

Part-I

Probability of events

Andras Vetier

2015 09 09

Contents

1	Introductory problems	3
2	Outcomes and events	10
3	Relative frequency and probability	12
4	Random numbers	14
5	Classical problems	16
6	Geometrical problems, uniform distributions	18
7	Basic properties of probability	22
8	Conditional relative frequency and conditional probability	24
9	Independence of events	29
10	*** Infinite sequences of events	31
11	*** Drawing with or without replacement. Permutations	35

Preface

Usual text-books in probability theory describe the laws of randomness by a text consisting of sentences, formulas, etc., and a collection of examples, problems, figures, etc. which are printed permanently in the book. The reader may read the text, study the examples, the problems, and look at the figures as many times as he or she wants to. That is OK. However, the laws of randomness can be experienced only if many experiments are performed many times. It is important to see the effect of the change of the parameters, as well. In a permanently printed text-book what is printed is printed, and cannot be changed. The reader cannot modify the parameters.

The main purpose of this electronic text-book is to make it possible to simulate the experiments as many times as the reader wishes, and to make it possible to change the parameters according to the wish of the reader. For the simulations, we use the program Excel, it is available in high-schools and universities, and most students know it to a certain level.

I am convinced that having experienced the real life meaning of the notions of probability theory, the mathematical notions and the mathematical proofs become more interesting and attractive for the reader. Since the mathematical proofs are available in many usual textbooks, we give only a few proofs in this textbook.

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

A list of suggested textbooks is available from the web-page:

<http://www.math.bme.hu/~vetier/Probability.htm>

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

Sections marked by *** may be skipped.

1 Introductory problems

Example 1. Coming home from Salzburg to Vac. My sister-in-law regularly plays the violin in an orchestra in Salzburg almost every Saturday evening, and comes home to Vac on Sundays. (Salzburg, hometown of W. A. Mozart is 600 km west of Budapest, and Vac, a little town next the Danube is 30 km north of Budapest.) Her friend brings her in his car to Nyugati railway station in Budapest, where she takes the train to Vac. The train leaves for Vac every hour. Sometimes she arrives to Nyugati railway station some minutes after the departure of the previous train, and has to wait almost an hour. Other times she arrives to Nyugati railway station some minutes before the departure of the next train, and she has to wait only some minutes. We may be interested in the amount of time passing after the departure of the previous train. Using the following file, you may study a simulation of the amount of time passing after the previous train.

Demonstration file: The amount of time after the departure of the previous train
ef-020-01-00

We may be interested in the amount of time she has to wait until the next train. It is natural to call this amount of time the waiting time until the next train. In the following file, the waiting time is also shown.

Demonstration file: Both the amount of time after the previous train and the waiting time until the next train are shown
ef-020-02-00

As you see the amount of time after the previous train is generated by the command `RAND ()`. This command gives a random number between 0 and 1, so the `RAND () * 60` command gives a random number between 0 and 60. Rounding is performed by the commands `ROUNDDOWN (_; _)` and `ROUNDUP (_; _)`.

Imagine that you observe the amount of time after the previous train on 10 occasions. You will get 10 real numbers between 0 and 60. In the following file, 10 experiments are simulated.

Demonstration file: 10 experiments for the amount of time after the previous train
ef-020-03-00

As you see, the same command, namely, `RAND () * 60` is used in all the 10 cells, but the numerical values returned are different.

If we make 1000 experiments, then - as you see in the next file - the 1000 corresponding dots overcrowd the line.

Demonstration file: 1000 experiments on a line
ef-020-04-00

This is why, for visualization purposes, we give each of the points a different second coordinate, as if the points were moved out from the line into a narrow horizontal strip. In the following file, where only 10 points are shown, you may check how the points jump out of the line.

Demonstration file: 10 experiments on a narrow horizontal strip
ef-020-05-00

When there are 1000 points, the points melt together on the overcrowded line, while the distribution of the points on the narrow horizontal strip is really expressive: whenever you press the F9-key, you may see that the points are uniformly distributed between 0 and 60, they constitute a uniformly distributed point-cloud.

Demonstration file: 1000 experiments on a narrow horizontal strip
ef-020-06-00

For my sister-in-law, it is a rather unpleasant event when she has to wait until the next train for more than 45 minutes. Waiting more than 45 minutes obviously means that the amount of time after the departure of the previous train is less than 15 minutes. In the next file, these points are identified, their number - the so called frequency of the event - is calculated, and then the relative frequency of the event, that is, the frequency divided by the total number of experiments is also calculated.

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

In order keep track of whether $\text{RAND}() * 60$ is less than 15 or not, in the simulation file, we use the `IF (_; _; _)` command. The structure of this command is very simple: the first argument is a condition, the second argument is the value of the command if the condition holds, the third argument is the value of the command if the condition does not hold.

Pressing the F9-key in the previous simulation file, you may be convinced that the relative frequency oscillates around a non-random value, in this problem, around 0.25. This value, around which the relative frequency oscillates, is an important characteristic of the event. We call this number the probability of the event, and we write:

$$P(\text{amount of time after the previous train} < 15) = 0.25$$

In the next file, the value of the probability is also visualized. You may see that the relative frequency oscillates around it.

Demonstration file: Probability of the unpleasant event
ef-020-08-00

Example 2. Random numbers. Random numbers generated by computers play an essential role in our simulation files. The basic property of a random number is that,

for any $0 \leq a \leq b \leq 1$, the probability of the event that the random number is between a and b is equal to the length of the interval $[a; b]$, which is $b - a$:

$$\mathbf{P}(a \leq \text{RND} \leq b) = \text{length of the interval } [a; b] = b - a$$

$$\mathbf{P}(a \leq \text{RND} < b) = \text{length of the interval } [a; b) = b - a$$

$$\mathbf{P}(a < \text{RND} \leq b) = \text{length of the interval } (a; b] = b - a$$

$$\mathbf{P}(a < \text{RND} < b) = \text{length of the interval } (a; b) = b - a$$

Whether the interval is closed or open, it does not make any difference in the value of the probability. In the following file, you may choose the values a and b , the left and right end-points of the interval. You will see that the relative frequency of the event $a \leq \text{RND} \leq b$ really oscillates around $b - a$.

Demonstration file: Probability of an interval for a random number generated by computer
ef-020-09-00

It is important to remember that, for any fixed number x which is between 0 and 1, we have that

$$\mathbf{P}(\text{RND} \leq x) = \text{length of the interval } [0; x] = x$$

$$\mathbf{P}(\text{RND} < x) = \text{length of the interval } [0; x) = x$$

Demonstration file: Probability of $\text{RND} < x$
ef-020-10-00

Example 3. Pairs of random numbers. Another basic property of the `RAND()` command is that using it twice, and putting the two random numbers RND_1 and RND_2 together to define a random point $(\text{RND}_1; \text{RND}_2)$, for this point it holds that, for any set A inside the unit square, it holds that

$$\mathbf{P}((\text{RND}_1; \text{RND}_2) \in A) = \text{area of } A$$

In order to see this fact, in the following file, A can be any triangle inside the unit square.

Demonstration file: Probability of a triangle
ef-020-11-00

In the following file, the relative frequencies of more complicated events are studied:

Demonstration file: Special triangle combined with a diamond-shaped region - unconditional ...
ef-020-12-00

In the following file not only frequencies and probabilities, but conditional frequencies and probabilities are involved. Playing with the file, you will discover the notion of conditional frequency and conditional probability.

Demonstration file: Special triangle combined with a diamond-shaped region - conditional ...
ef-020-13-00

Example 4. Non-uniform distributions. Just to see a point-cloud which is **not** uniformly distributed, let us replace the `RAND ()` command by the `POWER (RAND () ; 2)` command. The command `POWER (_; 2)` stands for taking the square.

Demonstration file: Non-uniformly distributed points using the square of a random number
ef-020-14-00

We get another non-uniformly distributed point-cloud if we apply the square-root function, `POWER (_; 1 / 2)`

Demonstration file: Non-uniformly distributed points using the square-root of a random number
ef-020-15-00

In the next file, relative frequencies related to non-uniform distributions are calculated.

Demonstration file: Relative frequency for non-uniform distribution
ef-020-16-00

In the next file, conditional relative frequencies related to non-uniform distributions are calculated.

Demonstration file: Conditional relative frequency for non-uniform distribution
ef-020-17-00

Example 5. Waiting time for the bus. My wife goes to work by bus every day. She waits for the bus no more than 10 minutes. The amount of time she waits for the bus is uniformly distributed between 0 and 10. In the next file, we simulate this waiting time, and we study the event that "the waiting time < 4". The probability of this event is obviously 0.4 The relative frequency of the event will clearly oscillate around 0.4.

Demonstration file: Waiting time for the bus
ef-020-18-00

Example 6. Traveling by bus and metro. My friend goes to work by bus and metro every day. He waits for the bus no more than 10 minutes. The amount of time he waits for the bus is uniformly distributed between 0 and 10. When he changes to the metro,

he waits for the metro no more than 5 minutes. No matter how much he waited for the bus, the amount of time he waits for the metro is uniformly distributed between 0 and 5.

This example involves two waiting times. As you will see in the next simulation file, the two waiting times together define a uniformly distributed random point in a rectangle.

*Demonstration file: Traveling by bus and metro: uniformly distributed waiting times
020-19-00*

Some events are visualized in the following files:

*Demonstration file: Waiting time for bus < 4 , using uniform distribution
ef-020-20-00*

*Demonstration file: Waiting time for metro > 3 , using uniform distribution
ef-020-21-00*

*Demonstration file: Waiting time for bus < 4 AND waiting time for metro > 3 , using uniform distribution
ef-020-22-00*

*Demonstration file: Waiting time for bus $<$ waiting time for metro, using uniform distribution
ef-020-23-00*

*Demonstration file: Total waiting time is more than 4, using uniform distribution
ef-020-24-00*

*Demonstration file: Waiting time for bus $<$ waiting time for metro, AND total waiting time > 4
ef-020-25-00*

*Demonstration file: Waiting time for bus $<$ waiting time for metro, OR total waiting time > 4 , using uniform distribution
ef-020-26-00*

Under certain conditions, the application of uniform distribution for the waiting times is justified, but under other conditions it is not. If the busses and metros follow "strict time-tables" and randomness is involved in the problem only because my friend does not follow a "strict time-table", then the application of uniform distribution for the waiting times gives a good model. However, if the busses and metros arrive to the stations where my friend gets on them, in a "chaotic" way, then - as we learn later - the application of a special non-uniform distribution - called "exponential" distribution - is more correct. You may see in the next file that exponentially distributed waiting times

are generated by using the `-LN (RAND ())` command, that is, taking the minus of the natural logarithm of a simple random number.

Demonstration file: Traveling by bus and metro, using exponential distribution
ef-020-27-00

The events studied above are visualized with exponentially distributed waiting times in the following files:

Demonstration file: Waiting time for bus < 4 , using exponential distribution
ef-020-28-00

Demonstration file: Waiting time for metro > 3
ef-020-29-00

Demonstration file: Waiting time for bus < 4 AND waiting time for metro > 3 , using exponential distribution
ef-020-30-00

Demonstration file: Waiting time for bus < waiting time for metro, using exponential distribution
ef-020-31-00

Demonstration file: Total waiting time > 4 , using exponential distribution
ef-020-32-00

Demonstration file: Waiting time for bus < waiting time for metro AND total waiting time > 4 , using exponential distribution
ef-020-33-00

Demonstration file: Waiting time for bus < waiting time for metro OR total waiting time > 4 , using exponential distribution
ef-020-34-00

Example 7. Dice. Toss a fair die and observe the number on top. This random number will be denoted here by X . It is easy to make 10 experiments for X . You may also make 100 experiments. But it would be boring to make 1000 experiments. This is why we will make - in the following file - a simulation of 1000 experiments. We will get 1000 integer numbers. The smallest possible value of X is 1, the largest is 6. We may count how many 1-s, 2-s, ... , 6-s we get. The numbers we get are the frequencies of the possible values. The frequencies divided by the total number of experiments are the relative frequencies.

In the following file, the frequencies are calculated by the `FREQUENCY (_ ; _)` command, which is a very useful but a little bit complicated command.

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

How to use the FREQUENCY command. The first argument of the FREQUENCY (_ ; _) command is the array of the data-set, the second argument is the array containing the list of the possible values. While entering the FREQUENCY (_ ; _) command, one must pay special attention to the following steps:

1. Type the FREQUENCY (_ ; _) command next to the first possible value with the correct arguments. You will get the frequency of the first possible value.
2. Mark - with the mouse - all the cells where the frequencies of the other possible values will be.
3. Press the F2-key.
4. Press the Ctrl-key, keep it pressed, and press the Shift-key, keep it pressed, too, and press the Enter-key. You will get the frequencies of all possible values. (You must not use the copy-paste command instead of the above sequence of commands. That would give false results.) The cells containing the frequencies will be stuck together, which means that later on they can be treated only together as a whole unit.

We see that each relative frequency is oscillating around $1/6$, so the probability of each possible value is $1/6$. This is shown in the next file.

Demonstration file: 1000 tosses with a fair die, relative frequencies and probabilities
ef-020-38-00

In the following file not only frequencies and probabilities, but conditional frequencies and probabilities are involved. Playing with the file, you may study what the notion of conditional frequency and probability mean. Because of the large size of the file, downloading it may take longer time.

Demonstration file: Conditional relative frequency and probability of events related to fair dice
ef-020-39-00

In the following two files, unfair dice are simulated. Because of the large size of the files, downloading them may take longer time.

Demonstration file: Unfair dice (larger values have larger probabilities)
ef-020-40-00

Demonstration file: Unfair dice (smaller values have larger probabilities)
ef-020-41-00

2 Outcomes and events

A **phenomenon** means that, under certain circumstances or conditions, something is happening, or we do something. When the conditions are fulfilled, we say that we perform a **valid experiment**. When the conditions are not fulfilled, we say that this is an invalid experiment. It will be important in our theory that for phenomenon (at least theoretically) the experiments can be repeated as many times as we want. When, related to the phenomenon, we decide or declare what we are interested in, what we observe, we define as an **observation**. The possible results of the observation are called the **outcomes** (or - in some text-books - **elementary events**). The set of all outcomes is the **sample space**. Here are some examples for phenomena and observations.

Example 1. Fair coin. Let the phenomenon mean tossing a fair coin on top of a table. Let an experiment be valid if one of the sides of the coin shows up (that is the coin does not stop on one of its edges). Here are some observations:

1. We observe where the center of the coin stops on a rectangular shaped table. Here the outcomes are the points of the top of the table. The sample space is the surface of the table, that is, a rectangle.
2. We observe how much time the coin rolls on the table before stopping. Here the outcomes are the positive real numbers. The sample space is the positive part of the real line.
3. We observe which side of the coin shows up when it stops. Now the outcomes are *heads* and *tails*. The sample space is the set $\{H, T\}$ consisting of two elements: H stands for *heads*, T stands for *tails*.

Example 2. Fair die. Let the phenomenon mean rolling a fair die on top of a table. Let an experiment be valid if the die remains on top of the table so that it stands clearly on one of its sides. Here are some observations:

1. We observe where the die stops. Here the outcomes are the points of the top of the table. The sample space is the surface of the table, that is, a rectangle.
2. We observe how much time the die rolls on the table before stopping. Here the outcomes are the positive real numbers. The sample space is the positive part of the real line.
3. We observe which side of the die shows up when it stops. Now the outcomes are 1, 2, 3, 4, 5, 6. The sample space is the set $\{1, 2, 3, 4, 5, 6\}$.
4. We observe whether we get 6 or we do not get 6. Here there are two outcomes: 6, not 6. The sample space is a set consisting of two elements: $\{6, \text{not } 6\}$.
5. We observe whether we get a number greater than 4 or not greater than 4. Here there are two outcomes again, namely: greater, not greater. The sample space is a set consisting of two elements: $\{\text{greater, not greater}\}$.

Example 3. Two fair dice. Let the phenomenon mean rolling two fair dice, a red and a blue, on top of a table. Let an experiment be valid if both dice remain on top of the table so that they stand clearly on one of their sides. Here are some observations:

1. We observe the pair of numbers we get. Let the first number in the pair be taken from the red die, the second from the blue. Here we have 36 outcomes, which can be arranged in a 6 by 6 table. The sample space may be represented as the set of the 36 cells of a 6 by 6 table.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

2. We observe the maximum of the two numbers we toss. Here the outcomes are again the numbers 1, 2, 3, 4, 5, 6. The sample space is the set $\{1, 2, 3, 4, 5, 6\}$.
3. We observe the sum of the two numbers we toss. Here there are 11 outcomes: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. The sample space is the set $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

Example 4. Toss a coin until the first head. Let the phenomenon mean tossing a fair die until the first time the head occurs. Here are some observations:

1. We observe the sequence of heads and tails we get. Now the outcomes are the possible sequences of heads and tails. The sample space is the set of all possible sequences of heads and tails.
2. We observe the number of tosses until the first time the head occurs. Now the outcomes are the positive integers: 1, 2, 3, ... and the symbol ∞ . The symbol ∞ means: we never ever get a head. The sample space is the set consisting of all positive integers and the symbol ∞ : $\{1, 2, 3, \dots, \infty\}$
3. We observe how many tails we get before the first head occurs. Now the outcomes are the non-negative integers: 0, 1, 2, ... and the symbol ∞ . The symbol ∞ means: we never get a head, that is why we get an infinite number of tails. The sample space is the set consisting of all non-negative integers and the symbol ∞ : $\{0, 1, 2, \dots, \infty\}$.

An **event** is a statement related to the phenomenon or to an observation so that whenever an experiment is performed we can decide whether the statement is *true* or *false*. When it is true we say that the event *occurs*, when it is not true, we say that the event *does not occur*. Instead of true and false, the words *yes* and *no* are also often used. We often write the number 1 for the occurrence, and the number 0 for the non-occurrence

of an event. An event, that is, a statement related to an observation obviously corresponds to a *subset* of the sample space taken for that observation. The subset consists of those outcomes for which the event occurs. For example, tossing a die and observing the number on the top, the event "greater than 4" corresponds to the subset $\{5, 6\}$.

It may happen that two different statements always occur at the same time. In this case we say that the two statements define the *same event*.

Now we list some **operations and relations on events**. We put the corresponding set-theoretical operations and relations into parentheses.

1. The **sure** or **certain** event always occurs. (Whole sample space.)
2. The **impossible** event never occurs. (Empty set.)
3. The **complement** of an event occurs if and only if the event does not occur. (Complementary set.)
4. The **intersection** or **product** of events is the logical *and*-operation, meaning that "each event occurs". (Intersection of sets.)
5. The **union** or **sum** of events is the logical *or*-operation, meaning that "at least one of the events occurs". (Union of sets.)
6. The **difference** of an event and another event means that the first event occurs, but the other event does not occur. (Difference of sets.)
7. Some events are said to be **exclusive** events, and we say that they **exclude** each other if the occurrence of one of them guarantees that the others do not occur. (Disjoint sets.)
8. An event is said to **imply** another event if the occurrence of the first event guarantees the occurrence of the other event. (A set is a subset of the other.)

Drawing a **Venn-diagram** is a possibility to visualize events, operations on events, etc. by sets drawn in the plain.

3 Relative frequency and probability

When we make experiments again and again for a phenomenon or an observation, then we get **sequence of experiments**. Assume now that we make a sequence of experiments for an event. We may take notes at each experiment whether the event occurs or does not occur, and we may count how many times the event occurs. This occurrence number is called the **frequency** of the event. The frequency divided by the number of experiments is the **relative frequency**. Since the occurrence of an event depends on randomness, both the frequency and the relative frequency depend on randomness.

Now the reader may study the following files again, which appeared among the introductory problems in Section 1.

Demonstration file: Waiting time for the bus
ef-020-18-00

Demonstration file: Traveling by bus and metro: uniformly distributed waiting times
ef-020-19-00

Demonstration file: Waiting time for bus < 4
ef-020-20-00

Demonstration file: Waiting time for metro > 3
ef-020-21-00

Demonstration file: Waiting time for bus < 4 AND waiting time for metro > 3
ef-020-22-00

Demonstration file: Waiting time for bus < waiting time for metro
ef-020-23-00

Demonstration file: Total waiting time > 4
ef-020-24-00

Demonstration file: Waiting time for bus is less than waiting time for metro AND total waiting time > 4
ef-020-25-00

Demonstration file: Waiting time for bus < waiting time for metro OR total waiting time > 4
ef-020-26-00

It is an important law, called the law of large numbers, that the relative frequencies of an event in a long sequence of experiments stabilize around a number, which does not depend on randomness, but it is a characteristic of the event itself. This number is called the **probability** of the event. The notion of probability can be interpreted like these:

1. Consider an interval around the probability value. If we make a large number of experiments of a (given) large length, then the great majority of relative frequencies (associated to this large length) will be in this interval.
2. If we could make an infinitely long sequence of experiments, then the sequence of relative frequencies would converge to the probability in the mathematical sense of convergence.

Probability theory deals, among others, with figuring out the probability values without performing any experiments, but using theoretical arguments.

In the following files you may learn how the relative frequencies stabilize around the probability. The first and the second are simpler, the third is a little-bit trickier.

Demonstration file: Event and relative frequency
ef-030-01-00

Demonstration file: Tossing a die - probability
ef-030-02-00

Demonstration file: Relative frequency with balls
ef-030-03-00

Playing with the next file, you may check your ability to guess a probability based on your impression when many experiments are performed. When you open it, choose the option "Don't Update".

Demonstration file: Probability guessed by impression
ef-030-04-00

If you want to change the hidden probability value in the previous file, then save the previous file (File A) and the following file (File B) into a folder, and close both. Then open the second file (File B), press F9 to regenerate a new hidden probability value, and open the first file (File A), and choose the option "Update", and close the second file (File B).

Demonstration file: Auxiliary file to generate a new hidden probability value
ef-030-05-00

4 Random numbers

Most calculators have a special key stroke and most computer programs have a simple command to generate random numbers. Calculators and computer programs are made so that the generated random number, let us denote it by RND, can be considered uniformly distributed between 0 and 1, which means that for any $0 \leq a \leq b \leq 1$, it is true that

$$\mathbf{P}(a < \text{RND} < b) = \text{length of } (a; b) = b - a$$

or, the same way,

$$\mathbf{P}(a \leq \text{RND} \leq b) = \text{length of } [a; b] = b - a$$

The following file illustrates this fact:

Demonstration file: Probability of an interval for a random number generated by computer
ef-040-01-00

Specifically, for any $0 \leq x \leq 1$ it is true that

$$\mathbf{P}(\text{RND} < x) = x$$

or, the same way,

$$\mathbf{P}(\text{RND} \leq x) = x$$

The following file illustrates this fact:

Demonstration file: Probability of $\text{RND} \leq x$
ef-020-10-00

The probability that a random number is exactly equal to a given number is equal to 0:

$$\mathbf{P}(\text{RND} = a) = \mathbf{P}(a \leq \text{RND} \leq a) = \text{length of } [a; a] = a - a = 0 \quad (\text{for all } a)$$

If two random numbers are generated, say RND_1 and RND_2 , then the random point $(\text{RND}_1, \text{RND}_2)$ is uniformly distributed in the unit square S which has the vertices $(0, 0)$, $(1, 0)$, $(1, 1)$, $(0, 1)$. This means that for any $A \subset S$, it is true that

$$P((\text{RND}_1, \text{RND}_2) \in A) = \text{area of } A$$

In order to illustrate this fact, in the following file, A can be a triangle inside the unit square.

Demonstration file: Probability of a triangle
ef-020-11-00

In the following file, the relative frequencies of more complicated events are studied.

Demonstration file: Special triangle combined with a diamond-shaped region
ef-020-12-00

If three random numbers are generated, say RND_1 , RND_2 and RND_3 , then the random point $(\text{RND}_1, \text{RND}_2, \text{RND}_3)$ is uniformly distributed in the unit cube S which has the vertices $(0, 0, 0)$, $(1, 0, 0)$, $(1, 1, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(1, 0, 1)$, $(1, 1, 1)$, $(0, 1, 1)$. This means that for any $A \subset S$, it is true that

$$P((\text{RND}_1, \text{RND}_2, \text{RND}_3) \in A) = \text{volume of } A \quad (A \subset S)$$

The following file deals with powers of random numbers.

Demonstration file: Powers of random numbers
ef-040-02-00

5 Classical problems

The simplest way of calculating a probability is when an observation has a finite number of outcomes so that, for some symmetry reasons, each outcome has the same probability. In this case the probability of an event is calculated by the **classical formula**:

$$\text{probability} = \frac{\text{number of favorable outcomes}}{\text{number of all outcomes}}$$

or, briefly:

$$\text{probability} = \frac{\text{favorable}}{\text{all}}$$

In the following files, we simply list all the outcomes, mark those which are favorable for the event in question, and then we use the classical formula to calculate the probability of the event.

*Demonstration file: 2 dice, P(Sum = 5)
ef-050-01-00*

*Demonstration file: 2 dice, P(Sum = k)
ef-050-02-00*

*Demonstration file: 5 coins, P(Number of heads = k)
ef-050-03-00*

*Demonstration file: 4 dice, on each dice: 1,2: red, 3,4,5,6: green, P(Number of red = k)
ef-050-04-00*

When the number of all outcomes is so large that we are unable to list them, or the problem contains not only numerical values but parameters as well, then combinatorics plays an important role in finding out the number of all outcomes and the number of favorable outcomes. The branch of mathematics dealing with calculating the number of certain cases is called **combinatorics**. It is assumed that the reader is familiar with the basic notions and techniques of elementary combinatorics. Here is only a list of some techniques and formulas we often use in combinatorics:

1. Listing - counting
2. Uniting - adding
3. Leaving off - subtracting
4. Tree-diagram, window technique - multiplication
5. Factorization (considering classes of equal size) - division

6. Permutations without repetition

$$n!$$

7. Permutations with repetition

$$\frac{n!}{k_1!k_2!\dots k_r!}$$

8. Variations without repetition

$$\frac{n!}{(n-k)!}$$

9. Variations with repetition

$$n^k$$

10. Combinations without repetition

$$\binom{n}{k}$$

Remember that the definition of the binomial coefficient $\binom{n}{k}$ is:

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

When we have to calculate the value of the binomial coefficient $\binom{n}{k}$ without a calculator, it is may be advantageous to use the following form of it:

$$\binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{1\ 2\ 3\dots k}$$

Notice that in the right side formula, both the numerator and the denominator are a product of k factors. In the numerator, the first factor is n , and the factors are decreasing. In the denominator the first factor is 1, and the factors are increasing. Simplification always reduces the fraction into an integer.

11. Combinations with repetition

$$\binom{n+k-1}{k}$$

12. Pascal triangle: if we arrange the binomial coefficients into a triangle-shaped table like this:

finite but not equal to 0, and the probability of any event, corresponding to some subset A of S , is equal to

$$\mathbf{P}(A) = \frac{\text{length of } A}{\text{length of } S}$$

in the one-dimensional case, or

$$\mathbf{P}(A) = \frac{\text{area of } A}{\text{area of } S}$$

in the two-dimensional case, or

$$\mathbf{P}(A) = \frac{\text{volume of } A}{\text{volume of } S}$$

in the three-dimensional case, or

$$\mathbf{P}(A) = \frac{n\text{-dimensional volume of } A}{n\text{-dimensional volume of } S}$$

in the n -dimensional case. Since the calculation of lengths, areas, volumes, first in the life of most students, is taught in geometry, such problems are called **geometrical problems**.

We also say that a random point is chosen in S according to **uniform distribution** if

$$P(\text{the point is in } A) = \frac{\text{length of } A}{\text{length of } S} \quad (A \subseteq S)$$

in the one-dimensional case, or

$$P(\text{the point is in } A) = \frac{\text{area of } A}{\text{area of } S} \quad (A \subseteq S)$$

in the two-dimensional case, or

$$P(\text{the point is in } A) = \frac{\text{volume of } A}{\text{volume of } S} \quad (A \subseteq S)$$

in the three-dimensional case, or

$$P(\text{the point is in } A) = \frac{n\text{-dimensional volume of } A}{n\text{-dimensional volume of } S} \quad (A \subseteq S)$$

in the n -dimensional case.

Now the reader may study the following files again, which appeared among the introductory problems in Section 1.

*Demonstration file: [Waiting time for the bus](#)
ef-020-18-00*

Demonstration file: Traveling by bus and metro: uniformly distributed waiting times
ef-020-19-00

Demonstration file: Waiting time for bus < 4
ef-020-20-00

Demonstration file: Waiting time for metro > 3
ef-020-21-00

Demonstration file: Waiting time for bus < 4 AND waiting time for metro > 3
ef-020-22-00

Demonstration file: Waiting time for bus < waiting time for metro
ef-020-23-00

Demonstration file: Total waiting time > 4
ef-020-24-00

Demonstration file: Waiting time for bus < waiting time for metro AND AND total waiting time > 4
ef-020-25-00

Demonstration file: Waiting time for bus is less than waiting time for metro OR total waiting time is more than 4
ef-020-26-00

The following example may surprise the reader, because the number π appears in the solution.

Example 1. Buffon's needle problem. Let us draw several long parallel lines onto a big paper so that the distance between adjacent lines is always D . Let us take a needle whose length is L . For simplicity, we assume that $L \leq D$. Let us drop the needle onto the paper "carelessly, in a random way" so that any direction or position is preferred for the needle the same way. When the needle stops jumping, it will either intersect a line (touching without intersection is included) or it will not touch lines at all. We may ask: what is the probability that the needle will intersect a line?

The following two files interpret Buffon's needle problem.

Demonstration file: Buffon's needle problem
ef-070-01-00

Demonstration file: Buffon's needle problem, more experiments
ef-070-02-00

Solution. The line of the needle and the given parallel lines define an acute angle, this is what we denote by X . The center of the needle and the closest line to it define

a distance, this is what we denote by Y . Obviously, $0 \leq X \leq \pi/2$ and $0 \leq Y \leq D/2$. The point (X, Y) is obviously a random point inside the rectangle defined by the intervals $(0; \pi/2)$ and $(0; D/2)$. Since X and Y follow uniform distribution and they are independent of each other, the random point (X, Y) follows uniform distribution on the rectangle. The needle intersects a line if and only if $Y \leq L/2 \sin(X)$, that is, the points in the rectangle corresponding to intersections constitute the range below the graph of the curve with equation $y = L/2 \sin(x)$. Thus, we get that

$$\mathbf{P}(\text{Intersection}) = \frac{\text{Area under the curve}}{\text{Area of the rectangle}} = \frac{\int_0^{\pi/2} \frac{L}{2} \sin(x) dx}{\frac{D}{2} \cdot \left(\frac{\pi}{2}\right)} = \frac{2L}{\pi D}$$

Remark. If $2L = D$, that is the distance between the parallel lines is twice the length of the needle, then we get the nice and surprising result:

$$\mathbf{P}(\text{Intersection}) = \frac{1}{\pi}$$

The following sequence of problems may seem a contradiction, because the (seemingly) same questions have different answers in the different solutions.

Example 2. Bertrand's paradox. Let us consider a circle. For the sake of Bertrand's paradox, a chord of the circle is called long, if it is longer than the length of a side of a regular triangle drawn into the circle. Let us Choose a chord "at random". We may ask: what is the probability that the chord is long? The following files interpret Bertrand's paradox.

Demonstration file: Bertrand paradox, introduction
ef-070-03-00

Demonstration file: Two points on the perimeter
ef-070-04-00

Demonstration file: One point inside
ef-070-05-00

Demonstration file: One point on a radius
ef-070-06-00

Demonstration file: Two points inside
ef-070-07-00

Demonstration file: One point on the perimeter, other point inside
ef-070-08-00

Demonstration file: Point and direction
ef-070-09-00

Demonstration file: Bertrand paradox, comparison
ef-070-10-00

7 Basic properties of probability

The following properties are formulated for probabilities. If we accept some of them as axioms, then the others can be proved. We shall not do so. Instead of such an approach, we emphasize that each of these formulas can be translated into a formula for relative frequencies by replacing the expression "probability of" by the expression "relative frequency of", or replacing the letter "**P**", which is an abbreviation of the expression "probability of", by the expression "relative frequency of". If you make this replacement, you will get properties for relative frequencies which are obviously true.

For example, the first three properties for relative frequencies sound like this:

1. Relative frequency of the **sure event** is 1.
2. Relative frequency of the **impossible event** is 0.
3. **Complement rule for relative frequencies:**

$$\text{relative frequency of } A + \text{relative frequency of } \bar{A} = 1$$

This is why it is easy to accept that the following properties for probabilities hold.

1. The probability of the **sure event** is 1.
2. The probability of the **impossible event** is 0.
3. **Complement rule:**

$$\mathbf{P}(A) + \mathbf{P}(\bar{A}) = 1$$

4. **Addition law of probability for exclusive events:**

If A, B are exclusive events, then

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$$

If A, B, C are exclusive events, then

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C)$$

If A_1, A_2, \dots, A_n are exclusive events, then

$$\mathbf{P}(A_1 \cup A_2 \cup \dots \cup A_n) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \dots + \mathbf{P}(A_n)$$

If A_1, A_2, \dots are exclusive events, then

$$\mathbf{P}(A_1 \cup A_2 \cup \dots) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \dots$$

5. Addition law of probability for arbitrary events:

If A, B are arbitrary events, then

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$

If A, B, C are arbitrary events, then

$$\begin{aligned} \mathbf{P}(A \cup B \cup C) = & +\mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) \\ & -\mathbf{P}(A \cap B) - \mathbf{P}(A \cap C) - \mathbf{P}(B \cap C) \\ & +\mathbf{P}(A \cap B \cap C) \end{aligned}$$

Remark. Notice that, on the right side

- in the 1st line, there are $\binom{3}{1} = 3$ terms, the probabilities of the individual events with "+" signs,

- in the 2nd line there are $\binom{3}{2} = 3$ terms, the probabilities of the intersections of two events with "-" signs,

- in the 3rd line there is $\binom{3}{3} = 1$ term, the probability of the intersection of all events with a "+" sign.

Poincaré formula: If A_1, A_2, \dots, A_n are arbitrary events, then

$$\begin{aligned} \mathbf{P}(A_1 \cup A_2 \cup \dots \cup A_n) = & \\ & +\mathbf{P}(A_1) + \mathbf{P}(A_2) + \dots + \mathbf{P}(A_n) \\ & -\mathbf{P}(A_1 \cap A_2) - \mathbf{P}(A_1 \cap A_3) - \dots - \mathbf{P}(A_{n-1} \cap A_n) \\ & +\mathbf{P}(A_1 \cap A_2 \cap A_3) + \mathbf{P}(A_1 \cap A_2 \cap A_4) + \dots + \mathbf{P}(A_{n-2} \cap A_{n-1} \cap A_n) \\ & \vdots \\ & +(-1)^{n+1}\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned}$$

Remark. Notice that, on the right side

- in the 1st line, there are $\binom{n}{1} = n$ terms, the probabilities of the individual

events with "+" signs,

- in the 2nd line there are $\binom{n}{2}$ terms, the probabilities of the intersections of two events with "-" signs,

- in the 3rd line there are $\binom{n}{3}$ terms, the probabilities of the intersections of two events, with "+" signs,

- in the n th line there is $\binom{n}{n} = 1$ term, the probability of the intersection of all events with a "+" or "-" sign depending on whether n is odd or even.

6. **Special subtraction rule:** If event B implies event A , that is, $B \subseteq A$, then

$$\mathbf{P}(A \setminus B) = P(A) - \mathbf{P}(B)$$

7. **General subtraction rule:** If A and B are arbitrary events, then

$$\mathbf{P}(A \setminus B) = P(A) - \mathbf{P}(A \cap B)$$

8 Conditional relative frequency and conditional probability

Let A and B denote events related to a phenomenon. Imagine that we make N experiments for the phenomenon. Let N_A denote the number of times that A occurs, and let $N_{A \cap B}$ denote the number of times that B occurs together with A . The **conditional relative frequency** is introduced by the fraction:

$$\frac{N_{A \cap B}}{N_A}$$

This fraction shows how often B occurs among the occurrences of A . Dividing both the numerator and the denominator by N , we get that, for large N , if $\mathbf{P}(A) \neq 0$, then

$$\frac{N_{A \cap B}}{N_A} = \frac{\frac{N_{A \cap B}}{N}}{\frac{N_A}{N}} \approx \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}$$

that is, for a large number of experiments, the conditional relative frequency stabilizes around

$$\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}$$

This value will be called the **conditional probability** of B on condition that A occurs, and will be denoted by $\mathbf{P}(B|A)$:

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

This formula is also named as the **division rule for probabilities**.

Remark. If event B implies event A , that is, $B \subseteq A$, then $A \cap B = B$, and thus the division rule for probabilities simplifies to

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

In the following files, not only frequencies and probabilities, but conditional frequencies and probabilities are involved.

Demonstration file: Special triangle combined with a diamond-shaped region
ef-020-13-00

Demonstration file: Circle and/or hyperbolas
ef-090-01-00

Multiplication rules. Rearranging the division rule, we get the **multiplication rule for two events**:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B|A)$$

which can be easily extended to the **multiplication rule for arbitrary events**:

$$\begin{aligned} \mathbf{P}(A_1 \cap A_2) &= \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \\ \mathbf{P}(A_1 \cap A_2 \cap A_3) &= \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \mathbf{P}(A_3|A_1 \cap A_2) \\ \mathbf{P}(A_1 \cap A_2 \cap A_3 \cap A_4) &= \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \mathbf{P}(A_3|A_1 \cap A_2) \mathbf{P}(A_4|A_1 \cap A_2 \cap A_3) \\ &\vdots \end{aligned}$$

As a special case, we get the **multiplication rule for a decreasing sequence of events**:

If

$$\begin{aligned} A_2 \text{ is implies } A_1, \text{ that is, } A_2 \subseteq A_1, \text{ or equivalently, } A_1 \cap A_2 = A_2, \\ A_3 \text{ is implies } A_2, \text{ that is, } A_3 \subseteq A_2, \text{ or equivalently, } A_2 \cap A_3 = A_3, \\ A_4 \text{ is implies } A_3, \text{ that is, } A_4 \subseteq A_3, \text{ or equivalently, } A_3 \cap A_4 = A_4, \\ \vdots \end{aligned}$$

then

$$\begin{aligned} \mathbf{P}(A_2) &= \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \\ \mathbf{P}(A_3) &= \mathbf{P}(A_2) \mathbf{P}(A_3|A_2) \\ \mathbf{P}(A_4) &= \mathbf{P}(A_3) \mathbf{P}(A_4|A_3) \\ &\vdots \end{aligned}$$

and, consequently

$$\begin{aligned}\mathbf{P}(A_2) &= \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \\ \mathbf{P}(A_3) &= \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \mathbf{P}(A_3|A_2) \\ \mathbf{P}(A_4) &= \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \mathbf{P}(A_3|A_2) \mathbf{P}(A_4|A_3) \\ &\vdots\end{aligned}$$

Example 1. Birthday paradox. Imagine that in a group of n people, everybody, one after the other, tells which day of the year he or she was born. (For simplicity, leap years are neglected, that is, there are only 365 days in a year.) It may happen that all the n people say different days, but it may happen that there will be one or more coincidences. The reader, in the future, at parties, may make experiments. Obviously, if n is small, then the probability that at least one coincidence occurs, is small. If n is larger, then this probability is larger. If $n \geq 366$, then the coincidence is sure. The following file simulates the problem:

*Demonstration file: Birthday paradox - simulation
ef-090-02-10*

We ask two questions:

1. For a given n ($n = 2, 3, 4, \dots, 366$), how much is the probability that at least one coincidence occurs?
2. Which is the smallest n for which $\mathbf{P}(\text{at least one coincidence occurs}) \geq 0.5$?

Remark. People often argue like this: the half of 365 is $365/2 = 182.5$, so the answer to the second question is 183. We shall see that this answer is very far from the truth. The correct answer is surprisingly small: 23. This means that when 23 people gather together, then the probability that at least one birthday coincidence occurs is more than half, and the probability that no birthday coincidence occurs is less than half. If you do not believe, then make experiments: if you make many experiments with groups consisting of at least 23 people, then the case that at least one birthday coincidence occurs will be more frequent than the case that no birthday coincidence occurs.

Solution. Let us define the event A_k like this:

$$A_k = \text{the first } k \text{ people have different birthdays} \quad (k = 1, 2, 3, \dots)$$

The complement of A_k is:

$$\overline{A_k} = \text{at least one coincidence occurs}$$

It is obvious that $\mathbf{P}(A_1) = 1$. The sequence of the events A_1, A_2, A_3, \dots clearly constitutes a decreasing sequence of events. In order to determine the conditional probability $\mathbf{P}(A_k|A_{k-1})$, let us assume that A_{k-1} occurs, that is, the first $k-1$ people have different birthdays. It is obvious that A_k occurs if and only if the k th person has a birthday

different from the previous $k - 1$ birthdays, that is, he or she was born on one of the remaining $365 - (k - 1)$ days. This is why

$$\mathbf{P}(A_k|A_{k-1}) = (365 - (k - 1))/365 \quad (k \geq 1)$$

that is

$$\mathbf{P}(A_2|A_1) = 364/365 = 0,9973$$

$$\mathbf{P}(A_3|A_2) = 363/365 = 0,9945$$

$$\mathbf{P}(A_4|A_3) = 362/365 = 0,9918$$

⋮

Now, using the multiplication rule for our decreasing sequence of events, we get:

$$\mathbf{P}(A_1) = 1$$

$$\mathbf{P}(A_2) = \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) = 1 \quad 0,9973 = 0,9973$$

$$\mathbf{P}(A_3) = \mathbf{P}(A_2) \mathbf{P}(A_3|A_2) = 0,9973 \quad 0,9945 = 0,9918$$

$$\mathbf{P}(A_4) = \mathbf{P}(A_3) \mathbf{P}(A_4|A_3) = 0,9918 \quad 0,9918 = 0,9836$$

⋮

Since the events A_n mean no coincidences, in order to get the probabilities of the birthday coincidences we need to find the probabilities of their complements :

$$\mathbf{P}(\overline{A_1}) = 1 - \mathbf{P}(A_1) = 1 - 1 = 0$$

$$\mathbf{P}(\overline{A_2}) = 1 - \mathbf{P}(A_2) = 1 - 0,9973 = 0,0027$$

$$\mathbf{P}(\overline{A_3}) = 1 - \mathbf{P}(A_3) = 1 - 0,9918 = 0,0082$$

$$\mathbf{P}(\overline{A_4}) = 1 - \mathbf{P}(A_4) = 1 - 0,9836 = 0,0164$$

⋮

The following file shows how such a table can be easily constructed and extended up to $n = 366$ in Excel:

[*Demonstration file: Birthday paradox - calculation ef-090-02-00*](#)

In this Excel table, we find the answer to our first question: the probability that at least one coincidence occurs is calculated for all $n = 1, 2, \dots, 366$. In order to get the answer to the second question, we must find where the first time the probability of the coincidence is larger than half in the table. We see that

$$\mathbf{P}(\overline{A_{22}}) = 0,4757$$

$$\mathbf{P}(\overline{A_{23}}) = 0,5073$$

which means that 23 is the smallest n for which the probability that at least one coincidence occurs is greater than half.

We say that the events A_1, A_2, \dots constitute a **total system** if they are exclusive, and their union is the sure event.

Total probability formula. If the events A_1, A_2, \dots have a probability different from zero, and they constitute a total system, then

$$\mathbf{P}(B) = \sum_i \mathbf{P}(A_i)\mathbf{P}(B|A_i)$$

The following example illustrates how the total probability formula may be used.

Example 2. Is it defective? There are three workshops in a factory: A_1, A_2, A_3 . Assume that

- workshop A_1 makes 30 percent,
- workshop A_2 makes 40 percent,
- workshop A_3 makes 30 percent of all production.

We assume that

- the probability that an item made in workshop A_1 is defective is 0,05,
- the probability that an item made in workshop A_2 is defective is 0.04,
- the probability that an item made in workshop A_3 is defective is 0.07.

Now taking an item made in the factory, what is the probability that it is defective?

Solution. The following file - using the total probability formula - gives the answer:

Demonstration file: Application of the total probability formula ef-090-02-50

The Bayes formula expresses a conditional probability in terms of other conditional and unconditional probabilities.

Bayes formula. If the events A_1, A_2, \dots have a probability different from zero, and they constitute a total system, then

$$\mathbf{P}(A_k|B) = \frac{\mathbf{P}(A_k)\mathbf{P}(B|A_k)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_k)\mathbf{P}(B|A_k)}{\sum_i \mathbf{P}(A_i)\mathbf{P}(B|A_i)}$$

Example 3. Which workshop made the defective item? Assuming that an item made in the factory in the previous problem is defective, we may ask: Which workshop made it? Obviously, any of them may make defective items. So, the good question consists of 3 questions, which may sound like this:

- What is the probability that the defective item was made in workshop A_1 ?
- What is the probability that the defective item was made in workshop A_2 ?
- What is the probability that the defective item was made in workshop A_3 ?

Solution. The following file - using the Bayes formula - gives the answer to these questions:

Demonstration file: Application of the Bayes formula
ef-090-03-00

Example 4. Is he sick or healthy? Assume that 0.001 part of people are infected by a certain bad illness, 0.999 part of people are healthy. Assume also that if a person is infected by the illness, then he or she will be correctly diagnosed sick with a probability 0.9, and he or she will be mistakenly diagnosed healthy with a probability 0.1. Moreover, if a person is healthy, then he or she will be correctly diagnosed healthy with a probability 0.8, and he or she will be mistakenly diagnosed sick with a probability 0.2. Now imagine that a person is examined, and the test says the person is sick. Knowing this fact what is the probability that this person is really sick?

Solution. The answer is surprising. Using the Bayes formula, it is given in the following file.

Demonstration file: Sick or healthy?
ef-090-04-00

9 Independence of events

Independence of two events. The event B and its complement \bar{B} are called to be **independent of** the event A and its complement \bar{A} if

$$\mathbf{P}(B|A) = \mathbf{P}(B|\bar{A}) = \mathbf{P}(B)$$

$$\mathbf{P}(\bar{B}|A) = \mathbf{P}(\bar{B}|\bar{A}) = \mathbf{P}(\bar{B})$$

It is easy to see that in order for these four equalities to hold it is enough that one of them holds, because the other three equalities are consequences of the chosen one. This is why many textbooks introduce the notion of independence so that the event B is called to be **independent of** the event A if

$$\mathbf{P}(B|A) = \mathbf{P}(B)$$

On the left side of this equality, replacing $\mathbf{P}(B|A)$ by $\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}$, we get that independence means that

$$\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \mathbf{P}(B)$$

or, equivalently,

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$$

Now dividing by $\mathbf{P}(B)$, we get that

$$\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \mathbf{P}(A)$$

that is

$$\mathbf{P}(A|B) = \mathbf{P}(A)$$

which means that event A is independent of the event B . Thus, we see that independence is a symmetrical relation, and we can simply say, that events A and B are independent of each other, or more generally the pair A, \bar{A} and the pair B, \bar{B} are **independent of each other**.

Independence of three events. The notion of independence of three events is introduced in the following way. The sequence of events A, B, C is called independent if

$$\mathbf{P}(B|A) = \mathbf{P}(B|\bar{A}) = \mathbf{P}(B)$$

$$\mathbf{P}(\bar{B}|A) = \mathbf{P}(\bar{B}|\bar{A}) = \mathbf{P}(\bar{B})$$

$$\mathbf{P}(C|A \cap B) = \mathbf{P}(C|A \cap \bar{B}) = \mathbf{P}(C|\bar{A} \cap B) = \mathbf{P}(C|\bar{A} \cap \bar{B}) = \mathbf{P}(C)$$

$$\mathbf{P}(\bar{C}|A \cap B) = \mathbf{P}(\bar{C}|A \cap \bar{B}) = \mathbf{P}(\bar{C}|\bar{A} \cap B) = \mathbf{P}(\bar{C}|\bar{A} \cap \bar{B}) = \mathbf{P}(\bar{C})$$

It can be shown (we omit the proof) that these equalities hold if and only if the following $2^3 = 8$ **multiplication rules** hold:

$$\begin{aligned} \mathbf{P}(A \cap B \cap C) &= \mathbf{P}(A) \mathbf{P}(B) \mathbf{P}(C) \\ \mathbf{P}(A \cap B \cap \bar{C}) &= \mathbf{P}(A) \mathbf{P}(B) \mathbf{P}(\bar{C}) \\ \mathbf{P}(A \cap \bar{B} \cap C) &= \mathbf{P}(A) \mathbf{P}(\bar{B}) \mathbf{P}(C) \\ \mathbf{P}(A \cap \bar{B} \cap \bar{C}) &= \mathbf{P}(A) \mathbf{P}(\bar{B}) \mathbf{P}(\bar{C}) \\ \mathbf{P}(\bar{A} \cap B \cap C) &= \mathbf{P}(\bar{A}) \mathbf{P}(B) \mathbf{P}(C) \\ \mathbf{P}(\bar{A} \cap B \cap \bar{C}) &= \mathbf{P}(\bar{A}) \mathbf{P}(B) \mathbf{P}(\bar{C}) \\ \mathbf{P}(\bar{A} \cap \bar{B} \cap C) &= \mathbf{P}(\bar{A}) \mathbf{P}(\bar{B}) \mathbf{P}(C) \\ \mathbf{P}(\bar{A} \cap \bar{B} \cap \bar{C}) &= \mathbf{P}(\bar{A}) \mathbf{P}(\bar{B}) \mathbf{P}(\bar{C}) \end{aligned}$$

The multiplication rules are symmetrical with respect to any permutation of the events A, B, C , which means that in the terminology we do not have to take into account the order of the events A, B, C , and we can just say that the events A, B, C are independent of each other.

Pairwise and total independence. It is important to keep in mind that it may happen that any two of the three events A, B, C are independent of each other, that is,

1. A and B are independent of each other,
2. A and C are independent of each other,
3. B and C are independent of each other,
4. but the three events A, B, C are not independent of each other.

If this is the case, then we say that the events A, B, C are **pairwise independent**, but they are not **(totally) independent**. So, pairwise independence does not imply (total) independence.

Independence of more events. The independence of n events can be introduced similarly to the independence of three events. It can be shown that the independence of n events can also be characterized by 2^n **multiplication rules**:

$$\begin{aligned} \mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) &= \mathbf{P}(A_1) \mathbf{P}(A_2) \dots \mathbf{P}(A_n) \\ \mathbf{P}(A_1 \cap A_2 \cap \dots \cap \bar{A}_n) &= \mathbf{P}(A_1) \mathbf{P}(A_2) \dots \mathbf{P}(\bar{A}_n) \\ &\vdots \\ \mathbf{P}(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) &= \mathbf{P}(\bar{A}_1) \mathbf{P}(\bar{A}_2) \dots \mathbf{P}(\bar{A}_n) \end{aligned}$$

The following files illustrate and use the multiplication rules for independent events.

Demonstration file: Multiplication rules for independent events
ef-100-01-00

Demonstration file: How many events occur?
ef-100-02-00

Playing with the following file, you may check your ability to decide - on the basis of performed experiments - whether two events are dependent or independent.

Demonstration file: Colors dependent or independent
ef-100-03-00

10 *** Infinite sequences of events

The following rule is a generalization of the addition law of the probability for a finite number of exclusive events, which was described among the basic properties of probability.

Addition law of probability for an infinite number of exclusive events: If A_1, A_2, \dots, A_n are exclusive events, then

$$\mathbf{P}(A_1 \cup A_2 \cup \dots) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \dots$$

Example 1. Odd or even? My friend I play with a fair coin. We toss it until the first time a head occurs. We agree that I win if the number of tosses is an odd number, that is, 1 or 3 or 5 . . . , and my friend wins if the number of tosses is an even number, that is, 2 or 4 or 6 What is the probability that I win? What is the probability that my friend wins?

Remark. You may think that odd and even numbers "balance" each other, so the asked probabilities are equal. However, if you play with the following simulation file for a couple of times, or you read the theoretical solution, then you will experience that this is not true:

Demonstration file: Odd or even?
ef-100-03-50

Solution.

$$\mathbf{P}(\text{I win}) =$$

$$\mathbf{P}(\text{First head occurs at the 1st toss or 3rd toss or 5th toss or } \dots) =$$

$$\mathbf{P}(\text{1st}) + \mathbf{P}(\text{3rd}) + \mathbf{P}(\text{5th}) + \dots =$$

$$0.5 + 0.5^3 + 0.5^5 + \dots = \frac{0.5}{1 - 0.5^2} = \frac{0.5}{1 - 0.25} = \frac{0.5}{0.75} = \frac{2}{3}$$

$$\mathbf{P}(\text{My friend wins}) =$$

$$\mathbf{P}(\text{First head occurs at the 2nd toss or 4th toss or 6th toss or } \dots) =$$

$$\mathbf{P}(\text{2nd}) + \mathbf{P}(\text{4th}) + \mathbf{P}(\text{6th}) + \dots =$$

$$0.5^2 + 0.5^4 + 0.5^6 + \dots = \frac{0.5^2}{1 - 0.5^2} = \frac{0.25}{1 - 0.25} = \frac{0.25}{0.75} = \frac{1}{3}$$

The following two properties are closely related to the addition law of probability for an infinite number of exclusive events.

Limit for an increasing sequence of events. Let A_1, A_2, \dots be an **increasing sequence of events**, that is, A_k implies A_{k+1} for all k . Let A be the union of the events A_1, A_2, \dots . A clearly means that one of the infinitely many events A_1, A_2, \dots occurs. The following limit relation holds.

$$\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$$

Limit for a decreasing sequence of events. Let A_1, A_2, \dots be a **decreasing sequence of events**, that is, A_k is implied by A_{k+1} for all k . Let A be the intersection of the events A_1, A_2, \dots . A clearly means the event that all the infinitely many events A_1, A_2, \dots occur. The following limit relation holds:

$$\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$$

The following example gives us an important message: if we consider an event which has a positive probability, and we make an unlimited number of independent experiments, then regardless of how small that probability of the event is, the event, sooner or later, will occur for sure.

Example 2. Unlimited number of exams. Let us assume that a student passes each of his exams, independently of the previous exams, with a positive probability p , and fails with a probability $q = 1 - p$. We will show that if p is positive and our student has an unlimited number of possibilities to take the exam in a course, then it is sure that the student, sooner or later, passes the course.

Solution. Let the event A_n mean that our student fails the first n exams. Because of the independence of the exams, the probability of A_n is:

$$\mathbf{P}(A_n) = q^n$$

Let the event A mean that the student fails all the infinite number of exams. Obviously A implies A_n for all n , so

$$\mathbf{P}(A) \leq \mathbf{P}(A_n) \quad \text{for all } n$$

Since $\mathbf{P}(A_n) \rightarrow 0$, when $n \rightarrow \infty$, the value of the probability $\mathbf{P}(A)$ cannot be positive. Thus, it is 0, which means that its complement has a probability 1, that is, the student, sooner or later, passes the course for sure.

In order to simulate the above problem see the following file. Whenever you press the F9 key, 10000 experiments are performed. Pressing the F9 key again and again, you will see that, regardless how small the probability of the success is, sooner or later success will occur.

Demonstration file: Many experiments for an event which has a small probability ef-100-04-00

The purpose of the following numerical example is to show that, if our student's knowledge is strongly declining, then, in spite of the fact that he or she has an infinite number of possibilities, the probability $\mathbf{P}(A)$ may be positive, that is, it is not sure at all that the student ever passes the course.

Example 3. Student's knowledge strongly declining. Let us assume that our student fails the first exam with a probability

$$\mathbf{P}(A_1) = 0.6 + 0.4/2 = 0.8000$$

and if our student fails the first n exams, then the probability of failing the next exam is:

$$\mathbf{P}(A_{n+1}|A_n) = \frac{0.6 + 0.4/(n+1)}{0.6 + 0.4/n} \quad \text{for all } n$$

meaning that

$$\mathbf{P}(A_2|A_1) = \frac{0.6 + 0.4/3}{0.6 + 0.4/2} = 0.9167$$

$$\mathbf{P}(A_3|A_2) = \frac{0.6 + 0.4/4}{0.6 + 0.4/3} = 0.9545$$

$$\mathbf{P}(A_4|A_3) = \frac{0.6 + 0.4/5}{0.6 + 0.4/4} = 0.9714$$

⋮

Using the multiplication rule for a decreasing sequence of events we get, for example, that the value of $\mathbf{P}(A_4)$ is:

$$\begin{aligned} \mathbf{P}(A_4) &= \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \mathbf{P}(A_3|A_2) \mathbf{P}(A_4|A_3) = \\ &= \frac{0.6 + 0.4/2}{1} \frac{0.6 + 0.4/3}{0.6 + 0.4/2} \frac{0.6 + 0.4/4}{0.6 + 0.4/3} \frac{0.6 + 0.4/5}{0.6 + 0.4/4} \\ &= 0.6 + 0.4/5 \end{aligned}$$

In a similar way, it can be shown that the value of $\mathbf{P}(A_n)$ is:

$$\mathbf{P}(A_n) = 0.6 + 0.4/(n + 1)$$

Since the events A_1, A_2, \dots constitute a decreasing sequence of events, we get:

$$\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \lim_{n \rightarrow \infty} 0.6 + 0.4/(n + 1) = 0.6,$$

which means that the student fails all the infinite number of exams with a probability 0.6.

Remark. Notice that in this numerical example the probability that the student fails all the infinite number of exams has a probability not only positive but greater than half. So, in spite of the fact that the student has an infinite number of possibilities, failure forever is more likely than a success ever.

Remark. Let us choose the positive numbers a and b so that $a + b = 1$. If in the above example, we replace the value 0.6 by a and the value 0.4 by b , then obviously

$$\mathbf{P}(A_n) = a + b/(n + 1)$$

and the probability that the student fails all the infinite number of exams is:

$$\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \lim_{n \rightarrow \infty} a + b/(n + 1) = a.$$

The following file illustrates this more general case.

Demonstration file: Student's knowledge strongly declining
ef-100-05-00

Remark. We may also calculate how much is the probability that the student fails the first $n - 1$ exams, but passes the n th exam:

$$\begin{aligned} & \mathbf{P}(\text{fails the first } (n - 1) \text{ exams, but passes the } n\text{th exam}) = \\ & \mathbf{P}(\text{fails the first } (n - 1) \text{ exams}) - \mathbf{P}(\text{fails the first } n \text{ exams}) = \\ & \left(a + \frac{b}{n}\right) - \left(a + \frac{b}{n + 1}\right) = \\ & \frac{b}{n(n + 1)} \end{aligned}$$

The following file show these probabilities.

Demonstration file: When does the student pass the exam?
ef-100-06-00

11 *** Drawing with or without replacement. Permutations

Drawing with replacement. Assume that a box contains 5 tickets with different letters on them, say A, B, C, D, E. Let us draw a ticket, write down the letter written on it, and let us put back the ticket into the box. Then, let us draw again, write down the letter on this ticket, and let us put back this ticket into the box, too. What we did is called drawing twice **with replacement**. Obviously, we may draw several times with replacement, as well. The following files illustrate what drawing with replacement means.

Demonstration file: Drawing with replacement from 10 elements
ef-110-01-00

Demonstration file: Drawing with replacement from 4 red balls and 6 blue balls
ef-110-02-00

Drawing without replacement. Now let us draw a ticket, write down the letter on it, and let us put aside the ticket. Then, let us draw another ticket from the box, write down the letter on this ticket, and let us put aside this ticket, too. What we did is called drawing twice **without replacement**. Obviously, if there are n tickets in the box, we may draw at most n times without replacement. The following files illustrate what drawing without replacement means.

Demonstration file: Drawing without replacement from 10 elements
ef-110-03-00

Demonstration file: Drawing without replacement from 4 red balls and 6 blue balls
ef-110-04-00

Permutations. If there are n tickets in the box, and we draw n times without replacement, then we get a permutation of the n tickets. Obviously, all possible permutations have the same probability. The following file gives a random permutation of the 10 given elements.

Demonstration file: Permutations of 10 elements
ef-110-03-05

The following problem entitled "Catching the Queen" may seem an artificial problem which is far from real life. But as you will see the solution of this problem will help us to find the optimal strategy of a typical real life problem which will be presented later under the title "Sinbad and the 100 beautiful girls".

Example 1. Catching the Queen. First, let us choose and fix a number c between 0 and 9. Take, for example, $c = 4$. Then let us consider a permutation of the numbers $1, 2, \dots, 10$, for example, $6, 5, 7, 4, 1, 2, 8, 10, 9, 3$. The largest number, that is, the 10 is called "the Queen", and the largest number before the Queen is called the Servant. In the above example, the Servant is the number 7. Let us denote the position of the Queen by X , and let us denote the position of the Servant by Y . In the above example, the position of the Queen is 8, that is, $X = 8$. and the position of the Servant is 3, that is, $Y = 3$. We are interested in the probability of the event that the position of the Queen is larger than c and the position of the Servant is smaller than or equal to c , that is, $X > c$ and $Y \leq c$. This probability obviously depends on c . We will express this probability in terms of c ,

Solution.

$$\begin{aligned} \mathbf{P}(X > c \text{ and } Y \leq c) &= \\ \sum_{k=c+1}^{10} \mathbf{P}(X = k \text{ and } Y \leq c) &= \\ \sum_{k=c+1}^{10} \mathbf{P}(X = k) \mathbf{P}(Y \leq c \mid X = k) &= \\ \sum_{k=c+1}^{10} \frac{1}{10} \frac{c}{k-1} &= \\ \frac{c}{10} \sum_{k=c+1}^{10} \frac{1}{k-1} \end{aligned}$$

Remark. Let us consider a permutation of the numbers $1, 2, \dots, 100$. The largest number, that is, the 100 is called "the Queen", and the largest number before the Queen is called the Servant. Let us choose a number c between 0 and 99. The probability of the event that the position of the Queen is larger than c and the position of the Servant is smaller than or equal to c , that is, $X > c$ and $Y \leq c$ can be calculated the same way as in the previous example:

$$\begin{aligned}
 \mathbf{P}(X > c \text{ and } Y \leq c) &= \\
 \sum_{k=c+1}^{100} \mathbf{P}(X = k \text{ and } Y \leq c) &= \\
 \sum_{k=c+1}^{100} \mathbf{P}(X = k) \mathbf{P}(Y \leq c \mid X = k) &= \\
 \sum_{k=c+1}^{100} \frac{1}{100} \frac{c}{k-1} &= \\
 \frac{c}{100} \sum_{k=c+1}^{100} \frac{1}{k-1} &
 \end{aligned}$$

For each number c between 0 and 99, the value of this probability can be calculated by Excel:

*Demonstration file: [Catching the Queen](#)
ef-110-03-08*

We see that the maximal probability occurs when $c = 37$, and the value of the maximal probability rounded to 6 decimal places is 0.371043, or rounded to 2 decimal places is 0.37. This fact will be used in the following example.

Example 2. Sinbad and the 100 beautiful girls. Imagine that the sultan offers Sinbad to choose one of the 100 beautiful girls in his harem. Sinbad has never seen the girls before. The method how Sinbad is allowed to make his choice is very strict: the girls show up for Sinbad separately, one after the other, in a random order, and Sinbad has the right to say "this is the girl I choose" only once. This means that when a girl shows up and then disappears because Sinbad does not choose her, then Sinbad has a no any chance to meet this girl again. The purpose of Sinbad is to catch the most beautiful girl, so when he makes his choice, he will ask whether he could catch the most beautiful girl. If he realizes that has caught the most beautiful girl, then he will be happy, otherwise he will be not. (We assume that "beauty" is well defined for each girls so that the girls could be arranged into a well defined order according to their beauties.)

You probably think that Sinbad has no much chance to become happy: 100 unknown girls are too many for Sinbad being able to catch one specific one, namely, the most beautiful one! If Sinbad picks a girl at random, then the probability to catch the most

beautiful one is 0.01. You will see soon that, applying a tricky strategy, Sinbad can catch the most beautiful girl with 0.37 probability, which is much larger than 0.01.

Solution. What Sinbad can actually do is to apply a strategy like this. First he observes, say c , girls, without choosing any of them, but keeping in mind the maximal beauty of them. Then he compares the beauty of each later girl to this maximal beauty. If one of the later girls is more beautiful than this maximal beauty, then he chooses that girl. If none of the later girls is more beautiful than this maximal beauty, then he will choose the last girl. Let us think out what is the probability that he can catch the most beautiful girl. Using the terminology and notation of the previous example, we may write:

$$\begin{aligned} \mathbf{P}(\text{Sinbad catches the Queen}) &= \\ \mathbf{P}(X > c \text{ and } Y \leq c) &= \\ \sum_{k=c+1}^{100} \mathbf{P}(X = k \text{ and } Y \leq c) &= \\ \sum_{k=c+1}^{100} \mathbf{P}(X = k) \mathbf{P}(Y \leq c \mid X = k) &= \\ \sum_{k=c+1}^{100} \frac{1}{100} \frac{c}{k-1} &= \\ \frac{c}{100} \sum_{k=c+1}^{100} \frac{1}{k-1} & \end{aligned}$$

The maximal value of this expression is calculated in the following file:

[Demonstration file: Catching the Queen
ef-110-03-08](#)

So we see that if $c = 37$, then $\mathbf{P}(\text{Sinbad catches the Queen}) = 0.37$. The best strategy for Sinbad is to observe 37 girls, remembering the maximal beauty of them, and then waiting for a girl who is more beautiful than this maximal beauty, and choosing this girl.

Remark. A very good approximation to the optimal strategy for Sinbad can be derived by the following analytical method. Using the approximation

$$\sum_{k=c+1}^{100} \frac{1}{k-1} \approx \ln(100) - \ln(c) = -\ln\left(\frac{c}{100}\right)$$

we may write

$$\mathbf{P}(\text{Sinbad catches the Queen}) =$$

$$\frac{c}{100} \sum_{k=c+1}^{100} \frac{1}{k-1} \approx -\frac{c}{100} \ln\left(\frac{c}{100}\right) = -x \ln(x)$$

where $x = \frac{c}{100}$. We know that the function

$$f(x) = -x \ln(x)$$

takes its maximum at $x = \frac{1}{e}$, and the maximal value of this function is $\frac{1}{e}$. This is how we get that the best strategy for Sinbad is to use approximately a c so that

$$\frac{c}{100} \approx \frac{1}{e} \approx 0.37$$

that is

$$c \approx \frac{100}{e} \approx 37$$

and then

$$\mathbf{P}(\text{Sinbad catches the Queen}) \approx \frac{1}{e} \approx 0.37$$

which is really an excellent approximation to the exact solution.

If the number of girls is not 100 but n , then this analytical method obviously gives the approximation

$$c \approx \frac{n}{e} \approx 0.37 n$$

and

$$\mathbf{P}(\text{Sinbad catches the Queen}) \approx \frac{1}{e} \approx 0.37$$

for the optimal solution. The best strategy for Sinbad is to observe approximately 37 percent of the girls, remembering the maximal beauty of them, and then waiting for a girl who is more beautiful than this maximal beauty, and choosing this girl.