = 6 + 0.05t^2$ ppm

b) $C(p(T)) = 0.5(10 + 0.1T^2) + 1 = 6.8$, $T = 4$ years

1.3.14

From the salary of an employee taxes and other contributions get subtracted, the resulting amount is the net income. The employer must pay some further forms of expenses. The subtractions and the taxes/contributions needed to be paid by the employer are listed below:

Gets subtracted	
Health insurance	7.00%
Contributions to the labour market	1.50%
Pension contribution	10.00%
Personal income tax	15.00%
Charged to the employer/company	
Social security	19.50%
Vocational training levy	1.50%

a) If the salary of an employee is 200 thousand Forints before taxes, then what is the net income? What is the total expense of the employer in this case? What percent of the net income do the subtracted taxes and contributions amount to?

b) If the net income equals 200 thousand Forints, then how much is the salary before taxes? How much is the total expense of the company in this case?

Solution:

a) net income: $200000 - 200000 \cdot (0,03 + 0,04 + 0,015 + 0,1 + 0,15) = 200000 \cdot 0,665 \ = 133000$ Ft

total expense: $200000 \cdot (19.5 + 1.5) \ = 200000 \cdot 21 \ = 42000$ Ft

$109000/133000 = 0,82 = 82\%$

b) salary: $200000/0,665 \ = 300752$ Ft

total expense: $300752 \cdot (19,5 + 1,5) \ = 63158$ Ft

2 Functions (2nd part)

If you have serious shortcomings in the field of geometric series, exponential function, logarithm, we recommend these two slightly longer but sufficiently detailed videos in Hungarian: geometric series, exponential function, logarithm and Exponential function and problems.

2.1 Motivating examples and introduction

Exponential functions, and their inverses, logarithmic functions, typically occur in interest, reproduction, and decomposition processes, and the "exponential growth" used in common parlance refers to processes that can be described by these functions. It is less well known that our sound perception is also logarithmic: a higher octave sound means twice the vibrational number (frequency), a decibel louder sound corresponds to a twice as strong (double amplitude) sound wave.

Examples of exponentially changing phenomena and problems:

- The spread of COVID-19 virus has shown an exponential increase in the world until the introduction of restrictive measures. According to forecasts of the spread of the epidemic, the number of people infected in a large city is increasing every day to 105% of the previous day's value. If the rate of growth were to be projected, in how many days would the total number of infected people rise from 0.2% to 1% of the total population?
- A new photocopier will cost 780,000 Forints. During use, the value of the machine decreases by 15% in one year. When the value of the machine falls below 150,000 Forints, it is scrapped. How many years from now will this machine be scrapped?

2.2 Theory

2.2.1 The exponential function

The exponential function is $f(x) = a^x$, where $a > 0$ is called the base. The most commonly occurring base in Business and Economics is $e \approx 2.72$ and the corresponding exponential function is the natural exponential function $f(x) = e^x = exp(x)$. The number e is defined as $e = lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$.
The most characteristic property of an exponential function is determined by the fact that the base is greater than 1 (then the function is strictly monotonically increasing) or less than 1 (then the function is strictly monotonically decreasing).

Compound interest: Compound interest is interest calculated on the initial principal, which also includes all of the accumulated interests of the previous periods. It can be thought of as "interest on interest," and will make a sum grow at a faster rate than simple interest, which is calculated only on the principal amount.

It is thought that Bernoulli discovered the mathematical constant e by studying a question about compound interest. He realized that if an account that starts with \$1.00 and pays 100% interest per year, at the end of the year, the value is \$2.00; but if the interest is computed (credited) and added twice in the year (the interest rate for each 6 months will be 50%), the initial \$1 will be $\$1 + \$0.5 + (\$1 + \$0.5)/2 = \$1 \cdot \left(1 + \frac{1}{2}\right)^2 = \$2.25$. Starting with \$1 initial value, compounding more frequently yields:

Compounded	Final amount (\$) after 1 year
yearly	$\left(1+\frac{1}{1}\right)^1 = 2$
semi-annually	$\left(1+\frac{1}{2}\right)^2 = 2.25$
quarterly	$\left(1+\frac{1}{4}\right)^4 = 2.4414...$
monthly	$\left(1+\frac{1}{12}\right)^{12} = 2.6130...$
weekly	$\left(1+\frac{1}{52}\right)^{52} = 2.6925...$
daily	$\left(1+\frac{1}{365}\right)^{365} = 2.7045...$
hourly	$\left(1+\frac{1}{8760}\right)^{8760} = 2.7181...$
every minute	$\left(1+\frac{1}{525600}\right)^{525600} = 2.7182...$
every second	$\left(1+\frac{1}{31536000}\right)^{31536000} = 2.7182...$

In general, if there are n compounding intervals, the interest for each interval will be $100\%/n$ and the value at the end of the year will be $\$1 \cdot \left(1+\frac{1}{n}\right)^n$.

Bernoulli noticed that if the frequency of compounding is increased without limit, this sequence can be modeled as follows:

For an interest rate r compounded at frequency n on an initial principal A, the value of the asset at time t is

$$V(n,t) = A\left(1+\frac{r}{n}\right)^{nt}$$

Continuous compounding is the mathematical limit that compound interest can reach if it's calculated and reinvested into an account's balance over a theoretically infinite number of periods. While this is not possible in practice, the concept of continuously compounded interest is important in finance. It is an extreme case of compounding, as most interest is compounded on a monthly, quarterly or semiannual basis. In theory, continuously compounded interest means that an account balance is constantly earning interest, as well as refeeding that interest back into the balance so that it also earns interest. Calculated as:

$$lim_{n\to\infty} V(n,t) = V(t) = Ae^{rt},$$

where

V = final amount

A = initial principal balance

r = interest rate

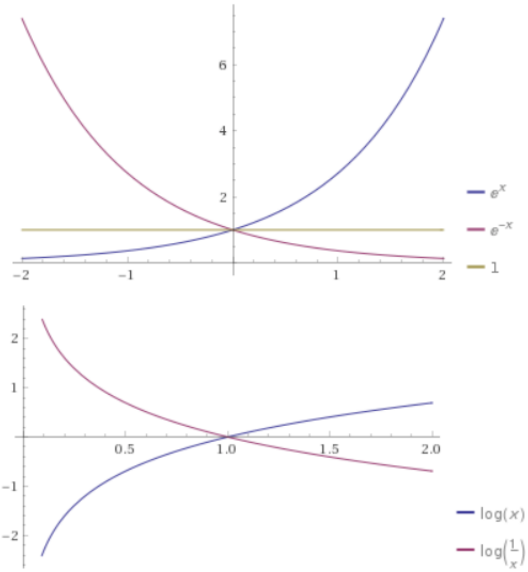
n = number of times interest applied per time period

t = number of time periods elapsed.

2.2.2 Logarithm

The logarithm of a given number x is the exponent to which the base a must be raised, to produce that number x, where $a > 0, a \neq 1, x > 0$

The logarithm function log in base a is the inverse function of the exponential function in base a: $y = \log_a x \iff x = a^y$, where $a > 0, a \neq 1, x > 0$. The natural logarithm, ln is the inverse function of the natural exponential function: $y = \ln x \iff x = e^y$. This means that $e^{\ln x} = \ln e^x = x$ and for any base a, $a^{\log_a x} = \log_a a^x = x$,



2.3 Exercises

2.3.1

We invested 2 million Forints with the annual interest rate being 4%.

a) What will our balance be in 3 years?

b) How many years do we have to wait in order for the balance to be twice the amount of the money invested?

c) When will we have 3 million Forints if the interest is compounded monthly?

d) What will be our balance after a year if the interest is compounded 4 times per year? (If the interest r is compounded k times per year, then the year is divided into k equal compounding periods and the interest rate in each period is $\frac{r}{k}$.)

Solution:

$A = 2 \cdot 10^6$, $r = 4\%$, $n = 1$, $t = 3$ years

a) $A\left(1 + \frac{r}{n}\right)^{nt} = 2 \cdot 10^6 \left(1 + \frac{0.04}{1}\right)^{1 \cdot 3} = 2 \cdot 10^6 \,(1.04)^3 \approx 2.25 \cdot 10^6$ Ft

b) $2 \cdot 10^6 \left(1 + \frac{0.04}{1}\right)^{1 \cdot T} = 2 \cdot 2 \cdot 10^6$, $T = ?$

$T = \dfrac{\ln(2)}{\ln 1.04} \approx 17.67$, therefore we have to wait 18 years.

c) $2 \cdot 10^6 \left(1 + \frac{0.04}{12}\right)^{12 \cdot T} = 2 \cdot 10^6 \left(1 + \frac{0.04}{12}\right)^{M} = 3 \cdot 10^6$, $M = ?$

$M = \dfrac{\ln(\frac{3}{2})}{\ln\left(1 + \frac{0.04}{12}\right)} \approx 121.8$, therefore we have to wait 122 months.

d) $2 \cdot 10^6 \left(1 + \frac{0.04}{4}\right)^{4 \cdot 1} \approx 2.08 \cdot 10^6$ Ft

2.3.2

We invest 1 Forint with the annual interest rate being 100%.

a) What will our balance be after a year, if interest is compounded once, twice, 4 times, 12 times or 100 times per year? What is the maximal possible amount that can be reached in theory, if the interest is compounded each moment? What mathematical constant do you recognize?

b) Given this scenario, what would be the balance after 3 years?

c) How much would we have to wait for the balance to be 10 Forints?

Solution:

a) see above, e

b) if for example $n = 100$, then the balance after 3 years is 1 Ft $\cdot \left(1 + \frac{1}{100}\right)^{100 \cdot 3} \approx 19.8$ Ft

c) if for example $n = 100$, then $1 \cdot \left(1 + \frac{1}{100}\right)^{100 \cdot T} = 10$, therefore $T = \dfrac{\ln 10}{\ln(1.01) \cdot 100} = 2.3$ years

2.3.3

* As in the previous problem, we invested 1 Forint. If the monthly interest rate is 100%, what is the rate of change for our balance between two consecutive months? What would be the rate of change for our balance between two consecutive days if the interest were compounded daily?

Solution: The average rate of change shows how much our balance changed over the given period (e.g., two consecutive months), but looking at how much the balance changed relative to what we have started out with at the beginning of the given period is more informative. For example, if our initial investment is 1 Ft with a monthly interest rate of 100%, the average rate of change between the second and third month would be $2^3 - 2^2 = 4$ Ft, while between the third and fourth month would be $2^4 - 2^3 = 8$ Ft, that is how much our balance grew within these given months. But how big was this change relative to what we have started out at the beginning of the given months? So in the example, this relative rate of change between the second and third month would be $\frac{2^3-2^2}{2^2} = 1$, while between the third and fourth month would also be $\frac{2^4-2^3}{2^3} = 1$, which means that our investment in both of the cases grew by 100%. In general, the relative rate of change is $\dfrac{(1+r)^{t+1} - (1+r)^t}{(1+r)^t} = r$, where r denotes the monthly interest rate, and t denotes the months. Notice that this formula does not depend upon the t time period.

If the monthly interest rate of 100% was compounded daily (i.e., the interest is $\dfrac{1}{30}$), then the relative rate of change would be

$$\frac{\left(1+\frac{1}{30}\right)^{t+1} - \left(1+\frac{1}{30}\right)^{t}}{\left(1+\frac{1}{30}\right)^{t}} = \frac{1}{30}$$

where t denotes the days.

2.3.4

We invested 400, 000 Forints, the annual interest rate is 5%, and the interest is compounded each moment.

a) What will the balance be after a year?

b) What will the balance be after 3 years?

c) How much time does it take for the balance to be twice as much as we have originally invested?

Solution:

a) $400000 \cdot e^{0.05\times 1} = 420508$ Ft

b) $400000 \cdot e^{0.05\times 3} = 464734$ Ft

c) $400000 \cdot e^{0.05 \times T} = 2 \cdot 400000$, then $T = \frac{\ln 2}{0.05} \approx 13.86$ year

2.3.5

Draw the graph of the exponential functions. What is the general shape of the graph of an exponential function? Plot the following functions: $f(x) = 2^x$, $g(x) = \left(\frac{1}{2}\right)^x$, $h(x) = 3^{2x}$, $i(x) = 2 \cdot 3^x$, $j(x) = 2 \cdot 3^{-x}$, $k(x) = 3^{-2x}$.

Draw the graph of the logarithmic functions as well. What can you tell about the shape? For example, plot:
$f(x) = \log_2(x)$, $g(x) = \log_{0.1}(x)$, $h(x) = \log_3 x - 1$.

2.3.6

How many times can you fold a piece of paper in half? What would be the thickness of the paper after folding it 20 times, if the original thickness was 0.1

Solution:

1-fold: 2 sheet of paper $\times 0.1$ mm = 0.2 mm

2-fold: 2^2 sheet of paper $\times 0.1$ mm = $2^2 \cdot 0.1$ mm = 0.4 mm

3-fold: 2^3 sheet of paper $\times 0.1$ mm = $2^3 \cdot 0.1$ mm = 0.8 mm

...

20-fold: 2^{20} sheet of paper $\times 0.1$ mm = $2^{20} \cdot 0.1$ mm = 104857.6 mm = 104.8576 m

2.3.7

The range of audible frequencies is 20 Hz to 20 kHz.

a) For a note with given frequency, the note which is an octave above it has double the frequency of the original note. How many octaves are there in the audible range?

b) The normal musical note A has a frequency of 440 Hz. What is the frequency of the other A notes?

Solution:

a)

Octave	Frequency
1st	$20 - 40$ Hz
2nd	$40 - 80$ Hz
...	...
nth	$2^n \cdot 10 - 2^{n+1} \cdot 10$ Hz
...	...
10th	$2^{10} \cdot 10 - 2^{11} \cdot 10$ Hz = $10240 - 20480$ Hz = $10.24 - 20.48$ kHz

b) $A_0 = 27.5$ Hz, $A_1 = 55$ Hz, $A_2 = 110$ Hz, $A_3 = 220$ Hz, $A_4 = 440$ Hz, $A_5 = 880$ Hz, $A_6 = 1760$ Hz, $A_7 = 3520$ Hz, $A_8 = 7040$ Hz, $A_9 = 14080$ Hz

2.3.8

The population of the Earth has reached 6.7 billion in 2007. At that time the annual growth rate was 1.2%. Assuming that this rate has not changed, what was the population in 2017?

Solution: $A = 6.7 \cdot 10^9$, $r = 1.2\%$, $n = 1$, (yearly rate), $t = 10$ (10 years from 2007 to 2017)

$$A\left(1+\tfrac{r}{n}\right)^{nt} = 6.7 \cdot 10^9 \left(1+\tfrac{0.012}{1}\right)^{1\cdot 10} = 6.7 \cdot 10^9\,(1.012)^{10} = 7.5 \cdot 10^9$$

2.3.9

The relationship between the demand for a given product and its unit price p can be described by a linear function: $q = D(p)$, where q means the demand for the product (i.e. the amount that will be bought) and p is the unit price of the product (in thousand Forints). For this particular product we know the following: if the unit price is increased by 5000 Forints, the demand will decrease by 2 (pieces). Furthermore, if the unit price is 550 thousand Forints, then the demand is 100 pieces.

a) Find the expression for the demand as a function of the selling price and plot the determined function.

b) If the supply is given as a function of the unit price p (in thousand Forints) as $S(p) = p^2+2p+2$, what will be the equilibrium unit price and the equilibrium amount produced?

Solution: p: unit price of the product (in thousand Forints)

$D(p)$: demand for the product (pieces)

a) linear relationship: $D(p) = a \cdot p + b$

slope: $a = \dfrac{-2}{5}$

$100 = a \cdot 550 + b = \dfrac{-2}{5}550 + b$, then $b = 320$, so $D(p) = \dfrac{-2}{5}p + 320 = -0.4p + 320$

b) $D(p) = S(p)$ (demand = supply), i.e., $-0.4p + 320 = p^2 + 2p + 2$, therefore $p \approx 16.673$, that is 16673 Ft.

In this case $D(16.673) = S(16.673) \approx 313$ pieces

2.3.10

The Carbon-14 (^{14}C) isotope has a half-life[1] of 5730 years.

[1] Amount of time that it takes half of the isotope in a sample to decay.

a) According to some source, the amount of non-decayed atoms after t time can be given as $N(t) = A \cdot 2^{-t/5730}$, where A denotes the original amount, while according to another source the formula is $N(t) = A/2^{t/5730}$, and there is a third formula as well: $N(t) = e^{-(\ln 2)t/5730+\ln A}$. Which of these is correct?

b) How long does it take for 3/4 of the nuclei to decay?

c) What percent of the nuclei is decayed in 4000 years?

Solution:

a) The formulas are the same.

b) $2^{-t/5730} = 1 - \frac{3}{4}$, therefore $t = 11460$ years

c) $1 - 2^{-4000/5730} = 1 - 0.616 = 0.384$, i.e., 38.4%

2.3.11

Producing q pieces of a product costs $C(q) = 0.2q^2 + q + 900$ dollar. After production is started, $q(t) = t^2 + 100t$ pieces of the product is produced within t hours. What is the cost of the production at the end of t hours? Calculate it for $t = 2$.

Solution: $C(q(t)) = 0.2q(t)^2 + q(t) + 900 = 0.2(t^2 + 100t)^2 + t^2 + 100t + 900$,

$C(q(2)) = 0.2(2^2 + 100 \cdot 2)^2 + 2^2 + 100 \cdot 2 + 900 \approx 9427$ dollar

2.3.12

On a unit price of p, $D(p) = \dfrac{40000}{p}$ pieces of product can be sold per month. According to the company's business plan, the unit price in t months will be $p(t) = 0.4t^{\frac{3}{2}} + 6.8$ dollar per piece.

a) How many pieces of products can be sold in month t? Calculate it for t = 4.

b) What is the revenue in month t?

Solution:

a) $D(p(t)) = \dfrac{40000}{p(t)} = \dfrac{40000}{0.4 \cdot t^{\frac{3}{2}} + 6.8}$

$D(p(4)) = \dfrac{40000}{p(4)} = \dfrac{40000}{0.4 \cdot 4^{\frac{3}{2}} + 6.8} = \dfrac{40000}{10} = 4000$ pieces

b) 40000 dollar

3 Linear algebra - Matrices and vectors

3.1 Motivating example and introduction

A car mechanic sells 4 different types of tires. The unit price and the number of tires in stock are listed in the following table.

	Tire 1	Tire 2	Tire 3	Tire 4
Unit price (Ft/pc)	13000	21000	43000	10000
In stock (pc)	10	5	2	8

This tabular storage of numbers is called a matrix.

In general, a matrix is a collection of numbers arranged into a fixed number of rows and columns. An important attribute of a matrix is its size or dimension, i.e., the number of rows and columns. A matrix is of size of $n \times k$ if it has n rows and k columns. For example, matrix

$$A = \begin{pmatrix} 2 & 3 & 1 \\ 0 & 1 & 2 \end{pmatrix} \text{ or } A = \begin{bmatrix} 2 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

has 2 rows and 3 columns. The individual items in an $n \times k$ matrix A, which are denoted by a_{ij} where i and j vary from 1 to n and 1 to k, respectively are called its elements or entries:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix}, \quad \text{or shortly} \quad \begin{pmatrix} & \vdots & \\ \cdots & a_{ij} & \cdots \\ & \vdots & \end{pmatrix} \quad \text{or} \quad \left(a_{ij}\right)$$

3.2 Notions and methods

3.2.1 Types of special matrices

1. If a matrix has only one row or only one column it is called a vector. A matrix of dimension $1 \times k$ consists of a single row, and is called a *row vector.* A matrix of dimension $n \times 1$ consists of a single column, and is called a *column vector.* For example, the row vector

$$v = \begin{pmatrix} 1 \,, & 3 \end{pmatrix}$$

represents a point in the Cartesian coordinate system, but also represents a (location) vector pointing there from the origin. Similarly, column vector

$$w = \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix}$$

corresponds to a (location) vector in the three dimensional space pointing from the origin to the point $(2, -1, 5)$.

2. A matrix is called a *square matrix* if it has an equal number of rows and columns.

3. A matrix is called a *null matrix* or *zero matrix* if all its elements are equal to zero. The zero matrix of dimension $n \times k$ is sometimes written as $0_{n\times k}$, but usually it is just denoted by 0. In this case the dimension of the zero matrix must be determined from the context.

4. The *main diagonal* of a square matrix is the set of all elements that lie on the imaginary line that runs from the top left corner to the bottom right corner of the matrix, in other words, those elements whose row and column indices are the same. The elements belonging to the diagonal are called *diagonal* elements, and all other elements are called *off-diagonal*.

5. A square matrix is *diagonal* if all the off-diagonal entries are zero.

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}.$$

6. The *identity matrix* is a matrix in which all the elements on the main diagonal are equal to 1 and all other elements are equal to 0. For matrix products (see Operations with Matrices) the role of 1 is played by the identity matrix. The $n \times n$ identity matrix is denoted by I_n. For example,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

7. A *permutation matrix* is a square matrix with exactly one entry of each row is being one, exactly one entry of each column is being one, and all other entries equal to zero, e.g.

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

3.2.2 Operations with matrices

1. The transpose of an $n \times k$ matrix A is the $k \times n$ matrix A^T formed by turning rows into columns and vice versa. For example, if

$$A = \begin{pmatrix} -1 & 3 & 0 \\ 0 & 2 & 1 \end{pmatrix} \text{ then } A^T = \begin{pmatrix} -1 & 0 \\ 3 & 2 \\ 0 & 1 \end{pmatrix}$$

2. Two matrices of the same dimension can be added together. The result is another matrix of the same dimension, obtained by adding the corresponding elements of the two matrices. For example,

$$\begin{pmatrix} -1 & 3 & 0 \\ 0 & 2 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 1 & 4 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1+2 & 3+1 & 0+4 \\ 0-1 & 2+0 & 1+1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 4 \\ -1 & 2 & 2 \end{pmatrix}$$

   Matrix subtraction is similar. As an example,

$$\begin{pmatrix} -1 & 3 \\ 0 & 2 \end{pmatrix} - I_2 = \begin{pmatrix} -1 & 3 \\ 0 & 2 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1-1 & 3-0 \\ 0-0 & 2-1 \end{pmatrix} = \begin{pmatrix} -2 & 3 \\ 0 & 1 \end{pmatrix}$$

   Adding the null matrix to any arbitrary matrix A, we get the matrix itself: $A + 0 = A$. Zero matrices play a similar role in operations with matrices as the number zero plays in operations with real numbers.

3. A matrix A and a number c can be multiplied. The resulting *scalar multiple of A with c* can be computed by multiplying every entry of A by c. For example,

$$A = \begin{pmatrix} -1 & 3 \\ 0 & 2 \end{pmatrix} \text{ and } c = 3, \text{ then } cA = \begin{pmatrix} -3 & 9 \\ 0 & 6 \end{pmatrix}$$

4. It is possible to multiply two matrices using matrix multiplication; however, matrix multiplication does not mean multiplying element by element. You can multiply two matrices A and B provided their dimensions are compatible, which means the number of columns of A equals the number of rows of B. Suppose A is an $m \times k$ and B is a $k \times n$ matrix, then the product matrix AB is a matrix of a dimension $m \times n$. To find the ijth element of this product, you need to use the ith row of A and the jth column of B. “Moving left to right along the ith row of A” while “moving top to bottom down the jth column of B”, you keep a running sum of the product of elements, one from A and one from B. The elements of the product matrix $C = (c_{ij})$, which is obtained by multiplying matrices $A = (a_{ij})$ and

$B = (b_{ij})$, are $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj}$, so that

$$\xrightarrow{i\text{th row}} \begin{pmatrix} \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & & \vdots \end{pmatrix} \begin{array}{c} j\text{th} \\ \text{column} \\ \downarrow \\ \begin{pmatrix} \dots & b_{1j} & \dots \\ \dots & b_{2j} & \dots \\ & \vdots & \\ \dots & b_{nj} & \dots \end{pmatrix} \end{array} = \begin{pmatrix} & \vdots & \\ \dots & c_{ij} & \dots \\ & \vdots & \end{pmatrix}.$$

For example,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1 \cdot 5 + 2 \cdot 7 & 1 \cdot 6 + 2 \cdot 8 \\ 3 \cdot 5 + 4 \cdot 7 & 3 \cdot 6 + 4 \cdot 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}, \text{ and}$$

$$\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 5 \cdot 1 + 6 \cdot 3 & 5 \cdot 2 + 6 \cdot 4 \\ 7 \cdot 1 + 8 \cdot 3 & 7 \cdot 2 + 8 \cdot 4 \end{pmatrix} = \begin{pmatrix} 23 & 34 \\ 31 & 46 \end{pmatrix}.$$

In general, matrix multiplication is not commutative: we do not (in general) have $AB = BA$. In fact, BA may not even make sense, or, if it makes sense, it may have different dimension than AB. For example, if A is 2×3 and B is 3×4, then AB makes sense (the dimensions are compatible), but BA does not even make sense (the dimensions are incompatible). Even when AB and BA both make sense and are the same dimension, i.e., when A and B are square matrices, we do not (in general) have $AB = BA$.

Even though matrix multiplication is not commutative, it is associative: if A is an $m \times p$ matrix, B is a $p \times q$ matrix, and C is a $q \times n$ matrix, then $A(BC) = (AB)C$.

We note that multiplying a matrix by the identity matrix gives the same matrix. The identity matrix plays a similar role in operations with matrices as the number 1 plays in operations with real numbers.

5. The inverse of an $n \times n$ square matrix A is the matrix A^{-1} for which $A \cdot A^{-1} = A^{-1} \cdot A = I_n$. Not all square matrices have an inverse. There are procedures to find the inverse matrix, but it can also be found with various software. In the case of a 2×2 matrix, the inverse can easily be calculated as follows. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then if $detA = ad - bc \neq 0$, the inverse of the matrix A is $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

   The determinant of a square matrix, $\det A$, will be defined later in general.

3.3 Exercises

3.3.1

Given the matrices below, evaluate the expressions, if possible. If it is not possible, explain why.

$$A = \begin{pmatrix} 4 & 6 \\ 2 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 3 & -2 \\ -1 & 6 \end{pmatrix} \qquad C = \begin{pmatrix} 2 & -1 & -3 \\ 6 & 0 & 2 \end{pmatrix} \qquad D = \begin{pmatrix} 4 & -1 & 3 \\ -5 & -1 & -3 \end{pmatrix}$$

1. dimensions of B
2. dimensions of D
3. dimensions of $A + B$
4. dimensions of $C - D$
5. a_{12}
6. b_{21}
7. c_{13}
8. d_{23}
9. $A + B$
10. $A - B$
11. $A + C$
12. $3D$
13. $-2B$
14. $C - D$
15. $D - C$
16. $3A + 2B$
17. A^{-1}

Solution:

1. dimensions of B 2×2.
2. dimensions of D 2×3.
3. dimensions of $A + B$ 2×2.
4. dimensions of $C - D$ 2×3.
5. $a_{12} = 6$
6. b_{21}=-1

7. c_{13}=-3
8. d_{23}=-3
9. $$A + B = \begin{pmatrix} 7 & 4 \\ 1 & 11 \end{pmatrix}$$
10. $$A - B = \begin{pmatrix} 1 & 8 \\ 3 & -1 \end{pmatrix}$$
11. $A + C$ not defined .
12. $$3D = \begin{pmatrix} 12 & -3 & 9 \\ -15 & -3 & -9 \end{pmatrix}$$
13. $$-2B = \begin{pmatrix} -6 & 4 \\ 2 & -12 \end{pmatrix}$$
14. $$C - D = \begin{pmatrix} -2 & 0 & -6 \\ 11 & 1 & 5 \end{pmatrix}$$
15. $$D - C = \begin{pmatrix} 2 & 0 & 6 \\ -11 & -1 & -5 \end{pmatrix}$$
16. $$3A + 2B = \begin{pmatrix} 18 & 14 \\ 4 & 27 \end{pmatrix}$$
17. $$A^{-1} = \begin{pmatrix} 5/8 & -6/8 \\ -2/8 & 4/8 \end{pmatrix}$$

3.3.2

Let

$$U = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} \qquad V = \begin{pmatrix} -1 & 4 \\ 0 & 2 \end{pmatrix} \qquad W = \begin{pmatrix} 5 & -5 \\ 4 & 7 \end{pmatrix}$$

Verify the following identities for the given matrices.

1. $U + W = W + U$
2. $2W + 3W = 5W$

3. $4(U+V) = 4U + 4V$
4. $2(3V) = 6V$
5. $V \cdot W^T = (W \cdot V^T)^T$
6. $(V \cdot W) \cdot U = V \cdot (W \cdot U)$

Solution:

1.
$$U + V = \begin{pmatrix} 2+(-1) & 3+4 \\ 1+0 & 2+2 \end{pmatrix} = \begin{pmatrix} 1 & 7 \\ 1 & 4 \end{pmatrix}$$
$$V + U = \begin{pmatrix} -1+2 & 4+3 \\ 0+1 & 2+2 \end{pmatrix} = \begin{pmatrix} 1 & 7 \\ 1 & 4 \end{pmatrix}$$

2.
$$2W + 3W = \begin{pmatrix} 2\cdot 5 & 2\cdot(-5) \\ 2\cdot 4 & 2\cdot 7 \end{pmatrix} + \begin{pmatrix} 3\cdot 5 & 3\cdot(-5) \\ 3\cdot 4 & 3\cdot 7 \end{pmatrix} = \begin{pmatrix} 25 & -25 \\ 20 & 35 \end{pmatrix}$$
$$5W = \begin{pmatrix} 5\cdot 5 & 5\cdot(-5) \\ 5\cdot 4 & 5\cdot 7 \end{pmatrix} = \begin{pmatrix} 25 & -25 \\ 20 & 35 \end{pmatrix}$$

3.
$$4(U+W) = 4\begin{pmatrix} 2+5 & 3+(-5) \\ 1+4 & 2+7 \end{pmatrix} = \begin{pmatrix} 28 & -8 \\ 20 & 36 \end{pmatrix}$$
$$4U + 4W = \begin{pmatrix} 8 & 12 \\ 4 & 8 \end{pmatrix} + \begin{pmatrix} 20 & -20 \\ 16 & 28 \end{pmatrix} = \begin{pmatrix} 28 & -8 \\ 20 & 36 \end{pmatrix}$$

4.
$$2(3V) = 2\begin{pmatrix} -3 & 12 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} -6 & 24 \\ 0 & 12 \end{pmatrix} = 6\begin{pmatrix} -1 & 4 \\ 0 & 2 \end{pmatrix} = 6V$$

5.
$$V \cdot W^T = \begin{pmatrix} -1 & 4 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 5 & 4 \\ -5 & 7 \end{pmatrix} = \begin{pmatrix} -25 & 24 \\ -10 & 14 \end{pmatrix}$$
$$(W \cdot V^T)^T = \left(\begin{pmatrix} 5 & -5 \\ 4 & 7 \end{pmatrix} \cdot \begin{pmatrix} -1 & 0 \\ 4 & 2 \end{pmatrix}\right)^T = \begin{pmatrix} -25 & -10 \\ 24 & 14 \end{pmatrix}^T = \begin{pmatrix} -25 & 24 \\ -10 & 14 \end{pmatrix}$$

6.
$$(V \cdot W) \cdot U = \left(\begin{pmatrix} -1 & 4 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 5 & -5 \\ 4 & 7 \end{pmatrix}\right) \cdot \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 11 & 33 \\ 8 & 14 \end{pmatrix} \cdot \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$$
$$= \begin{pmatrix} 55 & 99 \\ 30 & 52 \end{pmatrix}$$
$$V \cdot (W \cdot U) = \begin{pmatrix} -1 & 4 \\ 0 & 2 \end{pmatrix} \cdot \left(\begin{pmatrix} 5 & -5 \\ 4 & 7 \end{pmatrix} \cdot \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}\right) = \begin{pmatrix} -1 & 4 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 5 & 5 \\ 15 & 26 \end{pmatrix} =$$

$$= \begin{pmatrix} 55 & 99 \\ 30 & 52 \end{pmatrix}$$

3.3.3

In this problem R and M are matrices of mean SAT scores (SAT is a standardized test widely used for college admissions in the United States). The columns are average SAT reasoning scores for the years 2001-2008. The first row is scores for males and the second row is scores for females. Matrix R is the Critical Reading scores, and matrix M is the Mathematics scores.

$$R = \begin{pmatrix} 509 & 507 & 512 & 512 & 513 & 505 & 504 & 504 \\ 502 & 502 & 503 & 504 & 505 & 502 & 502 & 500 \end{pmatrix}$$

$$M = \begin{pmatrix} 533 & 534 & 537 & 537 & 538 & 536 & 533 & 533 \\ 498 & 500 & 503 & 501 & 504 & 502 & 499 & 500 \end{pmatrix}$$

1. Calculate $R + M$. What does this represent?
2. Calculate $M - R$. What does this represent?

Solution:

1. $$R + M = \begin{pmatrix} 1042 & 1041 & 1049 & 1049 & 1051 & 1041 & 1037 & 1037 \\ 1000 & 1002 & 1006 & 1005 & 1009 & 1004 & 1001 & 1000 \end{pmatrix}$$

   the average SAT scores for both reading and mathematics combined

2. $$M - R = \begin{pmatrix} 24 & 27 & 25 & 25 & 25 & 31 & 29 & 29 \\ -4 & -2 & 0 & -3 & -1 & 0 & -3 & 0 \end{pmatrix}$$

   the average SAT points the students scored more (or less if negative) on their mathematics than on their reading test

3.3.4

The number of meals, N, and the cost of the meals, C, for a weekend class reunion are given by the matrices

$$N = \begin{pmatrix} 20 & 35 & 70 \\ 30 & 35 & 50 \end{pmatrix} \text{ and } C = \begin{pmatrix} 8 \\ 12 \\ 50 \end{pmatrix}.$$

The first column of N is the number of breakfasts, the second is the number of lunches, and the third is the number of dinners. The first row of N is the meals needed on Saturday, the second on Sunday. The first row of C is the cost of breakfast, the second row is the cost of lunch, and the last row is the cost of dinner.

1. Calculate NC.
2. What is the practical meaning of NC?

Solution:

1. $$N \cdot C = \begin{pmatrix} 4080 \\ 3160 \end{pmatrix}$$
2. The entries of vector $N \cdot C$ are the total cost of meals on Saturday and on Sunday.

3.3.5

The vectors

$$P = (5, 22, 35, 18),\ \ S = (20, 33, 14, 40),\ \ F = (12, 28, 25, 20),\ \ W = (2, 19, 42, 12)$$

represent the average number of customers in the morning, early afternoon, late afternoon, and evening in a cafe during Spring (P), Summer (S), Fall (F), and Winter (W). For example, $P = (5, 22, 35, 18)$ means that, in Spring, the cafe has an average of 5 customers in the morning, 22 in the early afternoon, 35 in the late afternoon, and 18 in the evening.

1. In what season does the cafe have the most customers on an average day?
2. In what season does the cafe have the most customers on the average late afternoon?
3. At what time of day does the cafe have the most customers in an average Summer?
4. Find $Q = S - P$. What does Q represent?
5. Find $H = 91W$. What might H represent? [Hint: Assume there are 91 days in Winter.]
6. Find $A = 91(P + S + F + W)$. What might A represent? [Hint: Assume each season is 91 days long.]

Solution:

1. The total number of customers on a day is the sum of the four coordinates of the vector. It is $5+22+35+18 = 80$ during spring, $20+33+14+40 = 107$ in summer, $12+28+25+20 = 85$ in fall and $2 + 19 + 42 + 12 = 75$ in winter. Hence the most customers on an average day arrives in summer.
2. The number of customers in late afternoon is the third coordinate of the vector. This is the largest in winter, when it is 42.
3. The largest coordinate of vector S is the fourth one, hence the most customers in an average summer is in the evening.
4. $Q = S - P = (15, 11, -21, 22)$ yields the differences of the numbers of customers between summer and spring.

5. The vector $H = 91W = (182, 1729, 3822, 1092)$ yields the total number of customers in winter in each part of the day.
6. The vector $A = 91(P + S + F + W)$ yields the total number of customers in the whole year in each part of the day.

3.3.6

In the following problems the matrices refer to Olympic medal counts. The columns are the numbers of gold, silver, and bronze medals, and the rows are the number of medals for Australia, China, Germany, Russia and the United States.

$$E = 1996 \text{ Olympics} = \begin{pmatrix} 9 & 9 & 23 \\ 16 & 22 & 12 \\ 20 & 18 & 27 \\ 26 & 21 & 16 \\ 44 & 32 & 25 \end{pmatrix}$$

$$F = 2000 \text{ Olympics} = \begin{pmatrix} 16 & 25 & 17 \\ 28 & 16 & 15 \\ 13 & 17 & 26 \\ 32 & 28 & 28 \\ 40 & 24 & 33 \end{pmatrix}$$

$$G = 2004 \text{ Olympics} = \begin{pmatrix} 17 & 16 & 16 \\ 32 & 17 & 14 \\ 14 & 16 & 18 \\ 27 & 27 & 38 \\ 35 & 39 & 29 \end{pmatrix}$$

$$H = 2008 \text{ Olympics} = \begin{pmatrix} 14 & 15 & 17 \\ 51 & 21 & 28 \\ 16 & 10 & 15 \\ 23 & 21 & 28 \\ 36 & 38 & 36 \end{pmatrix}$$

1. Calculate $H - G$. What does this represent?
2. Calculate $E + F + G + H$. What does this represent?
3. Calculate $\frac{1}{4}(E + F + G + H)$. What does this represent?

Solution:

1. this many more (or less if negative) medals were received by the selected 5 countries in 2008 than in 1996

2. the total number of medals received during the four Olympics by these 5 countries
3. the average number of medals received based on the four Olympics

3.3.7

A car mechanic sells 4 different types of tires. The unit price and the number of tires in stock are listed in the following table.

	Tire 1	Tire 2	Tire 3	Tire 4
Unit price (Ft/pc)	13000	21000	43000	10000
In stock (pc)	10	5	2	8

Consider vectors v from the first row of the table, and vector w from the second row of the table. What is the value of the product $v \cdot w^T$? What is the interpretation of the product $v \cdot w^T$ in everyday terms?

Solution: The product $v \cdot w^T = 13000 \cdot 10 + 21000 \cdot 5 + 43000 \cdot 2 + 10000 \cdot 8 = 401000$ yields the total prize of tires in stock.

3.3.8

The tables below contain test scores of four students based on two written exams.

Test 1.	Problem 1.	Problem 2.	Problem 3.
Mary	0	4	5
Vicky	4	3	6
John	2	0	6
Mark	5	3	1

Test 2.	Problem 1.	Problem 2.	Problem 3.
Mary	2	2	4
Vicky	2	1	5
John	3	5	4
Mark	0	3	2

Let B be the matrix of the scores on the first test from table, and let C be the matrix of the scores on the second test.

1. What do the following matrices mean in everyday terms?

$$B + C, \quad C - B, \quad B \cdot v, \quad \frac{1}{4}\, u \cdot C,$$

where

$$u = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} ?$$

2. Write down and evaluate algebraic expressions involving the matrices B, C, u, v that represent the following quantities.
   a) The total score on the second exam for each student.
   b) The total score on two tests for each student.
   c) The average score for each problem on the first test.
   d) The average score on the second test.

Solution:

(a)

$$B + C = \begin{pmatrix} 0 & 4 & 5 \\ 4 & 3 & 6 \\ 2 & 0 & 6 \\ 5 & 3 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 2 & 4 \\ 2 & 1 & 5 \\ 3 & 5 & 4 \\ 0 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 6 & 9 \\ 6 & 4 & 11 \\ 5 & 5 & 10 \\ 5 & 6 & 3 \end{pmatrix}$$

the sum of scores for each student and for each exercise.

$$C - B = \begin{pmatrix} 2 & 2 & 4 \\ 2 & 1 & 5 \\ 3 & 5 & 4 \\ 0 & 3 & 2 \end{pmatrix} - \begin{pmatrix} 0 & 4 & 5 \\ 4 & 3 & 6 \\ 2 & 0 & 6 \\ 5 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -2 & -1 \\ -2 & -2 & -1 \\ 1 & 5 & -2 \\ -5 & 0 & 1 \end{pmatrix}$$

the entries yield the difference of scores on the second and first test for each student and for each exercise.

$$B \cdot v = \begin{pmatrix} 0 & 4 & 5 \\ 4 & 3 & 6 \\ 2 & 0 & 6 \\ 5 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 9 \\ 13 \\ 8 \\ 9 \end{pmatrix}$$

the total score of students on the first test.

$$u \cdot C = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 4 \\ 2 & 1 & 5 \\ 3 & 5 & 4 \\ 0 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 7 & 11 & 15 \end{pmatrix}$$

the total score for each exercise on the second test.

(b)

1. The total score on the second exam for each student:

$$C \cdot v = \begin{pmatrix} 2 & 2 & 4 \\ 2 & 1 & 5 \\ 3 & 5 & 4 \\ 0 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 8 \\ 12 \\ 5 \end{pmatrix}$$

.

2. The total score on two tests for each student:

$$(B + C) \cdot v = \begin{pmatrix} 2 & 6 & 9 \\ 6 & 4 & 11 \\ 5 & 5 & 10 \\ 5 & 6 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 17 \\ 21 \\ 20 \\ 14 \end{pmatrix}$$

.

3. The average score for each problem on the first test:

$$\frac{1}{4} u \cdot B = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \begin{pmatrix} 0 & 4 & 5 \\ 4 & 3 & 6 \\ 2 & 0 & 6 \\ 5 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 11/4 & 10/4 & 18/4 \end{pmatrix}$$

.

4. The average score on the second test:

$$\frac{1}{4} u \cdot C \cdot v = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \begin{pmatrix} 2 & 2 & 4 \\ 2 & 1 & 5 \\ 3 & 5 & 4 \\ 0 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 33/4$$

.

4 Linear algebra - Linear systems and determinant

4.1 Motivating example and introduction

Two families with four members have dinner and breakfast in a hotel. One member of the first family does not have breakfast (because of getting up too late) and one member of the second family does not have dinner (because going to bed too early). The first family pays 9000 HUF for the two meals, the other one pays 8500. How much does it cost a breakfast and a dinner for a single person?

This problem can be solved by setting up and solving a system of linear equations. Let us denote by r the price of a breakfast and by v the price of a dinner. Then the information given in the text can be described by the following two equations:

$$\begin{aligned} 3r + 4v &= 9000 \\ 4r + 3v &= 8500 \end{aligned}$$

Let us solve this system of equations. Subtracting the two equations we get $v - r = 500$, that is $v = r + 500$. Substituting this into the first equation leads to $3r + 4r + 2000 = 9000$, yielding $r = 1000$ and then $v = 1000 + 500 = 1500$. That is a breakfast costs 1000 HUF and a dinner costs 1500 HUF. Note that we would not have been able to solve the system if all persons had attended both meals, since the two equations would be identical in that case: $4r + 4v = 10000$.

In this section we will consider systems of linear equations in general. A system of linear equations is a collection of two or more linear equations involving the same set of variables. For example,

$$\begin{aligned} 3x - 4y + z &= 8 \\ x + y - 2z &= 5 \\ -x + 4z &= -1 \end{aligned}$$

is a system of three equations having three variables x, y and z. A solution to a system of linear equations is an assignment of values to the variables such that all the equations are simultaneously satisfied. A solution to the system above is given by

$$x = 5, \ y = 2, \ z = 1,$$

as it can be easily checked by substituting these values for x, y and z. We will show that a system of linear equations can be formulated in terms of matrices and vectors, enabling us to use matrix techniques to solve them. We will see that the solvability of a linear system is related to a special matrix function, the determinant.

4.2 Notions and methods

4.2.1 Elementary methods for solving linear systems

Three ways for solving linear systems as it was taught in high school:

1. *Expression – substitution.* We express one variable from one equation and substitute it to the other equations. This way we get a system of one less equations and one less variable. By repeating this procedure, we end up with a single equation, which we solve and then we substitute its value back into the other equations in order to obtain the remaining unknowns.
2. *Method of equal coefficients.* Multiples of one equation are added to the other equations; exactly as many times as it is needed for a selected variable to fall out of all the equations. This way we get a system of one less equation and one less variable. By repeating this procedure, we end up with a single equation, which we solve and then we substitute its value back into the other equations in order to obtain the remaining unknowns.
3. *Graphical solution.* If there are two variables and two equations, then each linear equation determines a straight line in the plane. Graph both equations in the same coordinate system. The solution set is the intersection of these lines (and is hence either a line, a single point, or the empty set). Geometrically, a linear equation in x, y and z is the equation of a plane. Solving a system of linear equations is equivalent to finding the intersection of the corresponding planes. (See Wikipedia's "System of linear equation" article.)

The above methods show and it can be generally proved that a system of linear equations can have a unique solution, no solution, or infinitely many solutions. (It cannot have, for example, exactly two or five solutions.)

4.2.2 Matrix form of a linear system

A system of linear equations can be represented by a matrix, where each row represents one equation in the system and each column represents a variable or the constant terms. For example, the system of linear equations

$$\begin{aligned} 3x_1 - 4x_2 + x_3 &= 8 \\ x_1 + x_2 - 2x_3 &= 5 \\ -x_1 + 4x_3 &= -1 \end{aligned}$$

can be written as

$$\begin{pmatrix} 3 & -4 & 1 \\ 1 & 1 & -2 \\ -1 & 0 & 4 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 5 \\ -1 \end{pmatrix}.$$

Therefore the matrix form for the system of linear equations is

$$A \cdot x = b,$$

where

- x is the column vector containing the unknowns,
- b is the "right hand side" of the system, in a column vector,
- A is the matrix of the coefficients of the unknowns, where rows correspond to equations, columns to variables.

We note that it was more convenient to denote the variables by x_1, x_2 and x_3 instead of x, y and z, because we use x for denoting the vector containing all the unknowns.

4.2.3 Solving a linear system by using the inverse matrix

There are numerous advantages of rephrasing the system of linear equations to matrix form, the most important being a change of viewpoint. We have now only one equation, although with weird new objects: matrices. We gain new insights from this approach to solve these equations. One of these new methods is the use of the inverse matrix for "square" systems.

In case of ordinary numbers the solution of $ax = b$ can be found after multiplying both sides by the reciprocal of a:

$$x = \frac{1}{a}b$$

This relies on the facts that $\frac{1}{a} \cdot a = 1$ for any non-zero a, and $1 \cdot x = x$ for any x.

If A^{-1} is known, then

$$A \cdot x = b$$

can be solved as follows. Multiply both sides from the left by A^{-1}. (It is important to do this from the left.)

$$A^{-1} \cdot Ax = A^{-1} \cdot b$$

Using that

$$A^{-1}(Ax) = (A^{-1}A)x = I_n x = x,$$

we have our solution

$$x = A^{-1}b.$$

This is in full analogy with the ordinary 1-variable equation case.

As an example, let us solve the system given in the Introduction

$$\begin{aligned} 3r + 4v &= 9000 \\ 4r + 3v &= 8500. \end{aligned}$$

The matrix is

$$A = \begin{pmatrix} 3 & 4 \\ 4 & 3 \end{pmatrix}$$

its determinant is $\det(A) = 9 - 16 = -7 \neq 0$, so the system has a unique solution. Using the determinant the inverse matrix can be easily given as

$$A^{-1} = -\frac{1}{7} \begin{pmatrix} 3 & -4 \\ -4 & 3 \end{pmatrix}$$

hence the solution is

$$\begin{pmatrix} r \\ v \end{pmatrix} = A^{-1} \cdot b = -\frac{1}{7} \begin{pmatrix} 3 & -4 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} 9000 \\ 8500 \end{pmatrix} = \begin{pmatrix} 1000 \\ 1500 \end{pmatrix}.$$

We note that computing the inverse of a matrix is difficult and takes a long time even with fast computers when the matrix is large. (In real life, systems with a million equations and unknowns often occur.) Therefore more effective algorithms for solving linear systems have been developed. The most important one is the so-called Gaussian elimination (which is also used for calculating an inverse matrix). The fundamental idea is to add multiples of one equation to the others in order to eliminate a variable and to continue this process until only one variable is left. Once this final variable is determined, its value is substituted back into the other equations in order to evaluate the remaining unknowns. This method, characterized by step by step elimination of the variables, is called Gaussian elimination. (During Gaussian elimination, the system is converted to so-called upper triangular form, that is below the main diagonal all elements will be zero.)

4.2.4 Determinant

As a motivation let us return to our introductory example with the two families having dinner and breakfast in the hotel. Our system of equations was

$$\begin{aligned} 3r + 4v &= 9000 \\ 4r + 3v &= 8500 \end{aligned}$$

As an other case, consider a family of three and a family of four and assume that everybody has breakfast and dinner as well. Then the system is modified as

$$\begin{aligned} 3r + 3v &= 7500 \\ 4r + 4v &= 10000 \end{aligned}$$

We can easily see that this system does not determine the values of r and v uniquely. We can only say that $r + v = 2500$, but there are infinitely many ways of choosing a value for r and for v, i.e. besides the original solution $r = 1000$, $v = 1500$ there are many alternatives, e.g. $r = 1100$,

$v = 1400$. How could we see from the matrices of the two systems above that the first system can be solved and the second cannot? The matrices are

$$A = \begin{pmatrix} 3 & 4 \\ 4 & 3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 3 & 3 \\ 4 & 4 \end{pmatrix}.$$

We can easily realize that the problem with the matrix B is that the two rows are constant multiples of each other. The second row is obtained by multiplying the first one by $4/3$. Alternatively, this can be reformulated as $b_{11}b_{22} = b_{12}b_{21}$, i.e. $3 \cdot 4 = 3 \cdot 4$ or $b_{11}b_{22} - b_{12}b_{21} = 0$. In the case of matrix A we have $a_{11}a_{22} - a_{12}a_{21} = 3 \cdot 3 - 4 \cdot 4 = -7 \neq 0$. We will see soon that this quantity determines the solvability of the linear system, hence it is called the determinant of the 2×2 matrix. The notation is:

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

The determinant of a matrix A can be alternatively denoted by $|A|$. If $\det(A) \neq 0$ then the system has a unique solution, and in the case $\det(A) = 0$ there is either no solution or there are infinitely many solutions. This can be immediately seen from the formula of the inverse matrix

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

If $\det(A) = 0$, then we have division by zero in the formula, hence the inverse does not exist.

Let us turn now to larger matrices. The determinant is a value associated with a square matrix, and can be computed from the elements of the matrix. Determinants have many uses, we list some important ones. They can be used to solve a system of n linear equations in n unknowns (this is called Cramer's rule). But they are also useful as they determine whether or not a matrix can be inverted. Later we will see that determinants are important in calculating eigenvalues and eigenvectors. Determinants are also a measure of the area (or volume) of the shape defined by the columns of the matrix (treated as vectors).

We explain here the computation of the determinant of a 3×3 matrix only. (Note that inventing this formula would take us a long time since many 3×3 matrices should have been studied to find the condition when the inverse exists.) For a 3×3 matrix A, its determinant is defined as

$$\det(A) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi.$$

The rules for computing the size 2 and 3 determinants can easily be memorized as follows:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$$

and

$$\left| \begin{matrix} a & b & c \\ d & e & f \\ g & h & i \end{matrix} \right| \begin{matrix} a & b \\ d & e \\ g & h \end{matrix} = aei + bfg + cdh - ceg - afh - bdi.$$

In the 3×3 case, write out the first 2 columns of the matrix to the right of the 3rd column, so that you have 5 columns in a row. Then add the products of the diagonals going from top to bottom (solid) and subtract the products of the diagonals going from bottom to top (dashed). The name of this procedure is called the Sarrus' rule.

An alternative way of memorizing the 3×3 case is the expansion rule:

$$\begin{aligned} \det(A) &= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg) \\ &= aei + bfg + cdh - ceg - afh - bdi. \end{aligned}$$

Here the entries of the first row are multiplied by the corresponding sub-determinants obtained by omitting the row and coloumn of the given entry. Then this products are summed up with alternating sign. We note that we would get the same result with using any other row, or even any coloumn of the matrix.

Let us compute the following examples.

$$det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1 \cdot 4 - 2 \cdot 3 = -2$$

$$det \begin{pmatrix} -1 & 0 & -1 \\ 1 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix} = \left| \begin{array}{ccc|cc} -1 & 0 & -1 & -1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 3 & 2 \end{array} \right. =$$

$$= (-1) \cdot 1 \cdot 1 + 0 \cdot 1 \cdot 3 + (-1) \cdot 1 \cdot 2 - (-1) \cdot 1 \cdot 3 - (-1) \cdot 1 \cdot 2 - 0 \cdot 1 \cdot 1 = 2$$

4.3 Exercises

4.3.1

Calculate the determinants of the following matrices.

1. $\begin{pmatrix} 1 & 5 \\ 1 & 1 \end{pmatrix}$
2. $\begin{pmatrix} 0 & -5 \\ 4 & 1 \end{pmatrix}$
3. $\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$
4. $\begin{pmatrix} 1 & 2 & 0 \\ -2 & 3 & 4 \\ 0 & -1 & 5 \end{pmatrix}$

5. $\begin{pmatrix} 1 & 2 & 0 \\ -2 & 3 & 0 \\ 0 & -1 & 0 \end{pmatrix}$

Solution:

1. $$\begin{vmatrix} 1 & 5 \\ 1 & 1 \end{vmatrix} = 1 \cdot 1 - 1 \cdot 5 = -4$$

2. $$\begin{vmatrix} 0 & -5 \\ 4 & 1 \end{vmatrix} = 0 \cdot 1 - 4 \cdot (-5) = 20$$

3. $$\begin{vmatrix} 1 & 2 \\ 3 & 6 \end{vmatrix} = 1 \cdot 6 - 3 \cdot 2 = 0$$

4. $$\begin{vmatrix} 1 & 2 & 0 \\ -2 & 3 & 4 \\ 0 & -1 & 5 \end{vmatrix} = 1 \cdot (3 \cdot 5 - 4 \cdot (-1)) - 2 \cdot ((-2) \cdot 5 - 0 \cdot 4) + 0 \cdot ((-2) \cdot (-1) - 3 \cdot 0) = 39$$

5. $$\begin{vmatrix} 1 & 2 & 0 \\ -2 & 3 & 0 \\ 0 & -1 & 0 \end{vmatrix} = 0 \cdot ((-2) \cdot (-1) - 0 \cdot 3) - 0 \cdot ((-1) \cdot 1 - 0 \cdot 2) + 0 \cdot (1 \cdot 3 - 2 \cdot (-2)) = 0$$

4.3.2

Rewrite the equations in matrix form, but solve the equations using methods learned in high school.

1. $\left.\begin{array}{rcr} -2x + 3y & = & 8 \\ x - y & = & -1 \end{array}\right\}$

2. $\left.\begin{array}{rcr} 3x - 4y + z & = & 8 \\ x + y - 2z & = & -1 \end{array}\right\}$

3. $\left.\begin{array}{rcr} -2x + 3y & = & 8 \\ x - y & = & -1 \\ -x + 2y & = & 7 \end{array}\right\}$

Solution: Matrix form: $Ax = b$.

1. $A = \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}, b = \begin{pmatrix} 8 \\ -1 \end{pmatrix}$

   Solution: $x = 5$, $y = 6$.

2. $A = \begin{pmatrix} 3 & -4 & 1 \\ 1 & 1 & -2 \end{pmatrix}$, $b = \begin{pmatrix} 8 \\ -1 \end{pmatrix}$

   Solution: $x = z + 4/7$, $y = z - 11/7$, z can be an arbitrary number.

3. $A = \begin{pmatrix} -2 & 3 \\ 1 & -1 \\ -1 & 2 \end{pmatrix}$, $b = \begin{pmatrix} 8 \\ -1 \\ 7 \end{pmatrix}$

   Solution: $x = 5$, $y = 6$.

4.3.3

Write the matrix equation $Ax = b$ as a system of linear equations.

1. $A = \begin{pmatrix} -2 & 3 \\ 4 & -6 \end{pmatrix}$, $b = \begin{pmatrix} 8 \\ -16 \end{pmatrix}$
2. $A = \begin{pmatrix} -2 & 3 \\ 4 & -6 \end{pmatrix}$, $b = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$
3. $A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 0 \\ -1 & 1 & 4 \end{pmatrix}$, $b = \begin{pmatrix} 5 \\ 4 \\ 10 \end{pmatrix}$
4. $A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 0 \\ -1 & 1 & 4 \end{pmatrix}$, $b = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$

Solution:

1. $\left.\begin{array}{rcr} -2x + 3y & = & 8 \\ 4x - 6y & = & -16 \end{array}\right\}$
2. $\left.\begin{array}{rcr} -2x + 3y & = & 4 \\ 4x - 6y & = & 3 \end{array}\right\}$
3. $\left.\begin{array}{rcr} x + 2z & = & 5 \\ 3x - y & = & 4 \\ -x + y + 4z & = & 10 \end{array}\right\}$
4. $\left.\begin{array}{rcr} x + 2z & = & 2 \\ 3x - y & = & 1 \\ -x + y + 4z & = & 3 \end{array}\right\}$

4.3.4

Check that the given matrices are proper inverses (i.e., they satisfy $A^{-1}A = I$).

1. Solve the system of equations using the inverse

$$\left.\begin{array}{rcl} x+2y &=& 5 \\ 3x+4y &=& 9 \end{array}\right\} \qquad A^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$$

2. Solve $Ax = b$, where

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \ b = \begin{pmatrix} -1 \\ 11 \end{pmatrix} \text{ and } A^{-1} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

3. Solve the equation

$$Ax = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 0 \\ -1 & -2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ -6 \\ 3 \end{pmatrix} \text{ using } A^{-1} = \begin{pmatrix} \frac{2}{9} & \frac{2}{9} & -\frac{1}{9} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ \frac{7}{18} & -\frac{1}{9} & \frac{1}{18} \end{pmatrix}$$

Solution:

1.
$$A^{-1}A = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

Solution: $\begin{pmatrix} x \\ y \end{pmatrix} = A^{-1}b = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 5 \\ 9 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$, that is $x = -1$, $y = 3$.

2.
$$A^{-1}A = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

Solution: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A^{-1}b = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} -1 \\ 11 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$, that is $x_1 = 3$, $x_2 = 4$.

3.
$$A^{-1}A = \begin{pmatrix} \frac{2}{9} & \frac{2}{9} & -\frac{1}{9} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ \frac{7}{18} & -\frac{1}{9} & \frac{1}{18} \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 0 \\ -1 & -2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I$$

Solution: $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = A^{-1}b = \begin{pmatrix} \frac{2}{9} & \frac{2}{9} & -\frac{1}{9} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ \frac{7}{18} & -\frac{1}{9} & \frac{1}{18} \end{pmatrix} \begin{pmatrix} 3 \\ -6 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix}$, that is $x_1 = -1$, $x_2 = 3$, $x_3 = 2$.

4.3.5

Write the following equations in matrix form. Solve them with any of the methods. (For practice, try to use as many methods as you can.)

1. $\left.\begin{array}{rcl} x+2y &=& 5 \\ 3x+4y &=& 9 \end{array}\right\}$

2. $$\left.\begin{array}{rcr} 3x - 4y + z & = & 8 \\ x + y - 2z & = & -1 \end{array}\right\}$$

4.3.6

Write the matrix equation $A \cdot x = b$ as a system of linear equations, and solve them.

1. $A = \begin{pmatrix} -2 & 3 \\ 4 & -6 \end{pmatrix}, b = \begin{pmatrix} 8 \\ -16 \end{pmatrix}$
2. $A = \begin{pmatrix} -2 & 3 \\ 4 & -6 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$
3. $A = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}, b = \begin{pmatrix} -1 \\ 11 \end{pmatrix}$

Solution:

1. $x_1 = \frac{3}{2}x_2 - 4$, x_2 can be an arbitrary number.
2. There is no solution.
3. $x_1 = 3$, $x_2 = 4$.

4.3.7 Software

Excel

WolframAlpha: *https://www.wolframalpha.com/*

Type the following and check the output:

1. $\{\{1,2\},\{3,4,5\}\}$ˆT
2. $\{\{1,2,3\},\{4,5,6\}\} + \{\{7,8,9\},\{10,11,12\}\}$
3. $\{\{1,2\},\{3,4\}\}\{\{5,6\},\{7,8\}\}$
4. inverse$\{\{-2,3\},\{4,6\}\}$
5. det$\{\{1,2\},\{3,4\}\}$
6. plot $8/3 + 2/3\,x$ and $x + 1$
7. $x - y = -1, x + 2y = 11$

5 Linear programming - Graphical approach

Linear programming (in the sequel LP) is a technique solving optimization problems. We show the main ideas and definition only for the two-variable case using graphical approach through a detailed example.

5.1 Motivating example - Giapetto's woodcarving company

Giapetto's woodcarving company manufactures two types of wooden toys: soldiers and trains. A soldier sells for \$27 and uses \$10 worth of raw materials. Each soldier that is manufactured increases Giapetto's variable labor and overhead costs by \$14. A train sells for \$21 and uses \$9 worth of raw materials. Each train built increases Giapetto's variable labor and overhead costs by \$10. The manufacture of wooden soldiers and trains requires two types of skilled labor: carpentry and finishing. A soldier requires 2 hours of finishing labor and 1 hour of carpentry labor. A train requires 1 hour of finishing and 1 hour of carpentry labor. Each week, Giapetto can obtain all the needed raw material but only 100 finishing hours and 80 carpentry hours. Demand for trains is unlimited, but at most 40 soldiers are bought each week. Give an advice to Giapetto how many soldiers and trains should be manufactured in order to maximize the weekly profit.

5.1.1 Designing LP problem

As a first step we would like to determine the LP problem and later on solve it. Giapetto has to decide how many soldiers and trains should be manufactured. Therefore, let us introduce the so-called ***decision variables***, i.e.

$$x_1 := \text{number of soldiers produced each week}$$

$$x_2 := \text{number of trains produced each week}$$

Giapetto's aim is maximizing the ***profit***. Mathematically it means that we have to optimize an objective function. Determining the ***objective function*** we optimize certain functions of decision variables. These mean typically maximizing the profit or minimizing the cost.

What is the profit? It has two components: the ***revenue*** and the ***cost***.

i, Revenue: Based on the problem's text we know that we can sell a soldier for \$27 and a train for \$21. So the weekly revenues using the decision variable leads to

$$27x_1 + 21x_2.$$

ii, Cost: In this example it has two components: raw material cost and manufacturing costs.

A soldier uses \$10 worth of raw materials and a train \$9 worth of raw materials. So the weekly raw material costs can be described by

$$10x_1 + 9x_2.$$

Similarly for the manufacturing cost we have

$$14x_1 + 10x_2.$$

So the weekly profit will be the difference between the weekly revenues and the weekly costs, i.e.

$$27x_1 + 21x_2 - [(10x_1 + 9x_2) + (14x_1 + 10x_2)] = 3x_1 + 2x_2.$$

Giapetto wants to maximize the above expression for the decision variables x_1 and x_2. Therefore the objective function in this example is

$$\max_{x_1,x_2} 3x_1 + 2x_2.$$

Now take into account that there are conditions (we will call them constraints) which influence the maximal profit. Based on the text there are three ***constraints***:

i, Each week at most 100 hours of finishing time may be used: 2 hours for a soldier and 1 hour for a train. It leads to

$$2x_1 + x_2 \leq 100.$$

ii, Each week at most 80 hours of carpentry time may be used: 1 hour for a soldier and 1 hour for a train. It leads to

$$x_1 + x_2 \leq 80.$$

iii, At most 40 soldiers should be manufactured each week:

$$x_1 \leq 40.$$

The above three constraints' right-hand side represent the quantity of a resource that is available. The coefficients of the decision variables 3 in the constraints are the so-called *technological coefficients* since they reflect the technology used to produce different products.

One can ask that whether the phrase "demand for trains is unlimited" means an actual constraint or not for the decision variable x_2. Since we have unlimited resources it doesn't mean an actual constraint for x_2.

There are more special kind of constraints. Namely Giapetto has to manufacture at least 0 soldier or train. In case of zero he won't manufacture the actual wooden toy. For positive value he is going to make the toys. We can express these by using the following inequalities:

$$x_1 \geq 0 \text{ and } x_2 \geq 0.$$

These are the so-called ***non-negativity constraints***. In other example for instance the decision variable x_1 can be negative (e.g. a firm owed more money than it had on hand). So, in general we have ***sign restricted*** and ***unrestricted in sign*** cases.

Now we are ready to give the LP problem corresponding to Giapetto's woodcarving company. As we have seen it has three big components:

i, Objective function

ii, Constraints

iii, Sign restriction/no restriction in sign

Here the LP problem is

$$\begin{aligned} \max_{x_1,x_2} \; & 3x_1+2x_2 \\ 2x_1 + x_2 &\leq 100 \\ x_1 + x_2 &\leq 80 \\ x_1 &\leq 40 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned}$$

Our aim is to solve this two-variable problem using the graphical approach. In order to do that, first we have to get acquainted with the relevant definitions and theoretical results.

5.2 Theoretical results

A function $f(x_1, x_2)$ of variables x_1 and x_2 is a **linear function** if and only if $f(x_1, x_2) = c_1x_1 + c_2x_2$ for some constants c_1 and c_2.

For any linear function $f(x_1, x_2)$ and arbitrary number b, the inequalities $f(x_1, x_2) \leq b$ and $f(x_1, x_2) \geq b$ are **linear inequalities**.

An **LP problems** is an optimization problem where

- We attempt to maximize (or minimize) a linear function of the decision variables.
- The values of the decision variables must satisfy a set of constraints. Each constraint must be a linear equation or linear inequality.
- A sign restriction (or no restriction in sign) is associated with each variable.

In the example of Giapetto's woodcarving company we have seen the above definitions are satisfied. Let's take a further step and define the feasible region and the optimal solution.

The **feasible region** for an LP is the set of all points that satisfies all the LP's constraints and sign restrictions. Any point that is not in an LP's feasible region is said to be an **infeasible point**.

For a maximization problem, an **optimal solution** to an LP is a point in the feasible region with the largest objective function value. Similarly, for a minimization problem, an optimal solution is a point in the feasible region with the smallest objective function value.

The following definitions play important roles related to theoretical results.

A set of points S is a **convex set** if the line segment joining any pair of points in S is wholly contained in S.

For any convex set S, a point P in S is an **extreme point** if each line segment that lies completely in S and contains the point P has P as an endpoint of the line segment.

We will understand the notions of convex set and extreme points through Giapetto's example and other problems.

The following theorem helps us determine the feasible region and the optimal solution.

Theorem 5.1: (Graphical approach)

i, The feasible region of an LP problem is convex.

ii, In a feasible region of an LP problem the number of extreme points is finite.

iii, For those LP problems which have optimal solution consist an extreme point which is optimal.

iv, One of the extreme points of feasible region is optimal.

Since it out of the scope of this course we don't deal with minimization problems, the cases of alternative optimal solutions, infeasible or unbounded LP problems.

If you are an eager Student and you are interested in more sophisticated techniques, general cases and application we highly recommend the book *Wayne L. Winston: Operations Research: Applications and Algorithms.*

5.3 Motivating example - continued and solution

As we have seen at the end of Section 5.1.1. the LP problem for Giapetto's woodcarving company is

$$\max_{x_1,x_2} 3x_1+2x_2$$
$$\begin{aligned} 2x_1 + x_2 &\leq 100 \\ x_1 + x_2 &\leq 80 \\ x_1 &\leq 40 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned}$$

As a first step we would like to plot the constraints and the non-negativity constraints. Using our secondary school knowledge it means making 5 figures.

First, we plot the 3 constraints keeping the order from the LP problem.

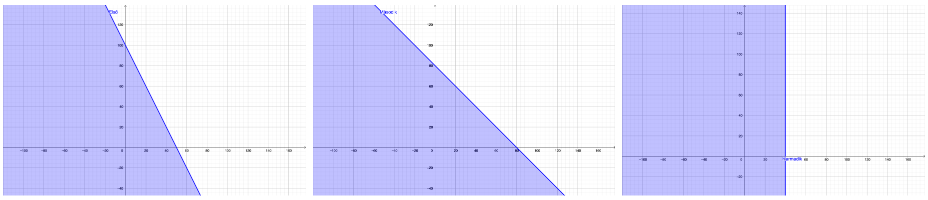


Figure 5.1: Left: $2x_1 + x_2 \leq 100$, Center: $x_1 + x_2 \leq 80$, Right: $x_1 \leq 40$

Second, we plot the 2 non-negativity constraints keeping the order from the LP problem.

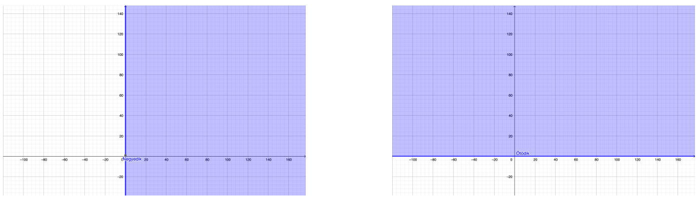


Figure 5.2: Left: $x_1 \geq 0$, Right: $x_2 \geq 0$

Satisfying all of the constraints means that we have to take the intersection of the above plotted figures. The intersection of Figures 5.1. and 5.2. consists all of the feasible solutions, i.e. we determine the feasible region.

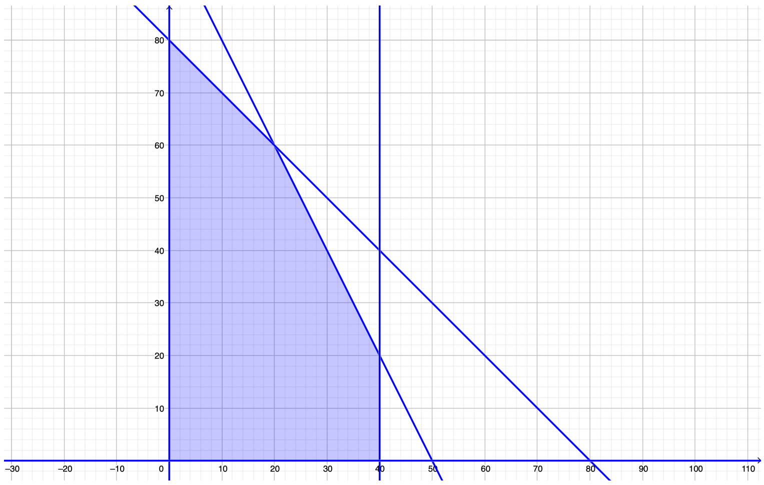


Figure 5.3: The feasible region

It is easy to see that the blue set, i.e. this polygon is a convex set. The extreme points are the vertices of this polygon. Taking into account Theorem 5.1 it is enough to evaluate the vertices in the objective function and this leads to the optimal solution.

The vertices of the polygon and the corresponding objective function values are

i, Vertex $(0, 0)$: The objective function value is $3 \cdot 0 + 2 \cdot 0 = 0$

ii, Vertex $(40, 0)$: The objective function value is $3 \cdot 40 + 2 \cdot 0 = 120$

iii, Vertex $(40, 20)$: The objective function value is $3 \cdot 40 + 2 \cdot 20 = 160$

iv, Vertex $(20, 60)$: The objective function value is $3 \cdot 20 + 2 \cdot 60 = 180$

v, Vertex $(0, 80)$: The objective function value is $3 \cdot 0 + 2 \cdot 80 = 160$

The biggest objective function value correspond to the vertex $(20, 60)$. It means that Giapetto's woodcarving company will make the highest \$180 profit if they sell 20 soldiers and 60 trains.

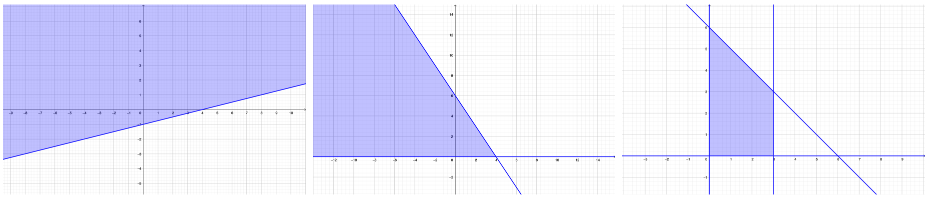
5.4 Exercises

5.4.1

Sketch the regions on the x-y plane that are given by the following inequalities:

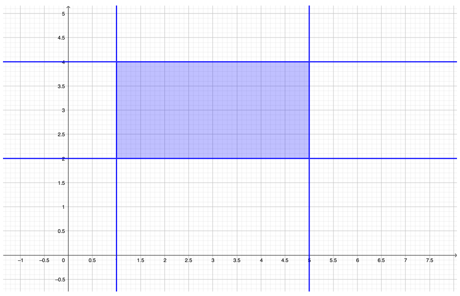
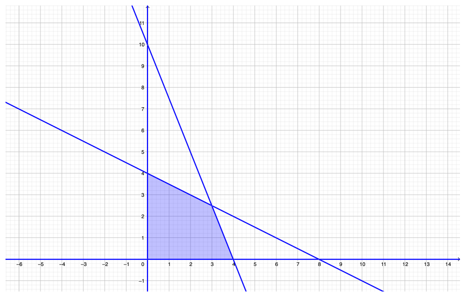
(a) $x - 4y \leq 4$ (b) $3x + 2y \leq 12,\ y \geq 0$ (c) $x + y \leq 6,\ 0 \leq x \leq 3,\ y \geq 0$

Solution:



5.4.2

Write down linear inequalities that describe the following regions

(a)  (b) 

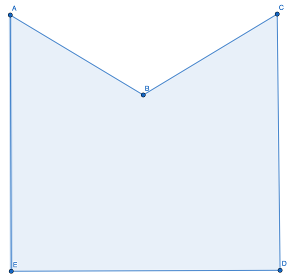
Solution:

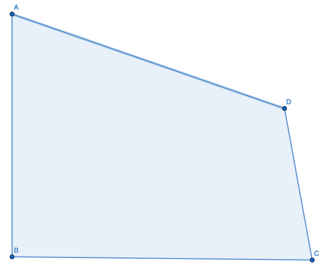
(a) $1 \leq x \leq 5,\ 2 \leq y \leq 4$

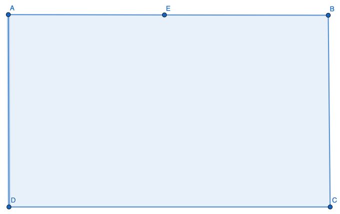
(b) $x \geq 0,\ y \geq 0,\ x + 2y \leq 8,\ 5x + 2y \leq 20$

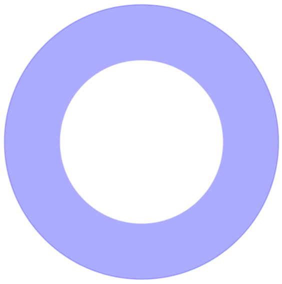
5.4.3

Determine whether the following regions are convex regions or not and determine the extreme points, too.

(a) 

(b) 

(c) 

(d) 

Solution:

(a) Not convex

(b) Convex; extreme point: the vertices

(c) Convex; extreme point: the vertices except point E

(d) Not convex

5.4.4

Solve the LP problem where the objective function $2x+y$ is to be maximized subject to constraints in Exercise 5.4.2.

Solution:

(a) $x = 5$ and $y = 4$. The objective function value is 14.

(b) $x = 3$ and $y = 2.5$. The objective function value is 8.5.

5.4.5

A small tailoring company manufactures costumes and suits. On each costume and suite they have \$4 and \$3 profit, respectively. Each costume and suit requires 1 hour tailoring. A tailored costume and a tailored suits can be sewed in 2 hours and 1 hour, respectively. Based on the weekly business plan the company has at most 40 hours for tailoring and 60 hours for sewing. They have unlimited materials. Give an advice to the tailoring company how many costumes and suits should be manufactured maximizing the profit.

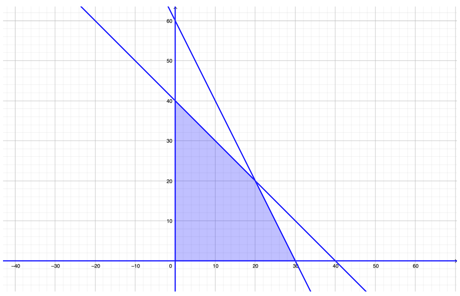
Solution:

x_1 := number of costumes should be manufactured

x_2 := number of suits should be manufactured

The corresponding LP problem is

$$\max_{x_1,x_2} 4x_1+3x_2$$
$$x_1 + x_2 \leq 40$$
$$2x_1 + x_2 \leq 60$$
$$x_1 \geq 0$$
$$x_2 \geq 0$$



The highest profit can be made in case of 20 costumes and 20 suit which is $4 \cdot 20 + 3 \cdot 20 = \$140$.

5.4.6

A furniture company manufactures tables made by solid wood and pressed wood. The profits are \$1 and \$2 for solid and pressed woods, respectively. The company expectation in terms of profit is at least \$6. A pressed wood table requires 1kg sawdust. Producing a solid wood table gives 1kg sawdust and the company has at most 3kg sawdust in store. Both table types requires 1 tube of glue and the company has at most 10 tubes of glue. According to the sales department every produced solid wood table increases the popularity of the company by 2 points. However, every produced pressed wood table decreases the popularity of the company by 3 points. Give an advice to company how many solid wood table and pressed wood table should be manufactured maximizing the popularity.

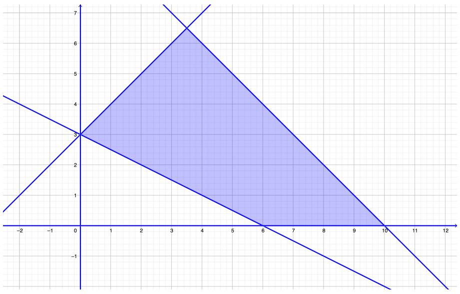
Solution:

x_1 := number of solid wood table should be manufactured

x_2 := number of pressed wood table should be manufactured

The corresponding LP problem is

$$\max_{x_1,x_2} 2x_1 - 3x_2$$
$$x_1 + 2x_2 \geq 6$$
$$-x_1 + x_2 \leq 3$$
$$x_1 + x_2 \leq 10$$
$$x_1 \geq 0$$
$$x_2 \geq 0$$



The highest popularity can be achieved if the company manufactures 10 solid wood tables and 0 pressed wood table. The the company popularity will be increased by $2 \cdot 10 - 3 \cdot 0 = 20$ points.

5.4.7

Eszter was looking for a part-time job and she received offers from two firms. She can do part-time at these places and decides to split her work at these firms. The following table shows how many hours should be spend with her colleagues and in front a computer in a working hour. The last column shows her capacity during a day.

	Firm 1	Firm 2	Eszter's capacity (hour/day)
Computer (hour/working hour)	0.2	0.5	3
Colleagues (hour/working hour)	0.2	0.1	1.4

(a) How should she split her working hours between the two firms if both of the firms are paying 2 thousand HUF per hour?

(b) How should she split her working hours between the two firms if Firm 1 pays 1 thousand HUF for an hour and Firm 2 pays 3 thousand HUF per hour?

(c) How should she split her working hours between the two firms if Firm 1 pays 3 thousand HUF for an hour and Firm 2 pays 1 thousand HUF per hour?

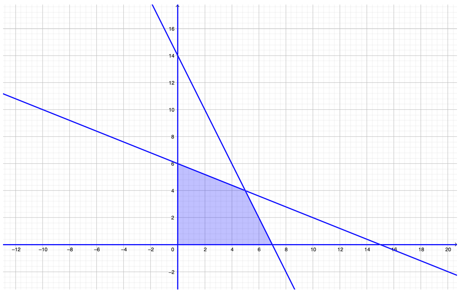
Solution:

$x_1 :=$ number of hours she should spend at Firm 1

$x_2 :=$ number of hours she should spend at Firm 2

The corresponding LP problem except from the objective function is

$$\begin{aligned} 0.2x_1 + 0.5x_2 &\leq 3 \\ 0.2x_1 + 0.1x_2 &\leq 1.4 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned}$$



The objective functions and the corresponding answers:

(a) The objective function is

$$\max_{x_1,x_2} 2000x_1 + 2000x_2$$

In this case Eszter should spend 5 hours at Firm 1 and 4 hours at Firm 2.

(b) The objective function is

$$\max_{x_1,x_2} 1000x_1 + 3000x_2$$

In this case Eszter should spend 0 hour at Firm 1 and 6 hours at Firm 2.

(c) The objective function is

$$\max_{x_1,x_2} 3000x_1 + 1000x_2$$

In this case Eszter should spend 7 hours at Firm 1 and 0 hour at Firm 2.

6 Linear algebra - Linear independence, eigenvalues, eigenvectors

In this section we deal with notions of linear independence, eigenvalues and eigenvectors.

6.1 Linear Independence

The vectors $v_1, \ldots, v_n$ are **linearly independent** if the equation

$$c_1 v_1 + \cdots + c_n v_n = 0$$

implies that

$$c_1 = 0, \ldots, c_n = 0. \quad (\textbf{trivial solution})$$

If the vectors are not linearly independent, they are linearly dependent.

The vectors $v_1, \ldots, v_n$ are **linearly dependent** if there exists a set of numbers $c_1, \ldots, c_n$ not all zero such that $c_1 v_1 + \cdots + c_n v_n = 0$.

In other words, the vectors $v_1, \ldots, v_n$ are linearly dependent if and only if at least one of them is a linear combination of the others, i.e., if at least one vector is equal to the sum of scalar multiples of the other vectors.

Example

$v_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$ are linearly dependent, since $2v_1 + (-1)v_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$v_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$ are linearly dependent, since there exists a set of constants $c_1 = 2,\ c_2 = -1$ and $c_3 = 1$ such that $c_1 v_1 + c_2 v_2 + c_3 v_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

Notice that the last equation can also be written in the form of

$$\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

How does one go about determining whether a set of vectors is linearly independent or linearly dependent?

The answer is rather simple, we can use the fact that a linear combination equation can be transformed into an equivalent matrix equation as seen in the example above. (Note: system of linear equations where the right hand side is a null vector, called homogeneous system of linear equations.)

Theorem 6.1: (Homogeneous linear system)

A set of n vectors of length n is linearly independent if the matrix with these vectors as columns, has a non-zero determinant. The set is, of course, dependent if the determinant is zero.

Example (continued) In the examples above, the determinants of the matrices with the vectors as columns are

$$\det\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} = \begin{vmatrix} 1 & 2 \\ 3 & 6 \end{vmatrix} = 1 \cdot 6 - 2 \cdot 3 = 0 \text{ and } \det\begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix} = \begin{vmatrix} 1 & 2 & 0 \\ -1 & 0 & 2 \\ 0 & 1 & 1 \end{vmatrix} = 0.$$

Linearly independent vectors are, for example, $v_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$, because $\det\begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix} = 1 \cdot 5 - 2 \cdot 3 = -1 \neq 0$.

6.2 Eigenvectors and Eigenvalues

Applications of mathematics sometimes encounter questions for example about the behavior of a sequence of vectors $Av, A^2v, \ldots$ (where A is a square matrix and v is a vector), where we need higher powers of a matrix, e.g., A^{100}. It turns out that finding A^{100} is easier by using the so-called ***eigenvalues*** of A, and not by multiplying 100 matrices. Eigenvalues and eigenvectors show a new insight into the matrices and are powerful enough to help solve intractable problems.

Almost all vectors change direction, when they are multiplied by a square matrix A.

The vectors v in the same direction as Av are called **eigenvectors**.

The basic equation is

$$Av = \lambda v.$$

The number λ in the basic equation is an **eigenvalue** of A.

The eigenvalue tells whether the special vector v is stretched or shrunk or its direction is reversed or left unchanged when it is multiplied by A. The goal is to find scalars λ for which there exists a nonzero vector v such that $Av = \lambda v$. The following are equivalent:

$$Av = \lambda v \Longleftrightarrow Av = \lambda I v \Longleftrightarrow (A - \lambda I)v = 0.$$

The theorem of homogeneous linear system says that $(A - \lambda I)v = \underline{0}$ has nontrivial ($v \neq \underline{0}$) solutions if and only if the determinant of $A - \lambda I$ is 0, i.e., λ is the eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

The first step is solving the $\det(A - \lambda I) = 0$ equation and finding all eigenvalues $\lambda_1, \ldots \lambda_n$. Evaluating this determinant yields an nth degree polynomial in λ. Since every algebraic equation of degree n has at most n roots, every $n \times n$ matrix has at most n eigenvalues.

Next, substituting each eigenvalue λ_i in the system of equations $(A - \lambda I)v = 0$, and solving it, we find the eigenvectors corresponding to the given eigenvalue λ_i.

Notice that the system of equations does not have a unique solution: if $v \neq 0$ is a solution, then any constant multiple of v is also a solution, so eigenvectors can be determined only to within a constant factor.

Example

Determine the eigenvalues and eigenvectors of $A = \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix}$.

i, Eigenvalues

$\det(A-\lambda I) = \begin{vmatrix} 1-\lambda & 2 \\ 3 & -4-\lambda \end{vmatrix} = (1-\lambda)(-4-\lambda) - 2\cdot 3 = \lambda^2 + 3\lambda - 10 = 0 \quad \Longrightarrow \quad \lambda_1 = 2$ and $\lambda_2 = -5$.

ii, Eigenvectors

Eigenvectors belonging to eigenvalue $\lambda_1 = 2$:

$$Av = \lambda_1 v \Longleftrightarrow \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 2\cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \Longleftrightarrow \left.\begin{aligned} v_1 + 2v_2 &= 2v_1 \\ 3v_1 - 4v_2 &= 2v_2 \end{aligned}\right\} \Longleftrightarrow \left.\begin{aligned} -v_1 &= -2v_2 \\ 3v_1 &= 6v_2 \end{aligned}\right\}.$$

This system of equations is equivalent to the equation $v_1 = 2v_2$, so the eigenvectors are

$$v = \begin{pmatrix} 2v_2 \\ v_2 \end{pmatrix}, \ v_2 \neq 0, \ \text{e.g.,} \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Eigenvectors belonging to eigenvalue $\lambda_2 = -5$:

$$Av = \lambda_1 v \Longleftrightarrow \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = -5\cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \Longleftrightarrow \left.\begin{aligned} v_1 + 2v_2 &= -5v_1 \\ 3v_1 - 4v_2 &= -5v_2 \end{aligned}\right\} \Longleftrightarrow \left.\begin{aligned} 6v_1 &= -2v_2 \\ 3v_1 &= -1v_2 \end{aligned}\right\}.$$

This system of equations is equivalent to the equation $v_2 = -3v_1$, so the eigenvectors are

$$v = \begin{pmatrix} v_1 \\ -3v_1 \end{pmatrix}, \; v_1 \neq 0, \; \text{e.g.,} \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

6.3 Exercises

6.3.1

Are the following vectors independent?

(a) $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$

(b) $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$

(c) $v_1 = \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 \\ 3 \\ 6 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$

Solution:

(a) $\det \begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix} = \begin{vmatrix} 1 & 2 \\ 0 & 4 \end{vmatrix} = 1 \cdot 4 - 0 \cdot 2 = 4 \neq 0$, so the vectors are independent

(b) $\det \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = 1 \cdot 4 - 2 \cdot 2 = 0$, so the vectors are dependent, $2v_1 - v_2 = 0$

(c) $\det \begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 2 \\ 5 & 6 & 1 \end{pmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 3 & 2 \\ 5 & 6 & 1 \end{vmatrix} = 1 \cdot 3 \cdot 1 + 0 \cdot 2 \cdot 5 + 0 \cdot 2 \cdot 6 - 0 \cdot 3 \cdot 5 - 1 \cdot 2 \cdot 6 - 0 \cdot 2 \cdot 1 = -9 \neq 0$, so the vectors are independent

6.3.2

Determine the eigenvalues and eigenvectors of the following matrices. Give a specific eigenvector.

(a) $\begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}$

(b) $\begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}$

(c) $\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$

(d) $\begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix}$

Solution:

(a) $\lambda_1 = 3, v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; $\quad \lambda_2 = 2, v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

i, Eigenvalues

$\det(A-\lambda I) = \begin{vmatrix} 1-\lambda & 1 \\ -2 & 4-\lambda \end{vmatrix} = (1-\lambda)(4-\lambda)-1\cdot(-2) = \lambda^2-5\lambda+6 = 0 \quad \Longrightarrow \quad \lambda_1 = 3$ and $\lambda_2 = 2$.

ii, Eigenvectors

Eigenvectors belonging to eigenvalue $\lambda_1 = 3$:

$$Av = \lambda_1 v \Longleftrightarrow \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 3\cdot\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \Longleftrightarrow \left.\begin{aligned} v_1 + v_2 &= 3v_1 \\ -2v_1 + 4v_2 &= 3v_2 \end{aligned}\right\} \Longleftrightarrow \left.\begin{aligned} v_2 &= 2v_1 \\ -2v_1 &= -v_2 \end{aligned}\right\}.$$

This system of equations is equivalent to the equation $v_2 = 2v_1$, so the eigenvectors are

$$v = \begin{pmatrix} v_1 \\ 2v_1 \end{pmatrix},\ v_1 \neq 0,\ \text{e.g.,} \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Eigenvectors belonging to eigenvalue $\lambda_2 = 2$:

$$Av = \lambda_1 v \Longleftrightarrow \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 2\cdot\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \Longleftrightarrow \left.\begin{aligned} v_1 + 2v_2 &= 2v_1 \\ -2v_1 + 4v_2 &= 2v_2 \end{aligned}\right\} \Longleftrightarrow \left.\begin{aligned} 2v_2 &= 2v_1 \\ -2v_1 &= -2v_2 \end{aligned}\right\}.$$

This system of equations is equivalent to the equation $v_1 = v_2$, so the eigenvectors are

$$v = \begin{pmatrix} v_1 \\ v_1 \end{pmatrix},\ v_1 \neq 0,\ \text{e.g.,} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

(b) $\lambda_1 = 3, v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; $\quad \lambda_2 = 2, v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

i, Eigenvalues

$\det(A-\lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = (3-\lambda)(2-\lambda)-0\cdot 1 = (3-\lambda)(2-\lambda) = 0 \implies \lambda_1 = 3$ and $\lambda_2 = 2$.

ii, Eigenvectors

Eigenvectors belonging to eigenvalue $\lambda_1 = 3$:

$$Av = \lambda_1 v \iff \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 3 \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \iff \left.\begin{aligned} 3v_1 + v_2 &= 3v_1 \\ 2v_2 &= 3v_2 \end{aligned}\right\} \iff \left.\begin{aligned} v_2 &= 0 \\ 0 &= v_2 \end{aligned}\right\}.$$

This system of equations is equivalent to the equation $v_2 = 0$, so the eigenvectors are

$$v = \begin{pmatrix} v_1 \\ 0 \end{pmatrix}, \ v_1 \neq 0, \text{ e.g.,} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Eigenvectors belonging to eigenvalue $\lambda_2 = 2$:

$$Av = \lambda_1 v \iff \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 2 \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \iff \left.\begin{aligned} 3v_1 + v_2 &= 2v_1 \\ 2v_2 &= 2v_2 \end{aligned}\right\} \iff \left.\begin{aligned} v_2 &= -v_1 \\ v_2 &= v_2 \end{aligned}\right\}.$$

This system of equations is equivalent to the equation $v_1 = -v_2$, so the eigenvectors are

$$v = \begin{pmatrix} -v_2 \\ v_2 \end{pmatrix}, \ v_2 \neq 0, \text{ e.g.,} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

(c) $\lambda_1 = 2, v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}; \quad \lambda_2 = 4, v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

i, Eigenvalues

$\det(A-\lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = (3-\lambda)(3-\lambda)-1\cdot 1 = \lambda^2 - 6\lambda + 8 = 0 \implies \lambda_1 = 2$ and $\lambda_2 = 4$.

ii, Eigenvectors

Eigenvectors belonging to eigenvalue $\lambda_1 = 2$:

$$Av = \lambda_1 v \iff \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 2 \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \iff \left.\begin{aligned} 3v_1 + v_2 &= 2v_1 \\ v_1 + 3v_2 &= 2v_2 \end{aligned}\right\} \iff \left.\begin{aligned} v_2 &= -v_1 \\ v_1 &= -v_2 \end{aligned}\right\}.$$

This system of equations is equivalent to the equation $v_1 = -v_2$, so the eigenvectors are

$$v = \begin{pmatrix} -v_2 \\ v_2 \end{pmatrix}, \; v_2 \neq 0, \; \text{e.g.,} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Eigenvectors belonging to eigenvalue $\lambda_2 = 4$:

$$Av = \lambda_1 v \Longleftrightarrow \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 4 \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \Longleftrightarrow \left. \begin{array}{r} 3v_1 + v_2 = 4v_1 \\ v_1 + 3v_2 = 4v_2 \end{array} \right\} \Longleftrightarrow \left. \begin{array}{r} v_2 = v_1 \\ v_1 = v_2 \end{array} \right\}.$$

This system of equations is equivalent to the equation $v_1 = v_2$, so the eigenvectors are

$$v = \begin{pmatrix} v_1 \\ v_1 \end{pmatrix}, \; v_1 \neq 0, \; \text{e.g.,} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

(d) $\lambda_1 = 1, v_1 = \begin{pmatrix} \frac{3}{2} \\ 1 \end{pmatrix}$; $\quad \lambda_2 = \dfrac{1}{2}, v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

i, Eigenvalues

$\det(A - \lambda I) = \begin{vmatrix} 0.8 - \lambda & 0.3 \\ 0.2 & 0.7 - \lambda \end{vmatrix} = (0.8 - \lambda)(0.7 - \lambda) - 0.2 \cdot 0.3 = \lambda^2 - 1.5\lambda + 0.5 = 0 \quad \Longrightarrow \quad \lambda_1 = 1$ and $\lambda_2 = \frac{1}{2}$.

ii, Eigenvectors

Eigenvectors belonging to eigenvalue $\lambda_1 = 1$:

$$Av = \lambda_1 v \Longleftrightarrow \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 1 \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \Longleftrightarrow \left. \begin{array}{r} 0.8v_1 + 0.3v_2 = v_1 \\ 0.2v_1 + 0.7v_2 = v_2 \end{array} \right\} \Longleftrightarrow \left. \begin{array}{r} 0.3v_2 = 0.2v_1 \\ 0.2v_1 = 0.3v_2 \end{array} \right\}.$$

This system of equations is equivalent to the equation $v_1 = \frac{3}{2}v_2$, so the eigenvectors are

$$v = \begin{pmatrix} \frac{3}{2}v_2 \\ v_2 \end{pmatrix}, \; v_2 \neq 0, \; \text{e.g.,} \begin{pmatrix} \frac{3}{2} \\ 1 \end{pmatrix}.$$

Eigenvectors belonging to eigenvalue $\lambda_2 = \frac{1}{2}$:

$$Av = \lambda_1 v \iff \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \iff \left. \begin{array}{r} 0.8v_1 + 0.3v_2 = \frac{1}{2}v_1 \\ 0.2v_1 + 0.7v_2 = \frac{1}{2}v_2 \end{array} \right\} \iff \left. \begin{array}{r} 0.3v_2 = -0.3v_1 \\ 0.2v_1 = -0.2v_2 \end{array} \right\}.$$

This system of equations is equivalent to the equation $v_1 = -v_2$, so the eigenvectors are

$$v = \begin{pmatrix} -v_2 \\ v_2 \end{pmatrix}, \ v_2 \neq 0, \ \text{e.g.,} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Software

WolframAlpha: https://www.wolframalpha.com/

Type the following and check the output:

independent$\{\{1, 3\}, \{2, 6\}\}$

independent$\{\{1, -1, 0\}, \{2, 0, 1\}, \{0, 2, 1\}\}$

eigenvalues$\{\{4, -1\}, \{2, 1\}\}$

eigenvectors$\{\{4, -1\}, \{2, 1\}\}$

7 Combinatorics and classical probability

In the second half of the course we will study basic probability theory. Many problems in probability theory require that we count the number of ways that a particular event can occur. Therefore in this chapter, we study the topics of permutations, variations and combinations. We also introduce probability.

7.1 Motivating examples

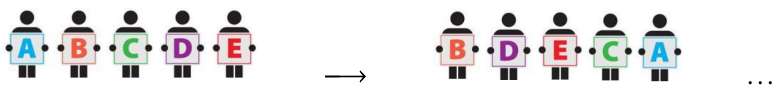
Every decision made in the business world has risk to it. Many businesses apply the understanding of uncertainty and probability in their business decision practices. Businesses can use probability theory in the calculation of long-term gains and losses, these probability models can greatly help them in increasing profitability and success, as well as in optimizing their policies.

For example, the optimization of a business's profit relies on how a business invests its resources. One important part of investing is knowing the risks involved with each type of investment. The only way a business can take these risks into account when making investment decisions is to use probability as a calculation method. After analyzing the probabilities of gain and loss associated with each investment decision, a business can apply probability models to calculate which investment or investment combinations yield the greatest expected profit.

7.2 Combinatorics

Permutation without replacement
Permutation of a set of n distinct elements is an ordering of the n elements. The number of permutations is: $n!$

> Example: Suppose we wish to arrange $n = 5$ people (A, B, C, D, E) for a portrait, standing side by side. How many such distinct portraits (permutations) are possible? (Note, that every different ordering counts as a distinct permutation. For instance, the ordering $ABCDE$ is distinct from $BDECA$, etc.) There are 5 possible choices who stands in the first position (eitherA, B, C, D or E). For each of these 5 possibilities,

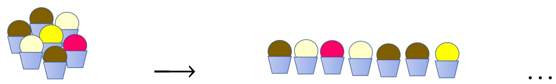


there are 4 possible choices left for the next position, and so on. Therefore, there are $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$ distinct permutations.

Permutation with replacement

Permutation with replacement of a set of n not necessarily distinct elements is an ordering of the n elements, when we do not differentiate among the same elements. The number of permutations is: $\dfrac{n!}{k_1!k_2!\dots k_r!}$ (where there are k_1 from the 1st type element, $\dots k_r$ from the rth type element ($k_1 + \dots + k_r = n$).)

Example: We have $n = 7$ ice cream cups with the following flavors: $k_1 = 2$ vanilla, $k_2 = 3$ chocolate, $k_3 = 1$ raspberry, $k_4 = 1$ lemon. In how many different orders could we eat them?

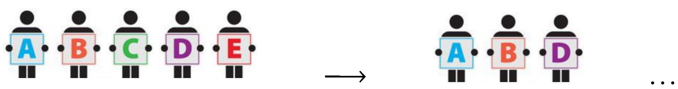


There are 7! permutations of 7 cups of ice cream, but since we do not differentiate between the same cups of ice cream, we need to divide by the permutations of these, i.e. by 2! (vanilla) and by 3! (chocolate). Therefore the number of permutations is: $\frac{7!}{2! \cdot 3!} = 420$.

***k*-variation without replacement**

k-variation of a set of n distinct elements is an ordering of a subset of k distinct elements. The number of k-variations is: $\dfrac{n!}{(n-k)!}$

Example: Now suppose we start with the same $n = 5$ people (A, B, C, D, E), but we wish to make portraits of only $k = 3$ of them at a time. How many such distinct portraits are possible? (Note, that every different ordering counts as a distinct permutation. For instance, the ordering ABD is distinct from DAB, etc.)

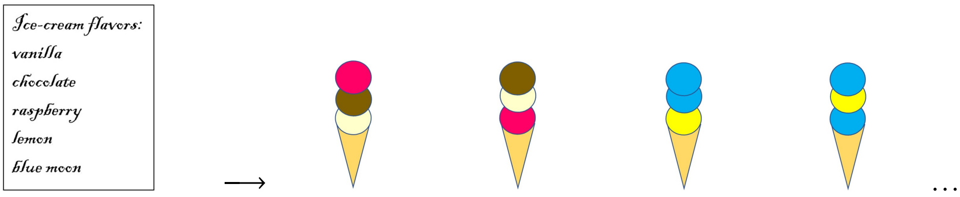


By using the reasoning above, there are $5 \cdot 4 \cdot 3 = \frac{5!}{2!} = 60$ permutations.

***k*-variation with replacement**

k-variation with replacement of a set of n distinct elements is an ordering of a subset of k elements, such that an element from the original set can be chosen more than once. The number of k-variations with repetitions is: n^k

Example: An ice cream shop sells ice creams in $n = 5$ different flavors: vanilla, chocolate, raspberry, lemon and blue moon. How many permutation of $k = 3$ scoops cone are possible? (Note: repetition of flavors is allowed, and the order in which they are chosen also matter.)



We have 5 flavors to choose from to a 3-scoop tall ice-cream. We can pick our first scoop from 5 flavors, the second and third scoops from also 5-5 flavors. Therefore, the total number of choices is 5^3.

k-combination without replacement

k-combination of a set of n distinct elements is a subset of k distinct elements. The number of k-combinations is equal to the binomial coefficient: $\binom{n}{k}$

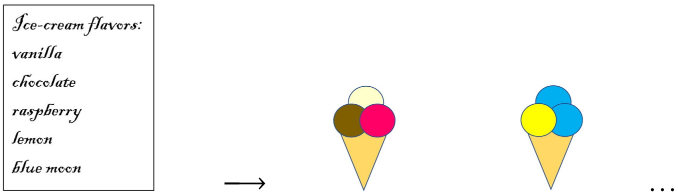
Example: Now suppose that instead of portraits (permutations), we wish to form committees (combinations) of $k = 3$ people from the original $n = 5$ people. How many such distinct committees are possible? (Note that different ordering does not count as a distinct combination. For instance, the committee $\{A, B, D\}$ is the same as the committee $\{B, D, A\}$, etc.)

From the calculation above, we know that the number of permutations of $k = 3$ people from $n = 5$ is equal to $\frac{5!}{2!} = 60$. But now, all the ordered permutations of any 3 people (there are 3! of these), will "collapse" into one single unordered combination. So we need to divide the possibilities by $3! = 6$. Therefore, there are $\frac{5!}{2! \cdot 3!} = \binom{60}{3} = 10$ combinations.

k-combination with replacement

k-combination with replacement of a set of n distinct elements is a subset of k elements, such that an element from the original set can be chosen more than once. The number of k-combinations with replacement is: $\binom{n+k-1}{k}$

Example: An ice cream shop sells ice creams in $n = 5$ different flavors: vanilla, chocolate, raspberry, lemon and blue moon. How many combinations of $k = 3$ scoops cone are possible? (Note: repetition of flavors is allowed, but the order in which they are chosen does not matter.)



Let's create 5 buckets of flavors: $\underline{\quad} \mid \underline{\quad} \mid \underline{\quad} \mid \underline{\quad} \mid \underline{\quad}$, then $5 - 1 = 4$ dividers (|) will be between them. Each bucket will have some scoops in them (o), altogether 3 scoops, e.g.: o | o | o | | or | | | o | o o. Now we solve this problem by sorting two types of objects, our 3 scoops and $5 - 1 = 4$ dividers. Thus you are permuting $3 + 5 - 1$ objects, 3 of which are same (the scoops) and $5 - 1$ of which are same (the dividers). Therefore, there are $\binom{5+3-1}{3} = \binom{7}{3} = 35$ choices of 3-scoop cones.

Note: The word "repetition" is commonly used instead of "replacement".

7.3 Classical probability

Probability is the mathematical term for the likelihood that something will occur, such as drawing an ace from a deck of cards or picking a red ball from a bag of balls with assorted colors. Classical probability is the concept that measures the likelihood of something happening, when every experiment has outcomes that are equally likely to happen, for example when rolling a fair die, it is equally likely to get a 1, 2, 3, 4, 5, or 6. The probability of a simple event happening is the number of desired outcomes, divided by the number of possible outcomes. Certainly, we will come across more complex cases later, but this simple model with finite number of possibilities is enough to understand the basic concepts.

7.4 Exercises

Combinatorics

7.4.1

How many anagrams (can be meaningless) does

a) MATH and

b) COMBINATORICS have?

Solution:

a) $4!$

b) Since there are two C's, two O's, and two I's: $\dfrac{13!}{2! \cdot 2! \cdot 2!}$

7.4.2

a) How many different 5-digit sequences can one make from $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ if no number is allowed to appear more than once (without repetition, all 5 digits will be different) or if a number can be used more than once (with repetition, digits can be the same)?

b) How many of these denote a 5-digit number?

c) ⋆ How many 5-digit numbers are even?

Solution:

a) without repetition: $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6$, with repetition: 10^5

b) without repetition: $9 \cdot 9 \cdot 8 \cdot 7 \cdot 6$, with repetition: $9 \cdot 10^4$

c) without repetition: $5 \cdot 8 \cdot 7 \cdot 6 \cdot 5 + 4 \cdot 8 \cdot 7 \cdot 6 \cdot 4 = 41 \cdot 8 \cdot 7 \cdot 6$ (odd + even start) or $9 \cdot 8 \cdot 7 \cdot 6 \cdot 1 + 8 \cdot 8 \cdot 7 \cdot 6 \cdot 4 = 41 \cdot 8 \cdot 7 \cdot 6$ (last digit equals to 0 + last digit can be 2, 4, 6, 8), with repetition: $9 \cdot 10^3 \cdot 5$

7.4.3

How many different license plates can one make from 26 letters and 10 digits

a) if the license plates contain 2 letters followed by 4 digits?

b) if the license plates contain 3 letters followed by 3 digits?

Solution:

a) $26^2 \cdot 10^4$

b) $26^3 \cdot 10^3$

7.4.4

How many different ways can a test consisting of 8 multiple choice questions be filled in, if for each question one has to choose from 4 answers?

Solution:

4^8

7.4.5

There are 15 people in a classroom.

a) If everyone shakes hand with everyone else exactly once, how many handshakes have taken place?

b) The students now decide to split people to work in three groups of size five each. How many ways can they make the groups?

Solution:

a) The first person shakes hands with everyone else, that will total 14 handshakes. If the second person now shakes hands with everyone, except for himself and the first person, with whom he has already shaken hands, this will add 13 handshakes to the total. Continuing along the line, each person will then add 12, 11, then 10 and so on until the final 15th person, who will already have shaken hands with everyone else and will add zero handshakes to the total. So the total number of handshakes are: $14 + \cdots + 2 + 1 + 0 = \frac{14\cdot(14+1)}{2} = 105$.
We can also get this result by counting how many ways we can choose a pair from 15 people: $\binom{15}{2} = 105$.

b) $\frac{\binom{15}{5}\cdot\binom{10}{5}\cdot\binom{5}{5}}{3!}$

7.4.6

As an intern at Morgan Stanley your job is to select 5 stocks from the S&P 500 index and order them according to their potential for growth. How many ways can this be done? If you can write down one selection in one second, how long would it take to write down all possible selections?

Solution:

The intern can select the indices $\binom{500}{5}$ way, then order the selection 5! ways, thus altogether $\binom{500}{5}\cdot 5! = \frac{500!}{495!}$ ways can the job be done. The time for writing down all possible selections is $\binom{500}{5} = 255244687600$ seconds.

7.4.7

How many ways can you put 8 rooks on a chessboard so they don't hit each other? (Rook is in Hungarian bástya.)

Solution:

The first rook could be placed onto 64 fields. Afterwards there can be no more rooks in the row and column of the covered field, so we can only place the next rook in the remaining 7 rows and 7 columns, which are 49 possibilities. With each new rook laid down, another row and column will be covered. So then there can be 36, 25, 16, 9, 4, then 1 possibility to put down the next rook. In this case, the order in which the rooks were laid was thus taken into account, even though all 8 rooks are identical, thus indistinguishable. So we have to divide by the number of permutations of the rooks put down, i.e., with 8!. Therefore we can put down the rooks in a total of $\frac{64\cdot49\cdot36\cdot25\cdot16\cdot9\cdot4\cdot1}{8!} = 40320 = 8!$ ways.

The final result can also be obtained directly by looking at the location of the rooks column by column (or row by row).

7.4.8

Let's suppose you are an investor at a micro loan group and you have \$10000 to invest in 4 companies (in \$1000 increments).

a) How many ways can you allocate the money?

b) What if you don't have to invest all \$10000, how many ways can you allocate the money?

c) You want to invest \$3000 in company 1 (and don't have to invest all \$10000), how many ways can you allocate the money?

Solution:

a) This is just like putting 10 balls ($k = 10$) into 4 urns ($n = 4$), where the balls are indistinguishable and the urns are distinguishable. Therefore, the number of ways to allocate the money is $\binom{4+10-1}{10} = \binom{13}{10} = 286$. You can also conclude this result from the following. Let's order the 4 urns, then $4 - 1 = 3$ dividers (|) will be between them. Each urn will have some balls in them (o), e.g.: o o o | o o | o o o o | o . Now you are going to solve this problem by sorting two types of objects, your 10 balls and $4 - 1$ dividers. Thus you are permuting $10 + 4 - 1$ objects, 10 of which are same (the balls) and $4 - 1$ of which are same (the dividers).

b) Simply imagine that you have an extra company – yourself. Now you are investing \$10000 in 5 companies. Thus the answer is the same as putting 10 balls into 5 urns. Therefore, the number of ways to allocate the money is $\binom{5+10-1}{10} = \binom{14}{10} = 1001$.

c) There is one way to give \$3000 to company 1. The number of ways of investing the remaining money is the same as putting 7 balls into 4 urns. Therefore, the number of ways to allocate the money is $\binom{4+7-1}{7} = \binom{10}{7} = 120$.

Classical probability

7.4.9

A 4-digit PIN number is selected. What is the probability that there are no repeated digits?

Solution:

$$\frac{10 \cdot 9 \cdot 8 \cdot 7}{10^4} = 0,504$$

7.4.10

A six-sided fair dice is rolled 6 times. What is the probability that each side shows up once?

Solution:

$$P(\text{each sides shows up once}) = \frac{\text{number of desired outcomes}}{\text{number of possible outcomes}} = \frac{6!}{6^6}$$

7.4.11

What is the probability that the digits of a randomly chosen 6-digit number are all different?

Solution:

The first digit can be chosen from $1, 2, \ldots, 9$, while the rest of the digits can be chosen from $0, 1, 2, \ldots, 9$. Thus the number of possible outcomes is $9 \cdot 10^5$. The number of desired outcomes is $9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5$, since we choose without replacement, the order of the outcomes counts, and we pay attention to the first digit not to be 0. Therefore the desired probability is $\frac{9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{9 \cdot 10^5} = \frac{136080}{900000} = 0{,}1512$.

7.4.12

In a store, 3 of the 10 seemingly identical computers have been refurbished, the rest are new. What is the probability that if we buy 5 machines for the lab, exactly 2 of them will be refurbished?

Solution:

Of the 10 machines, 3 have been refurbished and 7 are new. So we have to choose 2 of the 3 refurbished machines and choose the remaining 3 of the 7 new machines. The order of selection does not matter and the sampling is without replacement. The number of desired outcomes are $\binom{3}{2} \cdot \binom{7}{3} = 3 \cdot 35 = 105$. The total number of cases are $\binom{10}{5} = 252$. Therefore the desired probability is $\frac{105}{252} = 0{,}4167$. (We will learn in Chapter 11 that it follows a hypergeometric distribution with parameters $N = 10, M = 3, n = 5$.)

7.4.13

If we randomly choose our 6-digit password from the 10 numbers and 26 characters, then what is the probability that it will contain 3 numbers?

Solution:

We can choose the position of the 3 numbers of the 6-digits in $\binom{6}{3} = 20$ ways. Each of these three positions can be filled with a number in 10 ways, giving us the factor 10^3. Each of the remaining three positions can be filled with a character in 26 ways, giving us the factor 26^3. Hence, the number of favorable cases is $\binom{6}{3}10^3 26^3$. There are $10 + 26 = 36$ choices for each of the six digits in the password, giving us 36^6 possible outcomes. Therefore the desired probability is $\frac{\binom{6}{3}10^3 26^3}{36^6} = \binom{6}{3} \cdot (\frac{10}{36})^3 \cdot (\frac{26}{36})^3 = 0{,}1615$. (We will learn in Chapter 11 that it follows a binomial distribution with parameters $n = 60, p = \frac{10}{36}$.)

7.4.14

For the five-number draw lottery, give the probability that we win the jackpot (all five numbers win) when playing with a single lottery ticket, or that we will have at least four hits. What is the probability that all numbers drawn are even? (How does this relate to the case with replacement?)

Solution:

The probability that we win the jackpot: $\frac{\binom{5}{5}}{\binom{90}{5}} = \frac{1}{\binom{90}{5}}$.

The probability that we pick at least four numbers correctly: $\frac{\binom{5}{5}}{\binom{90}{5}} + \frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}}$.

The probability that all numbers drawn are even: $\frac{\binom{45}{5}}{\binom{90}{5}} \approx 0,028$.

In the case of drawing with replacement (i.e., when we can draw one number multiple times), the probability that we draw an even number is: $\left(\frac{45}{90}\right)^5 = \left(\frac{1}{2}\right)^5 \approx 0,031$. Though these two values are close to each other, the probability of drawing without replacement is smaller because in this case we have less and less even numbers during the selection.

7.4.15

Two six-sided fair dice are rolled. What is the probability that at least one of the dice shows a 4?

Solution:

There are altogether $6 \cdot 6 = 36$ possible outcomes. Outcomes with at least one of the dice showing a 4 are as follows: $\{14, 24, 34, 44, 54, 64, 41, 42, 43, 45, 46\}$, 6 + 6 - 1 = 11 cases. Thus P(at least one of the dice shows a 4) $= \frac{11}{36}$.

8 Events, probability of events

In this semester we will deal with experiments with finite and countable infinite random quantities, but the usability of these models is limited, as we would often not be able to answer simple questions if we only stayed in this circle. For example, the time of decay of a radioactive particle, the returns of stocks, etc. may be better modeled with noncountable values. In many cases continuous models can be used and are more manageable than the discrete ones. Therefore, as a preparation for next semester as well, we will be broaden our view of probability to a more general definition.

8.1 Events

An experiment (e.g., throwing a dice) is a process or activity that has several possible outcomes of which exactly one occurs when the experiment is performed (e.g., throwing a three: ⚂). An event is a set of one or more possible outcomes of an experiment (e.g., rolling a five: { ⚄ }, or rolling an even number: { ⚁, ⚃, ⚅ }). We can distinguish between two types of events: elementary (or simple) and compound events. A simple event consists of a possible outcome of the experiment (e.g., rolling a five: { ⚄ }), while a compound event is a combination of two or more simple events (e.g., throwing an even number: { ⚁, ⚃, ⚅ }). The collection of elementary events will be denoted by Ω, called *sample space* (e.g., throwing a dice: $\Omega =$ { ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ }). The subsets of Ω are events, usually denoted by $A, B, C \dots$. The empty set is always considered as an event, called empty event (or null event or impossible event) and is denoted by $\emptyset$. An event occurs when any of the elementary events that make the event up occurs. The probability is assigned to the events (e.g., the probability of throwing a 1 is 1/6 or the probability of throwing and even number is 1/2).

Operations of events

Events can be treated as sets, thus the usual logical operations on events correspond to operations on sets.

Union: $A \cup B$ or A *or* B is the event that happens when either A or B occurs
Intersection: $A \cap B$ or A *and* B is the event that happens when both A and B occur
Complement: $\overline{A}$ or A^C or *not* A is the event that happens when A does not occur
Difference: $B \backslash A$ is the event that happens when B occurs, but A does not occur

Mutually exclusive events

Two events are mutually exclusive if they cannot occur at the same time, i.e., when one event occurs the other cannot occur, and vice versa.

Examples of experiments and events:

Experiment	Collection of simple events (Ω)	Example of an event
Discrete outcomes:		
1. Rolling a die	$\{1, 2, 3, 4, 5, 6\}$	getting a roll less than 4, $A = \{1, 2, 3\}$
2. Flipping two coins	$\{HH, HT, TH, TT\}$	getting a H on the first coin, $A = \{HH, HT\}$
3. Drawing an odd natural number	$\{1, 3, 5, ...\}$	divisible by 5, $A = \{5, 15, ...\}$
Continuous outcomes:		
4. Typing "rand()" on an excel spreadsheet	$[0, 1]$	"rand()" gives a number larger than 0.5, $A = (0.5, 1]$
5. Stock price path from today to time T	a set of nonnegative continuous functions on $[0, T]$	stock price is above 2000 at time T,

Notice that experiments 3., 4. and 5. are examples of infinite Ω. However; in this semester in most of our cases, we will deal with finite Ω.

8.2 Probability of events

A random experiment can be characterized by the following 3 features:

- What are the possible outcomes of the experiment?
- What events can we observe? Alternatively, what information will be revealed to us at the end of the experiment?
- How do we assign probabilities to the events that we can observe?

The collection of certain events will be denoted by $\mathcal{A}$, called *event space.* For now, you can think of this event space consisting of every possible subset of Ω. (But not all subsets of Ω can be an event, for example if Ω is beyond countable.) We define probability on the elements of $\mathcal{A}$. Probability is a function $P : \mathcal{A} \to [0, 1]$, such that

1. $P(A) \geq 0$ for all $A \in \mathcal{A}$ events

2. $P(\Omega) = 1$

3. for (pairwise) mutually exclusive $(A_n) \subseteq \mathcal{A}$ sequence of events $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$

If $A \in \mathcal{A}$ is an event, then the number $P(A)$ assigned to it is called the probability of event A. Condition 3. can be applied also for a finite sequence of (pairwise) mutually exclusive events, for example if $n = 2$, then for two mutually exclusive events A and B the above condition gives $P(A \cup B) = P(A) + P(B)$. (This is sometimes called Addition rule.)

Basic properties of probability

$P(\emptyset) = 0$
$P(A) \leq 1$ for all $A \in \mathcal{A}$ events
$P(\overline{A}) = 1 - P(A)$ for all $A \in \mathcal{A}$ events
$P(A) \leq P(B)$ for $A \subset B \in \mathcal{A}$ events
$P(B \setminus A) = P(B) - P(A)$ for $A \subset B \in \mathcal{A}$ events

Inclusion-exclusion principle (Sieve formula) for two events: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for all $A, B \in \mathcal{A}$ events

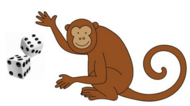
The illustrative equivalent of probability is the relative frequency. If a given event A occurred k times during n trials performed independently under the same conditions, then the relative frequency is k/n. For large n, the relative frequency fluctuates around a fixed number: this is called the probability of A.

8.3 Examples

Example 1: Let's flip a coin and observe its face. Then the above features are:

- The possible outcomes are: H, T, i.e., $\Omega = \{H, T\}$
- The collection of events we can observe: $\mathcal{A} = \{\emptyset, \{H\}, \{T\}, \Omega\}$
- Probabilities are computed on the basis that each outcome is equally likely, so we have the following:
  $P(\emptyset) = 0$, $P(\{H\}) = 0.5$, $P(\{T\}) = 0.5$, $P(\Omega) = 1$

Example 2: Let's suppose a monkey rolls two six-sided fair dice, but he only tells us the sum of the two rolls.



Then the above features are:

- The possible outcomes are: ⚀⚀, ⚀⚁, ⚀⚂, ⚀⚃, ⚀⚄, ⚀⚅, ⚁⚀, ⚁⚁, ⚁⚂, …, ⚅⚅ denoted by 11, 12, 13, 14, 15, 16, 21, 22, 23, . . . , 66.
  Sample space: $\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, \ldots, 66\}$

- The information available to us is the sum of the rolls which can be $2, 3, 4, \dots 12$. For example, we get 6 by observing the set $\{15, 24, 33, 42, 51\}$ (but we do not know which element of this set was the actual outcome of the monkey's experiment).
  Event space: $\mathcal{A} = \{\emptyset, \{11\}, \{12, 21\}, \{13, 22, 31\}, \dots \{66\}, \{11, 12, 21\}, \dots, \Omega\}$
- Probabilities are computed on the basis that each outcome is equally likely, so we have the following:
  $P(\emptyset) = 0$
  $P(\{11\}) = \frac{1}{36}$
  $P(\{12, 21\}) = \frac{2}{36}$
  $\dots$
  $P(\{15, 24, 33, 42, 51\}) = \frac{5}{36}$
  $\dots$
  $P(\{66\}) = \frac{1}{36}$
  $P(\{2, 3\}) = P(\{11, 12, 21\}) = \frac{3}{36}$
  $\dots$
  $P(\Omega) = 1$

8.4 Probability space

The probability space is essentially a collection of 3 mathematical objects $(\Omega, \mathcal{A}, P)$, representing the above 3 features of a random experiment. While the sample space describes the set of every possible outcome for an experiment, the event space is a family of subsets of the sample space whose elements are events. An outcome ω is an element in Ω ($\omega \in \Omega$), which we may or may not observe, but these elements are there, and we should know about their existence to determine the correct probability model. In Example 2., we do not observe this (only the monkey does). Therefore, the information available (sum of the rolls) can be characterized by the above partition of Ω.

Finite probability space

The triplet $(\Omega, \mathcal{A}, P)$ is called a finite probability space if the sample space Ω is a finite set, the $\mathcal{A}$ consists of every possible subset of Ω, and the probability function is as follows: if $\Omega = \{\omega_1, ..., \omega_n\}$, and let $p_i = P(\omega_i)$, then $p_1, \dots, p_n$ are non-negative real numbers which add up to 1. Then let the probability of A be the sum of the probabilities of the elementary events in A, i.e., if $A = \{\omega_{i_1}, \dots, \omega_{i_j}\} \subset \Omega$, then $P(A) = p_{i_1} + \cdots + p_{i_j}$.

The classical probability space is a special case of finite probability space. If Ω is finite and all elementary events are equally likely, then $(\Omega, \mathcal{A}, P)$ is a classical probability space, in which case for any event $A \in \mathcal{A}$,

$$P(A) = \frac{\text{number of elementary events in } A}{\text{number of elementary events in } \Omega} = \frac{|A|}{|\Omega|} = \frac{\text{number of desired outcomes}}{\text{number of possible outcomes}}.$$

We introduced this in the previous chapter. Long time ago this was considered the definition of probability. Although this can be used in many cases, it does not cover all cases and thus cannot be used as a general definition.

8.5 Exercises

8.5.1

In a small study there are 20 families. There are 4 families with one child, 8 families with 2 children, 5 families with 3 children, 2 families with 4 children and in one family there are 5 children.

a) Consider the experiment that a family is selected at random and the number of children is registered. What is the sample space, i.e., the set of all possible outcomes? What are the corresponding probability assignments?

b) Consider the experiment that a child is selected at random and the number of children in the family he/she comes from is registered. What is the sample space? What are the corresponding probability assignments?

Solution:

a) $\Omega = \{1, 2, 3, 4, 5\}, \quad P(1) = \frac{4}{20}, P(2) = \frac{8}{20}, P(3) = \frac{5}{20}, P(4) = \frac{2}{20}, P(5) = \frac{1}{20}$

b) $\Omega = \{1, 2, 3, 4, 5\}$
There are altogether $4 \cdot 1 + 8 \cdot 2 + 5 \cdot 3 + 2 \cdot 4 + 1 \cdot 5 = 48$ children, so $P(1) = \frac{4}{48}, P(2) = \frac{8 \cdot 2}{48}, P(3) = \frac{5 \cdot 3}{48}, P(4) = \frac{2 \cdot 4}{48}, P(5) = \frac{1 \cdot 5}{48}$

8.5.2

A coin is tossed twice.

a) What is the sample space, i.e., the set of all possible outcomes? (In English "fej" is Head, and "írás" is Tail.)

b) Assume the coin is fair, and both heads and tails are equally likely. What are the probabilities of the outcomes in the sample space above?

c) What outcomes correspond to the event that we tossed at least one head? What is the probability of this event?

Solution:

a) $\Omega = \{HH, HT, TH, TT\}$

b) $P(HH) = \frac{1}{4}, P(HT) = \frac{1}{4}, P(TH) = \frac{1}{4}, P(TT) = \frac{1}{4}$

c) $E = \{HH, HT, TH\}, P(E) = \frac{3}{4}$

8.5.3

We have two tetrahedral dice, one white and one black. (Tetrahedral dice were used since very early times. It is triangular pyramid shaped and has four triangular sides.) The dice are rolled simultaneously. Assume that the sides are labeled 1, 2, 3, 4 and that the dice are fair, i.e., each side is equally likely.

a) What is the sample space, the set of all possible outcomes? (You can describe this in words or list all of them.)

b) Let A be the event that the black die shows a smaller number than the white one. What is the probability of this event?

c) Let B be the event that at least one die shows a 1. What is the probability of this event?

Solution:

a) $\Omega = \{11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 33, 34, 41, 42, 43, 44\}$

b) $A = \{12, 13, 14, 23, 24, 34\}$ (assuming the first number refers to the black die), $P(A) = \frac{6}{16}$

c) $B = \{11, 12, 13, 14, 21, 31, 41\}, P(B) = \frac{7}{16}$

8.5.4

We toss 2 coins, then with as many more coins as many heads as we got with the first two coins. What will be the elements of the sample space?

Solution:

Let's denote the head with an H and the tail with T. Then the elements of the event space is $\Omega = \{HHHH, HHHT, HHTH, HHTT, HTH, HTT, THH, THT, TT\}$.

8.5.5

Let A, B, and C be three events. Using the above operation of events, write down the event that of these three events

a) exactly k event occurs,

b) at most k event occurs. ($k = 1, 2, 3$)

Solution:

a) (exactly 0 event occurs)= $(\overline{A} \cap \overline{B} \cap \overline{C}) = (\overline{A \cup B \cup C})$

(exactly 1 event occurs)= $(A \cap \overline{B} \cap \overline{C}) \cup (\overline{A} \cap B \cap \overline{C}) \cup (\overline{A} \cap \overline{B} \cap C)$

(exactly 2 events occur)= $(A \cap B \cap \overline{C}) \cup (A \cap \overline{B} \cap C) \cup (\overline{A} \cap B \cap C)$

(exactly 3 events occur)= $(A \cap B \cap C)$

b) (at most 1 event occurs)=(exactly 0 event occurs) $\cup$ (exactly 1 event occurs)

(at most 2 events occur)= $(\overline{A \cap B \cap C}) = (\overline{A} \cup \overline{B} \cup \overline{C})$

(at most 3 events occur)=this is always true, i.e., this is the entire Ω space

8.5.6

If we draw three cards with replacement from a German deck of cards (32 cards: hearts, leaves, acorns, bells), what is the probability that

a) we draw exactly one card with hearts on it?

b) we draw at least one card with hearts on it?

(The German deck of cards (hearts, leaves, acorns, bells) in Hungarian is Magyar kártya (piros, zöld, makk, tök.)

Solution:

a) From the 3 chosen cards we can decide $\binom{3}{1} = 3$-ways which should be the card with hearts on it. Afterwards we can assume that the first card drawn is the card with hearts on it, and the remaining two cards are of other suits. Since we choose with replacement, the probability of drawing a card with hearts on it is $\frac{8}{32}$, while the probability of drawing a card with other than hearts on it is $\frac{24}{32}$. Therefore the desired probability is $\binom{3}{1} \cdot (\frac{8}{32})^1 \cdot (\frac{24}{32})^2 = 3 \cdot \frac{1}{4} \cdot \frac{9}{16} = \frac{27}{64} = 0{,}4219$.

b) It is easier to calculate by subtracting the probability of the complement event from 1. The complement event is there is no card with hearts on among the cards drawn. The probability of this event is $\binom{3}{0} \cdot (\frac{8}{32})^0 \cdot (\frac{24}{32})^3 = \frac{27}{64}$. Therefore the desired probability is $1 - \frac{27}{64} = \frac{37}{64} = 0{,}5781$.

8.5.7

Assume that A and B are mutually exclusive events, and that $P(A) = 0.3$ and that $P(B) = 0.5$. What is the probability of the following events:

a) Either A or B happens.

b) A happens but B does not happen.

c) Both A and B happens.

Solution:

Since A and B are mutually exclusive events, $P(A \cap B) = 0$.

a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.5 = 0.8$

b) $P(A \cap \overline{B}) = P(A) = 0.3$

c) $P(A \cap B) = 0$

8.5.8

In a group of students everyone works out regularly: 68% of the students run, 41% of the students swim.

a) What is the probability that a randomly chosen student both runs and swims regularly?

b) What is the probability that a randomly chosen student runs but does not swim?

c) What is the probability that a randomly chosen student swims but does not run?

Solution:

Let R be the event that a student runs, and let S be the event that a student swims. Then $P(R) = 0.68$ and $P(S) = 0.41$. Since everyone works out $P(R \cup S) = 1$. Using the inclusion/exclusion principle for two events, $P(R \cup S) = P(R) + P(S) - P(R \cap S)$ we get:

a) $P(R \cap S) = P(R) + P(S) - 1 = 0.68 + 0.41 - 1 = 0.09$

b) $P(R \cap \overline{S}) = P(R \setminus S) = P(R) - P(R \cap S) = 0.68 - 0.09 = 0.59$

c) $P(\overline{R} \cap S) = P(S \setminus R) = P(S) - P(R \cap S) = 0.41 - 0.09 = 0.32$

8.5.9

Assume that in a family the probability of a boy or a girl is the same for each birth independently. What is the probability that in a family of three all children have the same sex?

Solution:

$P(BBB) + P(GGG) = \frac{1}{2^3} + \frac{1}{2^3} = \frac{1}{4}$

8.5.10

How many times a coin must be tossed in order to make the probability of at least one head greater than 90%?

Solution:

Let's suppose we toss the coin n times. Then $P(\text{at least 1 head}) = 1 - P(\text{no head tossed}) = 1 - P(\text{all tail tossed}) = 1 - \frac{1}{2^n} > 90\%$, thus $\frac{1}{2^n} < 1 - 0.9 = 0.1$, $n > \frac{\ln(0.1)}{\ln(0.5)} \approx 3.32$. Therefore at least 4 times should the coin be tossed.

8.5.11

Anna, Bob, Carol, David and Elizabeth all go to the Cabaret Voltaire once a month. Assume that there are 30 performances in a month, and that they select the date of going independent of each other.

a) What is the probability that they all go to the same performance?

b) What is the probability that they all go to different performances?

c) What is the probability that at least two of them meet at a performance?

Solution:

a) $P(\text{all go to the same performance}) = \frac{\text{number of desired outcomes}}{\text{number of possible outcomes}} = \frac{30}{30^5}$

b) $P(\text{all go to different performance}) = \frac{\binom{30}{5} \cdot 5!}{30^5} = \frac{\frac{30!}{25!}}{30^5}$

c) $P(\text{at least two of them meet at a performance}) = 1 - P(\text{all go to different performance}) = 1 - \frac{\frac{30!}{25!}}{30^5}$

8.5.12

Suppose k people are in a room. What is the probability that there is at least one shared birthday among these k people? ($k = 3, 5, 30$)

Solution:

$P(\text{at least one shared birthday}, k = 3) = 1 - P(\text{none, k=3}) = 1 - \frac{365 \cdot 364 \cdot 363}{365^3} = 0.0082$

$P(\text{at least one shared birthday}, k = 5) = 1 - P(\text{none, k=5}) = 1 - \frac{365 \cdot 364 \cdot 363 \cdot 362 \cdot 361}{365^5} = 0.027$

$P(\text{at least one shared birthday}, k = 30) = 1 - P(\text{none, k=30}) = 1 - \frac{365!}{335! \cdot 365^{30}} = 0.706$

$P(\text{at least one shared birthday, k}) = 1 - P(\text{none, k}) = 1 - \frac{365!}{(365-k)! \cdot 365^k} = 1 - \frac{365 \cdot 364 \cdot \ldots \cdot (365-k+1)}{365^k}$

8.5.13

There are 10 pairs of shoes in a bag. Selecting 4 pieces of shoes, what is the probability that there is a pair among them, if

a) the 10 pairs are the same?

b) the 10 pairs are different?

Solution:

a) There are 10 left and 10 right shoes. What is the probability that there is a left and a right shoe among the 4 draws? It is practical to subtract the complementary event from 1. The complementary event is the event that either 4 left or 4 right is drawn. The probability of this is $2 \cdot \frac{10\cdot 9\cdot 8\cdot 7}{20\cdot 19\cdot 18\cdot 17} = \frac{28}{323}$ or $2 \cdot \frac{\binom{10}{4}}{\binom{20}{4}} = \frac{28}{323}$. Therefore, the desired probability is $1 - \frac{28}{323} = 0{,}9133$.

b) It is worth calculating the probability of the complementary event again. The complementary event is the event that there is no pair among the 4 pieces of shoes. We can select the first shoe 20 ways, the second shoe 18 ways (the first and its pair cannot be selected), the third one 16 different ways, and the fourth one 14 ways. The total number of cases are $20 \cdot 19 \cdot 18 \cdot 17$. Therefore the probability of the complementary event is $\frac{20\cdot 18\cdot 16\cdot 14}{20\cdot 19\cdot 18\cdot 17} = \frac{224}{323}$. Or we can think of selecting the first 4 pairs from the 10 pairs, and then their left or right ones: $\frac{\binom{10}{4}\binom{2}{1}\binom{2}{1}\binom{2}{1}\binom{2}{1}}{\binom{20}{4}} = \frac{224}{323}$. Therefore the desired probability is $1 - \frac{224}{323} = 0{,}3065$.

8.5.14

⋆ We randomly place n distinguishable balls into n distinguishable boxes so that any number of balls can be placed in each box.

a) What is the probability that a ball will be placed in each box?

b) What is the probability that exactly one box will remain empty?

c) What are the above probabilities if we assume that the balls are indistinguishable?

Solution:

a) The balls are distinguishable (although this does not necessarily mean that the balls are physically different, it is also a distinction if we put them one after another into the boxes, thus we can talk about the first ball, second ball etc.). In this case, unless the task explicitly states otherwise, the interpretation of the word “random” is that each ball is placed into the boxes independently with the same $(1/n)$ probability.
For simplicity, consider the case $n = 2$. The probability space is naturally a product space, $\Omega = \{1, 2\} \times \{1, 2\}$, where the first component of the Descartes product encodes that the first ball is placed in a 1 or 2 box, the second component does the same with the second ball. For example, $\omega = (2, 1)$ means that the first ball is in box 2 and the second ball is in box 1. There can be a total of $2 \cdot 2 = 4$ possibilities, and due to the assumption of independence, each one is $1/2 \cdot 1/2 = 1/4$ likely.
In general: n distinguishable balls can be placed into n boxes n^n different ways (n-variation with replacement). The number of desirable outcomes is $n!$, i.e., the number of possible

permutations. Therefore, the desired probability is

$$P(\text{one ball in every box}) = \frac{n!}{n^n}.$$

b) If we distinguish the balls, then - as previously - we can put the n balls into n boxes in n^n ways (n-variation with replacement). The number of desired outcomes can be obtained as follows: we can select the empty box n ways, and the box containing 2 balls $n-1$ ways. The n balls can be chosen $n!$ ways, but we can get to this scenario in two ways, since any of the balls in the box with 2 balls could have come from the now empty box. Therefore the desired probability is

$$P(\text{exactly one box is left empty}) = \frac{n(n-1)\frac{n!}{2}}{n^n} = \frac{\binom{n}{2}n!}{n^n}.$$

c) If the balls are not distinguishable and the process of putting them into the boxes does not matter, then we can also reason that we see the endresult only and we define the probability space as the set of all possible outcomes. Note, however, that in the vast majority of cases, the above interpretation corresponds to the common concept of "random".

  a) Notice that in this case, contrary to the above, for $n = 2$ we only have 3 possibilities:
  (1) two balls in the first box and nothing in the second one
  (2) one ball in each box
  (3) nothing in the fist box and 2 balls in the second one
  In its structure, the probability space is very different than the previous one, not only the number of elements differ, but also there is no Descartes-product structure. In principle, the word "randomly" can be interpreted as the three outcomes are equally likely. For example, this way $1/3$ is the probability of a ball be placed in each box, while this probability is $1/2$ by the first interpretation. In general: n indistinguishable balls can be placed into n boxes in $\binom{2n-1}{n}$ ways (n-combination with replacement). [Let's arrange the n boxes in order, then there is $n-1$ divisor among them. The total number of outcomes is the number of sequences of the n balls and $n-1$ divisors, which is a permutation with replacement: $\frac{(n+(n-1))!}{n!\cdot(n-1)!} = \binom{2n-1}{n}$.] The number of the desired outcomes is 1, i.e., there is 1 ball places into each n box. Therefore the desired probability is

$$P(\text{one ball in every box}) = \frac{1}{\binom{2n-1}{n}}.$$

  b) If we do not distinguish the balls, then the n balls could be selected to n boxes $\binom{2n-1}{n}$ ways (n-combination with replacement). The number of desired outcomes can be obtained as follows: we can select the empty box n ways, and the box containing 2 balls $n-1$ ways. Therefore, the desired probability is

$$P(\text{exactly one box is left empty}) = \frac{n(n-1)}{\binom{2n-1}{n}}.$$

8.5.15

A gambling problem of Chevalier de Méré in the 17th century may have launched modern probability theory. De Méré turned to Pascal to explain his gambling losses. He suspected that (a) was higher than (b), but his mathematical skills were not good enough to demonstrate why this should be. Did Pascal prove de Méré was correct?

a) The probability of getting at least one six in 4 rolls of a single 6-sided fair die.

b) The probability of at least one double-six in 24 throws of two fair dice.

Solution:

a) Let A = getting at least one six in 4 rolls of one die, then $P(A) = 1 - P(A^C) = 1 - \left(\frac{5}{6}\right)^4 \approx 0,5177$

b) Let B = getting at least one double-six in 24 rolls of two dice, then $P(B) = 1 - P(B^C) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0,4914$

9 Independence of events, conditional probability

9.1 Independence of events, conditional probability

Example: Two dice are rolled. Let A be the event that the first die shows a number bigger then 4, let B be the event that the second die shows an even number. $P(A) = 1/3$, $P(B) = 1/2$, $P(A \cap B) = 1/6$. Hence $P(A \cap B) = P(A) \cdot P(B)$ holds for this pair of events. We say that the events A and B are *independent.*

Example: In a town 36% of the families own a dog, 30% own a cat. Out of those who own a dog 22% also own a cat.

a) What percentage of the families own both a cat and a dog?

b) Of those families who own a cat, what percentage also own a dog?

Solution: For simplicity assume that there are 10000 families in the town. Then 3600 families have a dog, 3000 families have a cat. $3600 \cdot 0.22 = 792$ families have both animals. That is $100 \cdot 792/3000 = 26.4$ % of the families which have a cat.

Let us analyze the calculation! Let C be the event that a random family has a cat, and D that a family has a dog. $P(D) = 0.36$, $P(C) = 0.3$. Then 0.22 is the probability of the event that a family has a cat *if we know that they have a dog*: $P(C|D) = 0.22$. That is called *conditional probability.* From the above calculation we get $P(C \cap D) = P(D) \cdot P(C|D) = 0.0792$ is the probability having both animals and $P(D|C) = P(C \cap D)/P(C) = 0.264$ is the probability that a family has a dog if we know that they have a cat.

Independence of events
A and B events are independent, if $P(A \cap B) = P(A) \cdot P(B)$, in other words, the occurrence of event A does not influence the occurrence of event B, and vice-versa.

Conditional probability
If event B occurred, what is the probability that event A occurs? $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, if $P(B) \neq 0$.

9.2 Law of total probability and Bayes theorem

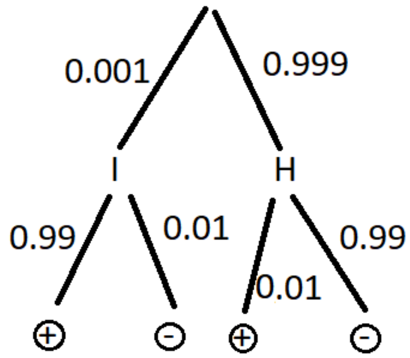
Example: A test of an illness shows false positive with 1 % probability for a healthy person and false negative with 1 % probability for an ill person. On average every 1000th people are ill. What is the probability of the illness of a man whose test is positive?

Solution: For simplicity let us make a table about 10000 random people.

	Positive test	Negative test	Together
Ill	$10 \cdot 0.99 = 9.9$	$10 \cdot 0.01 = 0.1$	10
Healthy	$9990 \cdot 0.01 = 99.9$	$9990 \cdot 0.99 = 9890.1$	9990
Together	109.8	9890.2	10000

Hence the probability of being ill if one has a positive test is $9,9/109,8 = 0,09$. The reason is that this sickness is very rare, therefore a person with a positive test has higher probability to be healthy than to be ill.

There is a more effective way to calculate this kind of problem. Let I, resp. H denote the event that a randomly chosen person is ill, resp. healthy. Let $\oplus$ denote that the test is positive, and $\ominus$ denotes the negative test.



The diagram shows that $P(I) = 0.001$, $P(H) = 0.999$, $P(\oplus|I) = 0.99$, $P(\oplus|H) = 0.01$ etc. From the diagram we can calculate $P(\oplus) = 0,001 \cdot 0,99 + 0,999 \cdot 0,01 = 0,01098$. (That is, $P(\oplus) = P(I) \cdot P(\oplus|I) + P(H) \cdot P(\oplus|H)$.) The answer is: $P(I|\oplus) = P(I \cap \oplus)/P(\oplus) = 0,001 \cdot 0,99/0,01098 = 0,09$.

Law of total probability:
Let A and B events, such that $P(B) \neq 0$ and $P(B) \neq 1$, then

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}).$$

In general:
Let $B_1, B_2, \dots$ be a system of mutually exclusive events whose union is the entire sample space, and $P(B_j) > 0$ for all j. Then for any event A

$$P(A) = \sum_{j=1}^{\infty} P(A|B_j)P(B_j).$$

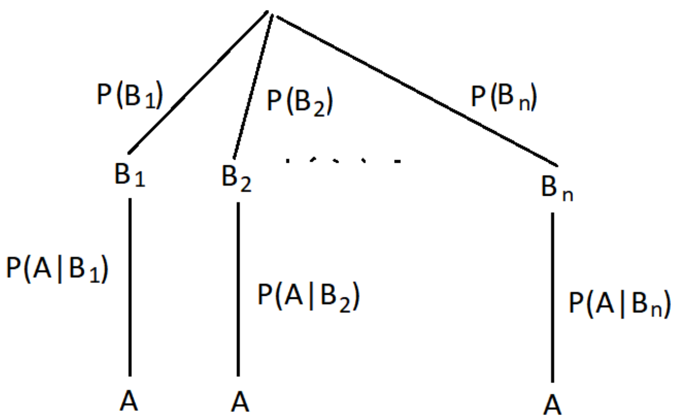
Bayes theorem: Let A and B be events, such that $P(A) \neq 0$, $P(B) \neq 0$ and $P(B) \neq 1$, then

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})}$$

In general:
Let $B_1, B_2, \ldots$ be a system of mutually exclusive events whose union is the entire sample space, and $P(B_j) > 0$ for all j. Let A be any event for which $P(A) > 0$, then

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum\limits_{j=1}^{\infty} P(A|B_j)P(B_j)}.$$

Instead of the formulas above we can use the following diagram to solve the exercise, like in the solution of the example.



9.3 Exercises

9.3.1

A and B are independent events, $P(A) = 0,5$, $P(B) = 0,3$. What is the probability of the following events?

(a) Both A and B happens.

(b) A happens, but B does not happen.

(c) A or B happens.

(d) What is the probability of A if we know that B happened?

Solution: (a) From the independence of the events we know $P(A \cap B) = P(A) \cdot P(B) = 0.5 \cdot 0.3 = 0.15$.

(b) $P(A \setminus B) = P(A) - P(A \cap B) = 0.35$.

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.65$.

(d) $P(A|B) = P(A \cap B)/P(B) = 0.5$. ($P(A|B) = P(A)$ holds for independent events.)

9.3.2

In a factory there are two machines A and B. They run independently of each other. A runs 60% of the time, and B runs 70% of the time. What are the probabilities of the following events?

a) Both machines run.

b) None of the machines run.

c) At least one of the machines run.

d) Exactly one machine runs.

Solution: Introduce the events A: machine A runs, B: machine B runs. $P(A) = 0.6$, $P(B) = 0.7$

(a) From the independence we get $P(A \cap B) = P(A) \cdot P(B) = 0.6 \cdot 0.7 = 0.42$.

(b) $P(\overline{A \cup B}) = 1 - P(A \cup B) = 0.12$.

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.7 - 0.42 = 0.88$.

(b) $P((A \setminus B) \cup (B \setminus A)) = P(A) + P(B) - 2P(A \cap B) = 0.46$.

9.3.3

Two six-sided fair dice are rolled, a red and a green one. Consider the following events:

$$A = \{\text{The sum of the numbers is 7.}\}, \quad B = \{\text{At least one of the die shows 6.}\},$$

$$C = \{\text{Both die shows an odd number.}\}, \quad D = \{\text{The dice show different numbers.}\},$$

$$E = \{\text{The green die shows 4.}\}, \quad F = \{\text{The red die shows 3.}\}.$$

(a) What is the probability of the above events?

(b) What is the probability of the common happening of the possible pairs (A and B, A and D etc.)? Which pairs of events are independent, which ones are mutually exclusive?

(c) Are the events A, E and F pairwise independent?

(d) Are the events A and $E \cap F$ independent? What can we conclude from this?

Solution: (a) $P(A) = 1/6$, $P(B) = 11/36$, $P(C) = 1/4$, $P(D) = 5/6$, $P(E) = 1/6$, $P(F) = 1/6$.

(b) Let's see some examples $P(A \cap B) = 1/18 \neq P(A) \cdot P(B)$, so A and B are not independent and not mutually exclusive.

$P(A \cap D) = P(A) = 1/6$, they are not independent and not mutually exclusive. B follows from A: if A happens, then B happens as well.

$P(A \cap C) = 0$, they are mutually exclusive.

$P(F \cap E) = P(F) \cdot P(E)$, they are independent.

$P(A \cap E) = P(A) \cdot P(E)$, they are independent.

(d) $P(A \cap E \cap F) = P(E \cap F) = 1/36$, so $P(A \cap E \cap F) \neq P(A) \cdot P(E \cap F)$, i.e. A and $E \cap F$ are not independent. Moreover $(E \cap F) \subset A$.

9.3.4

Is it possible that two events are independent and mutually exclusive at the same time?

Solution: $P(A \cap B) = P(A) \cdot P(B) = 0$, hence $P(A) = 0$ or $P(B) = 0$.

9.3.5

One die is rolled. What is the probability of the event that it shows 6, if we know that

(a) it shows an even number?

(b) it shows at least 3?

(c) it shows at most 5?

Solution: A: it shows 6, B: it shows even, C: it shows at least 3, D: it shows at most 5.

(a) $P(A|B) = P(A \cap B)/P(B) = |A \cap B|/|B| = 1/3$.

(b) $P(A|C) = P(A \cap C)/P(C) = |A \cap C|/|C| = 1/4$.

(c) $P(A|D) = 0$.

9.3.6

Two dice are rolled. What is the probability of the events that

(a) at least one of the dice shows 6 if we know that they show different numbers?

(b) at least one of the dice shows 2 if we know that the sum of the two numbers is 6?

Solution: A: at least one is 6, B: the two numbers are different, C: at least one is 2, D: the sum is 6.

(a) $P(A|B) = |A \cap B|/|B| = 10/30 = 1/3$.

(b) $P(C|D) = |C \cap D|/|D| = 2/5$.

9.3.7

There are 230 students and 30 teachers in a school. The doctor makes the following table about the pandemic:

	Sick	Healthy
Boy	50	60
Girl	40	80
Teacher	10	20

(a) We choose a medical card. What is the probability of the event that it is a medical card of a

1. boy?
2. sick person?
3. sick boy?

(b) We choose one of the cards of the girls. What is the probability of the event, that the chosen card is the card of a sick person?

(c) We choose one of the cards of the sick people. What is the probability of the event, that the chosen card is the card of a girl?

(d) We choose two of the cards of sick people. What is the probability of the event, that the first one is a teacher's card, the second one in a boy's card?

Solution: Let's summarize the tab:

	Sick	Healthy	Together
Boy	50	60	110
Girl	40	80	120
Teacher	10	20	30
Together	100	160	260

Let's denote the events with the initial letters B (boy), H (healthy) etc.

(a) (1): $P(B) = 110/260$, (2): $P(S) = 100/260$, (3): $P(S \cap B) = 50/260$.

(b) $P(S|G) = 40/120$.

(c) $P(G|S) = 40/100$.

(d) If we take the card back, the probability is $10 \cdot 50/100^2 = 0.05$.

If not, the probability is $10 \cdot 50/(100 \cdot 99)$.

9.3.8

János is a student at GTI. On a Wednesday he comes to the uni with probability $2/3$, and if he came in, he visits one of the 4 Math seminars with equal probability.

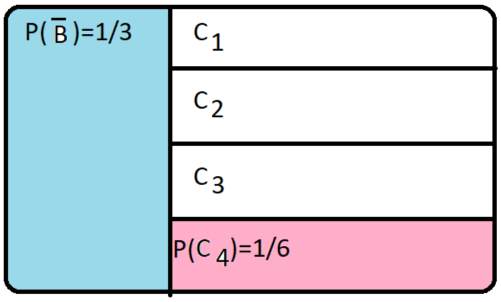
(a) What is the probability of the event that he visits the 3rd seminar?

(b) Somebody wants to find János. He was looking for him at the first 3 seminars, but she didn't find him. What is the probability of the event, that she will find János at the 4th seminar?

Solution: (a) Let B denote the event that János came in to the uni. $P(B) = 2/3$. C_i is the event that he went to the ith seminar. $P(C_1) = P(C_2) = P(C_3) = P(C_4) = 2/12 = 1/6$.

(b) If he is not at the first 3 seminars, two options remain: János is at the 4th seminar or he didn't came to the uni. That is $\overline{C_1 \cup C_2 \cup C_3} = C_4 \cup \overline{B}$. The answer is

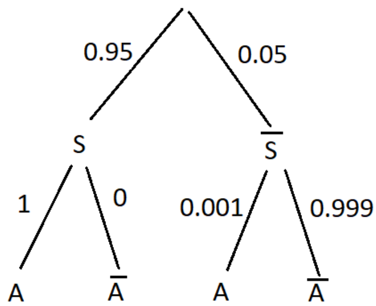
$$P(C_4|\overline{C_1 \cup C_2 \cup C_3}) = P(C_4|C_4 \cup \overline{B}) = \frac{P(C_4 \cap (C_4 \cup \overline{B}))}{P(C_4 \cup \overline{B})} = \frac{P(C_4)}{P(C_4 \cup \overline{B})} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{3}} = \frac{1}{3}.$$



9.3.9

Under the hill there is a crossing with flashing lights. 0.1 % of the drivers drive through the red sign. The sign is red on the average 5 % of the time. We spot from the hill that a car just drives through the crossing. What is the probability of the event, that the sign is free?

Solution: Let S be the event that the sign is free, $\bar{S}$ means that the sign is red. $P(\bar{S}) = 0.05$. Consider a random car. A is the event that the car drives through the crossing. Of course it always does, if the sign is free, hence $P(A|S) = 1$, and $P(A|\bar{S}) = 0.001$.



We have to calculate $P(S|A)$.

$$P(S|A) = \frac{P(S \cap A)}{P(A)}.$$

According to the diagram we have $P(S \cap A) = 0.95 \cdot 1 = 0.95$, $P(A) = 0.95 \cdot 1 + 0.05 \cdot 0.001$. $P(S|A) = 0.95/(0.95 + 0.001 \cdot 0.95) = 0.9999...$

9.3.10

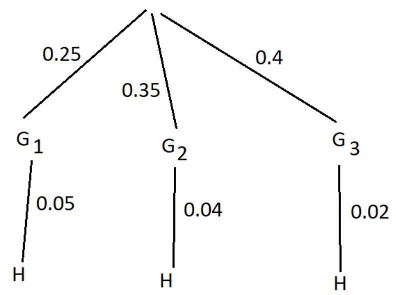
A factory makes doodads on three machines. 25 % of the doodads is made on the first machine, 35 % of them is made on the second machine. 5 % of the doodads made on the first machine, 4 % of the doodads made on the second machine, 2 % of the doodads made on the first machine are wrong.

(a) How many percent of all the doodads are wrong?

(b) The wrong doodads are collected in a container. We choose a doodads from this container. What is the probability of the event that this doodad was made on the first machine?

(c) We choose 5 from the doodads made on the second machine. What is the probability of the event that there is at least 1 wrong?

Solution: G_i denotes the event that a doodad is made on the i-th machine, and H means that a doodad is wrong.



(a) $P(H) = 0,25 \cdot 0,05 + 0.35 \cdot 0.04 + 0.4 \cdot 0.02 = 0.0345$.

(b) $P(G_1|H) = \frac{P(G_1 \cap H)}{P(H)} = \frac{0.25 \cdot 0.05}{0.0345} = 0.3623$.

(c) If we always take it back, the answer is $1 - 0,96^5$.

If we do no take it back, then we should know the number of all doodads. If this number is big, then the answer approximately agrees with the previous case.

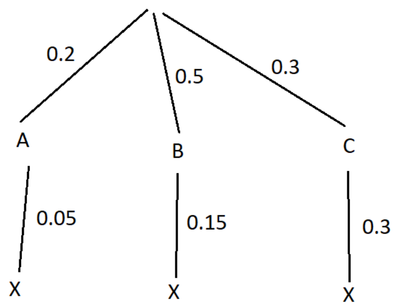
9.3.11

An auto insurance company classifies 20% of its drivers as low risks, 50% as medium risks, and 30% as high risks (called classes A, B, and C, respectively). The probability of at least one accident in a given year is 5 % for class A, 15 % for class B, and 30 % for class C.

(a) What percent of the costumers will have at least one accident this year?

(b) If one of their insured customer does not have an accident this year, what is the probability that he is a class A driver?

Solution:



Accident: X.

(a) $P(X) = 0.05 \cdot 0.2 + 0.15 \cdot 0.5 + 0.3 \cdot 0.3 = 0.175$.

(b)

$$P(A|\bar{X}) = \frac{P(A \cap \bar{X})}{P(\bar{X})} = \frac{0.2 \cdot 0.95}{1 - 0.175}.$$

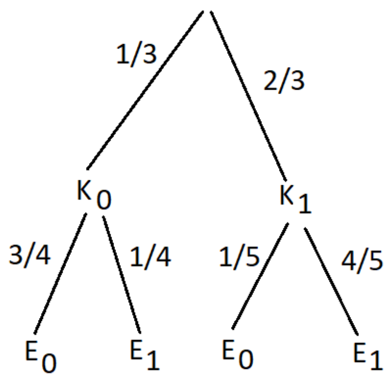
9.3.12

On a binary channel on average $1/3$ of the bits are 0, the rest are 1. The channel is noisy: when 0 is sent, then we get 1 with $1/4$ probability, and if 1 is sent, then we get 0 with $1/5$ probability.

(a) A bit is sent. What is the probability that we will get 0?

(b) We got 0. What is the probability that 0 were sent?

Solution: K_0 resp. K_1 are the events, that 0 resp. 1 is sent, and E_0 resp. E_1 denotes that we get 0 ot 1.



(a) $P(E_0) = \frac{1}{3} \cdot \frac{3}{4} + \frac{2}{3} \cdot \frac{1}{5} = \frac{23}{60}$.

(b) $P(K_0|E_0) = \frac{P(K_0 \cap E_0)}{P(E_0)} = \frac{1}{4} / \frac{23}{60} = \frac{15}{23}$.

10 Discrete random variables, distribution, expected value, standard deviation

10.1 Introduction

The result of an event, an experiment or a phenomenon usually scaled by real numbers, such as the result of rolling a dice, the winning numbers of a lottery or the speed of a car. For instance, the possible values of a die can be $1, 2, 3, 4, 5, 6$. We may ask the probability of getting 3 as the result of a roll, or the average of the results of two rolls. In these sentences the phrases 'the result of a roll' and 'the sum of two rolls' can be seen as *random variables.* The probability of the possible values of a random variable is given by its *distribution.* The expected value is a special average of a random variable which is important by its nature. It is also known as the expectation, mathematical expectation, mean or first moment. It might not be surprising that if we repeat an experiment n times independently under the same conditions and we take the arithmetic mean of the results, then it tends to the expected value as $n \to \infty$. Another important measure of a random variable is called *standard deviation* and its square called *variance.* Technically, it measures how far a set of numbers is spread out.

In the sequel we introduce notions and tools to be able to analyze and answer these type of questions.

10.2 Random variables

A *random variable* is a function that assigns a number to every outcome of an experiment. Random variables are generally denoted by capital Latin letters, preferring the letters at the end of the alphabet: X, Y etc., but lower case Greek letters are also common: ξ, η etc. The event that the value of X is equal to a number k is denoted by $\{X = k\}$ and its probability is by $P(X = k)$.

Examples for random variables:

1. When rolling a die, the number obtained can be thought of as a random variable:

$$X_1 = \begin{cases} 1, & \text{if rolling a 1} \\ 2, & \text{if rolling a 2} \\ 3, & \text{if rolling a 3} \\ 4, & \text{if rolling a 4} \\ 5, & \text{if rolling a 5} \\ 6, & \text{if rolling a 6} \end{cases}$$

   We can also define a random variable as follows:

$$X_2 = \begin{cases} 0, & \text{if the number rolled is even} \\ 1, & \text{if the number rolled is odd,} \end{cases}$$

   or

$$X_3 = 3.5, \text{ if the number rolled is } 1, 2, 3, 4, 5, 6.$$

2. We roll two dice. Let X denote the number of times we get a 4, then X is a random variable, with possible values of 0, 1 or 2. Let Y denote the sum of the numbers rolled, then Y is also a random variable, with possible values of $2, 3, \ldots, 12$. Let Z denote the number rolled with the first die, then Z is also a random variable, etc. Think of another random variable corresponding to this experiment!
3. For every event A there exists a so-called *indicator*, which is also a random variable:

$$I = \begin{cases} 1, \text{if event } A \text{ occurs,} \\ 0, \text{if event } A \text{ does not occur} \end{cases}$$

4. Sampling with or without replacement: let's suppose that there are balls is an urn/bag, of which some of them are red. We draw a few balls (w/ or w/o replacement). The number of red balls among the balls drawn is also a random variable.
5. The situation is the same as previously, but we draw balls one at a time (w/ replacement) until we get a red one. Let the random variable X be the step when we draw a red ball for the first time. This random variable has infinitely many possible values: $1, 2, 3, \ldots$.
6. A lottery game is a special case of sampling without replacement. The random variable, that corresponds to the previous example, is the number of the winning numbers chosen correctly. Another random variable can be the amount of money won.
7. Random experiments are often conducted repeatedly, independently, under the same conditions, so that the collective results may be subjected to statistical analysis. A fixed number of repetitions of the same experiment can be thought of as a composed experiment,

in which case the individual repetitions are called trials. For example, if one were to roll a die many times and record each result, each roll would be considered a trial within the experiment composed of all rolls. Or, for example, we draw a ball 100 times from an urn with replacement. If we observe an event A (related to the original experiment), the following two questions (random variables) commonly asked:

- Of n trials how many times does event A happen?
- How many trials are needed until the first time event A occurs?

Discrete random variables If a random variable X has finitely or infinitely, but countably many possible values we say that it is a *discrete random variable.* All of the above are examples of discrete random variables. In practice one often encounters continuous random variables, which can take values from an interval of the real numbers. For example:

- We generate a random number between 0 and 1. The random number is itself the random variable, its possible values are the elements of the interval $[0, 1]$.
- The bus is scheduled to go every 10 minutes, but we do not know exactly when it comes (it is not punctual). The random variable is the amount of time we have to wait at the bus stop. Its possible values are the elements of the interval $[0, 10)$ (in minutes).
- The length of the lifetime of a product (e.g. a laptop) or a creature (e.g. a human). The possible length is a positive real number in principle, i.e. this value is in the interval $(0, \infty)$. In practice, there can be set an upper bound, as for instance it is very rare that a human is 200 years old.
- A baker bakes 1 kg of bread, but of course it may not be exactly 1 kg. The actual weight depends on several random factors, therefore the weight of the bread is a random variable. In practice, it is reasonable to assume that the possible values of the weight of a bread is in interval $(0.99, 1.01)$.

10.3 Distributions of discrete random variables

In general, discrete random variables are given by their distribution, i.e., their possible values and the corresponding probabilities. If the random variable is finite, then its distribution can be given in *a table form* or by a *formula.* Example for a table form of a distribution:

- A random variable that has the value 1 or 0, according to whether a specified event occurs or not is called an indicator random variable for that event. Let A be an event with probability $P(A) = p$, then the indicator of A is the random variable X whose distribution can be given by the following table:

y	0	1
$P(X = y)$	$1 - p$	p

Further examples of discrete random variables given by formula will be introduced in the next chapter.

For every discrete random variable, the sum of the probabilities corresponding to each of its possible values equals to 1, i.e.,

$$\sum_{k} P(X = k) = 1.$$

For a random variable we can ask whether its value fall into a given interval or whether its value greater/smaller than a given number. Therefore, we can ask probabilities, such as: $P(X = 5)$, $P(X < 3)$, $P(-4 \leq X < 1)$ etc.

10.4 Expected value and standard deviation

Expected value of discrete random variables Let X be a (discrete) random variable with a finite number of outcomes $x_1, x_2, \ldots, x_n$ occurring with probabilities $p_1, p_2, \ldots, p_n$ respectively. Then the expected value of X is

$$E(X) = \sum_{k=1}^{n} P(X = x_k) \cdot x_k = P(X = x_1) \cdot x_1 + P(X = x_2) \cdot x_2 + \cdots + P(X = x_n) \cdot x_n = \sum_{k=1}^{n} p_k \cdot x_k$$

Since all probabilities add up to 1, the expected value is the weighted average of the possible values, with corresponding probabilities being the weights.

The above definition can be extended to random variables with infinitely but countably many outcomes $x_1, x_2, \ldots$ occurring with probabilities $p_1, p_2, \ldots$ respectively:

$$E(X) = \sum_{k=1}^{\infty} P(X = x_k) \cdot x_k = \sum_{k=1}^{\infty} p_k \cdot x_k \text{ (if this infinite sum is absolutely convergent).}$$

Standard deviation of discrete random variables The standard deviation is a measure of the spread/dispersion of a set of values, denoted by $D(X)$. By definition, the variance $D^2(X)$ is calculated as:

$$D^2(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2,$$

and the standard deviation is the square root of the variance:

$$D(X) = \sqrt{D^2(X)}.$$

A low standard deviation indicates that the values tend to be very close to the expected value, whereas high standard deviation indicates that the data are spread out over a large range of values.

Examples (cont'd):

$$E(X_1) = 1 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3.5 \text{ and } D(X_1) \approx 1.7$$

$$E(X_3) = 1 \cdot 3.5 = 3.5 \text{ and } D(X_3) = 0$$

10.5 Exercises

10.5.1

An event A has the probability $P(A) = 0.3$. Let X be the indicator of A, i.e.

$$X = \begin{cases} 0 & \text{if } A \text{ does not happen,} \\ 1 & \text{if } A \text{ does happen.} \end{cases}$$

(a) What is the expected value of the random variable X?

(b) What is its standard deviation?

Solution:

(a) $E(X) = 1 \cdot P(A) + 0 \cdot P(\overline{A}) = 0.3$

(b) $D^2(X) = E(X^2) - (E(X))^2 = 0.3 - (0.3)^2 = 0.21$, thus $D(X) = \sqrt{D^2(X)} \approx 0.46$

10.5.2

We roll a die. What is the expected value and standard deviation of the number we roll? Explain the results!

Solution: Let X = the number rolled with the die. Then

$$E(X) = 1 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3.5$$

$$E(X^2) = 1^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6} = \frac{91}{6} \approx 15.167$$

$$D(X) = \sqrt{E(X^2) - E(X)^2} = \sqrt{15.167 - (3.5)^2} \approx 1.708$$

10.5.3

We roll two dice. What is the expected value and standard deviation of the sum of the two numbers we roll? Explain the results!

Solution: Let X = the sum of the two numbers that we roll. The possible values of X are $k = 2, \ldots, 12$ and its distribution

$$P(X = k) = \begin{cases} 1/36, & \text{if } k = 2 \text{ or } k = 12 \\ 2/36, & \text{if } k = 3 \text{ or } k = 11 \\ 3/36, & \text{if } k = 4 \text{ or } k = 10 \\ 4/36, & \text{if } k = 5 \text{ or } k = 9 \\ 5/36, & \text{if } k = 6 \text{ or } k = 8 \\ 6/36, & \text{if } k = 7 \end{cases}$$

Therefore

$$E(X) = \sum_{k=2}^{12} k \cdot P(X = k) = 7$$

$$E(X^2) = \sum_{k=2}^{12} k^2 \cdot P(X = k) \approx 54.83$$

$$D(X) = \sqrt{E(X^2) - E(X)^2} \approx 2.415$$

10.5.4

Both Elemér and Huba roll a die. Huba will give an amount of money to Elemér which is equal to the sum of the numbers rolled, while Elemér will give an amount of money to Huba which is equal to the square of the difference of the two numbers. Which person is favored in this game?

Solution: Let X = sum of the numbers rolled, then $E(X) = 7$ (getting from the previous example). Let Y = squared difference of the two numbers, then

k	0^2	1^2	2^2	3^2	4^2	5^2
P(Y=k)	6/36	10/36	8/36	6/36	4/36	2/36

$E(Y) = 0 \cdot \frac{6}{36} + 1 \cdot \frac{10}{36} + 2^2 \cdot \frac{8}{36} + 3^2 \cdot \frac{6}{36} + 4^2 \cdot \frac{4}{36} + 5^2 \cdot \frac{2}{36} \approx 5.83$

10.5.5

The possible values of the random variable X and their respective probabilities are found in the table below.

(a) Calculate the value of $q = P(X = 3)$. What does it mean? Formulate your answer in a complete sentence.

y	-1	0	3	4
$P(X = y)$	0.1	0.3	q	0.4

(b) Calculate the following values: $P(X < 1), P(3 \leq X), P(-1 < X \leq 3)$. What do these mean?

(c) Calculate $E(X)$ and $D(X)$. Explain their meaning.

Solution:

(a) $q = 1 - 0.1 - 0.3 - 0.4 = 0.2$

(b) $P(X < 1) = 0.1 + 0.3 = 0.4,\ P(3 \leq X) = 0.2 + 0.4 = 0.6,\ P(-1 < X \leq 3) = 0.3 + 0.2 = 0.5$

(c) $E(X) = -1 \cdot 0.1 + 0 \cdot 0.3 + 3 \cdot 0.2 + 4 \cdot 0.4 = 2.1$
$E(X^2) = 1 \cdot 0.1 + 0 \cdot 0.3 + 9 \cdot 0.2 + 16 \cdot 0.4 = 8.3$
$D(X) = \sqrt{E(X^2) - E(X)^2} = 1.97$

10.5.6

In a gambling game, there are 40 thousand lottery tickets being sold. One of them is worth 1 million HUF, 10 of them are worth 50 thousand HUF and 100 of them are worth 5000 HUF. We buy one ticket. What is the expected amount of money we win?

Solution: $E = 10^6 \cdot \frac{1}{40000} + 50000 \cdot \frac{10}{40000} + 5000 \cdot \frac{100}{40000} = 50\text{HUF}$

10.5.7

The distribution of a random variable X is given by the expression $P(X = k) = k^2/30$, where $k = 1, 2, 3, 4$.

(a) Check that the above is indeed a probability distribution!

(b) Calculate the expected value and the standard deviation of X.

Solution:

(a) $\frac{1^2}{30} + \frac{2^2}{30} + \frac{3^2}{30} + \frac{4^2}{30} = 1$

(b) $E(X) = 1 \cdot \frac{1^2}{30} + 2 \cdot \frac{2^2}{30} + 3 \cdot \frac{3^2}{30} + 4 \cdot \frac{4^2}{30} = \frac{10}{3}$ and $D(X) = \sqrt{1^2 \cdot \frac{1^2}{30} + 2^2 \cdot \frac{2^2}{30} + 3^2 \cdot \frac{3^2}{30} + 4^2 \cdot \frac{4^2}{30} - (\frac{10}{3})^2} = \frac{1}{3}\sqrt{\frac{31}{5}}$.

10.5.8

We roll two dice. Let X be the greater or equal of the two numbers, and let Y be the smaller or equal of the two. What is the expected value of X and Y?

Solution: First we produce the table of distributions of X and Y

k	1	2	3	4	5	6
P(X=k)	1/36	3/36	5/36	7/36	9/36	11/36
P(Y=k)	11/36	9/36	7/36	5/36	3/36	1/36

We can also express these distributions by formulas, i.e. $P(X = k) = \frac{2k-1}{36}$, $P(Y = k) = \frac{13-2k}{36}$. Then the expected value of X is

$$E(X) = \sum_{k=1}^{6} \frac{k(2k-1)}{36} = \frac{161}{36} \approx 4.47.$$

Note that $P(X = k) + P(Y = k) = \frac{1}{3}$. Clearly

$$E(X + Y) = \sum_{k=1}^{6} k \cdot (P(X = k) + P(Y = k)) = \frac{\sum_{k=1}^{6} k}{3} = 7.$$

Thus $E(Y) = 7 - \frac{161}{36} = \frac{91}{36}$.

10.5.9

For a school excursion, the students are transported by 4 buses, which have room for 40, 33, 25 and 50 people, respectively. We randomly choose a student. Let X denote the number of passengers traveling in the bus the student is in (including the chosen student as well). We randomly choose a bus driver. Let Y denote the number of passengers traveling in the bus that the driver belongs to. Find the distribution of the two random variables, calculate their expected value and standard deviation. Before doing any calculation, make a guess: which expected value will be greater and why?

Solution: Altogether there are room for $40 + 33 + 25 + 50 = 148$ passengers on the 4 buses.
X: number of passengers on bus w/ chosen student Y: number of passengers on bus w/ chosen driver

$E(X) = 40 \cdot \frac{40}{148} + 33 \cdot \frac{33}{148} + 25 \cdot \frac{25}{148} + 50 \cdot \frac{50}{148} \approx 39.3$

$E(Y) = 40 \cdot \frac{1}{4} + 33 \cdot \frac{1}{4} + 25 \cdot \frac{1}{4} + 50 \cdot \frac{1}{4} = \frac{148}{4} = 37$

11 Special random distributions

11.1 Expected value and standard deviation of the sum and the average of random variables

- Let X and Y be real random variables[1] with finite expected values $E(X)$ and $E(Y)$, then

$$E(aX + bY) = aE(X) + bE(Y),$$

for all $a, b \in \mathbb{R}$.

- If X and Y are independent, then for the standard deviations $D(X)$ and $D(Y)$

$$D(aX + bY) = \sqrt{a^2(D(X))^2 + b^2 D(Y))^2}$$

holds.

- If we repeat an experiment n times independently under the same conditions, the trials are represented by the random variables $X_1, \dots X_n$. (These random variables have the same distribution.) The sum of these random variables are denoted by $S_n = X_1 + \cdots + X_n$. Then the expected value and standard deviation of S_n can be calculated as follows

$$E(S_n) = nE(X) \text{ and } D(S_n) = \sqrt{n}D(X).$$

where X is a random variable having the same distribution as each X_i ($i = 1, \dots n$).

- Now we take the average of $X_1, \dots X_n$, which is also a random variable that we denote by $\overline{X}$. Then the expected value and standard deviation of $\overline{X}$ can be calculated as follows

$$E(\overline{X}) = E(X) \text{ and } D(\overline{X}) = \frac{D(X)}{\sqrt{n}}.$$

11.2 Special probability distributions: indicator, binomial distribution, geometric and hypergeometric distribution

The following distributions of discrete random variables are called *special*, because they appear regularly in calculations and make the problem solving easier if we recognize them in the problem. These distributions can be given by a formula.

[1] Real random variable means that all of its possible values are real numbers

- In an urn there are N balls, of which M of them are red. We randomly draw n balls, and let X denote the number of red balls drawn.
  - If we drew without replacement, then
  $$P(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} \quad k = 0, 1, \ldots, n.$$
  In this case we say that the random variable X has a *hypergeometric distribution.* Notation: $X \sim Hypergeom(n, M, N)$.
  - If we drew with replacement, then
  $$P(X = k) = \binom{n}{k} \cdot \left(\frac{M}{N}\right)^k \cdot \left(\frac{N-M}{N}\right)^{n-k} \quad k = 0, 1, \ldots, n.$$
- We repeat an experiment n times independently of each other, and we count the number of times the event A happens. The probability of the event A is $P(A) = p$. Let X be the random variable of the number of times A happens during the n trials. Then
  $$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad k = 0, 1, \ldots, n,$$
  The random variable X has a *binomial distribution.* Notation: $X \sim Binom(n, p)$.

  Note: Sampling with replacement above is a special case of this, as each draw is independent of the others. Then the event A is drawing a red ball once, $P(A) = p = \frac{M}{N}$.
- We repeat an experiment independently of each other until the event A of interest happens ($P(A) = p$). Let the random variable X denote the number of trials needed for a successful experiment (i.e., when A happens). The random variable X has infinitely many possible values.
  $$P(X = k) = (1-p)^{k-1} \cdot p \quad k = 1, 2, \ldots$$
  In this case we say that the random variable X has a *geometric distribution.* Notation: $X \sim Geom(p)$.

11.3 Expected values and standard deviations of these special distributions

In the following table we collect the expected value and the variance of those special distributions that we discussed above.

	$E(X)$	$D(X)$
Indicator(p):	p	$p(1-p)$
Binomial(n,p):	np	$\sqrt{np(1-p)}$
Geometric(p):	$\frac{1}{p}$	$\sqrt{\frac{1-p}{p^2}}$
Hypergeometric(N,M,n):	$n\frac{M}{N}$	$\sqrt{n\frac{M}{N}(1-\frac{M}{N})(1-\frac{n-1}{N-1})}$
		(do not need to remember this one)

11.4 Law of large numbers

The Law of Large Numbers (LLN) is one of the single most important theorems in probability theory. It demonstrates the fundamental relationship between the concepts of probability and frequency.

If we roll a fair six-sided die, the mean of the number we get is 3.5. If we roll the die a large number of times (n times) and average the numbers we get (i.e., we compute $\overline{X}$), then we expect something close to 3.5. And we expect to get closer and closer as n gets larger and larger. In general, the LLN states that if an experiment is repeated independently a large number of times, the average of the results of the trials must be close to the expected value and the result becomes closer to the expected value as the number of trials increased.

As a special case, let X be the indicator random variable for an event A. It is easy to show that $E(X) = P(A)$, and $\overline{X}$ is the relative frequency of event A. Thus, from above we get that the relative frequency will get closer and closer to the probability. For example, flipping a regular coin many times results in approximately 50% heads and 50% tails frequency, since the probabilities of those outcomes are both 0.5.

11.5 Exercises

11.5.1

(a) We toss a coin until we get a head. What is the probability of getting a head for the first time at the 21st toss?

(b) For which value of n has the highest probability of getting a head for the first time at the nth toss?

(c) We toss a coin five times in a row. What is the probability of getting exactly 3 heads?

<u>Solution</u>:

(a) $(\frac{1}{2})^{21}$

(b) $n = 1$

(c) $\binom{5}{3}(\frac{1}{2})^3(\frac{1}{2})^2$

11.5.2

(a) We roll a die repeatedly. What is the probability of rolling a 2 for the first time at the 11th try?

(b) For which value of n has the highest probability of rolling a 2 for the first time at the nth try?

(c) We roll a die 6 times in a row. What is the probability of getting a 2 exactly 4 times?

Solution:

(a) $(\frac{5}{6})^{10}(\frac{1}{6})$

(b) $n = 1$

(c) $\binom{6}{4}(\frac{1}{6})^4(\frac{5}{6})^2$

11.5.3

There are 2 red and 3 blue balls in an urn. We draw a ball from the urn (with closed eyes), we look at the ball, then we put the ball back and repeat this process.

(a) What is the probability of drawing a red ball for the first time at the 5th draw?

(b) What is the probability of having drawn a red ball 3 times out of 8 draws?

Solution:

(a) $(\frac{3}{5})^4(\frac{2}{5})$

(b) $\binom{8}{3}(\frac{2}{5})^3(\frac{3}{5})^5$

11.5.4

There are 4 red and 6 blue balls in an urn. We draw a ball 3 times from the urn (with closed eyes). What is the expected number of red balls drawn, if

(a) we put back the ball after each draw?

(b) we do not put the ball back after any draw?

What is the standard deviation of the number of red balls drawn (in both cases)?

Solution:

(a) $X_1 \sim Binom(n = 3, p = \frac{4}{10})$, therefore $E(X_1) = 3 \cdot \frac{4}{10}$, $D(X_1) = \sqrt{3\frac{4}{10}(1 - \frac{4}{10})}$

(b) $X_2 \sim Hypergeom(n = 3, M = 4, N = 10)$, therefore
$E(X_2) = 3 \cdot \frac{4}{10}$, $D(X_2) = \sqrt{3\frac{4}{10}(1 - \frac{4}{10})(1 - \frac{3-1}{10-1})}$

11.5.5

We toss a fair coin 5 times in a row. What is the expected value and the standard deviation of the number of heads?

Solution: Let X = the number of heads, then $X \sim Binom(n = 5, p = \frac{1}{2},)$ therefore $E(X) = 5 \cdot \frac{1}{2}$ and $D(X) = \sqrt{5\frac{1}{2}(1-\frac{1}{2})}$.

11.5.6

We roll 4 dice. What is the expected value of the number of 3s? What is its standard deviation?

Solution: $E = 4 \cdot \frac{1}{6}, D = \sqrt{4\frac{1}{6}(1-\frac{1}{6})}$.

11.5.7

In a state lottery, 5 numbers are drawn out of 90 (without replacement). If we score 1 (or none of them) then we do not win anything. If we score 2, our prize is 700 HUF, for a score of 3 we get 10000 HUF, for a score of 4 we get 789 thousand HUF. For a perfect score we get 535 million HUF. What is the expected amount of money we win? A standard ticket costs 250 HUF. Is it worth playing the game?

Solution: $X \sim Hypergeom(n = 5, M = 5, N = 90)$.

number of scores	0	1	2	3	4	5
probability	$\frac{\binom{5}{0}\binom{85}{5}}{\binom{90}{5}}$	$\frac{\binom{5}{1}\binom{85}{4}}{\binom{90}{5}}$	$\frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}}$	$\frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}}$	$\frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}}$	$\frac{\binom{5}{5}\binom{85}{0}}{\binom{90}{5}}$
money (HUF) we win	0	0	700	10000	$789 \cdot 10^3$	$535 \cdot 10^6$

$$E = 0 \cdot \frac{\binom{5}{0}\binom{85}{5}}{\binom{90}{5}} + 0 \cdot \frac{\binom{5}{1}\binom{85}{4}}{\binom{90}{5}} + 700 \cdot \frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}} + 10000 \cdot \frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}} + 789000 \cdot \frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}} + 535000000 \cdot \frac{\binom{5}{5}\binom{85}{0}}{\binom{90}{5}}$$

11.5.8

In a survey, on average one person out of five will answer the question. The willingness of people answering does not depend on the others.

(a) If we ask 100 people, what is the probability that exactly 30 of them will answer the question?

(b) If we ask 100 people, what is the probability that none of them will answer the question?

(c) If we ask 100 people, what is the probability that the 10th person being asked will be the first to answer the question?

Solution: Let X = the number of people answer, then $X \sim Binom(n = 100, p = \frac{1}{5}$

(a) $P(X = 30) = \binom{100}{30}(\frac{1}{5})^{30}(\frac{4}{5})^{70}$

(b) $P(X = 0) = (\frac{4}{5})^{100}$

(c) $P(10th) = (\frac{4}{5})^9(\frac{1}{5})$

11.5.9

(Same as 10.5.2) We roll two dice. What is the expected value and standard deviation of the sum of the two numbers we roll? Explain the results!

Solution: Let X = the result of the first roll and Y = the result of the second rolls. In this case, $X + Y$ gives the sum of the two roll. Then

$$E(X + Y) = E(X) + E(Y) = 7$$

$$D(X + Y) = \sqrt{D(X)^2 + D(Y)^2} = \sqrt{2}D(X) \approx 2.415$$

11.5.10 ⋆

(a) Show that the binomial distribution is a distribution, i.e, the sum of the possible probabilities are 1. What could be the reason that it is called binomial?

(b) Show that the geometric distribution is a distribution, i.e, the sum of the possible probabilities are 1. What could be the reason that it is called geometric?

Solution:

(a) Let $X \sim Binom(n, p)$. Then $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ for $0 \le k \le n$. Binomial theorem says that

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

where $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$. Applying this for $a = p$ and $b = 1 - p$ we get that

$$1 = 1^n = (p + (1 - p))^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k}.$$

(b) Let $Y \sim Geom(p)$. Then $P(X = k) = p^{k-1}(1 - p)$ for $k = 1, 2, \ldots$. Therefore

$$\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} p^{k-1}(1 - p) = (1 - p) \sum_{l=0}^{\infty} p^l,$$

where $l = k - 1$. We an exclude the trivial cases when $p = 0$ or $p = 1$, since $p = 1$ has no meaning and $P(Y = 1) = 1$ trivially holds for $p = 0$. Thus $0 < p < 1$ can be assumed. In this case $(p^l)_{l=1}^{\infty}$ is the sequence of a geometric progression, hence $\sum_{l=0}^{n} p^l = \frac{1-p^{n+1}}{1-p}$ for every $n \in \mathbb{N}$ using $0 < p < 1$. If $p < 1$, then $p^{n+1} \to 0$ and hence $lim_{n\to\infty} \sum_{l=0}^{n} p^l = \sum_{l=0}^{\infty} p^l = \frac{1}{1-p}$. Thus $\sum_{k=1}^{\infty} P(X = k) = (1 - p) \sum_{l=0}^{\infty} p^l = \frac{1-p}{1-p} = 1$.

11.5.11 ⋆

Suppose there are $2N$ cards in a box, of which two cards have 1 on it, two cards have 2 on it, and so on. Let's choose m cards randomly. What is the expected number of pairs of cards left in the box?

Solution: Let X_i be the indicator that both card with an i written on it remains in the deck after removing m cards, i.e.,

$$X_i = \begin{cases} 1 & \text{if both cards with an } i \text{ written on it remains in the deck} \\ 0 & \text{otherwise} \end{cases},$$

then

$$p = E(X_i) = P(X_i = 1) = \frac{\binom{2N-2}{2m}}{\binom{2N}{m}},$$

with the convention that $\binom{2N-2}{2N-1} = \binom{2N-2}{2N} := 0$. Let X be the number of pairs remaining in the box after taking the m cards out. Then $X = X_1 + \cdot + X_N$, whose expected value is

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_N) = N \cdot p = N \cdot \frac{\binom{2N-2}{2m}}{\binom{2N}{m}} = \frac{(2N-m)(2N-1-m)}{2(2N-1)}.$$

The standard deviation of X can be expressed similarly, i.e.

$$D(X) = \frac{1}{\sqrt{N}}(D(X_1) + \cdots + D(X_N)) = \frac{p(1-p)}{\sqrt{N}} =$$
$$\frac{(2N-m)(2N-m-1)(2N(2N-1)-(2N-m)(2N-m-1))}{(2N(2N-1))^2} =$$
$$\frac{m(2N-m)(2N-m-1)(4N-m+1)}{(2N(2N-1))^2}$$

11.5.12 ⋆

Prove the following properties of the expected value and the standard deviation. (X and Y are random variables with finite expected values and standard deviations.)

- $E(X+Y) = E(X) + E(Y)$,
- $E(a \cdot X) = a \cdot E(X)$,
- $E(a \cdot X + b) = a \cdot E(X) + b$, where a and b are real numbers.
- In repeated experiments the expected value of the average:$E\left(\overline{X}\right) = E(X)$.
- $D^2(a \cdot X) = a^2 \cdot D^2(X)$.
- $D(a \cdot X) = |a| \cdot D(X)$,
- $D(a \cdot X + b) = |a| \cdot D(X)$, where a and b are real numbers.

- ** If X and Y are independent (i.e, every event depending only on X and only on Y, respectively, are independent), then $E(X \cdot Y) = E(X) \cdot E(Y)$.
- If X and Y independent, then $D^2(X + Y) = D^2(X) + D^2(Y)$.
- The expected value and the standard deviation of the average of independently repeated experiments are the following:

$$E\left(\overline{X}\right) = E(X) \text{ és } D\left(\overline{X}\right) = \frac{D(X)}{\sqrt{n}}.$$

12 Markov-chains

12.1 Markov chains – theory

12.1.1 Example

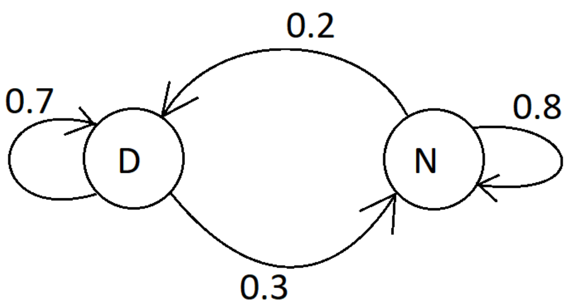
In a country the effect of the parents' education on the highest education level for their children was investigated. It was found that 70% of those whose parents had a university degree also got a university degree. Of those whose parents did not have a university degree only 20% got a university degree.

a) What is the probability that grandparents with a university degree have grandchildren with a university degree?

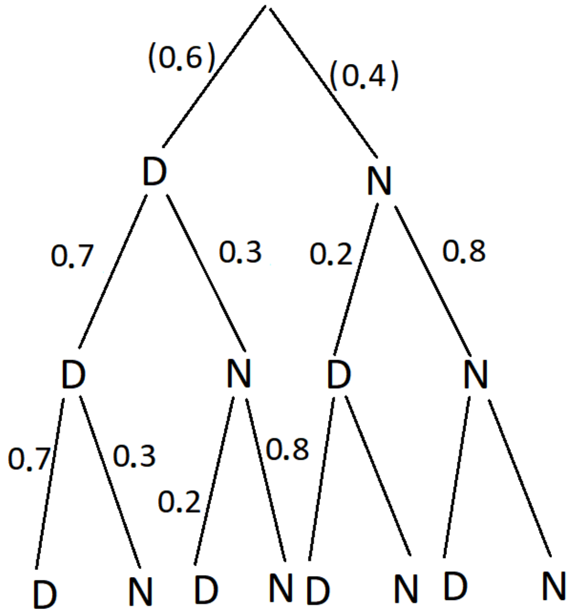
b) At present 60% of the population has a university degree. What percent of the next two generations will have a university degree?

c) What is the stationary distribution?

Solution: The mixture between the generations is shown by the diagram (D: has a degree, N: does not have a degree):



or by a tree known from Chapter 9:



From the diagram we can directly solve the questions a) and b) (as in Chapter 9):

a) $0.7 \cdot 0.7 + 0.3 \cdot 0.2 = 0.55$.

b) In the generation of the children $0,6 \cdot 0,7 + 0,4 \cdot 0,2 = 0,5$ is the ratio of the people who has a degree, therefore that ratio is $0,5 \cdot 0,7 + 0,5 \cdot 0,2 = 0,45$ in the generation of the grandchildren.

But how can we do the calculations if the question is about the 10th generation? Let's tabulate the data!

	Child (D)	Child (N)
Parent (D)	0.7	0.3
Parent (N)	0.2	0.8

Write it in a matrix form, called transition matrix:

$$M = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}.$$

(The order of D and N does not matter, but it is important to fix it once and for all.)

Consider the matrix $M^2 = M \cdot M$.

$$M^2 = \begin{pmatrix} 0.7 \cdot 0.7 + 0.3 \cdot 0.2 & 0.7 \cdot 0.3 + 0.3 \cdot 0.8 \\ 0.2 \cdot 0.7 + 0.8 \cdot 0.2 & 0.2 \cdot 0.3 + 0.8 \cdot 0.8 \end{pmatrix} = \begin{pmatrix} 0.55 & 0.45 \\ 0.3 & 0.7 \end{pmatrix}.$$

Comparing with the tree diagram we can see that M^2 is the transition matrix to the generation of the grandchildren. In general, M^n is matrix of the transition to the n-th generation.

c) The vector of the initial distribution is $v = (0.6; 0.4)$. The distribution of the children's generation is $v \cdot M$, the grandchildren's generation is $v \cdot M^2$, in general the distribution vector of the n-th generation is $v \cdot M^n$.

The stationary distribution is a distribution vector $w = (x; y)$ such that $w \cdot M = w$ holds. That is a linear equation:

$$\left.\begin{array}{rcl} 0.7x + 0.2y & = & x \\ 0.3x + 0.8y & = & y \end{array}\right\}$$

The equations are not independent, but we have an extra equation, which expresses that w is a distribution vector: $x + y = 1$. The unique solution of the system is $w = (0.4; 0.6)$. That means, if the initial distribution is $40\% - 60\%$ than it will be the same in every generation.

Why is the stationary distribution important? Because it tells us the long term behavior. Starting with any initial distribution, after several generations the distribution will be about $40\% - 60\%$.

12.1.2 Theory

We describe a Markov chain as follows: suppose we have a set of states, $S = \{S_1, S_2, \dots S_m\}$. The process starts in one of these states and moves successively from one state to another. Each move is called a *step*. If the chain is currently in state S_i, then after one unit of time it moves to state S_j at the next step with a probability denoted by p_{ij}, and the probability does not depend upon which states the chain was in before the current state, but it depends on the current state as the dependence of p_{ij} on i indicates. Even after one unit of time, the process can remain in the state it is in, and this occurs with probability p_{ii}. The probabilities p_{ij} are called *transition probabilities* and can be collected into a *transition matrix* P. (In this matrix, the nonzero matrix elements correspond to the arrows in the diagram. The rows represent inputs, the columns represent outputs.) P is an $m \times m$ square matrix with as many rows and columns as the number of states in the system. All elements of the matrix are nonnegative and the sum of each row equals to 1, i.e., $\sum_{j=1}^{m} p_{ij} = 1$, therefore these rows are probability distributions. Matrices of this kind will always have an eigenvalue of 1 with a corresponding eigenvector $(1, \dots 1)^T$ (and all other eigenvalues are in absolute value smaller or equal to 1).

Additionally, a Markov chain also has an *initial state vector*, u, represented as a $1 \times m$ row vector, that describes the probability distribution of starting at each of the m possible states. Entry i of the vector describes the probability of the chain beginning at state S_i.

Transitioning to the nth generation

We now know how to obtain the probability of transitioning from one state to another, but how about finding the probability of that transition occurring over multiple steps (assuming that the transition matrix remains constant with each generation)? To formalize this, we wish to determine the probability of moving from state S_i to state S_j over n steps. It turns out, this is relatively simple to find out. Given a transition matrix P, this can be determined by calculating the value of entry ij of the matrix obtained by raising P to the power of n. For small values of n, this can easily be done by hand with repeated multiplication. However, for large values of n, a more efficient way to raise a matrix to a power is to use the eigenvalues and eigenvectors of the matrix. If we start a Markov chain with initial state vector u, then the vector $u \cdot P^n$ gives the probabilities of being in the various states after n steps (n-state probabilities).

Stationary (stable) distribution

Recall that a matrix and its transpose have the same eigenvalues ($det(A^T - \lambda I) = det(A^T - \lambda I^T) = det((A - \lambda I)^T) = det((A - \lambda I))$. As already mentioned above, 1 is an eigenvalue of matrix P, therefore it is an eigenvalue of matrix P^T. Thus there is a column vector of size $n \times 1$, v, such that $P^T \cdot v = v$. With some additional conditions on P this vector is unique up to a constant multiple and all its entries are positive. We can thus normalize this vector so that it is a probability vector, i.e., its entries are positive and their sum is 1. Transposing this equation, we get that $v^T \cdot P = v^T$, in other words, there is a row vector $w(= v^T)$, such that $w \cdot P = w$. This w is called the left eigenvector of the matrix P. Then w describes the state of the system that will not change by time. This is called the stationary (steady) state. The uniqueness of the eigenvector is guaranteed if all elements of P are positive, i.e., there is a transition between any two states, in fact a weaker condition suffices, i.e., not all elements of P need to be positive, it is sufficient that all elements of P^k are positive for some $k \in \mathbb{N}$

Long-term behavior tends to the stationary distribution

The stationary state w also describes the long-term behavior of the system. It can be shown that starting from any state u, $u \cdot P^n$ gets closer and closer to w as n gets larger and larger.

12.2 Markov chains exercises

12.2.1

In a country the effect of the parents' education on the highest education level for their children was investigated. It was found that 3/4 of those whose parents had a university degree also got a university degree. Of those whose parents did not have a university degree only 1/3 got a university degree.

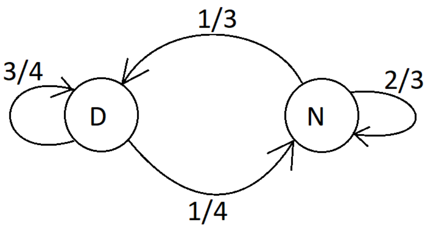
a) Draw the diagram that represents this process. Write down the transition matrix.

b) What is the probability that grandparents with a university degree have grandchildren without a university degree.

c) At present 2/3 of the population has a university degree. What percent of the next two generations will have a university degree?

d) What is the stationary distribution?

Solution: (a)



$$M = \begin{pmatrix} 3/4 & 1/4 \\ 1/3 & 2/3 \end{pmatrix}$$

(b) Second entry in the first row of M^2 is $17/48$.

(c) $v = (2/3, 1/3)$. The distribution of the children is $v \cdot M = (11/18; 7/18)$. Grandchildren: $v \cdot M^2 = (127/216; 89/216)$.

(d) The question is a $w = (x; y)$ distribution vector such that $w \cdot M = w$ holds. That is:

$$\begin{array}{rcl} x + y & = & 1 \\ 3/4x + 1/3y & = & x \\ 1/4x + 2/3y & = & y \end{array}$$

Solution: $w = (4/7; 3/7)$.

12.2.2

It is convenient to classify people by their income as lower, middle. or upper class. Sociologists have found that the strongest determinant of the income class of an individual is the income class of their parents. They found that 65% of children born into upper class will also become upper class, 28% becomes middle class and 7% becomes lower class. Of those born into middle class, 15% becomes upper class, 67% stay middle class and 18% becomes lower class. Finally, if someone is born into lower class income family, the probability that they remain lower class is 52%, the probability that they become middle class is 36% the probability that they become upper class is 12%.

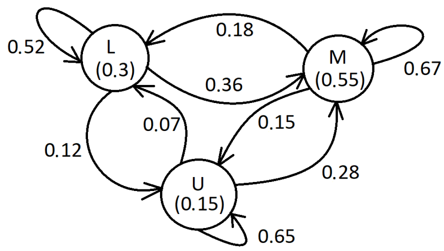
a) Draw a diagram representing the transition between these states.

b) If a couple is upper class, what is the probability that there grandchildren become lower class?

c) Currently 15% of people are upper class, 55% of people are middle class and 30% are lower class. What will be distribution of the social classes in the next generation, i.e. among the children. What will be the distribution among the grandchildren?

d) Find the stationary distribution. (This is the distribution that does not change from generation to generation.)

Megoldás: (a)



$$M = \begin{pmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{pmatrix}$$

(b) Third entry in the first row of M^2: 0.1323.

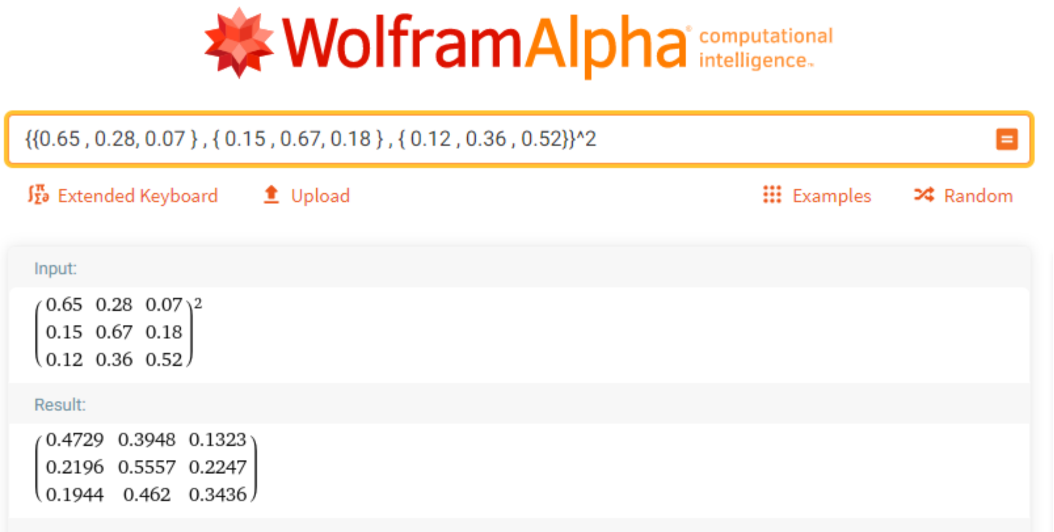
(c) $v = (0.15; 0.55; 0.3)$. The distribution of the children is $v \cdot M = (0.216; 0.5185; 0.2655)$. Grandchildren: $v \cdot M^2 = (0.25; 0.5; 0.246)$.

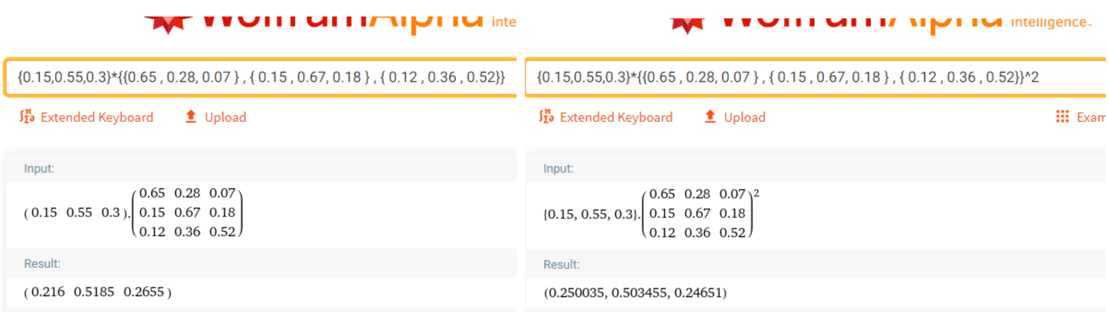
(d) The question is a distribution vector $w = (x; y; z)$ such that $w \cdot M = w$ holds. That is:

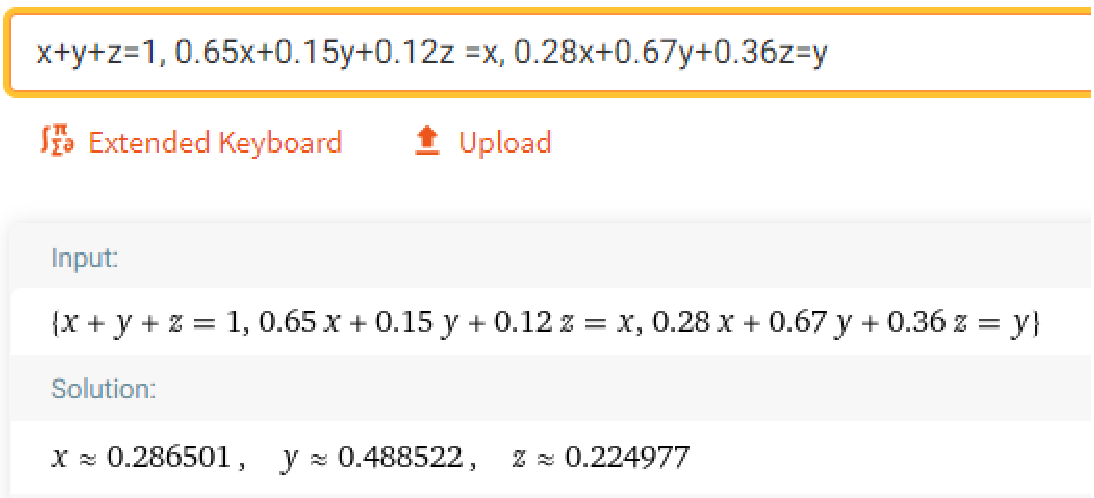
$$\begin{array}{rcl} x + y + z & = & 1 \\ 0.65x + 0.15y + 0.12z & = & x \\ 0.28x + 0.67y + 0.36z & = & y \\ 0.07x + 0.18y + 0.52z & = & z \end{array}$$

Solution: $w = (0.2865; 0.4885; 0.225)$.

WolframAlpha helps in calculation:







12.2.3

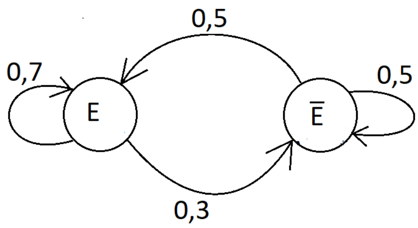
The weather in London behaves as follows. If it rains today then with 0.7 probability it will rain tomorrow as well. If it does not rain today then the probability that it does not rain tomorrow is 50%.

a) Draw the state transition diagram. Write down the matrix representing the state transitions.

b) It does not rain today. What is the probability that it will rain the day after tomorrow?

c) Find the stationary distribution and explain its everyday interpretation.

Solution: (a)



$$M = \begin{pmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{pmatrix}$$

(b) Firs entry of the second row of M^2: 0.6.

(c) The question is to find a distribution vector $w = (x; y)$ such that $w \cdot M = w$ holds. That is:

$$\begin{aligned} x + y &= 1 \\ 0.7x + 0.5y &= x \\ 0.3x + 0.5y &= y \end{aligned}$$

Solution: $w = (0.625; 0.325)$. That means an equilibrium, on average 62.5 % of the days are rainy, 32.5 % of them are dry.

12.2.4

When CJ is sad, which isn't very usual: she either goes for a run, gobbles down some ice cream or takes a nap. From historic data, if she spent sleeping a sad day away the next day it is 60% likely she will go for a run, 20% she will stay in bed the next day and 20% chance she will pig out on ice cream. When she is sad and goes for a run, there is a 60% chance she'll go for a run the next day, 30% she gorges on ice cream and only 10% chance she'll spend sleeping the next day. Finally, when she indulges on ice cream on a sad day, there is a mere 10% chance she continues to have ice cream the next day as well, 70% she is likely to go for a run and 20% chance that she spends sleeping the next day.

a) Write down the transition matrix.

b) If CJ chose to run today, what is the probability that she will sleep the day after tomorrow?

c) ⋆ What is the stationary distribution? How can it be interpreted?

Solution: (a) Fix an order: sleeping, running, ice cream.

$$M = \begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.7 & 0.1 \end{pmatrix}$$

(b) First entry of the second row of M^2: 0.14.

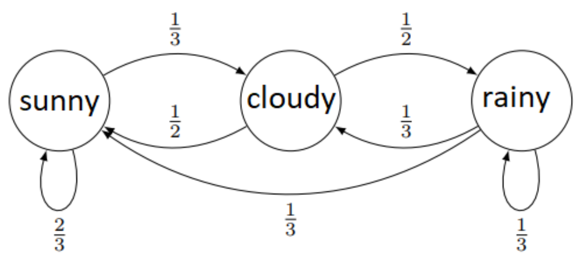
(d) $w = (x; y; z)$ distribution vector with the following propoerty: $w \cdot M = w$. That is:

$$\begin{array}{rcl} x + y + z & = & 1 \\ 0.2x + 0.1y + 0.2z & = & x \\ 0.6x + 0.6y + 0.7z & = & y \\ 0.2x + 0.3y + 0.1z & = & z \end{array}$$

Solution: $w = (0.137; 0.624; 0.238)$. That means an equilibrium, on avarage $13.7\ \%$ of the days she is sleeping, $62.4\ \%$ of the days she is running, and $23.38\ \%$ of the days she is eating ice cream.

12.2.5

In Gecsňopuszta people have observed weather patterns for centuries and noticed the following.



a) Interpret the diagram. What is transition matrix?

b) Today it's raining. What is the probability that day after tomorrow will be sunny?

c) Find the stationary distribution. What is the everyday interpretation of this vector?

Solution: (a) Fix an order: sunny, cloudy, rainy.

$$M = \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

(b) First entry of the third line of M^2: $1/2$.

(c) $w = (x; y; z)$ distribution vector, $w \cdot M = w$:

$$\begin{aligned} x + y + z &= 1 \\ 2/3x + 1/2y + 1/3z &= x \\ 1/3x + 1/3z &= y \\ 1/2y + 1/3z &= z \end{aligned}$$

Solution:: $w = (9/16; 1/4; 3/16)$.

12.2.6

A company has a new product: the super-K. According to a costumer research, 70% of the costumers who choose super-K on a week will choose super-K again on the next week. 20% of the costumers who choose the competitor product on this week will choose super-K on next week.

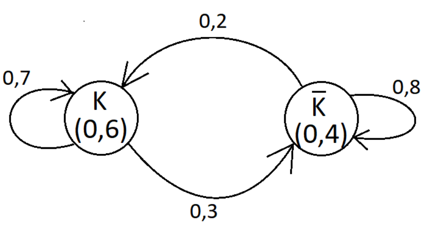
(a) Draw the diagram and write down the matrix of the transition.

(b) How many percent of those costumers who bought super-K on this week will buy a competitor product on the week after the next one?

(c) On this week 60 % of the costumers buy super-K. What is is this ratio one, two and three weeks later?

(d) Find the stationary distribution. What is the everyday interpretation of this vector?

Solution: (a)



$$M = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}$$

(b) Second element of the first row of M^2: $0,45$.

(c) $v = (0,6; 0,4)$. Next week: $v \cdot M = (0,5; 0,5)$. Two weeks later: $v \cdot M^2 = (0,45; 0,55)$. Three: $v \cdot M^3 = (0,425; 0,575)$.

(d) $w = (x; y)$ distribution vector, $w \cdot M = w$:

$$\begin{aligned} x + y &= 1 \\ 0,7x + 0,2y &= x \\ 0,3x + 0,8y &= y \end{aligned}$$

Solution: $w = (0,4; 0,6)$. After a long time on every week about 40% of the costumers will choose super-K.

13 Further Exercises

13.1 Linear programming - Graphical approach

13.1.1

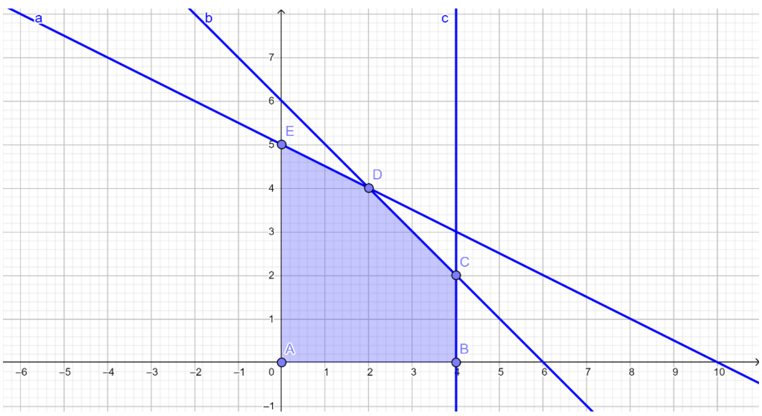
A Hungarian fashion company works for foreign order. They make jeans and jackets. It takes 1 hour to make a pair of jeans and 2 hours to make a jacket. Each pair of jeans and jacket require 2 meters of raw material, but they also need to use 2 meters of hem for the jeans, which they can only obtain from imports. The plant can produce 10 hours a day. 12 meters of raw material are available daily, and 8 meters of hem can be obtained daily from abroad. The company has a profit of 2000 Forints on a pair of jeans and 3000 Forints on a jacket. How many pairs of jeans and jackets should be made to maximize profits?

Solution:

x_1: number of pairs of jeans to be made
x_2: number of jackets to be made
The objective function is $z = 2000x_1 + 3000x_2$, which (in thousand Forints) can also be written in the form of $2x_1 + 3x_2$.
The constraints are:

$$x_1 + 2x_2 \leq 10;$$
$$2x_1 \leq 8;$$
$$2x_1 + 2x_2 \leq 12;$$
$$x_1 \geq 0, x_2 \geq 0, \text{ and } x_1, x_2 \text{ are integers}$$

The graphical representation of the constraints is:



Substituting the coordinates of the extreme points of the region denoting the set of possible solutions into the objective function (the coordinates are derived from the intersections of the straight lines):

Points	Value of the objective function (Forint)
$A(0;0)$	0
$B(4;0)$	8000
$C(4;2)$	14000
$D(2;4)$	16000
$E(0;5)$	15000

The highest value is obtained for point D; therefore, the company should make 2 pairs of jeans and 4 jackets per day in order to get the highest profit (16000 Forints).

13.1.2

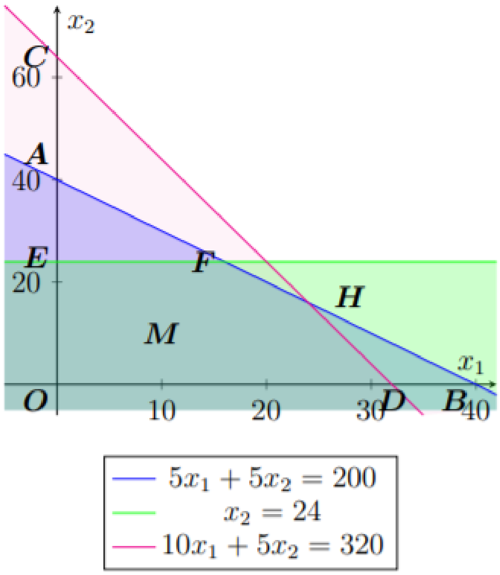
A company manufactures two products: T and D. One piece of T earns a profit of 25 euros, and one piece of D earns a profit of 20 euros. Next week, there will be 200 man-hours available to put the two products together. It takes 5 man-hours to assemble a piece of T and 5 man-hours to assemble a piece of D. Each D requires a special part, of which only 24 are in stock. The company has a storage space of 320 square meters, of which an area of 10 square meters is occupied by one T and an area of 5 square meters is occupied by one D. The company's management wants to maximize its profits. What production plan should they follow?

Solution:

x_1: number of T's to be made
x_2: number of D's to be made
The objective function is $z = 25x_1 + 20x_2$.
The constraints are:

$$5x_1 + 5x_2 \leq 200;$$
$$x_2 \leq 24;$$
$$10x_1 + 5x_2 \leq 320;$$
$$x_1 \geq 0, x_2 \geq 0, \text{ and } x_1, x_2 \text{ are integers}$$

The graphical representation of the constraints is:



Substituting the coordinates of the extreme points of the region denoting the set of possible solutions into the objective function (the coordinates are derived from the intersections of the straight lines):

Points	Value of the objective function (euro)
$O(\ 0;\ 0)$	0
$E(\ 0; 24)$	480
$F(16; 24)$	880
$H(24; 16)$	920
$D(32;\ 0)$	800

The highest value is obtained for point H; therefore, the company has to produce 24 T's and 16 D's to achieve the highest profit (920 euros).

13.1.3

A plant manufactures fingerprint readers for laptops. Two such devices are manufactured, they are U_1 and U_2. The assembly of the product line consists of three successive phases. The assembly operations corresponding to each phase are carried out by three different automatic production lines (denote these by A, B and C in the production order) so that all three production lines are suitable for the production of both types of devices, but under different conditions:
- Line A prepares 1 piece of U_1 in 5 minutes and 1 piece of U_2 in 10 minutes for the next production phase and can be operated for up to 16 hours per day.
- Line B prepares 1 piece of U_1 and U_2 for the last phase in 5 minutes, but can operate for up to 9 hours a day.
- Line C can operate non-stop, i.e., 24 hours a day, and it makes 1 piece of U_1 in 20 minutes and 1 piece of U_2 in 5 minutes.

After deducting the cost of expenses, U_1 generates a profit of 3 euros per piece, while U_2 generates 2 euros per piece. How to organize the production of the plant if the goal is to achieve the highest profit?

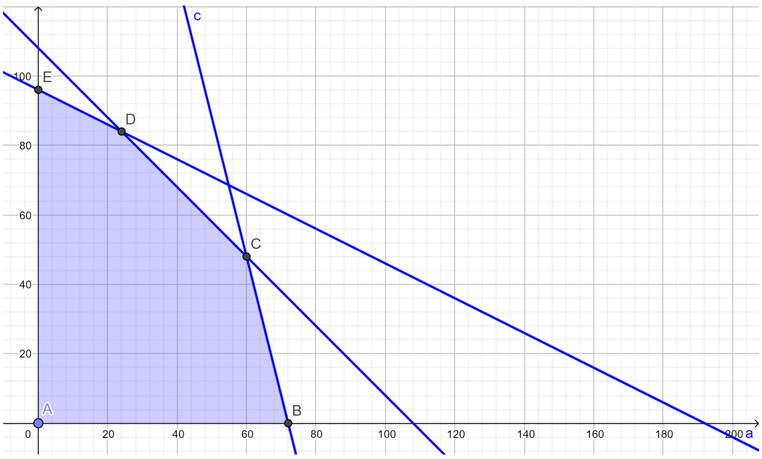
Solution:

x_1: number of U_1's to be manufactured
x_2: number of U_2's to be manufactured
The objective function is $z = 3x_1 + 2x_2$.
The constraints are:

$$x_1 + 2x_2 \leq 192;$$
$$x_1 + x_2 \leq 108;$$
$$4x_1 + x_2 \leq 288;$$
$$x_1 \geq 0, x_2 \geq 0, \text{ and } x_1, x_2 \text{ are integers}$$

(It is worth choosing the unit of time to be 5 minutes, otherwise 5 times the inequalities would be received.)

The graphical representation of the constraints is:



Substituting the coordinates of the extreme points of the region denoting the set of possible solutions into the objective function (the coordinates are derived from the intersections of the straight lines):

Points	Value of the objective function (euro)
$O(0; 0)$	0
$B(72; 0)$	216
$C(60; 48)$	276
$D(24; 84)$	240
$E(0; 96)$	192

The highest value is obtained for point C; therefore, the company has to produce 60 pieces of U_1 and 48 pieces of U_2 to achieve the highest profit (276 euros).

13.2 Combinatorics

13.2.1

In a 6-member company everyone shakes hand with everyone else exactly once. How many handshakes have been taken place? How many different ways can you calculate it?

Solution: $\binom{6}{2} = 6 \cdot 5/2 = 1 + 2 + 3 + 4 + 5 = 15$.

13.2.2

With the digits 1, 2, 3, 5, 7, 8, 9

(a) How many 7-digit numbers can be written?

(b) How many 7-digit numbers can be written, which contain every digit once?

(c) How many 4-digit numbers can be written?

(d) How many 4-digit numbers can be written such that all of its digits are different?

Solution: (a) 7^7, (b) $7!$, (c) 7^4, (d) $7 \cdot 6 \cdot 5 \cdot 4$.

13.2.3

(a) How many 5-digit numbers can be written using digits 0, 2, 4, 5, 7? How many even numbers are there among them?

(b) How many 5-digit numbers can be written with using digits 0, 2, 4, 5, 7, if we use each of them once? How many even numbers are there among them?

Solution: (a) Altogether $4 \cdot 5^4$ numbers can be constructed. The number of even numbers from these are $4 \cdot 5^3 \cdot 3$.

(b) $4 \cdot 4! = 96$. The answer for even numbers is more difficult, because 0 cannot be the first digit of a number and the last digit must be an even number. By these two conditions, we can distinguish two situations. Namely, if the last digit is 0, then for the first one we have 4 possible choices. On the other hand, if we select a nonzero even number for the last digit, then we have only 3 possible choice for the first digits. If we distinguish these two cases, then we get that the even numbers are $4! + 3 \cdot 3! \cdot 2 = 60$.

13.2.4

How many 7-digit numbers can be written using digits 1, 1, 1, 2, 3, 4, 4, if all digits can be used once? How many even numbers are there among them?

Solution: (a) The number of those numbers that can be constructed in this way is

$$\frac{7!}{2! \cdot 3!},$$

(b) For even numbers it is:

$$\frac{6! \cdot 3}{2! \cdot 3!}.$$

13.2.5

How many ways can 5 people sit next to each other? And if two of them want to sit next to each other?

Solution: *Megoldás:* (a) $5!$. (b) If two of them want to sit next to each other, then it is $4! \cdot 2!$.

Indeed, we can place 4 people in 4! ways. The last one who wants to sit next to another person can sit on the right or on the left to this person. So this step can be done in two ways.

13.2.6

How many ways can 5 people sit at a round table? And if two of them want to sit next to each other?

Solution: $4!$. If two of them want to sit next to each other $3! \cdot 2!$.

13.2.7

There are 20 people traveling on a bus. In the next five stops everyone gets off. How may ways can they do this?

Solution: For each person we have 5 options where they will get off. Therefore the result is 5^{20}.

13.2.8

In a tombola 5 prizes will be drawn among 20 people. How many ways is it possible if

(a) The prizes are different and one person can get only one of them?

(b) The prizes are different and one person can get more than one?

(c) The prizes are the same and one person can get only one of them?

(d) The prizes are the same and one person can get more than one?

Solution: (a) The first prize can be given to anyone from 20, the second prize can be given to one from 19, etc. Hence the result is $20 \cdot 19 \cdot 18 \cdot 17 \cdot 16$, this is the number of way that 5 prizes can be drawn among 20 people, if everyone can get at most one prize.

(b) The first prize can be given to anyone from 20, the second prize also, etc. Hence the result is $20 \cdot 20 \cdot 20 \cdot 20 \cdot 20 = 20^5$, this is the number of way that 5 prizes can be drawn among 20 people, if anyone can get more than one prize.

(c) Now we should select 5 people from 20, who will get prize. This is

$$\binom{20}{5} = \frac{20!}{5! \cdot 15!} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}$$

(d) This is combination with repetition, where we have 20 place where we can place the 5 prizes. This means that we have 5 balls (these represents the balls) and we have to place 20-1=19 dividers. The balls between two consecutive dividers represents the number of prize that the corresponding person get. Altogether we have 19+5=24 balls and dividers next to each other and we have to select the positions of the 5 balls. This can be done in

$$\binom{24}{19} = \binom{24}{5}$$

ways. This represents also the number of ways how we can distribute the prize in the exercise. Note that this is a combination with repetition.

13.2.9

An organization of 20 members wants to choose

(a) a leader, a financial expert and spokesman,

(b) three delegates to a larger assembly.

How many choices are possible?

Solution: (a) $20 \cdot 19 \cdot 18$. If the same person can be chosen multiple times, then it is 20^3.

(b) $\binom{20}{3}$.

13.2.10

How many ways can a TOTÓ coupon be filled? (A TOTÓ coupon contains 13 questions with A, B and X options.)

Solution: 3^{13}.

13.2.11

How many ways can a lottery ticket be filled? (In lottery ticket we can select 5 numbers from 1 to 90.)

Solution: $\binom{90}{5}$.

13.2.12

How many ways can a bus ticket be punched out if

(a) The system punches 3 holes?

(b) The system can punch any number of holes from 1 to 9?

Solution: (a) $\binom{9}{3}$.

(b) We can count it in two different ways:

I. solution: We can calculate according to how many punches are taken by the system. In this way we would get the following:

$$\binom{9}{1}+\binom{9}{2}+\cdots+\binom{9}{9}.$$

II. solution: Every punch of the ticket can be coded in such way we tell whether a given number from 1 to 9 was punched or not. This is two options for every numbers. Excluding that case when the system does not punch anything, we get that 2^9-1.

Thus we have the following identity:

$$2^9=\binom{9}{0}+\binom{9}{1}+\binom{9}{2}+\cdots+\binom{9}{9},$$

as $\binom{9}{0}=1$.

13.2.13

How many ways can the word 'bachelor' be read? (Moving right and down from the top left corner to the bottom right corner.) How many ways can you reach the letters in each position?

$$\begin{matrix} B & A & C & H & A \\ A & C & H & A & L \\ C & H & A & L & O \\ H & A & L & O & R \end{matrix}$$

Solution: In all ways we have to take 7 steps, 4 to the right and 3 downward. Thus the question is the following: how many ways can we select 4 steps downward from 7. It is

$$\binom{7}{4} = \frac{7!}{4! \cdot 3!} = 35.$$

If we apply this to any position we would get the following table:

$$
\begin{matrix}
\binom{0}{0} & \binom{1}{1} & \binom{2}{2} & \binom{3}{3} & \binom{4}{4} \\
\binom{1}{0} & \binom{2}{1} & \binom{3}{2} & \binom{4}{3} & \binom{5}{4} \\
\binom{2}{0} & \binom{3}{1} & \binom{4}{2} & \binom{5}{3} & \binom{6}{4} \\
\binom{3}{0} & \binom{4}{1} & \binom{5}{2} & \binom{6}{3} & \binom{7}{4}
\end{matrix}
$$