Basic Lecture 1 on Completeness Interpretation of Probability

Andrzej Piegat

Faculty of Computer Science and Information Systems

West Pomeranian University of Technology

Żołnierska 49, 71-062 Szczecin, Poland

apiegat@wi.zut.edu.pl, http://apiegat.strony.wi.ps.pl

Abstract

The lecture presents a new (according to the author's knowledge) interpretation of probability that is devoid a number of weaknesses characterizing the commonly used frequency interpretation. The lecture explains why the frequency interpretation is not credible in case of no or small data and mean number of data pieces. The main reason of this phenomenon is the nonlinear character of this interpretation, which causes the pieces of data of equal significance influence the probability value with non-equal strength. However, the limited credibility of the frequency interpretation at small number of data pieces does not mean that the method is generally incorrect, because its credibility increases with number of data pieces and at sufficiently high number of pieces the results become precise. The new completeness interpretation delivers credible results both at lack of data, at small, at mean, and at high number of data pieces. This interpretation is of linear character, in the sense that data pieces of equal significance influence the hypothesis probability with equal strength. Therefore its accuracy range is greater than the one of frequency interpretation.

Keywords

Probability, probability interpretations, the completeness interpretation of probability, evidential completeness, evidential complete set, uncertainty.

Contents

1 Introduction

Probability theory is the oldest and the most developed method of investigating uncertainty. Later also some other methods have been developed and these include fuzzy systems (Zadeh [12, 21]), Dempster-Shaffer's belief theory [19], possibility theory (Dubois and Prade [4]), info-gap theory (Yakov [20]).

The first interpretation of probability was formulated by Laplace in 1814 [13]. His interpretation is referred to as the classical one. Since then multiple books and papers on probability theory have been published. This theory is lectured on thousands of universities and widely used in practice. Therefore one could think that probability theory is a strongly based scientific method that rises no doubts. However, it appears that the truth is quite different. Uncertainty and disagreement among scientists as to the sense of probability are very large. There is a great number of questions, doubts and paradoxes concerning understanding probability. Some scientists are even of the opinion that probability theory is in deep crisis. An example of such opinion is shown in a book with a very meaningful title "The search for certainty- On the clash of philosophy of probability", [1], written by Professor K. Burdzy from University of Washington and published in 2009. This book has aroused a vivid discussion among scientists, mainly in the Internet [6, 10]. Some scientists gave whole hearted support to Professor Burdzy's opinion; other criticized it, but rather in a moderate way. Supporters of presently popular probability theory mainly underline the method's practical usefulness in statistics. Nevertheless, Professor Burdzy is not alone in his views. In literature some other, very strong opinions can also be found: "Probability does not exists" and "No matter how much information you have there is no scientific method to assign a probability to an event". These are opinions of a famous probabilist de Finetti [5]. Because of limited spatial extent of this lecture all critical remarks cannot be presented here. But an interested reader can easily find them in the book of Professor Burdzy [1] or on multiple web sites, e.g. [7, 8].

First of all, it is very surprising that although we live in the XXI century we don't know what probability really is, as even between specialists there is no agreement as to the sense of probability. Intuitively, in everyday life, everyone seems to understand probability. But the more one penetrates questions and problems it addresses, the more and more one begins to be aware how difficult the matter is. There are many scientific schools of interpreting (explaining) probability in various ways and usually one probability interpretation tries to improve weak points of other. Below, the main (not all) interpretations with short comments of Professor's Burdzy [1] are presented. Probability interpretations are also discussed e.g. in books of Khrennikov [11] and Rocchi [18], in the Internet [8] and in scientific encyclopedia [7].

1. The classical probability (Laplace, 1814 [13]),
   "which claims that probability is symmetry".

2. The logical probability (Carnap, 1950 [2]),
   "which claims that probability is 'weak' implication".

3. The frequency theory (von Mises, 1957 [15]),
   "which claims that probability is long run frequency".

4. The propensity theory (Popper, 1957, [17]),
   "which claims that probability is physical property".

5. The subjective theory (de Finetti, 1975, [5]),
   "which claims that probability is personal opinion".

These are not all probability interpretations but only the most known ones. Particular interpretations reveal large quantitative differences in explanation of probability and have their extreme advocates. There exists also an interesting opinion of Hajek [7], according to who each of the above interpretations accentuates one of many faces of probability: each of them being partially true. In the practice the most applied, and lectured on universities are the classical and frequency interpretation. Therefore these two interpretations will be shortly presented and discussed in the next chapter.

2 The classical and frequency interpretation of probability

The classical interpretation with its main representative Laplace [10], (1814) "assigns probabilities in the absence of any evidence or in the presence of symmetrically balanced evidence. The guiding idea is that in such circumstances probability is shared equally among all the possible outcomes, so that the classical probability of an event is simply the fraction of the total number of possibilities in which the event occurs", [7]. Mathematically this can be represented as follows: If a random experiment can result in N mutually exclusive and equally likely outcomes and if N_A of these outcomes result in the occurrence of the event A, the probability of A is defined by:

$$P(A) = \frac{N_A}{N}\,. \tag{1}$$

There are two clear limitations of the classical definition. Firstly, it is applicable only in situations in which there is only a 'finite' number of possible outcomes. But some important random experiments, such as tossing a coin until it rises heads, give rise to an infinite set of outcomes. And secondly, you need to determine in advance that all the possible outcomes are likely without relying on the notion of probability to avoid circularity – for instance by symmetry considerations, [7]. On the ground of classical interpretation many problems could not be explained. A trial of improvement of the classical interpretation and of removal of at least some weaknesses has been undertaken by 'frequentists with the main representative von Mises [15]. "Frequentists posit that the probability of an event is its relative frequency over time, i.e. its relative frequency of occurrence after repeating a process a large number of times under similar conditions .... If we denote by n_A the number of occurrences of an event A in n trials, then if:

$$\lim_{n\to\infty} \frac{n_A}{n} = p\,, \tag{2}$$

we say that $P(A) = p$", [7].

This interpretation is also called the long-run frequency interpretation. Because in practice a very large (infinite) number of experiments cannot be realized or the number of pieces of data (e.g. of statistical data) is limited we have to use the finite-frequency interpretation according to which the probability is calculated on the basis of data we have at disposal. The definition of probability according to finite-frequency interpretation is as follows: "the probability of an attribute A in a finite reference class B is the relative frequency of actual occurrence of A within B", Hajek in [7]. Thus:

$$p(A) = \frac{n_A}{n}\,, \tag{3}$$

where: n – a finite number.

3 Main objections to classical and frequency interpretations of probability

As mentioned in section 1 there exist more than 5 main interpretations of probability and there are also serious objections to each of them. Because of the volume limitation of this lecture only some of the objections to the classical and frequency interpretation are presented below.

1.
"Since the (classical) definition applies only to those situations in which all outcomes are equally 'possible' it does not apply to a single toss of a deformed coin", [1].

2.
The classical definition seems to be circular because it refers to "equally possible cases – and so probability is defined using the notion of probability", [1].

3.
"According to the finite frequentist, a coin that is never tossed and thus yields no actual outcomes whatever, lacks a probability for heads altogether; yet a coin that is never measured does not thereby lack a diameter", [1]. This problem can be called 'the zero-evidence case'.

4.
"According to the frequency theory one cannot apply the concept of probability to individual events", [1], such as a single coin tossing. "...a coin that is tossed exactly once yields a relative frequency of heads of either 0 or 1, whatever its bias. ...this is so called 'problem of the single case' ", [7]. From the fact that one coin toss yielded head should we conclude that the head probability equals 1? Such conclusion is suggested by the frequency theory. Such conclusion would be hasty and precipitate. Let us consider a second example, of a physician, who wants to determine the probability of cancer as result of smoking. This physician begins to collect data about patients. Let us assume that at the beginning his data basis consist of only one patient who has smoked cigars since many years but has no cancer. Direct conclusion suggested by frequency interpretation here is "probability of cancer at smoker is zero". Such conclusion would of course be nonsense. The above examples shown that the frequency theory cannot be applied to a single case. But what, when number of cases is higher, e.g. 2, 3, ..., 10? It appears that also for higher number of cases the frequency interpretation gives questionable or non-credible results. Let us assume that in 5 coin tosses the head was up. The frequency theory suggests in this case the head probability equal to 1 and the tail probability equal to zero. Let us assume that the physician from the previous example has 10 patients in his data basis who are smokers and have no cancer. The frequency interpretation also in

this case suggests directly probability 1 for hypothesis "Smoking does not cause cancer". Of course, no one of us would accept such scientific result.Thus, it is clear that in the case of small number of evidence pieces or when all evidence pieces support only one hypotheses and remaining hypotheses have no support (unilateral-evidence case), the frequency interpretation delivers doubtful or non-credible results. Therefore many scientists are of the opinion that "probability of single cases are nonsense", Hajek in [7]. However, it is not true. As the new probability interpretation presented in this lecture will show, also the single-case problem and the unilateral-evidence problem can credibly be solved.

5. Peculiarities of the frequency interpretation at unilateral evidence.
Let us reconsider the physician who wants to determine probability of the hypothesis h "smoking increase the risk of cancer". Anti-hypothesis $NOT\,h = \overline{h}$ is in this case "smoking does not increase the risk of cancer". After certain time the physician has in his database a number of $n = 50$ patients who are smokers but nobody of them has cancer. All these patients supports the anti-hypothesis $\overline{h}$ and no patient support the hypothesis h. Thus we have $n_h = 0$ and $n_{\overline{h}} = 50$. The frequency theory gives probability values as below.

$$p_h = n_h/n = 0/50 = 0 \qquad\qquad p_{\overline{h}} = n_{\overline{h}}/n = 50/50 = 1\,.$$

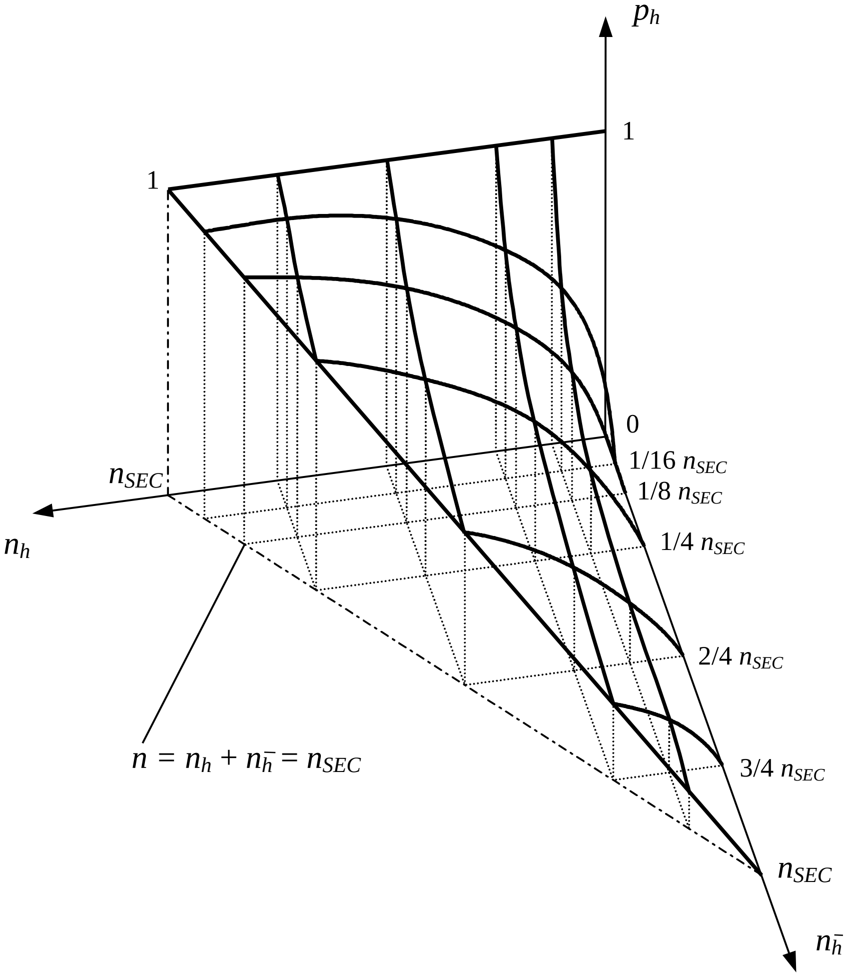
The above results suggest conclusion which is difficult to accept: "smoking does not increase the risk of cancer". A similar situation can be observed at tossing a usual coin. Let us assume, that in all $n = 10$ tosses the head was up. The direct conclusion suggested by the frequency theory concerning the head probability p_h is:

$$p_h = n_h/n = 10/10 = 1\,.$$

This probability value would mean the total domination of head in the coin which is non-credible. What is the reason of such incorrect results delivered by the frequency interpretation? The reason is the incorrect formula $p_h = n_h/n$ used in this interpretation for probability calculation. Illogicality of this formula is illustrated in Fig. 1.

The formula $p_h = n_h/n$ seemingly seems linear in relation to the n_h – number, because this number is in the nominator of the formula. Thus one could think that each piece of evidence supporting this hypotheses increases its probability at one and the same value $1/n$.

But it is not true, because the number n_h supporting the hypothesis h is also in the formula denominator $n = n_h + n_{\overline{h}}$ and thus we have $p_h = n_h/(n_h + n_{\overline{h}})$. This formula is nonlinear. The nonlinearity of the formula suggested by frequency theory is just the reason of observed illogicalities of this theory. Let us for example analyze a special, border situation when there is an evidential support of the hypothesis h ($n_h > 0$) but there does not exist any support of the anti-hypothesis $\overline{h}$ ($n_{\overline{h}} = 0$). This situation is modeled by the vertical, back face in Fig. 1. If we have no support both for the hypothesis h ($n_h = 0$) and no support for the anti-hypothesis $\overline{h}$ ($n_{\overline{h}} = 0$) then according to the frequency theory the probability p_h (and simultaneously $p_{\overline{h}}$) is undetermined (the zero-evidence case). If however even one single evidence piece supporting the hypothesis h would be achieved then the probability value p_h becomes at once known and has the highest possible value $p_h = 1$, what means certainty. If 10 evidence pieces for support of the hypothesis h have been obtained (and no support of the anti-hypothesis $\overline{h}$) then its probability does not change and still equals 1. Also if 100 pieces ($n_h = 100$ at $n_{\overline{h}} = 0$) or more would support the hypothesis, its probability does not change and keeps the same value($p_h = 1$,



$$p_h = \frac{n_h}{n} = \frac{n_h}{n_h + n_{\overline{h}}}$$

Figure 1: Functional surface of probability of the hypothesis h calculated with formula $p_h = n_h/n = n_h/(n_h + n_{\overline{h}})$ suggested by the frequency interpretation. Denotation: n – total number of evidence pieces being at disposal, n_h – number of pieces supporting the hypothesis h, $n_{\overline{h}}$ – number of pieces supporting the anti-hypotheses $\overline{h}$, n_{SEC} – a certain constant number of pieces ($n_{SEC} = \text{const}$).

$p_{\overline{h}} = 0$). It makes no difference, whether we have 1 or 100 or 1000000 evidence pieces supporting the hypothesis h (at no support of the anti-hypothesis $\overline{h}$). Its probability does not change. Let us consider yourselves, whether the number of evidence pieces supporting the hypothesis should or shouldnt influence the hypothesis probability or not?

According to the frequency theory, in certain situations, the number of evidence pieces has no meaning, in other situations it has meaning on probability value (when both $n_h > 0$ and $n_{\overline{h}} > 0$)! It is logical? Certainly not. It is illogical and, as it will be shown, the new, completeness interpretation proves in logical and convincing way that also in the extreme, special situations ($n_h = 0$ and $n_{\overline{h}} > 0$) or ($n_h > 0$ and $n_{\overline{h}} = 0$) the number of evidence pieces has an impact on probability value. It always has an impact! And let us consider another one question: If the frequency theory gives non-credible probability values in strictly extreme situations ($n_h = 0$, $n_{\overline{h}} > 0$ and $n_h > 0$, $n_{\overline{h}} = 0$) then will it give credible results in situations very close and similar to the extreme ones as e.g. when n_h is very small and $n_{\overline{h}}$ is not very small (for example $n_h = 1$ and $n_{\overline{h}} = 5$) and similar ones?

6. Fluctuations of the frequency probability at small number of evidence pieces.
The author had made an experiment of coin tossing and achieved the following result {T,H,T,T,T,H,T,H,H,T}, where H means head and T means tail. Table 1 shows values of the frequency probability $p_h = n_h/n$.

Toss result	n_h	n	$p_h = n_h/n$
T	0	1	0
H	1	2	$1/2 = 0.500$
T	1	3	$1/3 = 0.333$
T	1	4	$1/4 = 0.250$
T	1	5	$1/5 = 0.200$
H	2	6	$2/6 = 0.333$
T	2	7	$2/7 = 0.286$
H	3	8	$3/8 = 0.375$
H	4	9	$4/9 = 0.444$
T	4	10	$4/10 = 0.400$

Table 1: Values of the frequency probability $p_h = n_h/n$ calculated according to the frequency interpretation after each of succeeding coin tosses from the series of 10 tosses {T,H,T,T,T,H,T,H,H,T}, where: n_h – head number, n – number of tosses, H – head, T – tail.

Fig. 2 shows graph of probability estimates n_h/n of the hypothesis h (head domination) that are given in Table 1. Result of a single toss can be called confirmation of one of both possible hypotheses ("head domination" as hypothesis h or "tail domination" as anti-hypothesis $\overline{h}$).

As it can be seen on Fig. 2 the probability estimates considerably vary after each toss. Thus the estimate delivered by the frequency interpretation resembles a hesitant, undecided person, who too quickly and too hasty draws conclusions after achieving each single piece of information. Then, when new information pieces come she/he has to considerably correct her/his previous conclusions.

7. Fluctuations of frequency probability at high number of evidence pieces.
According to the frequency interpretation of probability based on a great number of evidence pieces (experiment results, samples) the true, exact value of probability can only be known after realizing an infinitely large number of experiments. Alas, experiments made by scientists have shown that even after a very large number of experiments the probability not always stabilizes and incessant fluctuations of probability are observed, Larose [14]. This phenomenon is illustrated by Fig. 3.

As analyses made by the author shown the main reason for the considerable fluctuations of probability defined as the relative frequency is the non-linear character of the formula ($p_h = n_h/n$) itself used by the frequency theory for probability calculation and not some objective reasons. It refers both to cases of a small and a large number of

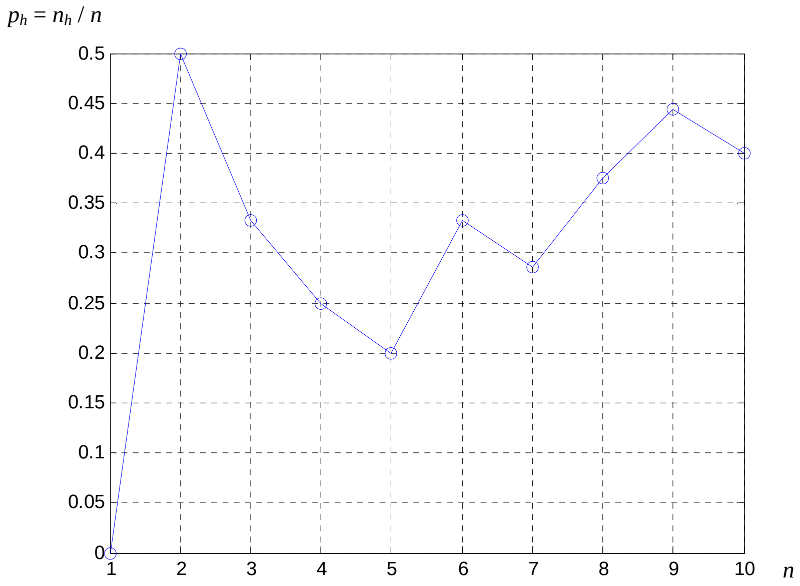


Figure 2: Graph of probability estimates $p_h = n_h/n$ of the hypothesis h (head domination) calculated after each succeeding toss of the series $\{$T,H,T,T,T,H,T,H,H,T$\}$ as example of considerable fluctuations if probability calculated on the basis of the frequency theory, where: n_h – number of heads after n tosses.

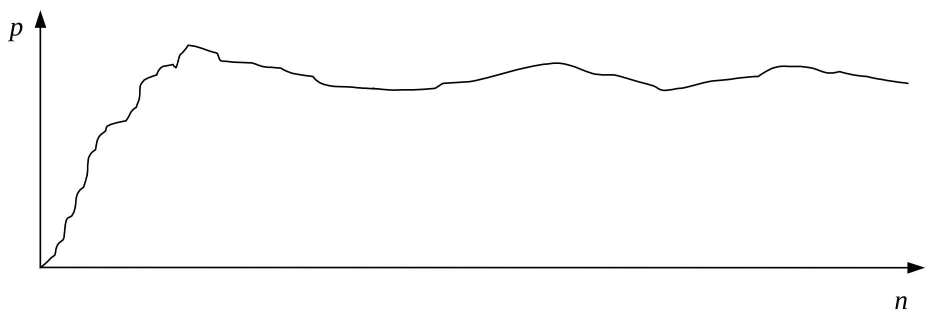


Figure 3: Phenomenon of fluctuations of frequency probability calculated according to frequency theory ($p_h = n_h/n$)observed after a great number of experiments.

experiments. In case of the new, completeness interpretation of probability fluctuations both at small and large number of experiments don't occur in practice and are considerably smaller than in case of the frequency interpretation. This confirms stabilization of probability at large number of evidence pieces.

8. "Injustice" and illogicality of the frequency interpretation of probability.
Let us assume, like previously in paragraph 6, that we have some results of coin tossing in the form of a sequence $\{$T,H,T,T,T,H,T,H,H,T$\}$. Table 2 presents values of probability estimates p_h (head domination) after each succeeding toss, after succeeding confirmations or negations of the hypothesis h.

As it can be seen in Table 2 the second toss gave head causing probability change $\Delta p_h = 0.5$. This head will be denoted as H2. The succeeding head was risen in the toss 6 (H6) and changed the head probability by $\Delta p_h == 0.133$, which is considerably smaller than the value 0.5 caused by H2. The third head H8 changed the head probability

Toss result	n_h	n	$p_h = n_h/n$	Change Δp_h
T	0	1	0	—
H	1	2	$1/2 = 0.500$	$+0.500$
T	1	3	$1/3 = 0.333$	-0.167
T	1	4	$1/4 = 0.250$	-0.083
T	1	5	$1/5 = 0.200$	-0.050
H	2	6	$2/6 = 0.333$	$+0.133$
T	2	7	$2/7 = 0.286$	-0.047
H	3	8	$3/8 = 0.375$	$+0.089$
H	4	9	$4/9 = 0.444$	$+0.069$
T	4	10	$4/10 = 0.400$	-0.044

Table 2: Values of probability $p_h = n_h/n$ of the hypothesis h about head domination and values of the probability changes Δp_h after each succeeding toss (in relation to the previous toss) in the sequence of 10 tosses {T,H,T,T,T,H,T,H,H,T} calculated on the basis of the frequency interpretation of probability, where n_h means the head number after n tosses, H means head and T means tail.

at a yet smaller value of $\Delta p_h = 0.089$ and the last head H9 at lowest value $\Delta p_h = 0.040$. Because each of the succeeding heads caused different changes of the probability it means that each of the heads had different evidential meaning for the probability evaluation. Is it logical and justified? **Why result of one coin toss should be more important for probability than result of another toss?** This phenomenon seems illogical and difficult to explain or justify. The phenomenon of different weight (significance) of particular evidence pieces is known in the literature as "sequence ordering problem" and has been remarked by many scientist, e. g [1, 7]. As it will be shown in next chapters, this phenomenon does not occur in case of the new completeness interpretation of probability, where each succeeding evidence piece change probability at the same value.

4 The new, completeness interpretation of probability

The completeness interpretation of probability [16] is according to the author's knowledge new and devoid of a series of weaknesses of the frequency interpretation. First of all, the formula $p_h = n_h/n$ proposed by the frequency theory is partly incorrect (though not fully incorrect). It allows for calculation of credible probability values only for large and very large number of samples. This fact explains its practical usefulness in statistics, where frequently a greater evidence sample is at disposal. On the other hand, at small number of evidence pieces, in case of unilateral evidence and at no evidence this formula cannot be used for probability estimation, thus it cannot model a certain class of problems. It also means that the formula $p_h = n_h/n$ is qualitatively incorrect. It seems that the reason for this qualitative incorrectness is a lack of an important element in the whole concept of the frequency interpretation. According to the author, the lacking element is 'evidential completeness'. Its meaning will be explained below. In Polish probabilityis called ‘prawdopodobieństwo’, which means ‘similarity to the truth’. In Latin also: ‘versimilitudo or probabilitas’ (veritas means the truth and probabilis means credible or probable). Perhaps probability has similar meaning in other languages too. Thus, if we want to determine probability of a given hypothesis h concerning an event on the basis of evidence pieces e_{hi}, $i = 1, \ldots, l$, that confirm the truth of h we should have an image of what would the complete set $EC = \{e_{h1}, \ldots, e_{hl}\}$ of such evidence pieces be, which would fully prove the truth (with certainty 1) of this hypothesis. Such evidential set is proposed to be called ‘evidential completeness’ (EC) or ‘evidence complete-set’ (ECS). Further into this chapter the case of a discrete random variable X will be discussed. Let us assume that the variable can take k possible values in a future event. Then we can formulate a hypothesis set H for this variable as below.

$$H = \{h_1^*, h_2^*, \ldots, h_k^*\}$$

It can easily be shown that each finite set of k hypotheses can be transformed in k binomial sets of the form:

$$H = \{h, NOT\, h\} = \{h, \overline{h}\}$$

where: $h = h_i^*$, $i = 1, \ldots, k$ and $\overline{h} = H - h_i^*$. E.g. in the case of a dice $h_i^* = h$ can mean ‘1’ and $\overline{h}$ – ‘NOT 1’. Thus, for each discrete variable we can apply the basic, binomial pattern: ‘hypothesis’, ‘anti-hypothesis’, that is well represented by coin tossing: $h =$ head domination, $h = NOT\, h = NOT$ head domination (tail domination). Then, we can successively investigate probability of the hypothesis $h_2^* =$ ‘2’ and of the anti-hypothesis $NOT\, h_2^*$ (NOT 2), etc. A problem with k hypotheses can be transformed in a problem with $k1$ binomial sub-problems of the type of coin tossing. A continuous random variable X can be discretized in k sub-intervals. This way we can also transform the task of identification of its probability density function while performing a task with a discrete variable for which $k1$ probabilities of sub-intervals have to be determined, Fig. 4.

Hypothesis h: $x \in [x_i, x_{i+1}]$

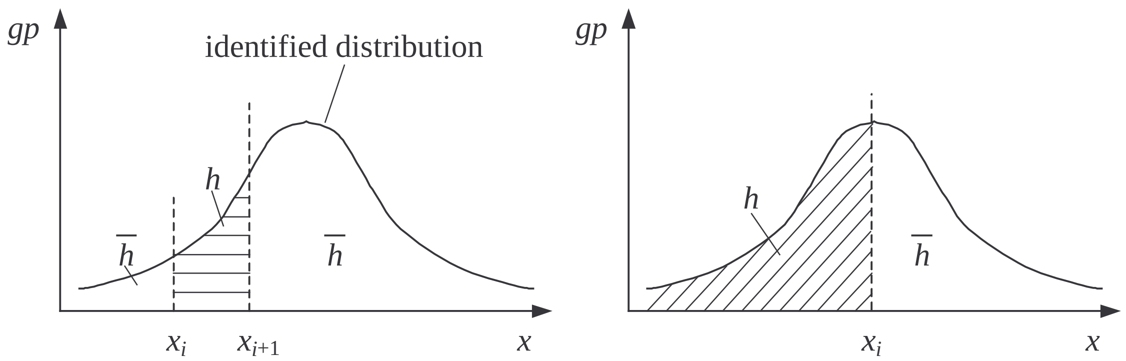


Figure 4: Transformation of the task of determining probability density function of a continuous random variable X in $k1$ tasks with the binary hypotheses set $H = \{h, NOT\, h\}$ of the type of coin tossing. Denotation: pd – probability density, $\overline{h}$ – $NOT\, h$.

Decreasing the granulation of subintervals allows for increasing the accuracy of approximation of the continuous variable by the discrete variable. The above shows that the binomial problem with 2 possible outcomes of the type of coin tossing is the basic probabilistic problem and its solution is a key to solving other more complicated problems with number of hypotheses larger than 2. Therefore the new probability interpretation will be explained on the example of the binomial problem. In the lecture it will be explained on an example of coin tossing to ensure the ease of understanding. However, **one should not draw a conclusion that an investigation of coin-tossing problem is this lecture's task. In here general problem with 2 hypotheses is analyzed. The example of coin tossing is only one of easiest possible illustrations of this general two-hypotheses problem**.

Below, there are given 3 examples of binomial problems.

- Hypothesis: the rate of people with diabetes among obese ones exceeds 25%.
  Anti-hypothesis: the rate of people with diabetes among obese ones does not exceed 25%.
- Hypothesis: the time of my today travel with a car from Szczecin to Berlin will exceed 2 hours.
  Anti-hypothesis: the time of my travel will not exceed 2 hours.
- Hypothesis: the average zloty/dollar-exchange rate in 2011 will drop below 2.5 zloty/dollar.
  Anti-hypothesis: the average zloty/dollar-exchange rate will not drop below 2.5 zloty/dollar.

According to the author, **probability should rather not be assigned to events or states but to hypotheses concerning possible outcomes of the events or forms of states**. For example, before coin tossing we can formulate 2 hypotheses concerning the outcome of this experiment: the hypothesis h (head domination) and the anti-hypothesis $NOT\,h = \overline{h}$ (tail domination) and assign their probabilities. After the coin tossing, we have to deal with its realization r = head or $\overline{r}$ = tail. However, their probabilities were assigned not by us but by the experiment. Probabilities of realizations can only have values 1 or 0. Fractional values are not possible. One coin tossing delivers only one confirmation: either it confirms the hypothesis h (head domination) or the anti-hypothesis h (NOT head domination). One coin tossing delivers only one piece of evidence. If n tosses were realized that ended with k heads and $n-k$ tails, then we have k confirmations of the hypothesis h and $n-k$ confirmations of the anti-hypothesis $\overline{h}$ at disposal. The number n of all confirmations can be different, e.g. 1 or 5 or 21 or 1000, etc. An important question arises: can we infer and assign probabilities to hypothesis on the basis on any number n of evidence pieces? If yes, then how accurate these probabilities will be? Does accuracy depend on the number n of evidence pieces or not? Are probabilities of one hypothesis determined on the basis of different numbers of confirmations equally credible? Thus, answering the question concerning the necessary cardinality n of the evidence set is essential. As proposed, the set of evidence pieces that fully proves the hypothesis h, makes it certain and excludes even minimal probability of the anti-hypothesis $NOT\,h$ ($p_h = 1$, $p_{NOT\,h} = 0$) will be called a complete evidential set (CES) or shortly evidential completeness (EC). In certain problems this set can have an ideal form, in other problems it will be impossible. As an example we can consider a crime, e.g. a murder. Let person A be suspect of murder (SP – suspected person). The person is not the only person suspected by police. The binomial set of hypotheses has in this case the form $H = \{h_A, NOT\,h_A\}$, where h means the hypothesis 'person A committed the murder' and the anti-hypothesis $NOT\,h_A$ means 'a person or persons other than A committed the murder'. The evidential completeness EC can be seen as below.

$EC = \{$ SP has no alibi for the murder time, SP had strong motives for the murder commission (e.g. large inheritance), SP was seen by few witnesses in time and on place of the murder in the course of the murder commission, on the murder place there was found genetic matter of A, on the knife which was the murder tool experts found some genetic matter of SP $\} = \{e_{CE1}, \ldots, e_{CE5}\}$

If we have such evidence against the person A as in the set EC we can be sure of the hypothesis h_A (A is the murderer) and probability of this hypothesis is equal to 1. However, if against A we only have an evidential set as below:

$E_A = \{$ A has no alibi for the murder time, A had strong motives for the murder commission $\}$

then we cannot be sure of the hypothesis and its probability is fractional. It should be noted that particular evidence pieces in the evidential set E_A are objective and confirmed facts determined by the police. Therefore their probabilities equals 1 and they don't have to be evaluated. Evaluation concerns only the probability degree of the hypothesis h_A. The probability degree of this hypothesis can approximately be evaluated by criminal experts. Similarly, the weight of particular evidence pieces also has to be evaluated by experts. **In some cases, the weights (significance degrees) of particular evidence pieces can be different.** Sometimes, as in the case of coin tossing weight of each single toss (of its result) is acknowledged to be equal.

In the murder problem the evidential completeness consisted of a finite number of pieces. Below a second example is shown, where the evidential completeness if finite. It is a binomial problem with following hypothesis:

The hypothesis: the first person I will meet today in my faculty will be a woman.

The anti-hypothesis: the first person I will meet today in my faculty will be a man.

In this problem the evidence completeness necessary for determining probability of the hypothesis is finite because it consists of all students studying in the faculty and of all workers of the faculty and this number is limited.

Let us go back now to the binomial hypothesis pattern $H = \{h, NOT\, h\}$. What will the evidential completeness EC in case of this pattern represented by coin tossing be? The ideal EC would consist in this case of an infinitely large number n of tossing results. But it is impossible to realize such number of trials. With similar situations we can also have to do in other cases. Therefore in many practical problems we will have to use not an ideal but an approximate evidence completeness which will be called **satisfactory evidential completeness, in short SEC**. It is such set evidence pieces, which as a matter of fact does not ensure the full truth of a given hypothesis, however, it insures this truth to a satisfactory degree, e.g. to 0.99 or 0.95, etc., in the scale [0,1].This degree should be determined by experts. In the case of coin tossing SEC will contain such number n_{SEC} of trail results that is sufficient for probability determining with a satisfactory accuracy. The number n_{SEC} can also be understood as a certain **model or representation of the infinity**, that is, a number which in the given specific problem well represents (replaces) the infinity. To find such number we can use the so called Chernoff bound [3, 9]. However, the problem of possibly precise determining of the n_{SEC}-number is very important, not easy and requires special and more advanced investigations for different hypothesis patterns (binomial, trinomial, ..., polynomial pattern). Some possibilities gives here Chernoff bound. Chernoff has derived a formula for the binomial hypothesis pattern which allows for determining the minimal, satisfactory number of evidence pieces (in the case of coin tossing the minimal number of tosses). On the basis of this number the probability of a given hypothesis h can be determined with an assumed, required credibility. The Chernoff formula (4) can be seen below, with a little changed denotation.

$$n_{SEC} \geq \frac{1}{(p_{hc} - 0.5)^2} \ln \frac{1}{\sqrt{\epsilon}} \tag{4}$$

Denotation:
ϵ – a certain, assumed, maximal probabilistic uncertainty (error) of the hypothesis proof (e.g. about head domination) which should not be exceeded. If we assume e.g. $\epsilon = 0.01$ then accuracy of the hypothesis proof based on the number n_{SEC} of evidence pieces will be equal to at least $1 - \epsilon = 0.99$. In practice, it means that if 100 series of coin tossing would have been made (in each series n_{SEC} tosses) then in only one of these 100 series ($\epsilon = 0.01$) its result would not confirm the hypothesis h but the anti-hypothesis $NOT\, h = \overline{h}$. In other 99 series (each consisting of n_{SEC} tosses) the hypothesis h (e.g. about head domination) would be confirmed.

p_{hc} – means an assumed probability of the hypothesis ($p_{hc} \geq 0.5$), which we want to check or prove on the basis of n_{SEC} evidence pieces. If in case of a coin we suspect (e.g. from introductory experiments) that the head domination (the coin asymmetry) is equal to about 0.55 then we assume $p_{hc} = 0.55$ and calculate the number n_{SEC} of tosses that is necessary for sufficiently credible prove of the hypothesis about the head domination (using Chernoff bound (4)). If after a series of n_{SEC} tosses the observed frequency p_h of head was higher than the assumed border p_{hc} (e.g. $p_h = 0.57$ at $p_{hc} = 0.55$) then it would mean that the realized trial number n_{SEC} was higher than necessary. If the observed frequency p_h of the head was lower than p_{hc} (e.g. $p_h = 0.53$ at $p_{hc} = 0.55$) then a new n_{SEC} value should be calculated for the new value of $p_{hc} = 0.53$. This new value would be higher than the old one, which would mean that additional trails would have to be realized. If no additional trials were made, then the new, real accuracy $(1 - \epsilon)$ of the hypothesis should be calculated. If the trials number was too low than required, then the accuracy would be also lower, not 0.99 but e.g. 0.91. Table 3 gives few examples of the suspected hypothesis probability p_{hc} and the number n_{SEC} of evidence pieces required for proving with accuracy 0.99.

p_{hc}	0.501	0.510	0.520	0.530	0.550	0.600
n_{SEC}	2 302 586	23 026	5 757	2 559	921	231

Table 3: Examples of satisfactory evidence completeness n_{SEC} (required number of evidence pieces) in the binomial problem (e.g. required number of coin tosses) allowing for hypothesis prove with accuracy $(1 - \epsilon) = 0.99$ at maximal error $\epsilon = 0.01$.

As examples in Table 3 show, generally the more symmetrical the coin is (the head probability approaches 0.5) the more difficult it becomes to prove the hypothesis because a greater number of tosses n_{SEC} is required. Likewise, the more asymmetrical the binomial problem is the less evidence pieces n_{SEC} are necessary. In some ways the task resembles identification of twins. In following paragraphs **definitions of probability for the binomial hypothesis pattern** will be presented. These definitions allow for showing, that in many cases it will not be possible to precisely determine the probability p_h of the hypothesis. However, as we will see the evidence set E_h we have at disposal, depending on its strength, will allow for more or less approximate determination of probability.

The minimal probability p_{hmin} **of the hypothesis** h concerning a given event or a state of matter is the degree of conformability (or similarity) of the evidence set E_h that we have at disposal for confirmation of the hypothesis h with the evidence required for full confirmation of the hypothesis, which is collected in the set EC_h of evidential completeness of hypothesis h.

The maximal probability p_{hmax} **of the hypothesis** h is equal to one minus the minimal probability $p_{NOT\,hmin}$ of the anti-hypothesis $NOT\,h$.

$$p_{hmax} = 1 - p_{NOT\,hmin} \tag{5}$$

The minimal probability $p_{NOT\,hmin}$ **of the anti-hypothesis** $NOT\,h$ is the degree of conformity (similarity) of the evidence set $E_{NOT\,h}$ containing evidence pieces confirming truth of the anti-hypothesis with the complete evidence set $EC_{NOT\,h}$ required for full confirmation of the anti-hypothesis truth.

The maximal probability $p_{NOT\,hmax}$ **of the anti-hypothesis** $NOT\,h$ is equal to one minus the minimal probability p_{hmin} of the hypothesis h, formula (6).

$$p_{NOT\,hmax} = 1 - p_{hmin} \tag{6}$$

The true (exact) probability value ph of the hypothesis h **(and also the true value of the anti-hypothesis probability)** can be determined only when the condition (7) will be satisfied, that is, when the sum of minimal probabilities of the hypothesis and the anti-hypothesis will be equal to one (7).

$$\text{IF } (p_{hmin} + p_{NOT\,hmin} = 1) \text{ THEN } [(p_{hmin} = p_h) \text{ AND } (p_{NOT\,hmin} = p_{NOT\,h})] \tag{7}$$

The true (exact) probability value of the hypothesis h **cannot be determined** and known if the minimal probabilities of the hypothesis and anti-hypothesis do not sum up to one, formula (8).

$$p_{hmin} + p_{NOT\,hmin} < 1 \tag{8}$$

The reason for the above situation is the insufficiency and scarcity of the evidence. Unfortunately, we frequently have to deal with such situations in real problems. Values of probability are then constrained by conditions given below.

$$0 \le p \le 1 \tag{9}$$

$$0 \le p_{hmin} + p_{hmax} \le 2 \tag{10}$$

$$0 \le p_{NOT\,hmin} + p_{NOT\,hmax} \le 2 \tag{11}$$

$$0 \le p_{hmin} + p_{NOT\,hmin} \le 1 \tag{12}$$

$$0 \le p_{hmax} + p_{NOT\,hmax} \le 2 \tag{13}$$

Correctness of all above can be easily proved with use of detailed formulas that will be given in next chapter. If the complete evidence set EC_h is not realizable in a given problem because of certain restrictions (e.g. it never can be collected), to approximately determine the probability of the hypothesis one can use the satisfactory evidence completeness SEC.

5 Uncertainty of probability

Let us now return to the binomial hypothesis pattern with the hypothesis set $H = \{h, NOT\,h\} = \{h, \overline{h}\}$ which is well presented by coin tossing. In this case h means the hypothesis of head domination, which also can be interpreted as "in the next toss a head will be up". Let us assume that we know from introductory experiments that head dominates and its probability should not be less than $p_{hc} = 0.55$. Using Chernoff bound method we calculate the number n_{SEC} of tosses (evidence pieces) required for determining of the hypothesis about head domination with accuracy $(1 - \epsilon) = 0.99$ which is equivalent to the maximal error $\epsilon = 0.01$ and achieve the result $n_{SEC} = 921$. This number means the satisfactory evidence completeness of the problem. Let's analyze **the first situation** and assume that in first experiments we got the following results: $n_h = 3$ heads and $n_{NOT\,h} = n_{\overline{h}} = 2$ tails (not heads). Thus the evidence set E_h contains 3 confirmations (evidence pieces) of head domination while the evidence set $E_{NOT\,h} = E_{\overline{h}}$ contains 2 confirmations of the anti-hypothesis about tail domination. The complete number n of all evidence pieces that we momentary have at disposal equals 5 and is considerably smaller than the satisfactory completeness $n_{SEC} = 921$. What can be done for determining probability of the hypothesis h in a situation of such large information insufficiency? The situation is illustrated by Fig. 5.

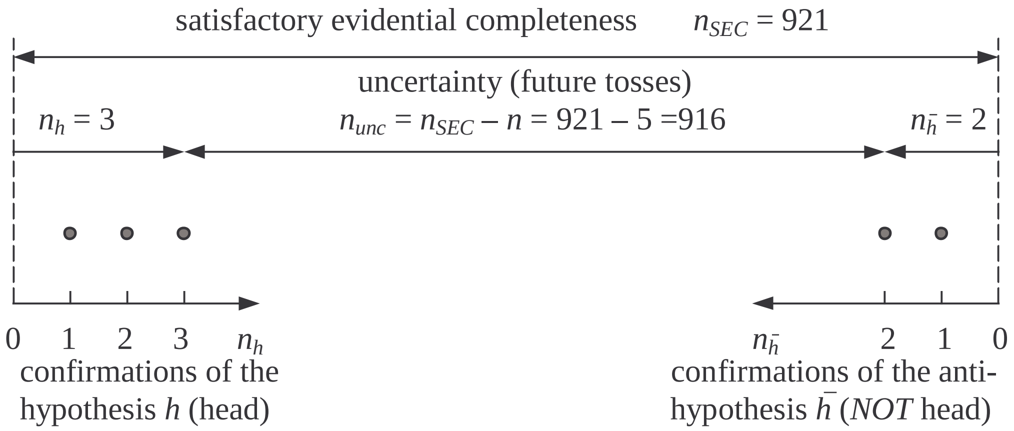


Figure 5: An illustration of evidence insufficiency (lack of 916 evidence pieces) and uncertainty caused by it in the task of determining the hypothesis h – probability of head domination in a coin tossing experiment. In this case we have the small number $n = 5$ of evidence pieces, which is considerably smaller than the number n_{SEC} required by the satisfactory evidence set SEC.

As Fig. 5 shows, to get the required number of 921 evidence pieces further 916 coin tosses would have to be made because only 5 toss results are at disposal now. Although momentary results indicate head domination ($n_h = 3$, $n_{\overline{h}} = 2$) next 916 tosses can change this situation in a fully unknown way, e.g., they may prove the tail domination. Therefore we are not allowed to formulate a categorical and hasty conclusion concerning the head or tail domination on a basis of only 5 evidence pieces. Because **the situation is very uncertain our conclusions should be very cautious**. The number of 5 evidence pieces is very low in comparison with the required number $n_{SEC} = 921$ however it delivers us certain knowledge about probability and this knowledge can be used, though we should not have too high expectations. In an extreme case, it is possible that all next 916 tosses will give heads. Then head would be supported by $3 + 916 = 919$ confirmations (evidence pieces). Thus the maximal possible head probability that can be concluded from 5 evidence pieces amounts to $p_{hmax} = 919/921$. And the minimal probability of head domination p_{hmin} secured by 3 possessed evidence pieces is equal to $p_{hmin} = 3/921$. Next, the minimal probability of the anti-hypothesis about tail domination equals to $2/921$, because at present we have 2 confirmations of the tail domination. Because it is possible that all next 916 tosses will give tail the maximal possible tail (anti-hypothesis) probability $p_{\overline{h}max}$ equals $(2 + 916)/921 = 918/921$. It can be easily seen that each tail confirmation increases its minimal probability $p_{\overline{h}min}$ at the value $1/921$ (supports the tail) and decreases the maximal possible probability p_{hmax} of "the opponent" head at the same value (acts to disadvantage of the opponent). Because, most probably, certain part of the future 916 tosses will give heads and the other part will give tails, the final, precise probability value p_h of the head will lie somewhere between the minimal p_{hmin} and the maximal possible value p_{hmax}.

$$p_{hmin} \leq p_h \leq p_{hmax} \qquad (3/921) \leq p_h \leq (919/921)$$

Correspondingly, the true probability value $p_{\overline{h}}$ of the anti-hypothesis $\overline{h}$ about the tail domination will lie somewhere between its minimal and maximal value.

$$p_{\overline{h}min} \leq p_{\overline{h}} \leq p_{\overline{h}max} \qquad (2/921) \leq p_{\overline{h}} \leq (918/921)$$

On a basis of the example analized, following formulas can be formulated.

$$p_{hmin} = n_h/n_{SEC}\,, \qquad p_{hmax} = 1 - p_{\overline{h}min} = 1 - n_{\overline{h}}/n_{SEC} \tag{14}$$

$$p_{\overline{h}min} = n_{\overline{h}}/n_{SEC}\,, \qquad p_{hmax} = 1 - p_{hmin} = 1 - n_h/n_{SEC} \tag{15}$$

The true and exact values of probability of the hypothesis h and of the anti-hypothesis $\overline{h}$ are correlated and satisfy condition (16).

$$p_h + p_{\overline{h}} = 1 \tag{16}$$

This condition means that the exact probabilities of the hypothesis and anti-hypothesis complement themselves to one. This situation is illustrated by Fig. 6.

If a commonly used formula (3) from the 5, suggested by frequency interpretation will be used for determining the probability of hypothesis h about head domination, results as below will be achieved.

$$p_h = n_h/n = 3/5\,, \qquad p_{\overline{h}} = n_{\overline{h}}/n = 2/5 \tag{17}$$

An important question arises: why only a value $p_h = 3/5$ is suggested as probability estimate of the hypothesis h and not some other value from the interval of possible values $[3/921, 919/921]$ shown in Fig. 6? Why the value $3/5$ should be distinguished from other values? Nevertheless, it is just as 'good' as any other value from the interval. The value $3/5$ is not the true probability of the hypothesis, it is only a momentary frequency f_r resulting from only 5 evidence pieces (samples) being at disposal in the moment. Next pieces could change it considerably.

Now, let us analyze the second situation in which we have not 5 but a considerably larger number $n = 700$ of evidence pieces consisting of 399 heads (confirmations

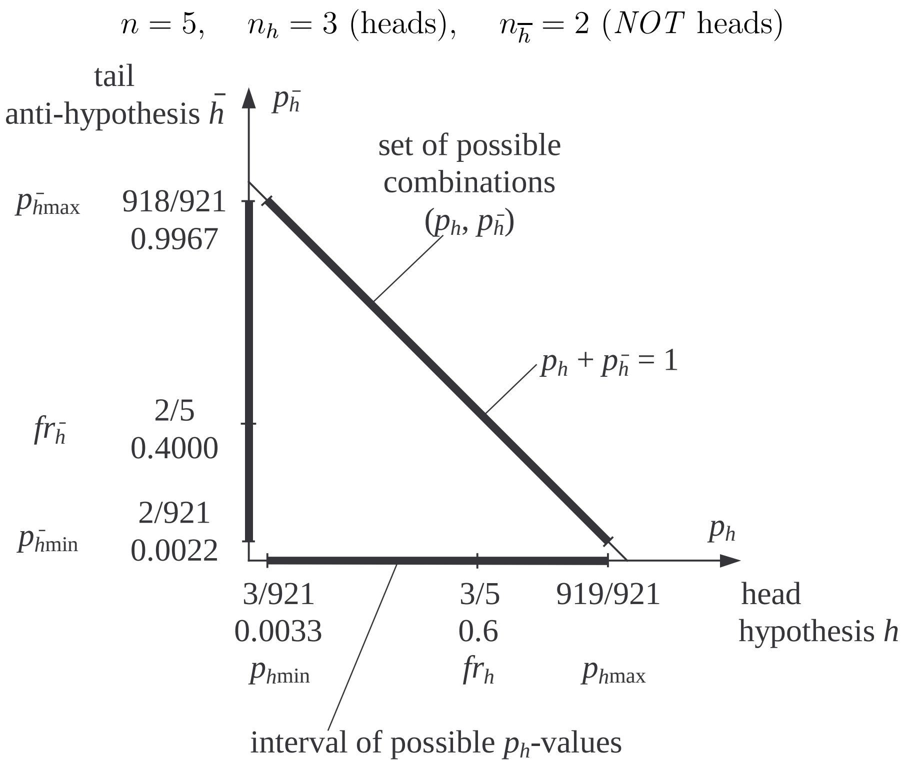


Figure 6: Illustration of uncertainty of probability p_h of the hypothesis h and $p_{\overline{h}}$ of the anti-hypothesis $\overline{h}$ identified on a basis of only 5 evidence pieces, what is considerably smaller than the number of pieces $n_{SEC} = 921$ required by the satisfactory evidence completeness SEC_h.

of the hypothesis about head domination) and 301 tails confirming the anti-hypothesis about tail domination. It means that the minimal head probability equals:

$$p_{hmin} = n_h/n_{SEC} = 399/921 = 0.433\,,$$

and the minimal tail probability equals:

$$p_{\overline{h}min} = n_{\overline{h}}/n_{SEC} = 301/921 = 0.430\,.$$

The maximally possible head probability is equal to:

$$p_{hmax} = 1 - p_{\overline{h}min} = 1 - (301/921) = 620/921 = 0.673\,,$$

and the maximally possible tail probability equals:

$$p_{\overline{h}max} = 1 - p_{hmin} = 1(399/921) = 522/921 = 0.567\,.$$

The true value of probability of the hypothesis h of head domination is not known. However, it is known that it is contained inside the determined limits. The same refers to the anti-hypothesis. Summary of knowledge achieved on the basis of the mentioned 700 evidence pieces is given below.

$$(399/921) \le p_h \le (620/921)\,, \qquad (301/921) \le p_{\overline{h}} \le (522/921)$$

Successive probabilities achieved after 100, 200, ..., 700 trials with coin are shown in Table 4.

n	0	100	200	300	400	500	600	700
n_h	0	59	117	172	226	292	349	409
$n_{\overline{h}}$	0	41	83	128	174	208	251	291
p_{hmin}	0	0.0641	0.1270	0.1868	0.2454	0.3170	0.3789	0.4441
p_{hmax}	1	0.9555	0.9099	0.8610	0.8111	0.7742	0.7275	0.6840

Table 4: Results of successive coin tosses. Denotation: n – number of trials, n_h – number of heads, $n_{\overline{h}}$ – number of tails, p_{hmin} – the lower limit of the head probability p_h, p_{hmax} – the upper limit of head probability p_h.

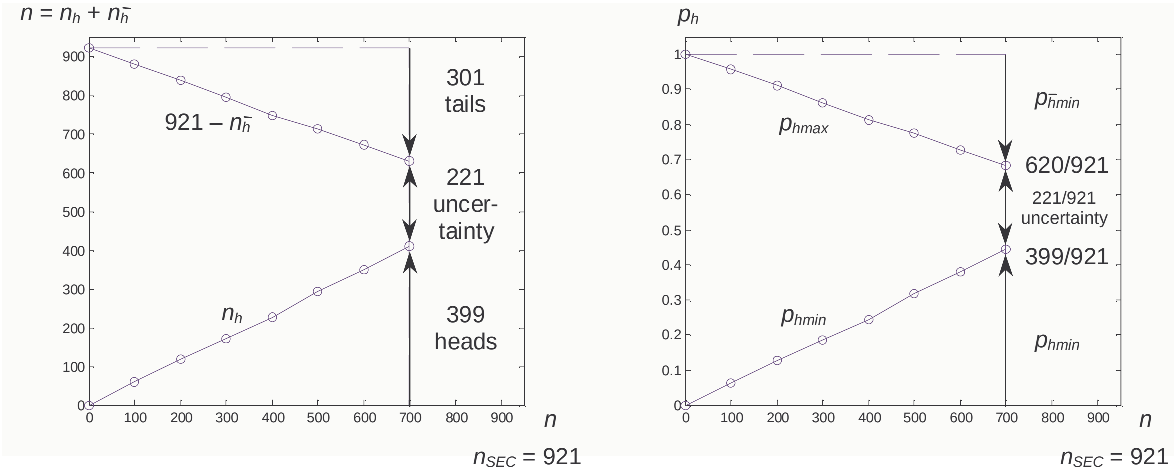


Figure 7: Illustration of uncertainty of probability p_h of the hypothesis h of head domination achieved on the basis of $n = 700$ coin tosses ($n_h = 399$ heads, $n_{\overline{h}} = 301$ tails) at the required number $n_{SEC} = 921$ of tosses, where: p_h – tail probability.

As Fig. 7 shows, the 700 evidence pieces being now at disposal considerably decreased uncertainty of probability p_h of the hypothesis about head domination in comparison with the first situation, where only 5 pieces were at disposal. Nevertheless, the uncertainty is still considerable because the 221 tosses lacking to the full number of 921 tosses required by satisfactory evidential completeness SEC may in different way change the present situation. At present the head is dominating ($p_{hmin} = 399/921$ at $p_{\overline{h}min} = 301/921$). But if an appropriately larger part of the lacking 221 tosses will give tails then the tail can become the dominating side of the coin. If these results of trials with a coin, $n = 700$, $n_h = 399$, $n_{\overline{h}} = 301$, were used for calculating the probability p_h on the basis of the frequency interpretation, results as the ones shown below would be achieved. Denotation fr means relative frequency.

$$p_h = fr_h = n_h/n = 399/700\,, \qquad p_{\overline{h}} = fr_{\overline{h}} = 301/700$$

Difference in results of both interpretations are considerable. They are shown in Fig. 8

Comparing Fig. 6 and 8, after an increase of trials from 5 to 700 (at the required satisfactory number $n_{SEC} = 921$) shows that a considerable decrease of the probability uncertainty took place. Fig. 8 also shows the position of the result calculated on the basis of the frequency interpretation of probability, that is of $p_h = 399/700 = 0.570$. It would not be easy to explain why just this value should represent the true but unknown (because

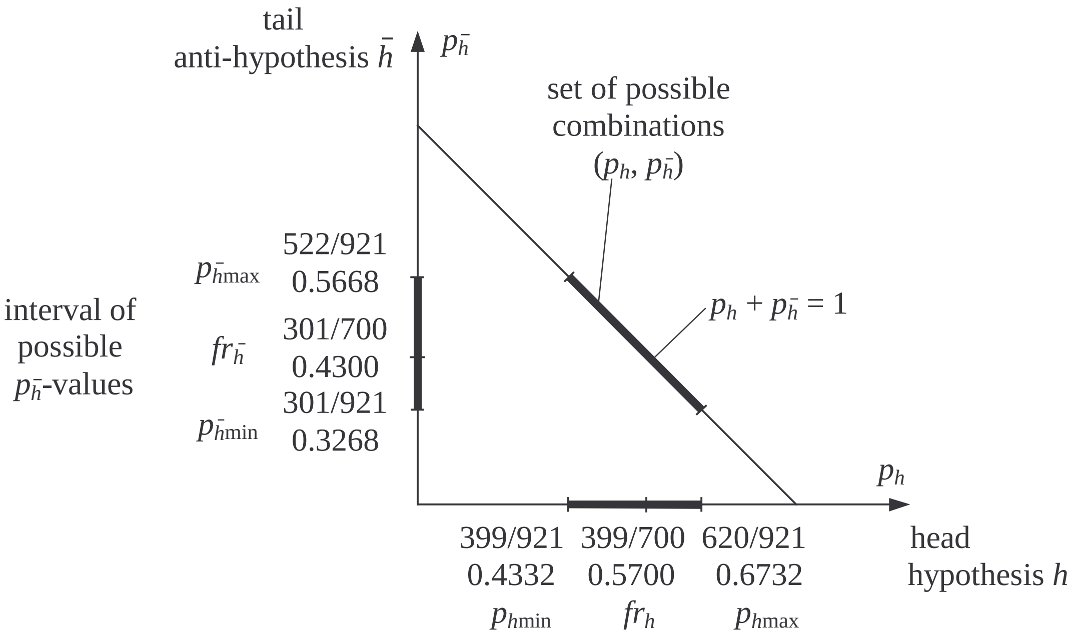


Figure 8: Comparison of results of determining the probability p_h of the hypothesis h about head domination and probability $p_{\overline{h}}$ of the anti-hypothesis about tail domination based on results of 700 trials (399 heads and 301 tails) calculated on the basis of the completeness and the frequency interpretation.

of insufficient evidence) value of the probability p_h. The value 399/700 suggested by the frequency interpretation is neither more or less credible than any other probability value from interval of possible values $p_h \in [399/921, 620/921]$.

Now, let us consider the third situation in which we possess the full satisfactory set of evidence pieces of cardinality $n = n_{SEC} = 921$, which consists of $n_h = 531$ heads and $n_{\overline{h}} = 390$ tails. Because we have the full evidence set, then on the basis of the completeness interpretation we get results as below.

$$p_h = n_h/n_{SEC} = 531/921 \text{ (head)}, \qquad p_{\overline{h}} = n_{\overline{h}}/n_{SEC} = 390/921 \text{ (tail)}$$

$$p_h + p_{\overline{h}} = (531/921) + (390/921) = 1$$

Identical probability values are in this case achieved on the basis of the frequency interpretation, because the value n of all trials which is used in the formula $fr_h = n_h/n$ of this representation is in this case exactly equal to the satisfactory number n_{SEC} (i.e. $n = n_{SEC} = 921$). The probability value $p_h = 531/921$ of the hypothesis about head domination is very close to the true value of this probability because the maximal error ϵ, according to the Chernoff bound (4) does not exceed 0.01. The ideally exact probability value cannot be determined, because it would require an infinitely large number of trials with the coin. The results achieved are shown in Fig. 9.

Table 5 and Fig. 10 show a history of increase of the number of evidence pieces from zero to the full number $n = n_{SEC} = 921$ required by the evidence completeness SEC.

Fig. 10 shows how both probabilities, minimal p_{hmin} and maximal p_{hmax} change as the number n of trials (and evidence pieces) increases and also how the gap $(p_{hmax} - p_{hmin})$ between them, being the uncertainty of the determined probability p_h, decreases. The

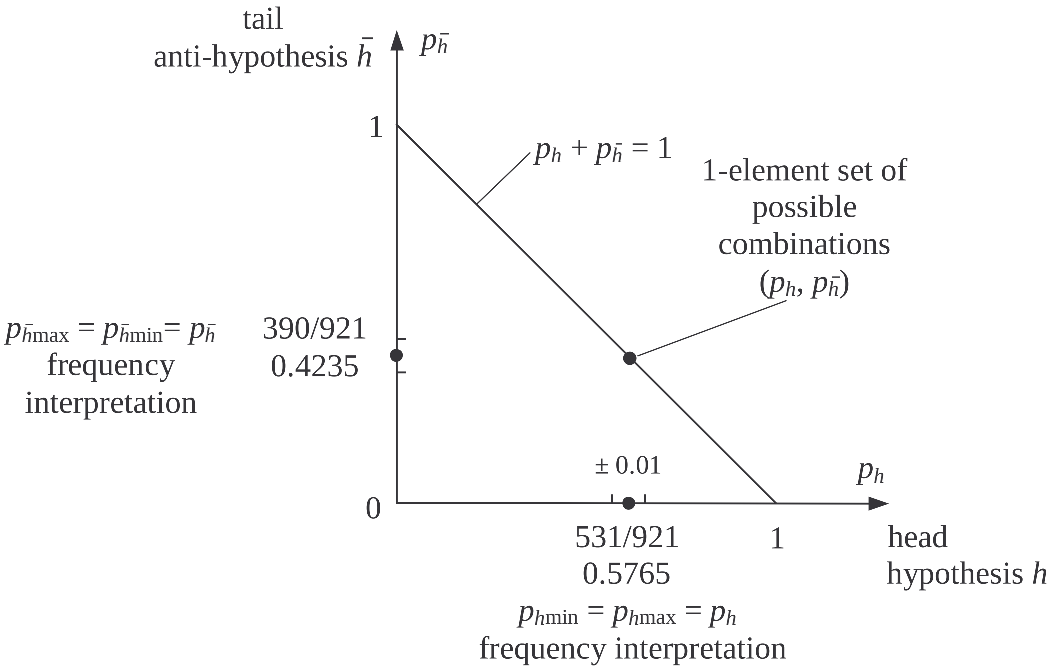


Figure 9: Illustration of results of determining probability p_h of the hypothesis about head domination and of probability $p_{\overline{h}}$ of the anti-hypothesis about tail domination in a situation of possession of the full satisfactory evidence set $n = n_{SEC} = 921$ (531 heads and 390 tails).

probability value $p_h = 531/921 = 0.5765$ shown in Fig. 10 is not the ideally exact value of the hypothesis probability p_h about head domination, because determining it would require an infinite number of trials. However, this value is a satisfactory approximation of the probability p_h. Its error, according to the Chernoff bound (4) does not exceeds $\epsilon = 0.01$.

n	0	100	200	300	400	500	600	700	800	900	$n_{SEC} = 921$
n_h	0	59	117	172	226	292	349	409	464	522	531
$n_{\overline{h}}$	0	41	83	128	174	208	251	291	336	378	390
p_{hmin}	0	0.0641	0.1270	0.1868	0.2454	0.3170	0.3789	0.4441	0.5038	0.5668	0.5765
p_{hmax}	1	0.9555	0.9099	0.8610	0.8111	0.7742	0.7275	0.6840	0.6352	0.5896	0.5765

Table 5: Results of the experiment of coin tossing until realizing the full number of trials $n_{SEC} = 921$ required by the satisfactory evidence completeness SEC. Denotation: n – the number of trials, n_h – head number, $n_{\overline{h}}$ – tail number, p_{hmin} – the minimal probability of the hypothesis about head domination, p_{hmax} – the maximal probability of the hypothesis.

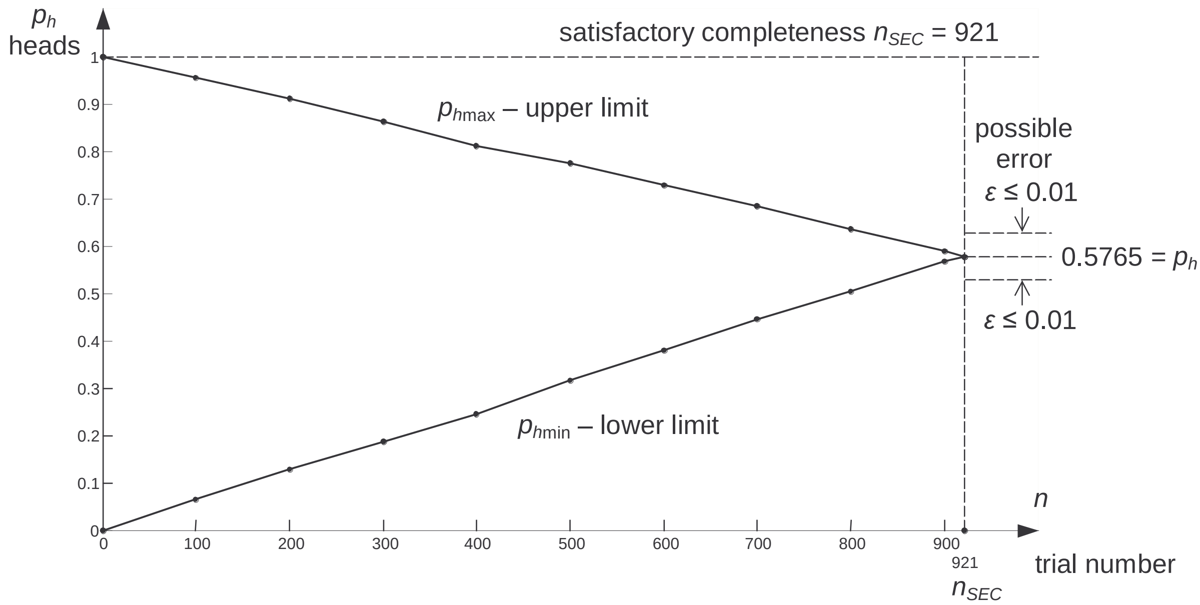


Figure 10: Changes of the lower boundary p_{hmin} and the upper boundary p_{hmax} of the probability p_h in the course of approaching the satisfactory number $n_{SEC} = 921$ of trials by the number n.

6 The optimal representation p_{hR} of the uncertainty interval $[p_{hmin}, p_{hmax}]$ of the hypothesis' probability p_h

In real decision making issues a simplified singleton-representation is often necessary (one, single number that is easily understandable for non-specialists). Therefore a question arises: "Which probability value from interval $[p_{hmin}, p_{hmax}]$ could fulfil this task the best?". To answer this question an optimality criterion has to be chosen. One of possible criteria is given by (18).

$$p_{hR} = \min[\max(p^*_{hR} - p_{hmin}, p_{hmax} - p^*_{hR})] \tag{18}$$

$$p^*_{hR} \in [p_{hmin},\, p_{hmax}]$$

It minimizes the maximal possible error of the representation p_{hR} in relation to the precise but unknown probability value p_h. Let us denote the best representation of the probability interval by p_{hR} among all possible representations p^*_{hR} contained in this interval. The best representation is given by (19).

$$p_{hR} = 0.5(p_{hmin} + p_{hmax}) \tag{19}$$

An important remark: the optimal representation p_{hR} usually is not the precise value of the true probability p_h (though sometimes it can be), because this probability cannot be determined precisely (it would require an infinite number of trials). The optimal representation is only the possibly best estimation of this value determined on the basis of such number of trials we have at disposal at present. It is only an estimation towards solving a problem and aiding decision making. Application of this representation in conditions of partial ignorance prevents large and very large errors in problem solutions. The representation p_{hR} is not the only possible representation of probability because also other representations can be proposed which can be generated by other criteria of optimality. The optimal representation p_{hR} can be shown in a more detailed form than (19) after joining in formulas (14) and (15).

$$p_{hR} = 0.5(p_{hmin} + p_{hmax}) = 0.5 + 0.5(n_h - n_{\overline{h}})/n_{SEC} \tag{20}$$

Analysis of the formula above allows for interesting conclusions. The completeness estimation p_{hR} of probability depends linearly both on the number of evidence pieces n_h confirming the hypothesis h and the number $n_{\overline{h}}$ of pieces confirming the anti-hypothesis $\overline{h}$. Each single confirmation of either the hypothesis or the anti-hypothesis changes the completeness estimation by the same value equal to $0.5/n_{SEC}$. Confirmation of the hypothesis increases its probability and confirmation of the anti-hypothesis decreases its probability). It means that value of each single evidence piece (result of a coin toss) is identical. But in case of the frequency estimate the situation is different. It results from formula (21) proposed by the frequency interpretation.

$$fr_h = n_h/n = n_h/(n_h + n_{\overline{h}}) \, , \tag{21}$$

In this formula fr_h means relative frequency of confirmations of the hypothesis h in the general number n of all evidence pieces. **Further on, denotation fr_h will be used for relative frequency as a distinction from the completeness estimate p_{hR} of probability and from the true probability p_h of the hypothesis.** Dependence of the frequency fr_h both from n_h (the number of confirmations of the hypothesis h) and $n_{\overline{h}}$ (the number of confirmations of the anti-hypothesis $\overline{h}$) is non-linear (both n_h and $n_{\overline{h}}$ occur in the denominator of formula (21)). It means that a single confirmation of the hypothesis h changes the frequency estimation fr_h at different absolute value than a single confirmation of the anti-hypothesis $\overline{h}$. Similarly, in case of two successive confirmations of the hypothesis h the first confirmation changes the frequency estimation at a different value than the second confirmation. It means that significance of single confirmations (evidence pieces as results of coin tosses) is not alike and diverse, what is not logical and difficult to explain. Attention to this fact has been drawn earlier by known scientists as Hajek [7] and Burdzy [1].

An interesting thing is how the optimal representation p_{hR} changes with increasing number n of evidence pieces (trial results) for $n \leq n_{SEC}$. Table 6 gives results of the coin tossing experiment. Fig. 11 shows changes of the lower and upper limit of the probability p_h and of the optimal representation p_{hR} of the uncertainty interval of the completeness estimate.

As Fig. 11 shows, at small number n of evidence pieces (trial results) the uncertainty $(p_{hmax} - p_{hmin})$ of probability p_h is very large. However, with increase of the pieces number n the uncertainty successively decreases to a minimum (to the possible error $\epsilon = 0.01$ in the sense of Chernoff bound (4). Also the completeness estimation p_{hR} of probability p_h, with increasing number n of evidence pieces, successively and without fluctuations approaches its end value $531/921 = 0.5765$. Credibility of the calculated probability $p_h = 531/921$ equals 0.99, what means that if we would repeat the series of 921 tosses 100 times, then in only one series (of the 100 series) the value p_h calculated from a single series will differ from the true p_h-value more than at $\epsilon = 0.01$. Table 6 and Fig. 6 present changes of the completeness estimation p_{hR} at large number n of evidence pieces till $n = 921$.

Following Table 7 presents successive results of an example series of 10 coin tosses and calculated results of probability estimates. Fig. 12 shows changes of the compared estimates p_{hR} and fr_h for small number of evidence pieces $n \leq 10$ given in Table 7.

n	0	100	200	300	400	500	600	700	800	900	$n_{SEC} = 921$
n_h	0	59	117	172	226	292	349	409	464	522	531
$n_{\overline{h}}$	0	41	83	128	174	208	251	291	336	378	390
p_{hmin}	0	0.0641	0.1270	0.1868	0.2454	0.3170	0.3789	0.4441	0.5038	0.5668	0.5765
p_{hr}	0.5	0.5098	0.5185	0.5239	0.5282	0.5456	0.5532	0.5641	0.5695	0.5782	0.5765
p_{hmax}	1	0.9555	0.9099	0.8610	0.8111	0.7742	0.7275	0.6840	0.6352	0.5896	0.5765
$\frac{n_h}{n}$	–	0.5900	0.5850	0.5733	0.5650	0.5840	0.5817	0.5843	0.5800	0.5800	0.5765
$\|\Delta\|$	–	0.0802	0.0665	0.0494	0.0368	0.0384	0.0285	0.0202	0.0105	0.0018	0

Table 6: Example results of coin tosses. Denotation: n – number of all tosses, n_h – head number, $n_{\overline{h}}$ – tail number, p_{hmin} – the lower limit of probability p_h of the hypothesis, p_{hmax} – the upper limit of the probability p_h of the hypothesis, p_{hR} – the optimal representation of the uncertainty interval of p_h (estimate of this probability), $n_h/n = fr_h$ – hypothesis probability calculated according to the frequency interpretation, $\Delta = p_{hR} - n_h/n$ – difference between probability estimates determined according to the completeness and the frequency interpretation.

Figure 11: Illustration of the decreasing process of the uncertainty interval $[p_{hmin}, p_{hmax}]$ of the probability p_h of the hypothesis and of concurrent changes of the completeness estimate p_{hR} of this probability with increasing number of evidence pieces (results of coin tosses).

n	0	1	2	3	4	5	6	7	8	9	10
n_h	0	1	2	2	3	3	4	4	5	5	6
$n_{\overline{h}}$	0	0	0	1	1	2	2	3	3	4	4
p_{hmin}	0	0.0011	0.0022	0.0022	0.0033	0.0033	0.0043	0.0043	0.0054	0.0054	0.0065
p_{hr}	0.5	0.5385	0.5011	0.5005	0.5011	0.5005	0.5011	0.5005	0.5011	0.5005	0.5011
p_{hmax}	1	1	1	0.9989	0.9989	0.9978	0.9978	0.9967	0.9967	0.9956	0.9956
$\frac{n_h}{n}$	–	1	1	0.6667	0.7500	0.6000	0.6667	0.5714	0.6250	0.5555	0.6000
$\|\Delta\|$	–	0.4615	0.4999	0.1662	0.2498	0.0995	0.1656	0.0709	0.1239	0.0550	0.0989

Table 7: Results of successive 10 coin tosses and corresponding values of the lower limit p_{hmin} and the upper limit p_{hmax} of the estimated probability p_h, the completeness estimate p_{hR}, and the frequency estimate $fr_h = n_h/n$, and of the absolute difference of both estimates $|\Delta| = p_{hR} - n_h/n|$.

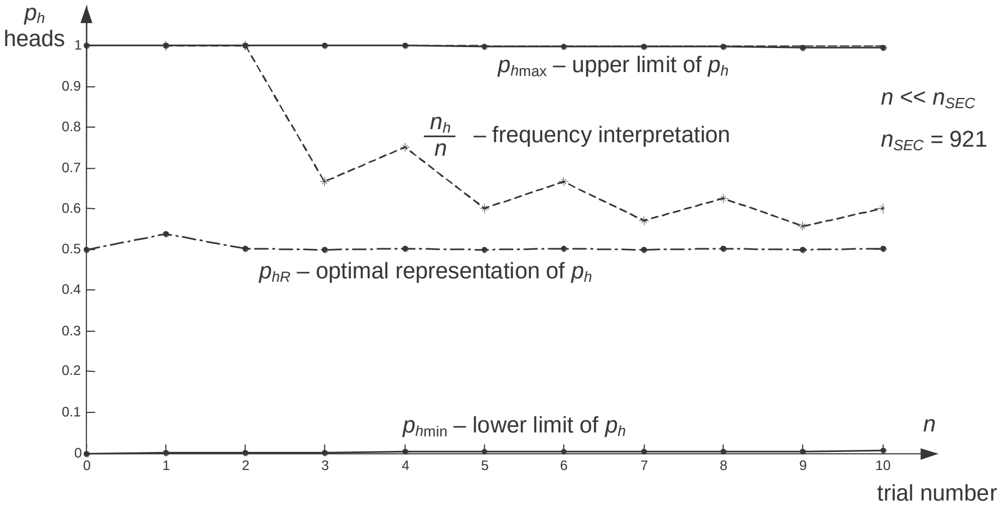
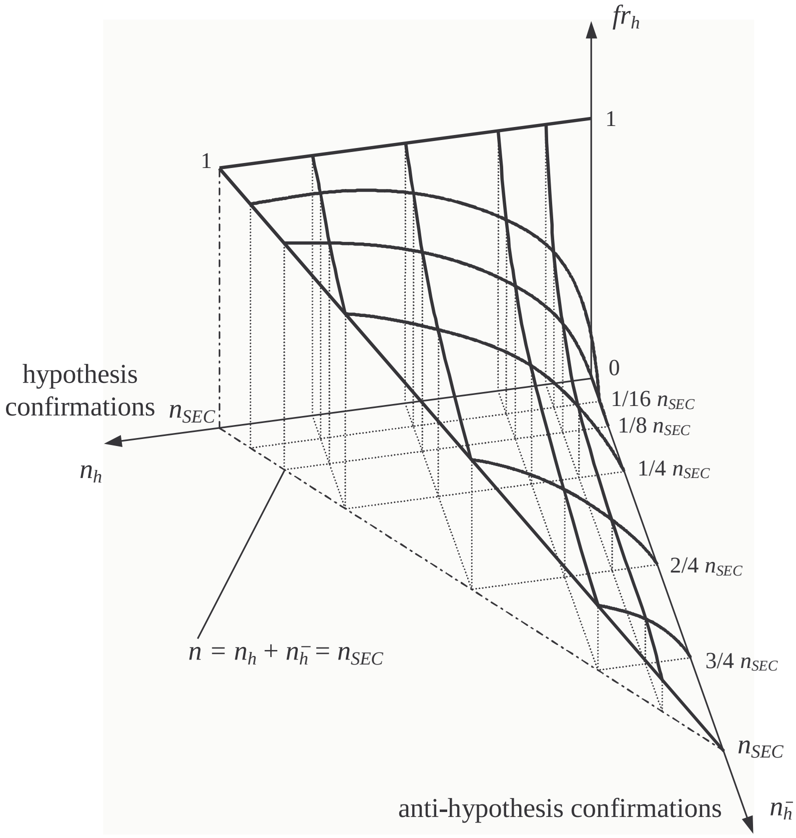


Figure 12: Changes of the completeness estimate of probability p_{hR} and of the frequency estimate $fr_h = n_h/n$ for small number of evidence pieces $n \leq 10$.

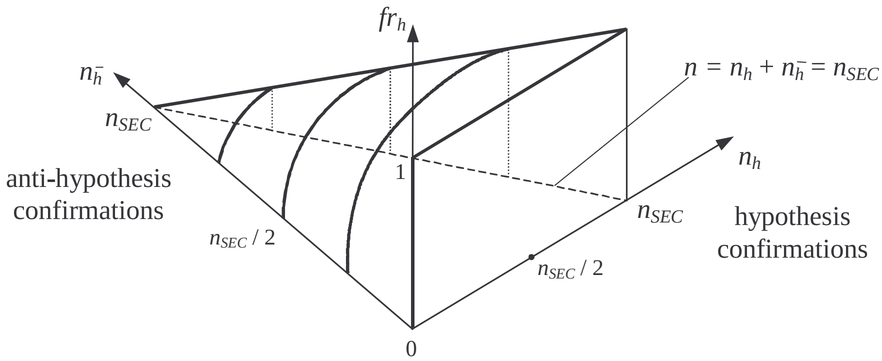
Fig. 12 makes us aware how non-credible the frequency estimation is at a small number n of evidence pieces. If we only have one piece ($n = 1$, the single case problem) then the frequency estimate can have either a value of 1, as in the example in Fig. 12, or a zero-value. It means that this estimate, based only on one evidence piece, suggests extreme values of probability (certainty). Such strong conclusions inferred from only one piece of evidence are exaggerated and hasty. To be sure, at increasing number n of evidence pieces the frequency estimate step by step approaches the true value of probability p_h of the hypothesis, but its course shows considerable fluctuations, which means that successive evidence pieces (results of coin tosses) considerably change estimated probability value. The frequency interpretation is also not able to determine any probability value for zero number of evidence pieces (in situation of lack of any evidence). We have then, following the formula $fr_h = n_h/n$, to deal with division by zero. Result of such operation cannot be determined. Instead, the completeness interpretation gives credible result both for zero-number of evidence pieces ($p_{hR} = 0.5$) and for one piece. Similarly, credible results can be obtained for any number of pieces. The course of this estimate shows very small fluctuations, considerably smaller than in case of the frequency estimate. The completeness estimate can be compared to the person that does not change strongly her/his opinions after achievement of every following evidence piece but makes it carefully and with consideration. This estimate step by step approaches the true value of probability p_h, which is shown in Fig. 11. At a large number n of evidence pieces both estimates are similar. However, the frequency estimate fr_h can show distinct fluctuations also at a large number of pieces whereas the completeness estimation shows imperceptible small fluctuations. Fig. 13 shows, in front view, the functional surface of the dependence of the hypothesis probability understood in the sense of the frequency interpretation as a relative frequency $fr_h = n_h/(n_h + n_{\overline{h}})$.



$$p_h = \frac{n_h}{n} = \frac{n_h}{n_h + n_{\overline{h}}} \qquad p_{\overline{h}} = \frac{n_{\overline{h}}}{n} = \frac{n_{\overline{h}}}{n_h + n_{\overline{h}}}$$

Figure 13: The functional surface of the dependence of the relative frequency $fr_h = n_h/n = n_h/(n_h + n_{\overline{h}})$ from the confirmation number n_h of the hypothesis h and the confirmation number $n_{\overline{h}}$ of the anti-hypothesis $\overline{h}$ in rear view.

The curved, nonlinear surface of the dependence $fr_h = n_h/(n_h+n_{\overline{h}})$ explains well why successive hypothesis confirmations (successive heads) do not have the same evidential significance (weight). Depending on the current evidential situation, (i.e. current number n_h of the hypothesis and of the number $n_{\overline{h}}$ of the anti-hypothesis confirmations), an addition of a single evidence piece causes not equal but different changes of the frequency fr_h, which testifies the uneven significance of identical evidence pieces in the frequency representation of probability. As we will see, the completeness interpretation of probability treats all evidence pieces in the same way and assigns an identical weight to each of them, independently of current evidential state, i.e. current n_h and $n_{\overline{h}}$ values.



$$fr_h = \frac{n_h}{n} = \frac{n_h}{n_h + n_{\overline{h}}} \qquad fr_{\overline{h}} = \frac{n_{\overline{h}}}{n} = \frac{n_{\overline{h}}}{n_h + n_{\overline{h}}}$$

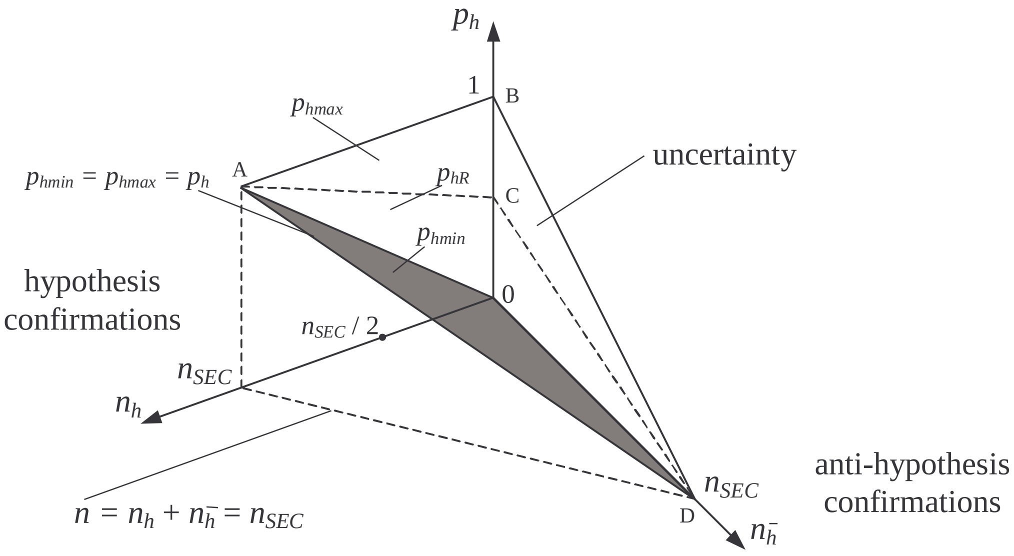
Figure 14: Functional surface of dependence of the relative frequency $fr_h = n_h/n = n_h/(n_h + n_{\overline{h}})$ of the hypothesis h on the number n_h of the hypothesis confirmations and the number $n_{\overline{h}}$ of the anti-hypothesis confirmations, in rear view, where n_{SEC} is the number of confirmations required by the satisfactory evidential completeness SEC of the hypothesis.

Figure 14 shows well the reasons for illogicality of the results delivered by the frequency interpretation in certain situations. The functional surface of the dependence $fr_h = f(n_h, n_{\overline{h}})$ is a nonlinear surface that possess insensitivity zones at its borders. And so, if there are no confirmations of the anti-hypothesis $\overline{h}$ ($n_{\overline{h}} = 0$) but there are confirmations of the hypothesis h ($n_h > 0$) then, independently of how large the number n_h of these pieces is whether it equals 1 or 2 or 10 or 100 or 1000000), the frequency estimate of probability fr_h is always equal to 1 and does not change with increasing number n_h of evidence pieces. In this case there exist no dependence between the confirmation number n_h of the hypothesis (the strength of its evidence) and its probability. Yet it is logical that with a change of the number of the hypothesis confirmations its probability also should change instead of being constant and invariable. Thus, the relative frequency fr_h does not map the hypothesis probability p_h at the lack of the anti-hypothesis confirmations ($n_{\overline{h}} = 0$). As we will see further on, the completeness interpretation of probability is devoid this fault and in logical and convincing way informs about the hypothesis probability.

Figure 15 presents 3 functional dependencies resulting from the completeness interpretation of probability: $p_{hmin} = f_1(n_h, n_{\overline{h}})$, $p_{hmax} = f_3(n_h, n_{\overline{h}})$ and $p_{hR} = f_2(n_h, n_{\overline{h}})$. This figure shows particularly well the convergence of the lower and upper border of the hypothesis probability and its estimate. The convergence of these 3 functional surfaces into one straight line takes place at the number of pieces $n = n_h + n_{\overline{h}} = n_{SEC}$ required by the satisfactory evidence set SEC. Then the lower and upper border of the hypothesis probability sums up to one independently of the proportion of the hypothesis and anti-hypothesis confirmation numbers $n_h/n_{\overline{h}}$.

$$p_{hmin} + p_{hmax} = 1$$

If the total number of all evidence pieces $n_h + n_{\overline{h}} = n < n_{SEC}$ then the lower and upper limit of the probability does not sum up to one and their difference $(p_{hmax} p_{hmin})$ represents uncertainty of our knowledge about the true value of the hypothesis probability p_h. Figure 16 shows the same functional surfaces in the rear view.



$$p_{hmin} = \frac{n_h}{n_{SEC}} \qquad p_{hmax} = 1 - \frac{n_{\overline{h}}}{n_{SEC}}$$

$$p_{hR} = \frac{p_{hmin} + p_{hmax}}{2} = \frac{1}{2} + \frac{1}{2}\frac{n_h - n_{\overline{h}}}{n_{SEC}}$$

Figure 15: Three functional surfaces representing the lower border of the hypothesis probability $p_{hmin} = n_h/n_{SEC}$, $p_{hmax} = 1n_{\overline{h}}/n_{SEC}$ and $p_{hR} = 0.5 + 0.5(n_h - n_{\overline{h}})/n_{SEC}$ resulting from the completeness interpretation of probability, in front view.

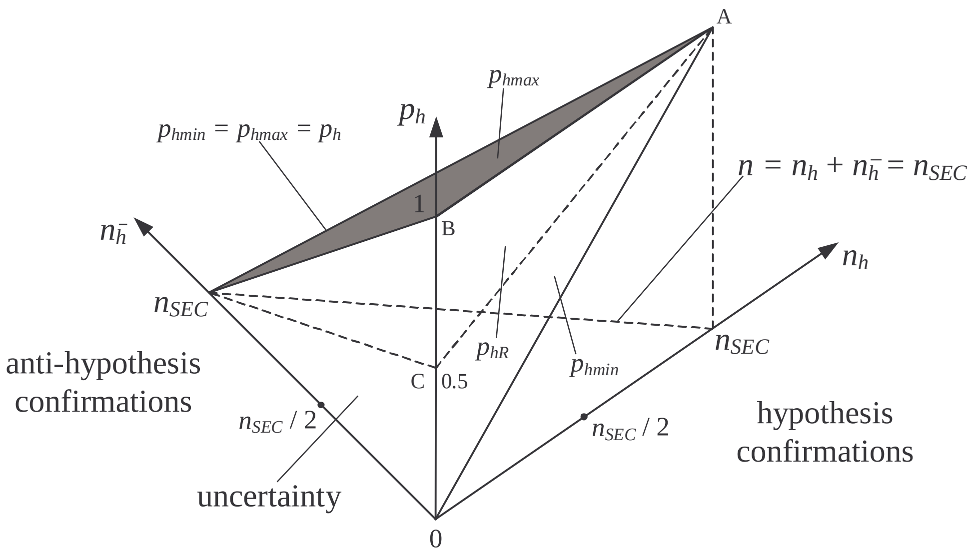
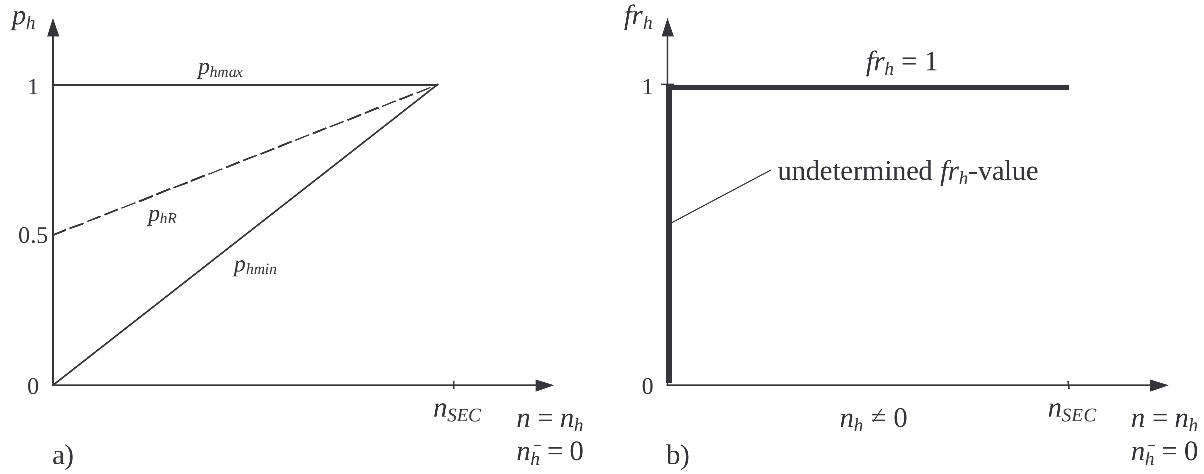


Figure 16: Functional surfaces resulting from the completeness interpretation of probability: $p_{hmin} = n_h/n_{SEC}$, $p_{hmax} = 1n_{\overline{h}}/n_{SEC}$ and $p_{hR} = 0.5 + 0.5(n_h - n_{\overline{h}})/n_{SEC}$, in the rear view.

Figure 16 shows well the evaluation of the hypothesis probability p_h by the completeness interpretation at a lack of anti-hypothesis confirmations $n_{\overline{h}} = 0$ and at existing (or not) confirmations of the hypothesis $n_h \geq 0$. With the increase of the number n_h of hypothesis confirmations the minimal value p_{hmin} of the hypothesis probability also increases, which is logical. But the upper limit p_{hmax} does not increase what logically results from the fact that the number of the anti-hypotheses confirmations does not change, it is constant and equal to zero ($n_{\overline{h}} = 0$). The border, special situations are still better presented and explained by Fig. 17 and Fig. 18. Fig. 17 shows a comparison of probability p_h of the hypothesis resulting from the frequency and completeness interpretations in the special situation when there are evidence pieces confirming the hypothesis ($n_h \geq 0$) but there are no pieces confirming the anti-hypothesis ($n_{\overline{h}} = 0$). An example of such situation is the situation of a physician who investigates probability of the following hypothesis h: "regular sport practicing prevents obesity" and who possess in his data base data of 10 persons that practice sport regularly and are slim. Thus, they confirm the hypothesis h of the physician ($n_h = 10$). However,at present he does not have examples of persons that confirm the anti-hypothesis $NOT\, h = \overline{h}$, i.e. of persons that regularly practice sport but are not slim. Thus, the number of the anti-hypothesis confirmations is $n_h = 0$. Should the physician infer from this data basis that regular sport practicing in any case and always secures slimness ($p_h = 1$ and $p_{\overline{h}} = 0$)? It would be good to have the above example in mind when reading the lecture further on.

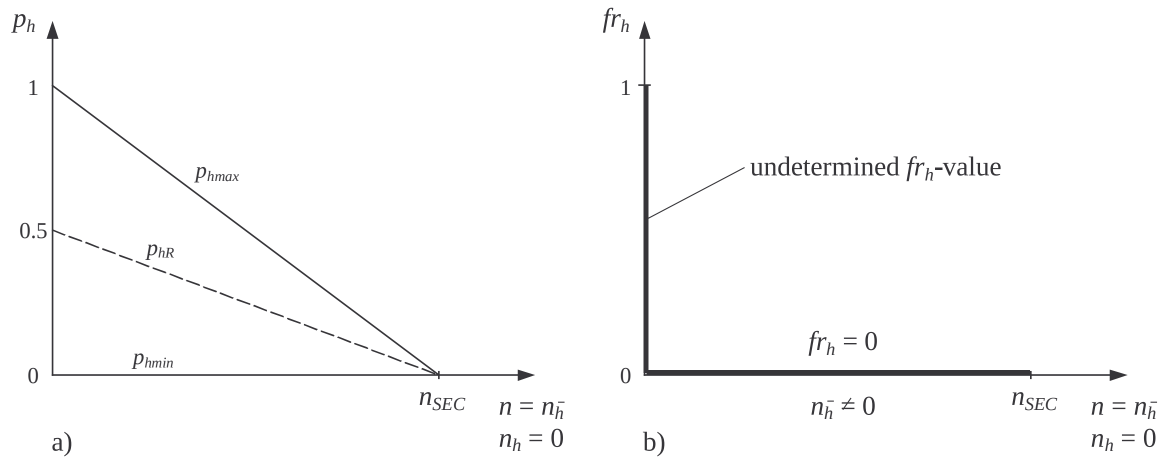


$p_{hmin} = n_h/n_{SEC} \quad p_{hmax} = 1 - (n_{\overline{h}}/n_{SEC})$
$p_{hR} = 0.5(p_{hmin} + p_{hmax}) = 0.5 + 0.5(n_h - n_{\overline{h}})/n_{SEC}$
$fr_h = n_h/n = n_h/(n_h + n_{\overline{h}})$

Figure 17: Comparison of probabilities p_h of the hypothesis h generated by the completeness (a) and the frequency interpretation of probability (b) in a situation of existence of hypothesis confirmations ($n_h \geq 0$) and non-existence of the anti-hypothesis confirmations ($n_{\overline{h}} = 0$).

As Fig. 17a shows, with an increasing number n_h of the hypothesis confirmations the completeness interpretation of probability generates higher values of the lower limit p_{hmin} of the hypothesis probability without changing the upper probability limit p_{hmax} which depends on the confirmation number $n_{\overline{h}}$ of the anti-hypothesis. Because this number is constant and equal to zero the value of p_{hmax} is also constant and equal to 1, according to the formula given in the figure. Instead, the frequency interpretation shown in Fig. 17b does not react at all to an increase of the number n_h of the hypothesis confirmations and generates a constant value of the probability estimate $fr_h = 1$, which is not logical. Additionally, the frequency estimate is not able to assign any probability value for the evidential state $n_h = 0$ and $n_{\overline{h}} = 0$ because we would have to deal with the undetermined state of division by zero. Instead, in this situation the completeness interpretation generates a credible estimate of probability equal to 0.5 (in the sense of the optimal representation of the uncertainty interval that is here maximal and equal to 1).

Fig. 18 shows a comparison of probabilities generated by the completeness and by the frequency interpretation for situation that is inverse to the one presented in Fig. 17, where there are only evidence pieces confirming the anti-hypothesis (tail domination) i.e. $n_{\overline{h}} \geq 0$, $n_h = 0$ (there are no confirmations of the head domination). Reading the explanations given further on it would be good to keep in mind the example of a policeman who investigates probability of the following hypothesis h: "A driver who drives very fast causes accidents". Let us assume that the policeman at present has in his data base only 10 drivers who drive very fast but haven't had any accidents. These are examples confirming the anti-hypothesis "A driver who drives very fast does not cause accidents". Thus $n_{\overline{h}} = 10$ and $n_h = 0$ because the policeman has no examples confirming the hypothesis h, that is, he has not examples of drivers who drive very fast and had accidents. Should the policeman infer on the basis of his present evidence (as the frequency interpretation suggests) that, with probability 1, very fast driving does not cause accidents?



$p_{hmin} = n_h/n_{SEC} \quad p_{hmax} = 1 - (n_{\overline{h}}/n_{SEC})$
$p_{hR} = 0.5(p_{hmin} + p_{hmax}) = 0.5 + 0.5(n_h - n_{\overline{h}})/n_{SEC}$
$fr_h = n_h/n = n_h/(n_h + n_{\overline{h}})$

Figure 18: Comparison of probabilities generated by the completeness (a) and the frequency (b) representation of probability in a special case when only evidence pieces that confirm the anti-hypothesis are at disposal ($n_{\overline{h}} \geq 0$) and no pieces that confirm the hypothesis ($n_h = 0$) exist.

Fig. 18a shows that because the number of hypothesis confirmations n_h is constant and equal to zero, the minimal probability p_{hmin} of this hypothesis is also constant and equal to zero ($p_{hmin} = 0$). At the beginning, at lack of confirmations or at small number $n_{\overline{h}}$ of the anti-hypothesis confirmations the maximal probability p_{hmax} of the hypothesis is high. However, along with the increase of the confirmation number $n_{\overline{h}}$ of anti-hypothesis, the possible maximal probability of the hypothesis decreases, which is logical, because confirmations of the anti-hypothesis decrease potential chances of the hypothesis. Instead, in case of the frequency interpretation presented in Fig. 18b, independently of how large is the number $n_{\overline{h}}$ of the anti-hypothesis confirmations is, which still decreases chances of the hypothesis, this interpretation suggests a constant and equal to zero value as the estimate fr_h of the hypothesis probability. This phenomenon can be interpreted as follows: "because I have no evidence pieces that confirm the hypothesis I fully exclude its truth fulness, also the minimal one".

7 Summary

The lecture presented a new interpretation of probability that does not possess a number of faults and weaknesses of the frequency interpretation. In particular, the completeness interpretation in a sensible and credible way models probability in certain critical, special situations such as at small number of evidence pieces and at a lack of pieces confirming the truth of one of the hypotheses. The frequency interpretation does not give credible values of probability in this situations. And a case of a small number of samples is the one that we meet often in real problems. Thus, the application range of the completeness interpretation can be much wider than the range of the frequency interpretation, which gives correct results only at a great number of evidence pieces. The completeness interpretation allows us to visualize that in many cases the true and precise value of probability cannot be learned and that we can only determine the lower and upper limit of probability. That represents the uncertainty and vagueness of probability. In this lecture the completeness interpretation was presented for the simplest, binomial hypothesis set, i.e. $H = \{h, NOT\,h\}$. Leaving this particular case, the completeness interpretation can be widened for the trinomial-, tetranomial-, pentanomial-, ..., multinomial case. The lecture presented the first investigations on the completeness interpretation of probability. This subject is quite new and anybody interested in problems of uncertainty is invited to investigate it. The author will publish his new results in next lectures and papers.

References

[1] Burdzy K. (2009) *The search for certainty. On the clash of science and philosophy of probability.* World Scientific, New Jersey.

[2] Carnap R. (1950) *Logical foundations of probability.* University Press, Chicago.

[3] Chernoff H. (1952) *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.* Annals of Mathematical Statistics, 23 (4), pp. 493-507.

[4] Dubois D., Prade H. (1988) *Possibility theory.* Plenum Press, New York and London.

[5] de Finetti B. (1975) *Theory of probability. A critical introductory treatment.* Wiley, London.

[6] http://www.math.washington.edu/burdzy

[7] Stanford Encyclopedia of Philosophy. Free access by http://plato.stanford.edu/entries/probability-interpret/

[8] http://en.wikipedia.org./wiki/Category:Probability_interpretations

[9] http://en.wikipedia.org/wiki/Chernoff_bound

[10] http://search4certainty.blogspot.com/

[11] Khrennikov A. (1999) *Interpretations of probability.* Brill Academic Pub., Utrecht, Boston.

[12] Klir G.J, Bo Yuan (editors) (1996) *Fuzzy sets, fuzzy logic, and fuzzy systems.* Selected papers by Lofti A. Zadeh. World Scientific, Singapore, New Jersey, London, Hong Kong.

[13] Laplace P.S. (1951) *A philosophical essay on probabilities.* English edition, New York, Dover Publication Inc.

[14] Larose D.T. (2010) *Discovering statistics.* W.H. Freeman and Company, New York.

[15] von Mises R. (1957) *Probability, statistics and the truth.* Dover, New York, Macmillan, second revised English edition.

[16] Piegat A. (2010) *Uncertainty of probability.* Plenary lecture on the Ninth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets, Institute of System Sciences of Polish Academy of Sciences, Warsaw, October 8, 2010, conference proceedings will be published in 2011.

[17] Popper K.R. (1957) *The propensity interpretation of the calculus of probability and the quantum theory.* In S. Koerner (editor), The Colston Papers, pp. 65-70.

[18] Rocchi P. (2003) *The structural theory of probability: new ideas from computer science on the ancient problem of probability interpretation.* Kluwer Academic/Plenum Publishers, New York.

[19] Shafer G. (1976) *A mathematical theory of evidence.* Princetown University Press, Princetown and London.

[20] Yakov B.H. (2006) *Info-gap decision theory.* Second edition. Elsevier, Oxford, Amsterdam.

[21] Zadeh L. A. (1965) *Fuzzy sets.* Information and Control, vol. 8, pp. 338-353.