Statistics - handout
Part-I: Probability of events

Andras Vetier

January 13, 2017

Contents

I am sure you will find mistakes in this text-book. I ask you to let me know them so that I could correct them. Send an email to vetiera@gmail.com Let the subject of the letter be "Mistake found". Thanks for your cooperation.

1 Introduction

The world is full with randomness. Instead of giving the Reader examples, the Reader is kindly asked to find many many examples in his/her life.

When I was a young boy I liked mathematics because of its precision, correctness, clarity. I enjoyed that mathematical statements as soon as I understood them, they remained clear and true forever and everywhere. I could not imagine how randomness can fit to mathematics. Randomness and mathematics seemed to me to contradict to each other. It was a wonderful experience for me when I learnt probability theory later. The laws of randomness are precisely described by mathematics.

When I became a teacher of probability, I realised that, for most students, it is rather difficult to understand the laws of randomness. Not only the mathematics is difficult for the students, but they have not experienced the laws of randomness in their life. A 5 year child feels what the velocity is: my grandson encourages me to drive faster or slow down when I drive fast. However, he has no idea what the law of large numbers is. I suspect that most of my MSc students have not ever made even the simplest experiment: rolling a fair die 1000 times and checking the average of the tossed numbers.

In spite of the "mess" of the randomness, the average will be close to 3.5. Not exactly 3.5, but close to it. You may say that this is obvious because with a fair die, the numbers 1, 2, 3, 4, 5, 6 are equally probable, and 3.5 is at the center of the numbers 1, 2, 3, 4, 5, 6.

Consequently, when we learn probability theory, we have to work not only with mathematics, but we have to make real life experiments, as well. When the actual performance of a real life experiment is not possible, we shall make a simulation. We shall use the computer and make it work for us. And the computer will be obedient! You are curious, aren't you?

The world is full with randomness. Instead of giving you examples, I kindly ask you to find many many examples in your own life.

When I was a young boy I liked mathematics because of its precision, correctness, clarity. I enjoyed that mathematical statements as soon as I understood them, they remained clear and true forever and everywhere. I could not imagine how randomness can fit to mathematics. Randomness and mathematics seemed to me to contradict to each other. It was a wonderful experience for me when I learnt probability theory later. The laws of randomness are precisely described by mathematics.

When I became a teacher of probability, I realised that, for most students, it is rather difficult to understand the laws of randomness. Not only the mathematics is difficult for the students, but they have not experienced the laws of randomness in their life. A 5 year child feels what the velocity is: my grandson encourages me to drive faster or slow down when I drive fast. However, he has no idea what the law of large numbers

is. I suspect that most of my students have not ever made even the simplest experiment: rolling a fair die 1000 times and checking the average of the tossed numbers.

In spite of the "mess" of the randomness, the average will be close to 3.5. Not exactly 3.5, but close to it. You may say that this is obvious because with a fair die, the numbers 1, 2, 3, 4, 5, 6 are equally probable, and 3.5 is at the center of the numbers 1, 2, 3, 4, 5, 6. Consequently, when we learn probability theory, we have to work not only with mathematics, but we have to make real life experiments, as well. When the actual performance of a real life experiment is not possible, we shall make a simulation. We shall use the computer and make it work for us.

And the computer will be obedient! You will see!

2 Outcomes and events

We start with some basic notions of the theory.

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 a phenomenon (at least theoretically), the experiments can be repeated as many times as we want. When, related to the phenomenon, we declare what we are interested in, what we observe, we define 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: $\{$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 arranges 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, $\ldots$ 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 ∞: $\{1, 2, 3, \ldots, \infty\}$

3. We observe how many tails we get before the first head occurs. Now the outcomes are the non-negative integers: 0, 1, 2, $\ldots$ 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 ∞: $\{0, 1, 2, \ldots, \infty\}$.

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*.

3 Operations and relations on events

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.

4 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.

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.

5 Integer valued random numbers

In Excel there is the very simple `RANDBETWEEN` random command, which generates integer numbers between a lower and an upper value. For example, the `=RANDBETWEEN(1;6)` command gives a random number as if we tossed a fair die. The numbers 1, 2, 3, 4, 5, 6 are all equally probable. As an other example, the `=RANDBETWEEN(0;9)` command gives a random number so that the numbers 0, 1, 2, 3, 4, 5, 6, 7, 8 ,9 may occur, each with the same, 1/10 probability.

6 Random numbers between 0 and 1

The command `RAND()` simulates a random real number so that the smallest possible value is 0, the largest is 1. (TO BE CONTINUED)

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 $0 \le a \le b \le 1$, it is true that

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

or, the same way,

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

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

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

or, the same way,

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

The following file illustrates this fact:

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

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

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 Swhich has the vertices $(0,0)$, $(1,0)$, $(1,1)$, $(0,1)$. This means that for any $A \subset S$, it is true that

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

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

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

7 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:

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

8 Combinatorics

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:

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

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:

$$\begin{array}{ccccccccccccc}
 & & & & & & \binom{0}{0} & & & & & & \\
 & & & & & \binom{1}{0} & & \binom{1}{1} & & & & & \\
 & & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & & \\
 & & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & & \\
 & & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & & \\
 & \binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} & \\
\ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots
\end{array}$$

and calculate the numerical value of each binomial coefficient in this triangle-shaped table, we get the following array:

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
\ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots
\end{array}$$

The numbers in this triangle-shaped table satisfy the following two simple rules:

(a) The elements at the edges of each row are equal to 1.

(b) Addition rule: Elements which are not at the edges are equal to the sum of the two numbers which stand above that element.

Based on these rules one can easily construct the table and find out the numerical values of the binomial coefficients.

9 Geometrical problems, uniform distributions

Another simple way of calculating a probability is when the outcomes can be identified by an interval S of the (one-dimensional) real line or by a subset S of the (two-dimensional) plane or of the (three-dimensional) space or of an n-dimensional Euclidean space so that the length or area or volume or n-dimensional volume of S is 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

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

in the one-dimensional case, or

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

in the two-dimensional case, or

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

in the three-dimensional case, or

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

in the n-dimensional case.

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?

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

$$\mathbf{P}\left(\text{Intersection}\right) = \frac{\text{Area under the curve}}{\text{Area of the rectangle}} = \frac{\int\limits_0^{\frac{\pi}{2}} \frac{L}{2} \sin(x)\ dx}{\frac{D}{2} \cdot \left(\frac{\pi}{2}\right)} = \frac{2L}{\pi D}$$

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:

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

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.

10 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**:

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

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**:

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

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 $A_1, A_2, \ldots, A_n$ are exclusive events, then

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

   If $A_1, A_2, \ldots$ are exclusive events, then

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

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

$$\begin{aligned}\mathbf{P}(A \cup B \cup C) &= +\mathbf{P}(A_1) + \mathbf{P}(B) + \mathbf{P}(C) \\ &\quad -\mathbf{P}(A_1 \cap B) - \mathbf{P}(A_1 \cap C) - \mathbf{P}(B \cap C) \\ &\quad +\mathbf{P}(A_1 \cap B \cap C)\end{aligned}$$

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 $A_1, A_2, \ldots, A_n$ are arbitrary events, then

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

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 nth 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

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

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

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

11 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

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

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**:

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

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

If

A_2 is implies A_1, that is, $A_2 \subseteq A_1$, or equivalently, $A_1 \cap A_2 = A_2$,
A_3 is implies A_2, that is, $A_3 \subseteq A_2$, or equivalently, $A_2 \cap A_3 = A_3$,
A_4 is implies A_3, that is, $A_4 \subseteq A_3$, or equivalently, $A_3 \cap A_4 = A_4$,
$\vdots$

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, leep 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 ore 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:

We ask two questions:

1. For a given n ($n = 2, 3, 4, \ldots, 366$), how much is the probability that at least one coincidence occurs?

2. Which is the smallest n for which **P**(at least one coincidence occurs) ≥ 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:

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

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 $A_1, A_2, A_3, \ldots$ 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 kth 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

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

that is

$$\begin{aligned}
&\mathbf{P}(A_2|A_1) = 364/365 = 0,9973\\
&\mathbf{P}(A_3|A_2) = 363/365 = 0,9945\\
&\mathbf{P}(A_4|A_3) = 362/365 = 0,9918\\
&\vdots
\end{aligned}$$

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

$$\begin{array}{lllllll}
\mathbf{P}(A_1) & = & 1 & & & & \\
\mathbf{P}(A_2) & = & \mathbf{P}(A_1)\ \mathbf{P}(A_2|A_1) & = & 1 & 0,9973 & = & 0,9973\\
\mathbf{P}(A_3) & = & \mathbf{P}(A_2)\ \mathbf{P}(A_3|A_2) & = & 0,9973 & 0,9945 & = & 0,9918\\
\mathbf{P}(A_4) & = & \mathbf{P}(A_3)\ \mathbf{P}(A_4|A_3) & = & 0,9918 & 0,9918 & = & 0,9836\\
\vdots & & & & & & &
\end{array}$$

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

$$\begin{aligned}
&\mathbf{P}\left(\overline{A_1}\right) = 1 - \mathbf{P}(A_1) = 1 - 1 &&= 0\\
&\mathbf{P}\left(\overline{A_2}\right) = 1 - \mathbf{P}(A_2) = 1 - 0,9973 &&= 0,0027\\
&\mathbf{P}\left(\overline{A_3}\right) = 1 - \mathbf{P}(A_3) = 1 - 0,9918 &&= 0,0082\\
&\mathbf{P}\left(\overline{A_4}\right) = 1 - \mathbf{P}(A_4) = 1 - 0,9836 &&= 0,0164\\
&\vdots
\end{aligned}$$

In this Excel table, we find the answer to our first question: the probability that at least one coincidence occurs is calculated for all $n = 1, 2, \ldots, 366$. 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. Wee see that

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

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

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, ... 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, ... 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?

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

Bayes formula. If the events A_1, A_2, ... 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.

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.

12 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

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

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

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

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

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

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

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**:

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

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.