PROBABILITY THEORY

COMBINATORICS

Combinatorics has many applications within computer science for solving complex problems. However, it is under-represented in literature since there is little application of combinatorics in business applications. Fortunately, the science behind it has been studied by mathematicians for centuries, and is well understood and well documented. However, mathematicians are focused on in how many elements will exist within a combinatorics problem, and have little interest in actually going through the work of creating those lists. The complete list of combinatorial collections is:

- permutations,
- permutations with repetitions,
- combinations,
- combinations with repetitions,
- variations,
- variations with repetitions.

The principle of multiplication

The principle of multiplication

If the activity consists in k successive steps and
the first step may be performed n_1 ways,
the second step may be performed n_2 ways,
$\vdots$
the k-th step may be performed n_k ways,
then number of different ways of performing an action is $n_1 \cdot n_2 \dots n_k$.

Example

Mary is going to school. She has 3 blouses, 4 skirts and 2 sweaters. How many ways can he combine a blouse, a skirt and a sweater?

Solution. Mary has 3 options for choosing a blouse, 4 options for choosing a skirt and 2 options for choosing a sweater. The number of options for combining these 3 pieces of clothing is

$$3 \times 4 \times 2 = 24.$$

Factorial function

In the following, we shall need the factorial function defined as follows

$$0! = 1, \qquad (n+1)! = (n+1)n! \tag{1}$$

and the binomial coefficients, defined by the formula

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}, \qquad (n, k \text{ are nonnegative integers}, n \geq k) \tag{2}$$

From (1) and (2), we obtain

$$\binom{n}{0} = 1, \quad \text{in particular,} \quad \binom{0}{0} = 1.$$

Example

Calculate $4!$, $7!$, $\binom{6}{4}$, $\binom{8}{5}$.

Solution:

$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$

$7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040 \text{ or equivalently } 7! = 7 \cdot 6 \cdot 5 \cdot 4! = 210 \cdot 24 = 5040$

$\binom{6}{4} = \frac{6!}{2! \cdot 4!} = \frac{6 \cdot 5 \cdot 4!}{2 \cdot 1 \cdot 4!} = 15$

$\binom{8}{5} = \frac{8!}{3! \cdot 5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{3 \cdot 2 \cdot 1 \cdot 5!} = 8 \cdot 7 = 56$ $\surd$

Permutations

Permutations are all possible orderings of a given input set. Each ordering of the input is called a permutation. When each item in the input set is different, there is only one way to generate the permutations.

Theorem

The number of permutations of n different items taken all at a time is

$$P(n) = n! = 1 \cdot 2 \cdot \ \ldots \ \cdot n. \tag{3}$$

Example

In how many ways can you pair up 8 boys and 8 girls?

Solution:

If sounds like there are 2 sets of 8 items to consider permuting, but in reality we are only permuting one set of 8 items. Think of the boys as in a fixed order: boy 1, boy 2, ..., boy 8. Each arrangement of girls corresponds to one pairing with the boys: girl 1 in the arrangement with boy 1, girl 2 in the arrangement with boy 2, etc. The girls can be arranged in $P(8) = 8!$ ways. $\surd$

Permutations with Repetition

Permutations with Repetition sets give allowance for repetitive items in the input set that reduce the numbers of orderings.

Theorem

If n given items can be divided into c classes of alike items differing from class to class, then the number of permutations of these items taken all at a time is

$$P'_{n_1,n_2,\ldots,n_c}(n) = \frac{n!}{n_1!n_2!\ldots n_c!} \qquad (n_1 + n_2 + \cdots + n_c = n) \tag{4}$$

where n_j is the number of items in the j-th class.

Example

What is the number of distinguishable arrangements that can be made from the word “repetition”?

Solution:

Since the letters e, t, i occur twice, we have permutations with repetition. The number is

$$P_{2,2,2}(10) = \frac{10!}{2!2!2!}. \qquad \checkmark$$

Combinations

Combinations are subsets of a given size taken from a given input set. The size of the set is known as the Upper Index (n) and the size of the subset is known as the Lower Index (k). Unlike permutations, combinations do not have any order in the output set.

Theorem

The number of different combinations of n different items, k at time is

$$C(n,k) = \binom{n}{k}. \tag{5}$$

Combinations with Repetition are determined by looking at a set of items, and selecting a subset while allowing repetition. We will not deal with them.

Example

The board of directors of a corporation comprises 10 members. An executive board, formed by 4 directors, needs to be selected. In how many possible ways are there to form the executive board?

Solution:

We choose 4 directors from 10 people, whereby ordering is not important. So we have combinations $C(10, 4) = \frac{10!}{6! \cdot 4!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6!}{6! \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 210$. $\surd$

Example

In how many ways can you choose 4 groups of 4 people from 16 people, assuming the groups are distinct?

Solution:

Choosing the members of the first group we choose 4 people from 16, which are combinations $C(16, 4)$. The members of the second and the third group are chosen from 12 and 8 people, respectively. The number of ways, we can choose 4 groups of 4 people from 16 people is $\binom{16}{4} \cdot \binom{12}{4} \cdot \binom{8}{4} = 63063000$. $\surd$

Variations and Variations with Repetitions

Variations combine features of combinations and permutations, they are the sets of all ordered combinations of items to make up a subset. Like combinations, the size of the set is known as the Upper Index (n) and the size of the subset is known as the Lower Index (k). The generation of variations can be based on the repeating of output items. These are called Variations and Variations with Repetition. Variations are permutations of combinations. That is, a variation of a set of n items choose k, is the ordered subset of size k.

Theorem (Variations)

The number of different variations of n different items, taken k at time is

$$V(n,k) = \frac{n!}{(n-k)!}. \tag{6}$$

Example

There is a basket of fruits containing an apple, a banana and an orange and there are five girls who want each to eat one fruit. In how many ways are there to give three of the five girls one fruit each and leave two of them without a fruit to eat?

Solution:

Giving the 3 fruits to 3 of the 5 girls is a sequential problem. We first give the apple to one of the girls. There are 5 possible ways to do this. Then we give the banana to one of the remaining girls. There are 4 possible ways to do this, because one girl has already been given a fruit. Finally, we give the orange to one of the remaining girls. There are 3 possible ways to do this, because two girls have already been given a fruit. Hence there are $5 \cdot 4 \cdot 3 = 60$ ways to give them fruits. (We use the principle of multiplication)
In fact, the number of ways to assign the three fruits is equal to the number of 3-variations of 5 objects (without repetition). If we denote it by $V(5,3)$, then we get

$$V(5,3) = \frac{5!}{(5-3)!} = 60. \qquad \surd$$

Theorem (Variation with repetitions)

The number of different variations of n different items, taken k at time with repetitions is

$$V'(n,k) = n^k. \tag{7}$$

Example

How many three-digit numbers can we create from digits 1, 5, 6 and 8, if

a) the digits can be repeated;

b) the digits can not be repeated.

Solution:

a) We have variations with repetition for $n = 4,\ k = 3$. We obtain $V'(4,3) = 4^3 = 64$.

b) We have variations without repetition, i. e., $V(4,3) = \frac{4!}{(4-3)!} = 24$.

Probability theory

The probability theory has the purpose of providing mathematical models of situations affected or governed by "chance effect" for instance, in weather forecasting, life insurance, quality of technical products, traffic problems, and, of course, games of chance with cards or dice.

An experiment is a process of measurement or observation, in a laboratory, in a factory, on the street, in nature, or wherever: so "experiment " is used in rather general case. Our interest is in experiments that involve randomness, chance effects, so that we cannot predict a result exactly. Experiments are for examples: throwing dice, flipping coin, pulling out the card,. . . .

A trial is a single performance of an experiment. Its result is called a random event. We will denote events with capital letters $A, B, C, \ldots$.

In throwing dice, events are for example $A = \{1, 3, 5\}$ ("odd number"), $B = \{2, 4, 6\}$ ("even number"), $C = \{5, 6\}$ ("number greater than 4"), etc.

Event that after the experiment will never be is called an impossible event, we write $\emptyset$ – for example "number greater than 6". Event that after the experiment will always occur is called a certain event, we write I – for example "number least than 7".

Basic knowledge of events

Definition

1. *Event A is a subset of event B, if it follows from the occurrence of event A, that event B has occurred.*
2. *Events A and B are equivalent, if event A occurs if and only if event B occurs, we write $A = B$.*
3. *The complement of A is an event, which occurs if and only if event A does not occur, we write $\overline{A}$.*
4. *The union of events A and B is an event, which occurs if occurs at least one of events A and B, we write $A \cup B$.*
5. *The intersection of events A and B is an event, which occurs if occur both of events A and B, we write $A \cap B$.*

Throwing a dice once, let event A be that the number on the face is odd and event B be that the number on the face is greater than 4. We denote events as the sets. We have $A = \{1, 3, 5\}$ and $B = \{5, 6\}$.
Then $\overline{A} = \{2, 4, 6\}$, $\overline{B} = \{1, 2, 3, 4\}$, $A \cap B = \{5\}$, $A \cup B = \{1, 3, 5, 6\}$.

Properties of events

Let A, B, Cbe arbitrary events. Then

1. $A \subset A, \quad \emptyset \subset A, \quad A \subset I$;
2. if $A \subset B$ a $B \subset C$, then $A \subset C$;
3. $\bar{I} = \emptyset$, $\bar{\emptyset} = I$, $\quad \overline{(\overline{A})} = A$;
4. $A \cap A = A, \quad A \cup A = A$;
5. $A \cap B = B \cap A, \quad A \cup B = B \cup A$;
6. $A \cap (B \cap C) = (A \cap B) \cap C, \quad A \cup (B \cup C) = (A \cup B) \cup C$;
7. $\overline{A} \cap A = \emptyset$, $A \cap \emptyset = \emptyset$, $A \cap I = A$, $A \cap B \subset A$;
8. $\overline{A} \cup A = I$, $A \cup \emptyset = A$, $A \cup I = I$, $A \subset A \cup B$;
9. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
10. $\overline{(A \cup B)} = \overline{A} \cap \overline{B}, \quad \overline{(A \cap B)} = \overline{A} \cup \overline{B}$ (de Morgan rules).

Disjoint events and hypothesis

Definition

1. *Events A and B are called disjoint if they can not occur at the same time, i.e., $A \cap B = \emptyset$.*
2. *Events $H_1, H_2, \ldots, H_n$ are disjoint, if $H_1 \cap H_j = \emptyset$ for each $i \neq j$, $i, j \in \{1, 2, \ldots, n\}$.*

Definition

The system of disjoint events $H_1, H_2, \ldots, H_n$ is called complete, if its union is a certain event. The complete system of disjoint events is called the system of hypothesis.

When throwing dice, let H_1 denotes the event that the face shows an even number, i.e., $H_1 = \{2, 4, 6\}$ and H_2 denotes the event that the face shows an odd number, i.e., $H_2 = \{1, 3, 5\}$. Since $H_1 \cup H_2 = \{1, 2, 3, 4, 5, 6\}$-certain event and $H_1 \cap H_2 = \emptyset$, the events H_1, H_2 represent hypothesis.

Simple and composite events

Definition

1. *An event E is called* simple*, if there does not exist events A_1, A_2 different from E such that $E = A_1 \cup A_2$.*
2. *Each event, which is not elementary, is called a* composite event*.*
3. *The set of all elementary events which can occur as the output of a random experiment is called a* space of elementary events*, i. e., $\gamma = \{E_1, E_2, \dots, E_n, \dots\}$.*
4. *The set of all events is called a* sample space *and it is denoted by τ.*

When throwing dice, $\gamma = \{E_1, E_2, \dots, E_6\}$, where E_k denotes the event that the face shows the number k. An event A can be equivalent to some of elementary events but also can include several elementary events. For example, showing an odd number on the face consists of three elementary events E_1, E_2, E_3.
The operations with random events are reduced to the operations with sets and they follow the same rules. The empty set corresponds to the impossible event and the space of elementary events corresponds to the certain event.

Classical definition of probability

Definition

Let $\gamma = \{E_1, E_2, \dots, E_n\}$ *be a space of elementary events. Let each elementary event be "equally possible". For each event* $A \in \tau$ *we define its* *probability* *as follows:*

$$P(A) = \frac{m}{n}, \tag{8}$$

where m *is the number of elementary events from which* A *consists, i. e.,* $A = E_{i_1} \cup E_{i_2} \cup \cdots \cup E_{i_m}$.

We can equality (8) interpret as follows:

$$P(A) = \frac{\text{the number of favorable outcomes of event } A}{\text{the number of possible outcomes of event } A}.$$

Example

Suppose a coin is flipped 3 times. What is the probability of getting two tails and one head? Note that each flipping of coin has two possible outcomes H (heads) and T (Tails).

Solution. For this experiment, the space of elementary events consists of 8 events:

$$\gamma = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

Each event is equally likely to occur. The event A "getting two tails and one head" consists of the following elementary events: $A = \{TTH, THT, HTT\}$. Since $m = 3$ and $n = 8$, the probability of event A is

$$P(A) = \frac{3}{8}.$$

Basic properties of classical probability

Theorem

The following assertions hold:

1. $0 \leq P(A) \leq 1$ *for each event* A*;*
2. $P(\emptyset) = 0$, $P(I) = 0$*;*
3. $P(\overline{A}) = 1 - P(A)$ *for each* $A \in \tau$*;*
4. *if* $A \subseteq B$ *then* $P(A) \leq P(B)$*;*
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *for each two events* A, B*;*
6. *for each events* A, B, C *the following equality holds:*

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

Example

In a math class of 30 students, 17 are boys and 13 are girls. On a unit test, 4 boys and 5 girls made an "A" grade. If a student is chosen at random from the class, what is the probability of choosing a girl or an "A" student?

Solution:

Let A be the event that chosen student made "A" and B be the event that chosen student is a girl. We want to compute $P(B \cup A)$. We get

$$P(B \cup A) = P(B) + P(A) - P(B \cap A) = \frac{13}{30} + \frac{9}{30} - \frac{5}{30} = \frac{17}{30}.$$

$\surd$

Conditional Probability

The conditional probability of an event A given B is the probability that the event A will occur given the knowledge that the event B has already occurred. This probability is written $P(A|B)$, notation for the probability of A given B. In the case where events A and B are independent (where event B has no effect on the probability of event A), the conditional probability of event A given B is simply the probability of event A, that is $P(A)$.

Definition (**Conditional probability**)

The probability of A given B is defined as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{9}$$

If $P(B) = 0$ then $P(A|B)$ is not defined.

Example

A new MasterCard has been issued to 2000 customers. Of these customers, 1500 hold a Visa card, 600 hold an American Express card. Find the probability that a customer chosen at random holds a Visa card, given that the customer holds an American Express card.

Solution:

Let A be the event that a customer holds a Visa card and B be the event that a customer holds a American Express card. We have

$$P(A) = \tfrac{1500}{2000} = \tfrac{3}{4}, \qquad P(B) = \tfrac{600}{2000} = \tfrac{3}{10}, \qquad P(A \cap B) = \tfrac{100}{2000} = \tfrac{1}{20}.$$

Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{20}}{\frac{3}{10}} = \frac{10}{60} = \frac{1}{6}.$$

$\surd$

If $P(A) \neq 0$ and $P(B) \neq 0$ then $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$. Using this equality together with (9) we get $P(A \cap B) = P(A) \cdot P(B|A)$ and $P(A \cap B) = P(B) \cdot P(A|B)$ which implies

$$P(A \cap B) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A). \tag{10}$$

Example

Suppose that five good fuses and two defective ones have been mixed up. To find the defective fuses, we test them one-by-one, at random and without replacement. What is the probability that we are lucky and find both of the defective fuses in the first two tests?

Solution:

Let A be the event that we find a defective fuse in the first test and B be the event that we find a defective fuse in the second test. We are told that $P(A) = \frac{2}{7}$ and $P(B|A) = \frac{1}{6}$. We want to compute $P(A \cap B)$. We get

$$P(A \cap B) = P(A) \cdot P(B|A) = \frac{2}{7} \cdot \frac{1}{6} = \frac{1}{21} = 0,047619.$$

$\surd$

The following theorem is a generalization of the previous reasoning.

Theorem

Let $A_1, A_2, \ldots, A_n$ be events. Then

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \cdots \cdot P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \quad (11)$$

Example

A school survey found that 7 out of 30 students walk to school. If four students are selected at random without replacement, what is the probability that the first and the second walk to school, but the third and the fourth do not walk to school?

Solution:

Let A_i denotes the event that ith selected student walks to school.

$$P(A_1 \cap A_2 \cap \overline{A_3} \cap \overline{A_4}) = P(A_1) \cdot P(A_2|A_1) \cdot P(\overline{A_3}|A_1 \cap A_2) \cdot P(\overline{A_4}|A_1 \cap A_2 \cap \overline{A_3}) =$$

$$\frac{7}{30} \cdot \frac{6}{29} \cdot \frac{23}{28} \cdot \frac{22}{27} = \frac{21252}{657\,720} = 0,03231$$

$\surd$

Total Probability Law

The relationship between the hypothesis and the conditional probability is expressed by the following theorem.

Theorem

Let $H_1,\ H_2, \ldots, H_n$ be a complete system of hypothesis. Then for each event A the following equality holds:

$$P(A) = P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) + \cdots + P(H_n) \cdot P(A|H_n) =$$

$$= \sum_{i=1}^{n} P(H_i) \cdot P(A|H_i). \tag{12}$$

Example

A grocery store obtains 35% of its products from vendor A, and 65% of its produce from vendor B. It is expected that 12% of the production from vendor A and 17 % of the products to the vendor B should be discarded. What is the probability that a randomly selected product should be discarded?

Solution:

Let A be the event that the randomly selected product should be discarded. Let H_1 be the event that the product is from vendor A and H_2 be the event that the product is from vendor B. We have

$$P(H_1) = 0.35, \quad P(H_2) = 0.65, \quad P(A/H_1) = 0.12, \quad P(A/H_2) = 0.17.$$

By using the total law probability we obtain:

$$P(A) = P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) = 0.12 \cdot 0.35 + 0.17 \cdot 0.65 = 0.1525.$$

$\surd$

Example

Two cards from an ordinary deck of 52 cards are missing. What is the probability that a random card drawn from this 50-card deck is a spade?

Solution:

Let A be the event that the randomly drawn card is a spade. Let H_i be the event that i spades are missing from the 50-card (defective) deck for $i = 0, 1, 2$. We want to compute $P(A)$, which we compute by conditioning on how many spades are missing from the original (good) deck. We have

$$P(H_0) = \frac{\binom{13}{0} \cdot \binom{39}{2}}{\binom{52}{2}}, \quad P(H_1) = \frac{\binom{13}{1} \cdot \binom{39}{1}}{\binom{52}{2}}, \quad P(H_2) = \frac{\binom{13}{2} \cdot \binom{39}{0}}{\binom{52}{2}}.$$

and conditional probabilities are

$$P(A|H_0) = \frac{13}{50}, \qquad P(A|H_1) = \frac{12}{50}, \qquad P(A|H_2) = \frac{11}{50}.$$

By using the total law probability we obtain:

$$P(A) = P(A|H_0) \cdot P(H_0) + P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) =$$
$$= \frac{13}{50} \cdot \frac{\binom{13}{0} \cdot \binom{39}{2}}{\binom{52}{2}} + \frac{12}{50} \cdot \frac{\binom{13}{1} \cdot \binom{39}{1}}{\binom{52}{2}} + \frac{11}{50} \cdot \frac{\binom{13}{2} \cdot \binom{39}{0}}{\binom{52}{2}} = 1/4.$$

Bayes' theorem

Theorem (**Bayes' theorem**)

Let H_1, $H_2, \ldots, H_n$ *be hypothesis and* A *be such that* $P(A) \neq 0$*. Then for each* H_k, $k = 1, 2, \ldots, n$ *the following equality holds:*

$$P(H_k|A) = \frac{P(H_k) \cdot P(A|H_k)}{\sum_{i=1}^{n} P(H_i) \cdot P(A|H_i)} = \frac{P(H_k) \cdot P(A|H_k)}{P(A)}. \tag{13}$$

Example

Urn 1 contains 5 white balls and 7 black balls. Urn 2 contains 3 white and 12 black balls. A fair coin is flipped; if it is head, a ball is drawn from Urn 1, and if it is tail, a ball is drawn from Urn 2. Suppose that this experiment is done and you learn that a white ball was selected. What is the probability that this ball was in fact taken from Urn 2?

Solution:

Let H_1 be the event that the coin flip was heads and H_2 be the event that the coin flip was tails. Let A be the event that a white ball is selected. From the given data, we know that $P(A|H_1) = 5/12$ and that $P(A|H_2) = 3/15 = 1/5$. Since the coin is fair, we know that $P(H_1) = P(H_2) = 1/2$.
We want to compute $P(H_2|A)$, which we do using the Bayes Formula:

$$P(H_2|A) = \frac{(1/5) \cdot (1/2)}{(1/5) \cdot (1/2) + (5/12) \cdot (1/2)} = \frac{12}{37} = 0,3243.$$

$\surd$

Independent Events

Definition

Two events A, B are independent if the occurrence of one does not affect the probability of the other or if the probability of one of them equals zero, i.e., one of the following possibilities occurs:

$$P(A|B) = P(A) \quad or \quad P(B) = 0 \quad or \quad P(B|A) = P(B) \quad or \quad P(A) = 0. \tag{14}$$

Two events are dependent if the outcome or occurrence of the first affects the outcome or occurrence of the second so that the probability is changed.

Theorem

Events A, $B \in \tau$ are independent if and only if

$$P(A \cap B) = P(A) \cdot P(B). \tag{15}$$

Remark

The independence of events A and B implies the independence of these pairs of events:

$$A \text{ and } \overline{B}, \qquad \overline{A} \text{ and } B, \qquad \overline{A} \text{ and } \overline{B}.$$

Example

Determine whether events A and B are independent or dependent.
Rolling two dice, with
Event A: Rolling 1 on the first die.
Event B: The sum of the values on the dice is 7.

Solution:

We have $P(B) = 6/36 = 1/6, \quad P(B|A) = 1/6.$
Since $P(B|A) = P(B)$, the events are independent. $\surd$

Example

Determine whether events A and B are independent or dependent.
Flip three coins, with
Event A: The first two coins are heads.
Event B: There are at least two heads among the three coins.

Solution:

We have

$$P(A) = \frac{2}{8} = \frac{1}{4}, \quad P(B) = \frac{4}{8} = \frac{1}{2}, \quad P(A \cap B) = \frac{2}{8} = \frac{1}{4}.$$

Since $P(A) \cdot P(B) \neq P(A \cap B)$, the events are dependent. $\surd$

Overall independent events

Definition

The system of events $A_1, A_2, \ldots, A_n$ is overall independent, if the probability of occurrence of one of them does not change upon the occurrence of any group of other events or if the probability of one of them equals zero.

Theorem (**Probability of intersection of overall independent events**)

Let $[\gamma, \tau, P]$ be a probability space. The system of events $A_1, A_2, \ldots, A_n$ is overall independent, if the following condition is satisfied:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \ \ldots \ \cdot P(A_n). \tag{16}$$

Theorem (**Probability of union of overall independent events**)

If the system of events $A_1, A_2, \ldots, A_n$ is overall independent then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot \cdots \cdot P(\overline{A}_n). \tag{17}$$

Example

Four basketball players throw the ball to the basket. The probabilities of successful attempts of individual basketball players are 0.8; 0.7; 0.85 a 0.9. Determine the probability that

a) *all four hit to the basket,*
b) *none hits to the basket,*
c) *at least one hits to the basket,*
d) *at least one does not hit to the basket,*
e) *exactly one hits to the basket.*

Solution:

Denote by A_i the event, that i-th basketball player hits to the basket. We have

$$P(A_1) = 0.8, \quad P(A_2) = 0.7, \quad P(A_3) = 0.85, \quad P(A_4) = 0.9.$$

Hence

a) $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot P(A_4) = 0.8 \cdot 0.7 \cdot 0.85 \cdot 0.9 = 0.4284$
b) $P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3 \cap \overline{A}_4) = P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot P(\overline{A}_3) \cdot P(\overline{A}_4) = 0.2 \cdot 0.3 \cdot 0.15 \cdot 0.1 = 0.0009$
c) $P(A_1 \cup A_2 \cup A_3 \cup A_4) = 1 - P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot P(\overline{A}_3) \cdot P(\overline{A}_4) = 0.9991$
d) $P(\overline{A}_1 \cup \overline{A}_2 \cup \overline{A}_3 \cup \overline{A}_4) = 1 - P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot P(A_4) = 0.5716$

Probability of Repeated Independent Trials

Many experiments share the common element that their outcomes can be classified into one of two events, e.g., a coin can come up head or tail; a child can be male or female; a person can be employed or unemployed. These outcomes are often labeled as "success" or "failure". The usual notation is p is probability of success, q is probability of failure. Note that $p + q = 1$. We are often interested in the number of successes in repeated trials.

Example

Determine the probability of getting two numbers "6" in 5 die rollings.

At first, we have to determine the probability of one possible way the event can occur, and then determine the number of different ways the event can occur. That is,

$$P(\text{event}) = (\text{number of ways event can occur}) \cdot P(\text{one occurrence}).$$

We will call getting a "6" a "success." The number of repetitions is $n = 5$, the number of successes is $r = 2$, and the number of failures is $n - r = 5 - 2 = 3$. One way this can occur is if the first 2 rollings are "6" and the last three are not "6". The probability is $(1/6)^2 \cdot (5/6)^3$. The number of ways event can occur is $\binom{5}{2}$, so the probability of this event is $P = \binom{5}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^3 = \frac{3750}{6^5} = 0.48335.$

Theorem (**Bernoulli theorem**)

Let p be the probability that the outcome of trial is "success" considering event A and $P_{n,p}(k)$ be the probability that repeating a trial n times the number of "successöutcomes is k. Then

$$P_{n,p}(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad \textit{for } k = 0.1, \ldots, n. \tag{18}$$

Example

An apartment building has residents living on the second and third floors. The residents use an elevator to get to their apartments. There are 20 people living on the second floor and 6 people living on the third floor. Six random residents use the elevator to get to their apartments. What is the probability that exactly 4 of the people exit on the second floor and that 2 people exit on the third floor?

Solution:

We will call an exiting on the second floor a "success." So the probability of success is $p = 20/26 = 10/13$. We have $n = 6$, $k = 4$ and by (18) we obtain

$$P_{6,\frac{10}{13}}(4) = \binom{6}{4}\left(\frac{10}{13}\right)^4 \cdot \left(\frac{3}{13}\right)^2 = 0.194.$$

Random Variable

A probability distribution or, briefly, distribution, shows the probabilities of events in an experiment. In probability theory and statistics, a random variable is a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense). A random variable can take on a set of possible different values (similarly to other mathematical variables), each with an associated probability, in contrast to other mathematical variables.
A random variable's possible values represent the possible outcomes of a yet-to-be-performed experiment.

Definition

A random variable is each mapping $X : \gamma \to \mathbb{R}$, where γ is the set of elementary events. For each elementary event E is $X(E)$ some real number, called a value of a random variable X for the event E.

Random variables will be denoted by capital letters X, Y, X_1, X_2, ... and values of random variables will be denoted by small letters x, y, x_1, x_2, ... We suppose that for each $a \in \mathbb{R}$ we can determine the probabilities of types:

1. $P(X = a)$, i. e., the probability that the value of the random variable X is equal to the number a;
2. $P(X \leq a)$, i. e., the probability that the value of the random variable X is not greater than the number a;
3. $P(X \in I)$, i. e., the probability that the value of the random variable X take values from the interval I.

Random variables can be

1. discrete, that is, taking any of a specified finite or countable list of values,
2. continuous, taking any numerical value in an interval or collection of intervals.

Example

Which of the following random variables are discrete and which are continuous?

a) *The number of students in a section of a statistics course.*
b) *The air pressure in an car tire.*
c) *The sum of the values on the two dice*
d) *The height of students at TUKE.*
e) *The speed of randomly selected vehicles on a highway.*
f) *The time it takes a student to register for spring semester.*
g) *The number of points earned on the credit.*

Solution.
discrete – a), c), g)
continuous – b), d), e), f)

Discrete random variables and their distributions

Definition

A random variable X has a distribution of discrete type if X takes on only finitely many or at most countably many values $x_1, x_2, x_3, \dots$ called the possible values of X, with probabilities $p_i = P(X = x_i)$ whereas $P(X \in I) = 0$ for any interval I containing no possible values.

Obviously, the discrete distribution is also determined by the probability function $f(x)$ of X, defined by

$$f(x) = \begin{cases} p_j & \text{for } x = x_j, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

The set $\mathcal{H}(X) = \{x_1, x_2, x_3, \dots\}$ is called a range of values of X.

We can describe a random variable X by the probability table of random variable X.

$$\begin{array}{c||c|c|c|c} x_i & x_1 & x_2 & \cdots & x_n \\ \hline P(X = x_i) = p_i & p_1 & p_2 & \cdots & p_n \end{array}, \tag{20}$$

It is easy to see, that for the probabilities so-called normalization condition holds:

$$\sum_{i=1}^{n} p_i = 1. \tag{21}$$

Cumulative Distribution Function

Definition

Cumulative Distribution Function of random variable X is the function defined for each $x \in \mathbb{R}$ as follows

$$F(x) = P(X \leq x). \tag{22}$$

We get the values of distribution function $F(x)$ by taking sums,

$$F(x) = \sum_{x_j \leq x} f(x_j) = \sum_{j=1}^{j: x_j \leq x} p_j. \tag{23}$$

Theorem (The properties of the distribution function)

1. *For each $x \in \mathbb{R}$ we have $0 \leq F(x) \leq 1$;*
2. $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$;
3. *F is nondecreasing, i. e., for each $a < b$ is $F(a) \leq F(b)$;*
4. *if $a < b$, then*

$$P(X \in (a,\, b\rangle) = P(a < X \leq b) = F(b) - F(a). \tag{24}$$

Example

The student passes examinations at the ordinary term with probability 0.7 and event of failure, the probability of passing the test by him on the resit exam increases still by 0.1. There are two resists. We determine the probability table of a random variable that takes the value of completed terms on student test.

Solution. The student pass exam to the ordinary term with probability 0.7 so $P(X = 1) = 0.7$.
The student completes two terms, if student does not pass exam to the ordinary term, but he pass term in the first resist. Hence $P(X = 2) = 0.3 \cdot 0.8 = 0.24$.
The student will participate in three tests in the event that on the first two terms he was unsuccessful, i. e., $P(X = 3) = 0.3 \cdot 0.2 = 0.06$.
We obtain the probability table of random variable X:

x_i	1	2	3
p_i	0.7	0.24	0.06

.

Numerical Parameters of Discrete Random Variable

For each random variable we assign so-called numerical parameters, which will give us some information on character of the studied random variable.

Definition

Let X be a discrete random variable and let the law of the probability distribution of random variable X be given.

1. *An expected value of random variable X is the number $E(X)$ defined as follows:*
$$E(X) = \sum_{i=1}^{n} x_i \, p_i. \tag{25}$$
2. *The variance of random variable X is the number $D(X)$ defined as follows:*
$$D(X) = \sum_{i=1}^{n} (x_i - E(X))^2 \, p_i. \tag{26}$$
3. *The standard deviation of random variable X is the number $\sigma(X)$ defined as follows:* $\sigma(X) = \sqrt{D(X)}$

Theorem (Properties of the expected value and variance)

Let X and Y be random variables and let a and b be constants. Then

1. *for of the constant random variable A taking value a is $E(A) = a$; $D(a) = 0$;*
2. $E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y)$;
3. $D(a \cdot X) = a^2 \cdot D(X)$;
4. $D(a \cdot X + b) = a^2 \cdot D(X)$;
5. $D(X) = E(X^2) - [E(X)]^2$.

Example

We roll two dice. Let X be a random variable that takes on the value of the maximum of the thrown values. Determine:

a) *the probability table for X;*

b) *the distribution function of random variable X.*

c) $P(X < 4)$, $P(X \geq 3)$, $P(2 < X \leq 5)$;

d) *an expected value and the variance of X;*

Solution. a) The maximum of the thrown values can be 1,2,3,4,5 or 6, so $\mathcal{H}(X) = \{1, 2, 3, 4, 5, 6\}$. For each value from the set $H(X)$ we calculate the probability with which the given value is reached.
To calculate $P(X = 1)$, it is necessary to realize that the maximum is equal to 1 only if we roll the number 1 on both dice, i.e., $m = 1$, $n = 36$. Hence $P(X = 1) = \frac{1}{36}$.
The maximum of the thrown values equals 2 for pairs (1,2), (2,1), (2,2), so $P(X = 2) = \frac{3}{36}$.

We calculate the remaining probabilities in the same way. We write the result in the table:

x_i	1	2	3	4	5	6
$P(X = x_i) = p_i$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

b) The distribution function is defined by the formula $F(x) = P(X \leq x)$.

- For each $x \in (-\infty, 1)$ is $F(x) = P(X \leq x) = 0$ (the value of X is never less than any number from the interval $(-\infty, 1)$).
- For each value $x \in \langle 1, 2)$ we get $F(x) = P(X \leq x) = P(X = 1) = \frac{1}{36}$.
- for $x \in \langle 2, 3)$ is $F(x) = P(X \leq x) = P(X = 1) + P(X = 2) = \frac{1}{36} + \frac{3}{36} = \frac{4}{36}$.
- In the same way, we calculate the value of $F(x)$ at other intervals.

We obtain

$$F(x) = \begin{cases} 0, & \text{pre } x \in (-\infty, 1), \\ \frac{1}{36}, & \text{pre } x \in \langle 1, 2), \\ \frac{4}{36}, & \text{pre } x \in \langle 2, 3), \\ \frac{9}{36}, & \text{pre } x \in \langle 3, 4), \\ \frac{16}{36}, & \text{pre } x \in \langle 4, 5), \\ \frac{25}{36}, & \text{pre } x \in \langle 5, 6), \\ 1 & \text{pre } x \in \langle 6, \infty). \end{cases}$$

c) $P(X < 4) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{9}{36} = \frac{1}{4}$;
$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = \frac{32}{36} = \frac{8}{9}$;
$P(2 < X \leq 5) = P(X = 3) + P(X = 4) + P(X = 5) = \frac{21}{36} = \frac{7}{12}$;
or equivalently by (24):
$P(2 < X \leq 5) = F(5) - F(2) = \frac{25}{36} - \frac{4}{36} = \frac{21}{36} = \frac{7}{12}$.

d) $E(X) = \sum_{i=1}^{6} x_i \cdot p_i = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36} = \frac{161}{36}$;

$D(X) = \sum_{i=1}^{6} (x_i - E(X))^2 p_i = (1 - \frac{161}{36})^2 \cdot \frac{1}{36} + (2 - \frac{161}{36})^2 \cdot \frac{3}{36} + (3 - \frac{161}{36})^2 \cdot \frac{5}{36} +$
$(4 - \frac{161}{36})^2 \cdot \frac{7}{36} + (5 - \frac{161}{36})^2 \cdot \frac{9}{36} + (6 - \frac{161}{36})^2 \cdot \frac{11}{36} = \frac{2555}{1296}$.

Otherwise, we can calculate $D(X)$ according to the formula
$D(X) = E(X^2) - [E(X)]^2$. We have
$E(X^2) = \sum_{i=1}^{6} x_i^2 \cdot p_i = 1 \cdot \frac{1}{36} + 4 \cdot \frac{3}{36} + 9 \cdot \frac{5}{36} + 16 \cdot \frac{7}{36} + 25 \cdot \frac{9}{36} + 36 \cdot \frac{11}{36} = \frac{791}{36}$ and

$D(X) = E(X^2) - [E(X)]^2 = \frac{791}{36} - (\frac{161}{36})^2 = \frac{2555}{1296}$

Poisson distribution

The Poisson distribution is a discrete probability distribution of a random variable X that has these characteristics:

- The experiment consists of counting the number of times x, and event occurs in a given interval. The interval can be an interval of time, space, area, or volume.
- The probability of the event occurring is the same for each interval (of time, space, area, or volume).
- The number of occurrences of the event in one interval is independent of the number of occurrences in other intervals.
- The mean number of successes, denoted λ, is known over the interval. That is, λ is the expected value over the given interval.

Definition

A random variable X has the Poisson distribution with parameter λ if and only if

1. *its range of values is* $\mathcal{H}(X) = \{0, 1, 2, \dots\} = \mathbb{N} \cup \{0\}$;
2. *the probability function is*

$$f(x) = P(X = x) = \frac{\lambda^x \cdot \mathrm{e}^{-\lambda}}{x!} \quad \text{for each } x \in \mathcal{H}(X). \tag{27}$$

We denote $X \sim \mathrm{poiss}(\lambda)$.

Theorem

If $X \sim \mathrm{poiss}(\lambda)$, *then*

$$E(X) = \lambda \quad \text{and} \quad D(X) = \lambda. \tag{28}$$

Example

Certain website is visited for the period for one hour in average by 30 guests. We determine:

a) *the probability that during four minutes visit this page one guest;*
b) *the probability that during four minutes visit this page at least one guest;*
c) *the probability that during four minutes visit this page at least three, but less than eleven guests.*

Solution:

We denote by X the random variable, which takes the value of the number of visitors of website during four minutes. The expected value of X is $E(X) = \lambda_1 = 4 \cdot 30/60 = 2$.

a) According to (27) we have $P(X = 1) = \frac{2^1 \cdot \mathrm{e}^{-2}}{1!} = 0.2707.$

b) We use the opposite event:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{2^0 \cdot \mathrm{e}^{-2}}{0!} = 0.8647.$$

c) We have to compute $P(3 \leq X < 11)$:

$$P(3 \leq X < 11) = \sum_{x=3}^{10} P(X = x) = \sum_{x=3}^{10} \frac{2^x \cdot \mathrm{e}^{-2}}{x!} = 0.3233.$$

Continuous random variables and density function

Definition

A random variable X is called continuous, if there exists a non-negative and on the set $\mathbb{R}$ integrable function f such that the following assertions hold:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \quad x \in (-\infty, \infty), \tag{29}$$

where F is the distribution function of variable X.
Function f satisfying (29) is called a probability density function of random variable X.

Properties of probability density function

Theorem

Let F be the distribution function and let f be the density function .Then the folowing assertions hold:

1. *if $f(x)$ exists, then $f(x) \geq 0$;*
2. *if there exists $F'(x)$, then $F'(x) = f(x)$, $x \in \mathbb{R}$;*
3. *the normalization condition for the density function is*

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1; \tag{30}$$

4. *for $a < b$ we have $F(a) \leq F(b)$;*
5. *the distribution function is continuous on $\mathbb{R}$;*
6. *for each $a \in \mathbb{R}$ is $P(X = a) = 0$;*
7. $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x.$

Definition

Let X be a continuous random variable and let the law of the probability distribution of random variable X be given.

1. *The expected value of random variable X is the number $E(X)$ defined as follows*
$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x \tag{31}$$
2. *The variance and standard deviation of random variable X are defined as follows*
$$D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)\,\mathrm{d}x \qquad \sigma(X) = \sqrt{D(X)}. \tag{32}$$

Example

Let the function $f(x) = \begin{cases} ax^{-3} & \text{for } x \in \langle 1, 2\rangle, \\ 0 & \text{for } x \notin \langle 1, 2\rangle \end{cases}$ *be given. We determine:*

a) *the number* $a \in \mathbb{R}$ *such that* $f(x)$ *is the density function of some random variable* X;

b) *the expected value* $E(X)$;

c) *the variance* $D(X)$.

Solution.

a) We shall use the normalization condition:

$$1 = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_{1}^{2} \frac{a}{x^3}\,\mathrm{d}x = \left[\frac{-a}{2x^2}\right]_1^2 = -\frac{a}{8} + \frac{a}{2} = \frac{3a}{8}, \quad \text{which implies } a = \frac{8}{3}.$$

b) $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x = \int_{1}^{2} x \cdot \frac{8}{3x^3}\,\mathrm{d}x = \left[\frac{-8}{3x}\right]_1^2 = -\frac{8}{6} + \frac{8}{3} = \frac{8}{6} = \frac{4}{3}.$

c) We shall use formula $E(X) = E(X^2) - [E(X)^2]$.

$$E(X^2) = \int\limits_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x = \int\limits_{1}^{2} x^2 \cdot \frac{8}{3x^3}\,\mathrm{d}x = \left[\frac{8}{3}\ln x\right]_1^2 = \frac{8}{3}\ln 2$$

$$D(X) = E(X^2) - [E(X)]^2 = \tfrac{8}{3}\ln 2 - (\tfrac{4}{3})^2 \approx 0.0706$$

Example

Let us have the function $F(x) = \begin{cases} a & \text{for } x < 1, \\ bx + cx^2 & \text{for } 1 \leq x < 3, \\ d & \text{for } 3 \leq x, \end{cases}$

where a, b, c, d *are real constants. We determine:*

a) *the values of* a, b, c, d, *such that* F *can be a distribution function of some random variable* X*;*

b) *the density function* f *of those random variable.*

a) For the distribution function holds

$$\lim_{x \to -\infty} F(x) = a \qquad \text{and} \qquad \lim_{x \to \infty} F(x) = d$$

We get $a = 0$ and $d = 1$. Since X is a continuous random variable, F is a continuous function on $\mathbb{R}$, so F has to be continuous at points 1 and 3. Since

$$\begin{array}{llll} \lim\limits_{x \to 1^-} F(x) = 0. & \lim\limits_{x \to 1^+} F(x) = b + c, & \text{we get} & b + c = 0 \\ \lim\limits_{x \to 3^-} F(x) = 3b + 9c, & \lim\limits_{x \to 3^+} F(x) = 1 & \text{we get} & 3b + 9c = 1. \end{array}$$

The solution is $b = -1/6$, $c = 1/6$.

We have

$$F(x) = \begin{cases} 0 & \text{for } x < 1, \\ (-x + x^2)/6 & \text{for } 1 \leq x < 3, \\ 1 & \text{for } 3 \leq x, \end{cases}$$

which is the distribution function of some random variable X.
b) We obtain the density function f using formula $f(x) = F'(x)$:

$$f(x) = \begin{cases} 0 & \text{for } x < 1, \\ (2x - 1)/6 & \text{for } 1 \leq x < 3, \\ 0 & \text{for } 3 \leq x. \end{cases}$$

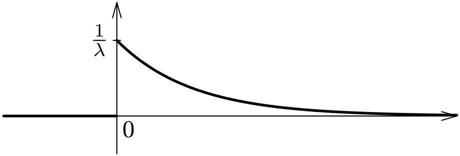
Exponential Distribution

Definition

A random variable X has the exponential distribution with parameter λ if the density function f is

$$f(x) = \begin{cases} \frac{1}{\lambda}\,\mathrm{e}^{-x/\lambda} & \textit{for } x \geq 0, \\ 0 & \textit{for } x < 0. \end{cases} \tag{33}$$

We denote the exponential distribution by $X \sim \exp(\lambda)$.



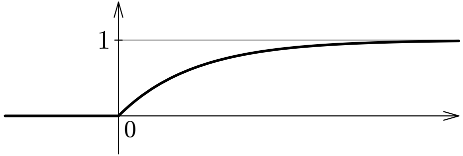
Obr.: Density function of exponential distribution

It is easy to see that $F(x) = 0$ for $x < 0$, for $x \geq 0$ we obtain

$$F(x) = \int_0^x 1/\lambda \cdot \mathrm{e}^{-\frac{t}{\lambda}}\,\mathrm{d}t = \Big[-\mathrm{e}^{-\frac{t}{\lambda}}\Big]_0^x = 1 - \mathrm{e}^{-x/\lambda}.$$

Hence the distribution function is

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 - \mathrm{e}^{-x/\lambda} & \text{for } x \geq 0. \end{cases} \tag{34}$$



Obr.: Distribution function of exponential distribution

Theorem

If $X \sim \exp(\lambda)$, then

$$E(X) = \lambda \quad \textit{and} \quad D(X) = \lambda^2. \tag{35}$$

Example

The lifetime of the product has an exponential distribution with an expected value 200 hours. We determine:

a) *the probability that the product is functional at least 300 hours;*
b) *the probability that the product will not be functional in excess of its average lifetime;*
c) *the maximum warranty period to be guaranteed if the manufacturer allows a maximum of 5% of complaints about the product.*

Solution. Let T be the random variable, which takes the value of the product lifetime. T has an exponential distribution, so $E(T) = 200 = \lambda$ and $T \sim \exp(200)$.

a) We are required to compute $P(T \geq 300)$. We obtain

$$P(T \geq 300) = 1 - P(T < 300) = 1 - P(T \leq 300) = 1 - F(300) =$$

$$1 - \left(1 - \mathrm{e}^{-1.5}\right) = 0.2231.$$

b) The average lifetime of the product is actually the expected value of the random variable T. We have

$$P(T \leq E(T)) = P(T \leq 200) = F(200) = 1 - \mathrm{e}^{-1} = 0.6321.$$

c) It is necessary to determine such a maximum value z, that the inequality $P(T \leq z) \leq 0.05$ is satisfied, so $F(z) \leq 0.05$. We get

$$1 - \mathrm{e}^{-\dfrac{z}{200}} \leq 0.05 \Rightarrow \mathrm{e}^{-\dfrac{z}{200}} \geq 0.95 \Rightarrow -\frac{z}{200} \geq \ln 0.95 \Rightarrow$$

$$z \leq -200 \cdot \ln 0.95 = 10.2587.$$

The manufacturer would probably give warranty for 10 hours.

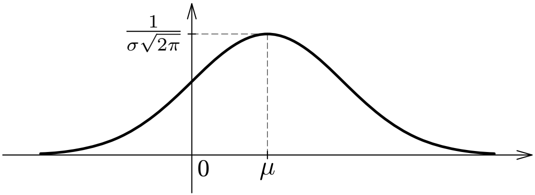
Normal (Gauss) Distribution

Definition

A random variable X has the normal (the Gauss) distribution with parameters μ and $\sigma > 0$ if the density function f is determined by formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for each } x \in \mathbb{R}. \tag{36}$$

We denote $X \sim \mathrm{norm}(\mu, \sigma)$ or $X \sim \mathrm{N}(\mu, \sigma)$.



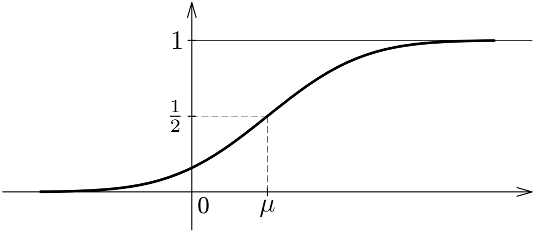
Obr.: Density function of normal distribution.

Distribution function

The distribution function is

$$F(x) = P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}} \,\mathrm{d}t \qquad \text{for each } x \in \mathbb{R}. \tag{37}$$

The graph is shown on Figure 4.



Obr.: Distribution function of normal distribution.

Theorem

If $X \sim \mathrm{norm}(\mu, \sigma)$, *then*

$$E(X) = \mu \quad \textit{and} \quad D(X) = \sigma^2. \tag{38}$$

By scaling of the random variable $X \sim \mathrm{norm}(\mu, \sigma)$ we get the random variable

$$Y = \frac{X - \mu}{\sigma}$$

such that $E(Y) = 0$ and $D(Y) = 1$, so $Y \sim \mathrm{norm}(0, 1)$. The density function φ of random variable Y is given by

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{y^2}{2}} \qquad \text{for each } y \in \mathbb{R} \tag{39}$$

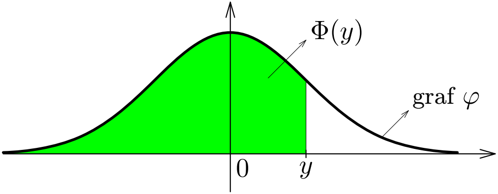
and the distribution function Φ by

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \mathrm{e}^{-\frac{u^2}{2}} \, \mathrm{d}u \qquad \text{for each } y \in \mathbb{R}. \tag{40}$$

We can derive the relation between $F(x)$ and $\Phi(y)$ in the following way:

$$F(x) = P(X \le x) = P(X - \mu \le x - \mu) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) =$$
$$P\left(Y \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

The number $\Phi(y)$ determines the area of the marked surface



The function φ is even. This follows that

$$\Phi(-y) = 1 - \Phi(y) \quad \text{for each } y \in \mathbb{R}, \quad \left(\text{specially } \Phi(0) = \frac{1}{2}\right). \tag{41}$$

Theorem

If $X \sim \mathrm{norm}(\mu, \sigma)$, then for each $a < b$ we have

$$P(a < X < b) = P(a < X \le b) = P(a \le X \le b) = P(a \le X < b) =$$
$$= F(b) - F(a) = \Phi\left(\tfrac{b-\mu}{\sigma}\right) - \Phi\left(\tfrac{a-\mu}{\sigma}\right). \tag{42}$$

Example

The mass of produced weights has a normal probability distribution with mean values of 10 grams. The manufacturer provides a standard deviation of 0.02 g. We determine the probability that randomly bought weights will have a real mass

a) *greater than 10.03 g;*

b) *less than 9.99 g;*

c) *at least 10 g but not more than 10.05 g.*

Solution:

Let X be a random variable, which takes the value of real mass of bought weights. Obviously $X \sim \text{norm}(10, 0.02)$.

a) We want to determine $P(X > 10.03)$. We get

$$P(X > 10.03) = 1 - P(X \leq 10.03) = 1 - F(10.03) =$$

$$= 1 - \Phi\left(\frac{10.03 - 10}{0.02}\right) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668.$$

$\surd$

b) We are required to compute $P(X < 9.99)$. We have

$$P(X < 9.99) = P(X \leq 9.99) = F(9.99) = \Phi\left(\frac{9.99 - 10}{0.02}\right) =$$

$$\Phi(-0.5) = 1 - \Phi(0.5) = 0.3085.$$

c) According to (42) we have

$$P(10 \leq X \leq 10.05) = \Phi\left(\frac{10.05 - 10}{0.02}\right) - \Phi\left(\frac{10 - 10}{0.02}\right) =$$

$$= \Phi(2.5) - \Phi(0) = 0.4938.$$

Numerical Methods

In this chapter we will show some basic numerical methods for solving of the equation of the real variable $f(x) = 0$.
We have to find such points $\alpha \in \mathbb{R}$ for which $f(\alpha) = 0$. These points α are called the roots of the equation $f(x) = 0$.
When solving the equation $f(x) = 0$, we try to determine the number of roots and we determine the intervals in which there is exactly one root of the equation. The process of finding these intervals is called the separation of the roots of the equation $f(x) = 0$.
We say that the roots of the equation $f(x) = 0$ are separated, if there are the intervals $\langle a_1, b_1 \rangle$, $\langle a_2, b_2 \rangle, \dots, \langle a_n, b_n \rangle$ such that

(i) $(a_i, b_i) \cap (a_j, b_j) = \emptyset$ for $i, j \in \{1, 2, \dots, n\}$, $i \neq j$;

(ii) Each interval $\langle a_i, b_i \rangle$ contains exactly one root of the equation $f(x) = 0$.

Sufficient condition

Theorem

Let a function $f : y = f(x)$ be given and let $\langle a, b\rangle \subseteq \mathcal{D}(f)$. Let $f(x)$ be a continuous on an interval $\langle a, b\rangle$. If

$$f(a) \cdot f(b) \le 0, \tag{43}$$

then there is at least one root of the equation $f(x) = 0$ on the interval $\langle a, b\rangle$.

Corollary

Let x_n be such that $f(x_n - \varepsilon) \cdot f(x_n + \varepsilon) < 0$. Then there is a root $\alpha \in (x_n - \varepsilon, x_n + \varepsilon)$, so x_n is a solution of the equation $f(x) = 0$ with accurancy ε. This test is called a $\pm\varepsilon$-test.

On the interval $\langle a, b\rangle$ can also be more than one root.

- If (43) is satisfied, then the interval $\langle a, b\rangle$ contains an odd number of roots.
- If (43) is not satisfied, then the interval $\langle a, b\rangle$ either does not contain a root or it contains an even number of roots.

Example

Bisection Method

The Bisection method is the simplest numerical method for solving non-linear equations.

Let a continuous function $f : y = f(x)$ contains exactly one root α on an interval $\langle a, b\rangle$. Our task is to find such approximative c_n, which is sufficiently close to the root α.

An algorithm:

We define the sequence of intervals $\langle a_n, b_n\rangle$, $n = 0, 1, 2, \dots$ given as follows:

1. $\langle a_0, b_0\rangle = \langle a, b\rangle$
2. Let the interval $\langle a_n, b_n\rangle$ be defined, whereby $f(a_n) \cdot f(b_n) < 0$. We put
$$c_n = (a_n + b_n)/2. \tag{44}$$
3. If $f(c_n) = 0$, then c_n is a root. If $f(c_n) \neq 0$ and $f(a_n) \cdot f(c_n) < 0$, then we put $a_{n+1} = a_n$, $b_{n+1} = c_n$. If $f(a_n) \cdot f(c_n) > 0$, then we put $a_{n+1} = c_n$, $b_{n+1} = b_n$. If the sequence $\{c_n\}_{n=1}^{\infty}$ is finite, then the last member is a root of $f(x) = 0$. If the sequence $\{c_n\}_{n=1}^{\infty}$ is infinite, then it has a limit
$$\lim_{n\to\infty} c_n = \alpha.$$
4. go to Step 2

Error Estimate

We know that the root α lies inside the interval $\langle a_n, b_n \rangle$. Therefore, the approximate value c_k can be from the exact value α at a distance at most half of the length of the interval $\langle a_n, b_n \rangle$. Therefore,

$$|c_n - \alpha| < \frac{b-a}{2^{n+1}}.$$

Let us calculate the root of the equation with the given accuracy $\varepsilon > 0$, we end the dividing process if the inequality $|b_n - a_n| < 2\varepsilon$ holds and put a root equal to the number

$$\alpha \approx (a_n + b_n)/2.$$

Example

Using Bisection method solve the equation $x^3 - x - 1 = 0$ with accurancy $\varepsilon = 0.005$.

We rewrite the given equation into the form $x^3 = x + 1$. Denote $h(x) = x + 1$, $g(x) = x^3$. The graphs of the functions f, g intersect at one point, whose x-coordinate is within interval $(1; 2)$. Indeed, $f(1) \cdot f(2) < 0$, so $\alpha \in (1; 2)$. We put $a_0 = 1$, $b_0 = 2$ and will end when $|a_n - b_n| < 2 \cdot \varepsilon = 0.01$.

n	$a_n\,(f(a_n))$	$b_n\,(f(b_n))$	$c_n\,(f(c_n))$	$f(a_n) \cdot f(c_n)$
0	1 (–)	2(+)	1.5 (+)	+
1	1 (–)	1.5 (+)	1.25 (–)	+
2	1.25 (–)	1.5 (+)	1.375 (+)	–
3	1.25 (–)	1.375 (+)	1.3125 (–)	+
4	1.3125 (–)	1.375 (+)	1.34375 (+)	–
5	1.3125 (–)	1.34375 (+)	1.32813 (+)	–
6	1.3125 (–)	1.32813 (+)	1.32032 (–)	+
7	1.32032 (–)	1.32813 (+)	1.32423	

Since $|1.32813 - 1.32032| < 0.01$ a $f(1.32813) \cdot f(1.32032) < 0$, we can approximate the root with the value c_7. Therefore $\alpha \approx 1.32423$.

Newton's Method

Newton's Method, also called tangential method, is the method that approximates the solution of the equation $f(x) = 0$ using the tangents to the graph of the function f. Suppose that there is exactly one root on interval $\langle a, b\rangle$. Let the first and the second derivative of the function $f(x)$ be continuous and $f''(x)$ does not change sign on $\langle a, b\rangle$, i. e., f is either convex on $\langle a, b\rangle$ or concave on $\langle a, b\rangle$. We choose the initial approximation x_0 such that $f(x_0) \cdot f''(x_0) > 0$ We construct a tangent line t_0 to the graph of the function f at the point $[x_0, f(x_0)]$. We denote the intersection of t_0 with x-axis as the point x_1. In this way we obtain the sequence of approximations $x_0,\ x_1,\ x_2, \dots$

The equation of the tangent t_n at the point $[x_n, f(x_n)]$ is

$$t_n :\ \ y - f(x_n) = f'(x_n)(x - x_n).$$

The intersection of t_n with x-axis $(y = 0)$ is x_{n+1} such that

$$-f(x_n) = f'(x_n)(x_{n+1} - x_n).$$

We get the formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots. \tag{45}$$

Theorem

Let the following conditions hold at the interval $\langle a, b\rangle$:

1. $f(a) \cdot f(b) < 0$.
2. $|f''(x)| > 0$ for each $x \in \langle a, b\rangle$, i. e., $f''(x)$ does not change a sign on $\langle a, b\rangle$.
3. $f(x_0) \cdot f''(x_0) > 0$, $x_0 \in \langle a, b\rangle$.

Then the sequence given by formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \ldots . \tag{46}$$

converges to the root α of the equation $f(x) = 0$, that $\lim_{n\to\infty} x_n = \alpha$ is true. In addition, when the condition

4. $|f'(x)| > 0$ for each $x \in \langle a, b\rangle$ is true, , i. e., $f'(x)$ does not change a sign on $\langle a, b\rangle$, we can estimate the accuracy by formula $|x_n - \alpha| \leq \frac{|f(x_n)|}{m}$, where $m = \min_{x\in\langle a,b\rangle} |f'(x)|$.

STOP test:

(i) if $|f'(x)| > 0$ for each $x \in \langle a, b\rangle$ and $\frac{|f(x_n)|}{m} < \varepsilon \Rightarrow \alpha \approx x_n$

(ii) If $f(x_n - \varepsilon) \cdot f(x_n + \varepsilon) < 0 \Rightarrow \alpha \approx x_n$

Example

Using Newton's method solve the equation $x^3 - x - 1 = 0$ with accuracy $\varepsilon = 0,005$.

We have
$f(1) = -1 < 0,\ f(1.5) = 0.875 > 0 \Rightarrow \alpha \in \langle 1, 1.5 \rangle$.
$f'(x) = 3\,x^2 - 1 > 0$ for each $x \in \langle 1, 1.5 \rangle$; $m = \min\limits_{x \in \langle 1;1.5 \rangle} |f'(x)| = 2$.
$f''(x) = 6\,x > 0$ for each $x \in \langle 1, 1.5 \rangle$ and $f(1.5) > 0 \Rightarrow x_0 = 1.5$.

n	x_n	$f(x_n)$	$f'(x_n)$	$\frac{\vert f(x_n)\vert}{m}$
0	1.5	0, 875	5.75	0.4375
1	1.34783	0.10068	4.44990	0.05034
2	1.32520	0.00205	4.26846	$0.001025 < \varepsilon$

Since

$$|x_2 - \alpha| \le \frac{|f(x_2)|}{m} < \varepsilon,$$

we obtain $\alpha \approx 1.32520$.

The approximation of functions

It is often the case, that the function is not specified by any functional expression and we know only its functional values in some of its points, which are often received from different measurements. For such functions it is difficult to obtain the functional values in other than the given points, find its derivative, or integrate it. Therefore it is appropriate to replace this function by another function, which is similar to given function and it has the functional expression and it is easy to do calculations with it. The most commonly used in approximation as the polynomial function (polynomial of the n-th degree), which is defined on the set $\mathbb{R}$. A polynomial function is easily differentiable and integrable on the set $\mathbb{R}$.
The requirements on the function, by which we want to approximate the given function may be various.

- If we use interpolation for approximation of function, we demand that approximate function has the same functional values at selected points as original function.
- When using the least squares method is not necessary for approximative function to directly pass through given points, just that in a certain sense, to be as close as possible to given points of original function.

Interpolation

Let the function f be given by $n+1$ different points $x_0, x_1, \dots, x_n$. These points are called nodes of interpolation. We denote the function values at those points as $y_0, y_1, \dots, y_n$, where $y_0 = f(x_0)$, $y_1 = f(x_1)$, $y_2 = f(x_2), \dots, y_n = f(x_n)$.
Thus, the function f is given by the table of values

x	x_0	x_1	x_2	$\dots$	x_n
y	y_0	y_1	y_2	$\dots$	y_n

Let us find the polynomial $P_n(x)$ of degree at most n, such that the node points take the same functional values as the function f, i. e.,

$$P_n(x_i) = y_i \text{ for } i = 0, 1, 2, \dots, n.$$

Theorem

Let nodes $[x_i, y_i]$ for $i = 0, 1, 2, \dots, n$ be given and $x_i \neq x_j$ for $i \neq j$. Then there exists a polynomial $P_n(x)$ of degree at most n such that $P_n(x_i) = y_i$ for $i = 0, 1, 2, \dots, n$.

Lagrange polynomial

We approximate a function f given by functional values in $n+1$ points with polynomial in the form

$$L_n(x) = \sum_{i=0}^{n} y_i \cdot g_i(x) \tag{47}$$

such that

$$L_n(x_i) = y_i, \quad i = 0, 1, \dots, n. \tag{48}$$

We will call the points x_i nodal points.

The functions g_i are chosen so that $g_i(x_j) = \begin{cases} 1, & \text{pre } i = j, \\ 0, & \text{pre } i \neq j. \end{cases}$ Since $g_i(x_j) = 0$ for $j \neq i$, we obtain

$$g_i(x) = C_i(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n),$$

where C_i is a real constant, calculated by the conditions of $g_i(x_i) = 1$, i. e.,

$$C_i = \frac{1}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}$$

Consequently, the polynomial $L_n(x)$ gets the form

$$L_n(x) = \sum_{i=0}^{n} \frac{(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)} \cdot y_i.$$

Example

Replace the given function with a polynomial $L_2(x)$ and calculate the approximate value of the function f at point $x = 3$ with $L_2(3)$ if the specified function values $f : \langle 1, 4\rangle \to \mathbb{R}$ are given by the table:

x	1	2	4
y	2	3	11

.

For $n = 2$ we have

$$L_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} \cdot y_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} \cdot y_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \cdot y_2.$$

According to the given function values we obtain

$$L_2(x) = \frac{(x-2)(x-4)}{(1-2)(1-4)} \cdot 2 + \frac{(x-1)(x-4)}{(2-1)2-4)} \cdot 3 + \frac{(x-1)(x-2)}{(4-1)(4-2)} \cdot 11 = x^2 - 2x + 3.$$

For $x = 3$ we obtain

$$L_2(3) = \frac{(3-2)(3-4)}{(1-2)(1-4)} \cdot 2 + \frac{(3-1)(3-4)}{(2-1)(2-4)} \cdot 3 + \frac{(3-1)(3-2)}{(4-1)(4-2)} \cdot 11 = 6.$$

Least squares method

In experiments measurements are often carried out several times under the same conditions, which is contrary to the assumptions of interpolation, requiring that all nodes are different from each other. Also, in measurements we receive data containing errors. This data are therefore not appropriate to interpolating as this would propagate these errors. Therefore, if we know at least a little about the functional relationship (linear, quadratic, logarithmic, exponential, etc.), we can approximate that function so that from the supposed type of functional dependency (the set of all linear function or the set of all quadratic functions, etc.) we choose such a function, which is as close as possible to the real situation.

Formulation of the problem

Let nodes $x_0, x_1, x_2, \ldots, x_n$ and corresponding function values $y_0, y_1, y_2, ..., y_n$ be given. Let functions $\varphi_0, \varphi_1, \ldots, \varphi_k$, where $k < n+1$ be given.
Then, from all functions of the form

$$\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_k\varphi_k(x)$$

we are looking for such a function, for which the quadratic error

$$S(a_0, a_1, \ldots, a_k) = \sum_{i=0}^{n}[\varphi(x_i) - y_i]^2 = \sum_{i=0}^{n}[a_0\varphi_0(x_i) + a_1\varphi_1(x_i) + \cdots + a_k\varphi_k(x_i) - y_i]^2$$

is minimal.
The points $[x_i, y_i]$, $i = 0, 1, 2, \ldots, n$ and functions $\varphi_0, \varphi_1, \ldots, \varphi_k$ are given, therefore the quadratic error depends only on the coefficients $a_0, a_1, \ldots, a_k$ of the function $\varphi(x)$.
From the differential calculus of functions of several variables we know, that a necessary condition for the function $S(a_0, a_1, \ldots, a_k)$ to obtain a minimum, is fulfilment of the conditions expressed by

$$\frac{\partial S}{\partial a_j} = 0, \qquad j = 0, 1, \ldots, k.$$

We will search an approximate function in the form $\varphi(x) = a_0 \cdot \varphi_0(x) + a_1 \cdot \varphi_1(x)$. Suppose that the values $f(x_i)$, $i = 0, 1, \ldots, n$ were measured with approximately same precision. We calculate the coefficients a_0, a_1 so that the sum of the squares d_i^2, $i = 0, 1, \ldots, n$ was the smallest, i. e., we are looking for the local minimum of the function

$$S(a_0, a_1) = \sum_{i=0}^{n} [a_0 \cdot \varphi_0(x_i) + a_1 \cdot \varphi_1(x_i) - y_i]^2.$$

We have partial derivatives

$$\frac{\partial S}{\partial a_0} = 2 \sum_{i=0}^{n} [a_0 \varphi_0(x_i) + a_1 \varphi_1(x_i) - y_i] \cdot \varphi_0(x_i),$$

$$\frac{\partial S}{\partial a_1} = 2 \sum_{i=0}^{n} [a_0 \varphi_0(x_i) + a_1 \varphi_1(x_i) - y_i] \cdot \varphi_1(x_i),$$

From the conditions $\frac{\partial S}{\partial a_0} = 0$, $\frac{\partial S}{\partial a_1} = 0$ we obtain

$$a_0 \cdot \sum_{i=0}^{n} \varphi_0(x_i) \varphi_0(x_i) + a_1 \cdot \sum_{i=0}^{n} \varphi_1(x_i) \varphi_0(x_i) = \sum_{i=0}^{n} y_i \varphi_0(x_i),$$

$$a_0 \cdot \sum_{i=0}^{n} \varphi_0(x_i) \varphi_1(x_i) + a_1 \cdot \sum_{i=0}^{n} \varphi_1(x_i) \varphi_1(x_i) = \sum_{i=0}^{n} y_i \varphi_1(x_i).$$

Example

We approximate the given function by a polynomial of degree one:

x_i	0.2	0.3	0.5	0.7	0.8	1	1.1
y_i	3.45	3.54	4.1	4.35	4.6	5.05	5.14

.

We have

$$\varphi(x) = a_0 + a_1 x,$$

therefore $\varphi_0(x) = 1$, $\varphi_1(x) = x$. The theoretical system of equations has the form

$$a_0 \sum_{i=0}^{n} 1 \cdot 1 + a_1 \sum_{i=0}^{n} 1 \cdot x_i = \sum_{i=0}^{n} 1 \cdot y_i,$$

$$a_0 \sum_{i=0}^{n} x_i \cdot 1 + a_1 \sum_{i=0}^{n} x_i \cdot x_i = \sum_{i=0}^{n} x_i \cdot y_i.$$

Substituting the values from the table we receive

$$7a_0 + 4.6a_1 = 30.23.$$

$$4.6a_0 + 3.72a_1 = 21.231.$$

The solution is $a_0 = 3.03135$, $a_1 = 1.9588$. Hence

$$\varphi(x) = 3.03135 + 1.9588x.$$

Numerical integration

It is well known that not every function must have a primitive function (anti-derivative), or to find the primitive function can be very complicated. Unfamiliarity of primitive function does not allow us to use the Newton-Leibniz formula to calculate the definite integral.
In numerical analysis, numerical integration constitutes a broad family of algorithms for calculating the numerical value of a definite integral of the form

$$\int_a^b f(x)\,\mathrm{d}x,$$

where $a < b$ are real numbers. A large class of quadrature rules can be derived by constructing interpolating functions that are easy to integrate. Typically these interpolating functions are polynomials. In practice, only polynomials of low degree are used, typically linear and quadratic.
Suppose that a function $f : \langle a, b\rangle \to \mathbb{R}$ is integrable on $\langle a, b\rangle$. We divide the interval $I = \langle a, b\rangle$ to n intervals of the same length with nodal points

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b,$$

where $x_i = x_0 + i \cdot h$, $i = 0, 1, \ldots, n$ and $h = \dfrac{x_n - x_0}{n} = \dfrac{b - a}{n}$ is a constant step.

Trapezoidal method

At each interval $\langle x_i, x_{i+1} \rangle$ for $i = 0, 1, 2, \ldots, n-1$, we approximate the function f by a linear function, which passes through the points $[x_i, f(x_i)]$ and $[x_{i+1}, f(x_{i+1})]$. We approximate the surface of the original area over the interval $\langle x_i, x_{i+1} \rangle$ using the area of trapezoid over the interval $\langle x_i, x_{i+1} \rangle$ which can be expressed by the formula:

$$S_i = \frac{h}{2} \cdot [f(x_i) + f(x_{i+1})].$$

This is an elementary formula for trapezoidal method.
Summing of all areas we obtain:

$$\int_a^b f(x)\,\mathrm{d}x \approx T(n) = \frac{h}{2} \cdot (f(a) + f(x_1)) + \frac{h}{2} \cdot (f(x_1) + f(x_2)) + \cdots +$$

$$\frac{h}{2} \cdot (f(x_{n-2}) + f(x_{n-1}) + \frac{h}{2} \cdot (f(x_{n-1}) + f(b)) = \frac{h}{2} \cdot \left[f(a) + f(b) + 2 \cdot \sum_{i=1}^{n-1} f(x_i) \right].$$

The calculation contains an error which is bounded by

$$|R_T(n)| \leq \frac{(b-a)^3 M_2}{12n^2}, \quad \text{where} \quad M_2 \geq \max_{x \in \langle a,b \rangle} |f''(x)|.$$

Example

Using the trapezoidal method, calculate $\int_4^{6.4} \ln x \, dx$ *for* $n = 6$ *and estimate the calculation error.*

Solution:

We have $f(x) = \ln x$, $a = x_0 = 4$, $b = x_6 = 6.4$. We compute $h = \frac{6.4-4}{6} = 0.4$

and create a table:

i	x_i	$f(x_i)$
0	4	1.3862
1	4.4	1.4816
2	4.8	1.5686
3	5.2	1.6486
4	5.6	1.7227
5	6	1.7919
6	6.4	1.8562

We obtain $\int_4^{6.4} \ln x \, dx \approx$
$\frac{0.4}{2} \cdot (1.3862 + 1.8562 + 2 \cdot (1.4816 + 1.5686 + 1.6486 + 1.7227 + 1.7919)) = 3.93388$
To estimate the calculation error we differentiate: $f'(x) = \frac{1}{x}$, $f''(x) = \frac{-1}{x^2}$ and calculate $M_2 = \max\limits_{x \in \langle 4;6.4 \rangle} |\frac{-1}{x^2}| = \frac{1}{16}$. Using formula for calculation error we obtain $R_T(6) \leq \frac{2.4^3 \cdot \frac{1}{16}}{12 \cdot 6^2} = 0.002.$

Simpson's method

Simpson's rule is is based on a polynomial of order 2. Let us divide the interval $\langle a, b\rangle$ to the even number of intervals. We approximate the function f by the quadratic function which passes through the points $[x_{2i}, f(x_{2i})]$, $[x_{2i+1}, f(x_{2i+1})]$, and $[x_{2i+2}, f(x_{2i+2})]$ at each interval $[x_{2i}, x_{2i+2}]$. We replace the area of the original region over the interval $[x_{2i}, x_{2i+2}]$ by the area of a curvilinear trapezoid, which can be expressed by formula

$$S_{2i} = \frac{h}{3} \cdot [f(x_{2i}) + 4\, f(x_{2i+1}) + f(x_{2i+2})].$$

This is an elementary formula for Simpson's method. Summing of elementary formulas we obtain

$$\int_a^b f(x)\,\mathrm{d}x \approx S(n) = \frac{h}{3} \cdot \Big[f(a) + f(b) + 4\big(f(x_1) + f(x_3) + \ldots f(x_{n-1})\big) +$$

$$+2 \cdot \big(f(x_2) + f(x_4) + \cdots + f(x_{n-2})\big)\Big]\,.$$

The calculation contains an error which is bounded by

$$|R_S(n)| \leq \frac{(b-a)^5 M_4}{180 n^4}, \quad \text{where} \quad M_4 \geq \max_{x \in \langle a,b\rangle} |f^{(4)}(x)|. \qquad (49)$$

Example

Using the Simpson's method, calculate $\int_4^{6.4} \ln x\, dx$ *for* $n = 6$ *and estimate the calculation error.*

Solution:

i	x_i	$f(x_i)$
0	4	1.3862
1	4.4	1.4816
2	4.8	1.5686
3	5.2	1.6486
4	5.6	1.7227
5	6	1.7919
6	6.4	1.8562

We obtain $\int_4^{6.4} \ln x\, dx \approx$
$\frac{0.4}{3} \cdot (1.3862 + 1.8562 + 4 \cdot (1.4816 + 1.6486 + 1.7919) + 2 \cdot (1.5686 + 1.7227)) = 3.93512.$
To estimate the calculation error we differentiate: $f'''(x) = \frac{2}{x^3}$, $f''''(x) = \frac{-6}{x^4}$ and calculate $M_4 = \max\limits_{x \in \langle 4; 6.4 \rangle} |\frac{-6}{x^4}| = \frac{3}{128}$. Using formula for calculation error we obtain $R_S(6) \leq \frac{2.4^5 \cdot \frac{3}{128}}{180 \cdot 6^4} = 8 \cdot 10^{-6}$.

$\surd$