Uncertainty of Probability

Andrzej Piegat
Faculty of Computer Science and Information Systems,
West Pomeranian University of Technology,
Żołnierska 49, 71-062 Szczecin
apiegat@wi.zut.edu.pl, http://apiegat.strony.wi.ps.pl

Abstract

Human intelligence is able to solve problems with high amount of uncertainty. Also artificial intelligence tries to solve similar problems. Towards realizing this aim it uses probability theory (PrTh), fuzzy set theory, possibility theory, Dempster-Shafer theory, info-gap theory, etc. PrTh as the oldest one (XVII century) seems to be a ripe and well grounded scientific method. However, according to many opinions, it is not true. In this paper the author shows that the basic and commonly used formula for calculation probability of an event A, $p(A) = n_A/n$, is both qualitatively and quantitatively rather incorrect. This formula was suggested by the frequency interpretation of probability. Furthermore, the author presents the evidential-completeness interpretation of probability that seems better suited to describe uncertainty. This interpretation explains why in most cases probability cannot be determined precisely and that only an uncertainty interval of probability can be found.

Keywords: Uncertainty, probability, probability interpretations, the completeness interpretation of probability, evidential completeness.

1 Introduction

Probability theory (PrTh) is the oldest and the most developed scientific method of investigating uncertainty. The first definition of probability was formulated by Laplace in 1814 [10]. His definition is referred to as the classical one. Since then multiple of books and papers on PrTh have been published. Therefore one could think that PrTh is a very strongly based science. However, it appears that the truth is quite different. Uncertainty and disagreement among scientist as to the sense o probability is very large. There is a great number of questions, doubts and paradoxes concerning understanding probability. Some scientists are even of the opinion that PrTh has met with a repulse. 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, see e.g. [5]. Some scientists gave whole hearted support to Professor Burdzys opinion; other criticized it, but rather in a moderate way. Defenders of the present PrTh mainly quote practical usefulness of PrTh in statistics. Views of Professor Burdzy are not at all individual ones. In literature some very strong opinions can also be found: "Probability does not exist" and "No matter how much information you have there is no scientific method to assign a probability to an event", de Finetti in [4]. Because of limited volume of this paper not all critical opinions can be quoted.

However, these can be easily found in [1], [5], and [6]. There are at least 5 main interpretations of probability that result from various understanding of probability [6]. These are presented below, with comments of Professor Burdzy.

1. The classical probability (Laplace [10], 1814]),
   "which claims that probability is symmetry".

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

3. The frequency theory (von Mises [12], 1957),
   "which claims that probability is long run frequency".

4. The subjective theory (de Finetti [4], 1975),
   "which claims that probability is personal opinion".

5. The propensity theory (Popper [13], 1957),
   "which claims that probability is physical property".

Particular interpretations reveal large qualitative differences in explanation of probability and try to remove weaknesses of other interpretations. There exists also an interesting opinion of Hajek [6]: "...there is still much work to be done regarding the interpretation of probability. Each interpretation that we have canvassed seems doing complete justice to it. Perhaps the full story about probability is something of a patchwork, with overlapping pieces. In that sense, the above interpretations might be regarded as complementary, ...” .

2 Classical and frequency interpretations 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”, [6]. “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 (1).

$$p(A) = \frac{N_A}{N} \tag{1}$$

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 their main representative von Mises [12]. “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 $P(A) = p$”, [6].

This interpretation is also called the long-run frequency interpretation (LRFr-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 (FFr-interpretation) according to which the probability is calculated on the basis of data we have at disposal. The definition of probability according to FFr-interpretation is as follows: "the probability of an attribute A in a finite reference class B is the relative frequency of actual occurrences of A within B", Hajek in [6]. Thus:

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

where: n – a finite number.

3 Main objections to classical and frequency interpretations of probability

The number of all objections and questions is very large [1, 6]. Only few of them 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 or multiple 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 whatsoever, lacks a probability for heads altogether; yet a coin that is never measured does not thereby lack a diameter", [6]. This problem can be called 'the zero-evidence problem' or 'the zero-case problem'.

4. "According to the frequency theory one can not 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', [6].

5. The 'small number of data pieces'-problem. In many real problems we have only small or very small number of data pieces. It strongly constraints applicability and credibility of the frequency interpretation of probability.

6. The fluctuation problem. Even if in some problems we can realize a large number of experiments, frequently no stable convergence of probability occurs and its fluctuations in a long-run calculations are observed, [1, 11].

4 Proposed 'completeness interpretation' of probability

In this chapter the completeness interpretation of probability will be presented that according to the authors knowledge is new and lacks certain faults of the frequency interpretation. The completeness interpretation was discussed on a scientific seminar in Faculty of Computer Science and Information Systems, where the author works.

Before all, the author believes that mathematical formulas (2) and (3) of the type $p_A = n_A/n$ for calculation of probabilities are generally incorrect. The fact, that these formulas, suggested by the frequency interpretation, are not able to give reasonable results for a single case problem or at small number of data pieces means that they are qualitatively incorrect (their mathematical form is incorrect). It seems that reason of this qualitative incorrectness is lack of some important element in the whole frequency-interpretation concept. According to the author, the lacking element is 'evidential completeness'. Its meaning will be explained below. In Polish probability is called 'prawdopodobieństwo', which means 'similarity to the truth'. In Latin also : 'verisimilitudo 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} , $i = 1, \ldots, k$, that confirm the truth of h we should have an image of what would be the complete set EC (EC=$\{e_{h1}, \ldots, e_{hk}\}$) of such evidence pieces, 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).

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

It can be easily 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 $h_1^* = h$ 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, $\overline{h} = NOT$ head domination. Also continuous random variable can be treated this way after their discretization. Therefore the completeness interpretation of probability will now be explained on an example of the binomial hypothesis-pattern $\{h, NOT\ h\}$ as the basic one. In authors opinion, probability should not be assigned to (future or past) events but rather to hypotheses concerning the events. Thus, before coin tossing we can formulate the hypothesis h (head domination) and the anti-hypothesis $NOT\ h = \overline{h}$ (tail domination) and assign their probabilities. After coin tossing, we have to do with realization $r =$ (head) or $r =$ (tail). However, their probabilities were assigned not by us but by the experiment. Probabilities 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 $\overline{h}$ (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, etc. An important question arises: can we infer and assign probabilities to hypotheses on the basis of any number n of evidence pieces? If yes, then how accurate these probabilities will be? In that case, does accuracy depend on the number n of evidence pieces or not? These questions should be answered. As proposed, the set of evidence pieces that fully proves the hypothesis h and makes it certain ($p_h = 1$ and $p_{\overline{h}} = 0$) has been called evidential completeness and denoted by EC. In certain problems an ideal EC can be determined by experts. As an example we can use a crime, e.g. a murder. Let person A be suspect of murder (SP-suspected person). The evidential completeness EC can be 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 of the murder, on the knife which was the murder tool experts found some genetic matter of SP} = $\{e_{EC1}, \ldots, e_{EC4}\}$

If we have such evidence against the person A we can be sure of the hypothesis h_A (A is the murderer). However, if against A we only have an evidential set E_A 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 this hypothesis. Then it is only probable. Its approximate probability can be assigned by criminal experts as conformability degree of the evidence set E_A with the completeness EC. The above example shows that probability is (perhaps not always) of a mixed, objective-subjective character. Evidence pieces are most often objective. However, their meaning and weighing, their evaluation and aggregation to determine the probability value must be done by experts. In case of coin tossing each trial delivers one piece of evidence. Significance of each evidence piece has to be evaluated too. As long as we have no reason for different weighing of particular tosses they will be assigned the same weight. However, what will be evidential completeness EC in the case of coin tossing? The ideal EC would consist of an infinitely large number n of tossing results. But, it is impossible to realize such number of trials. With similar situation we may have to deal also in other cases. Sometimes no ideal evidential completeness is possible. Therefore, to solve some real problems we will have to use 'satisfactory evidential completeness', (SEC). It is such set of 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 to 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 trial results that is sufficient for probability determining with a satisfactory accuracy. To find this number we can use the so called Chernoff bound [3, 8]. However, other mathematical tools, if suited, can also be used. Chernoff derived a formula that is given below with a slightly changed denotation (4):

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

where: ϵ represents the maximal error of the result, $\epsilon = 1 -$ accuracy. E.g.: $\epsilon = 0.01$ means accuracy $= 0.99$. p_{hc} means an assumed, higher limit of the excluded probability of head, e.g.: if we suspect on the basis of introductory experiments with the coin that its asymmetry is so large, that the head probability $p_h \notin [0.4, 0.6]$, then we assume $p_{hc} = 0.6$. The smaller the coin asymmetry (bias) the more trials are necessary for accurate determining the p_h-probability. If this probability lies outside the interval $[0.45, 0.55]$ then at the minimal accuracy 0.99, Chernoff bound (4) delivers the number of required trials $n_{SEC} = 921$. If the head probability is outside the interval $[0.499, 0.501]$, then $n_{SEC} = 2\,302\,586$ trials. In case of a continuous random variable, the asymmetry of probability exists between e.g. two parts in which the probability density function can be partitioned. Therefore, also here the binomial pattern hypothesis, anti-hypothesis as coin tossing can be applied. Further on, the following definition of probability will be proposed.

> The minimal probability p_{hmin} of the hypothesis h concerning a given event is the conformability degree of the evidence collected in 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 contained in the set EC_h of evidential completeness of the h-hypothesis.

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

$$p_{hmax} = 1 - p_{NOThmin}$$

The exact probability p_h of the hypothesis h can be determined only if the following condition is satisfied:

$$\text{IF } (p_{hmin} + p_{NOThmin} = 1) \text{ THEN } (p_{hmin} = p_h)\,.$$

If the above condition is not satisfied then the exact probability p_h can not be determined.

However, if collecting the evidence required by the ideal evidential completeness would for a given problem be impossible a satisfactory evidential completeness can be used that allows to prove the hypothesis not fully but to a certain, satisfactory degree of accuracy.

5 Uncertainty of probability

Let us now come back to the binomial problem with the hypothesis set $H = \{h, NOT\ h\} = \{h, \overline{h}\}$ which is well represented by coin tossing. Let us assume, that we know from introductory experiments that the coin is biased so, that assumption can be made concerning the excluded p_h-interval: $p_h \notin [0.45, 0.55]$. It means $p_{hc} = 0.55$ and with Chernoff bound (4) the number $n_{SEC} = 921$ trials required by SEC for calculation accuracy 0.99 has been determined. Let us further on assume that only 5 coin tosses were realized ($n = 5$): 3 tosses gave heads ($n_h = 3$) and 2 tosses gave tails ($n_{\overline{h}} = 2$). The number $n = 5$ of evidence pieces is considerably smaller than the number $n_{SEC} = 921$ of trials required by SEC. The situation is illustrated in Fig. 1.

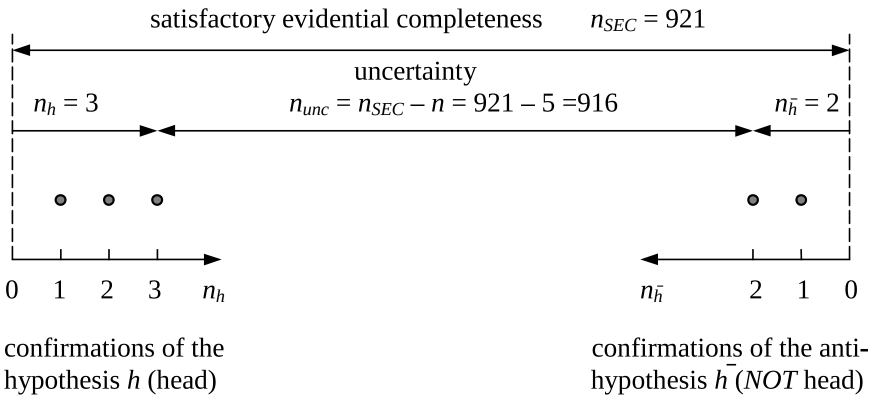

Figure 1: Illustration of the uncertainty existing in determining probability p_h of the hypothesis h from the binomial hypothesis set $H = \{h, \overline{h}\}$ in the example of coin tossing.

As Fig. 1 shows, probability p_h cant be precisely determined because 5 trails are not enough and further 916 trials are necessary for sufficiently precise determination. In an extreme case, all these 916 trials can give heads. Then the hypothesis h (head domination) would have 919 confirmations ($n_h = 919$). It is also possible that all lacking 916 trials will give tails. Then the anti-hypothesis $\overline{h}$ would have 918 confirmations ($n_{\overline{h}} = 918$). Because we dont know what next 916 trials will give we can understand them as uncertainty of probability. However, because we have 3 confirmations for head, we can be sure that the head probability p_h will be not less than $n_h/n_{SEC} = 3/921$ and not higher than $(1 - n_h/n_{SEC}) = (1 - 2/921) = 919/921$.

$$p_h \in \left[\frac{n_h}{n_{SEC}}, 1 - \frac{n_{\overline{h}}}{n_{SEC}}\right] = \left[\frac{3}{921}, \frac{919}{921}\right] = [0.0033, 0.9978] \qquad (5)$$

$$p_{\overline{h}} = 1 - p_h \qquad p_{\overline{h}} \in \left[\frac{2}{921}, \frac{918}{921}\right]$$

The result given in (5) is the one and only knowledge about the head probability after 5 trials. The value of this probability is still unknown. It can be any value lying between the borders $p_{hmin} = 3/921$ and $p_{hmax} = 919/921$. However, these 5 trials decreased the uncertainty from the initial one $921/921 = 1$ to $916/921$. Thus, the trials were useful. Let us see now, what result we will get if we apply the frequency interpretation of probability, formula (6).

$$p_h = \frac{n_h}{n} = \frac{3}{5} \tag{6}$$

The result delivered by the FFr-interpretation and by the completeness interpretation are shown in Fig. 2.

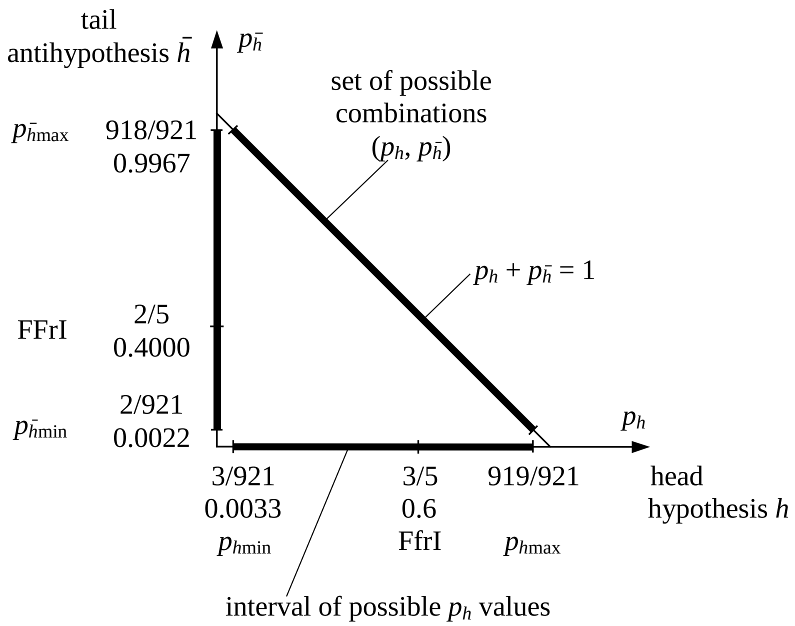

Figure 2: Illustration of uncertainty of probability p_h of the hypothesis h (head domination) after only 5 trials. The result $p_h = 3/5$ is delivered by the FFr-interpretation of probability.

The example in Fig. 2 shows that the estimation of probability delivered by the commonly used formula (6) suggested by the FFr-interpretation is slightly doubtful. Is there a reason to assume that this result $p_h = 3/5 = 0.6$

is more credible than any other value from interval $[3/921, 919/921]$, e.g. $1/5$ or $2/5$ or $4/5$ etc. Now, let us assume, that we have some evidence in form of $n = 700$ trials from which 399 trials gave heads ($n_h = 399$) and 301 gave tails ($n_{\overline{h}} = 301$). These results mean, that the lowest limit of probability of head equals $p_{hmin} = n_h/n_{SEC} = 399/921 = 0.4332$ and the highest limit $p_{hmax} = 1 - n_h/n_{SEC} = 620/921$. The FFr-representation gives a 'precise' result $p_h = n_h/n = 399/700 = 0.5700$. Results of both interpretations are shown in Fig. 3.

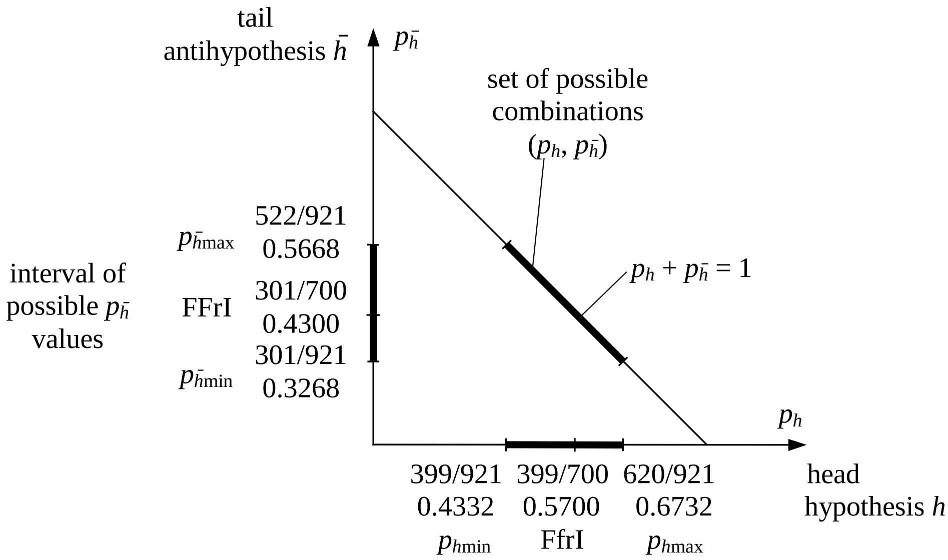

Figure 3: Illustration of uncertainty of probability p_h of the hypothesis h (head domination) after 700 trials. The result $p_h = 399/700$ was delivered by the FFr-interpretation of probability.

Let us assume now that the full complete set of $n = 921$ trials required by the satisfactory evidential completeness was realized with 531 heads ($n_h = 531$) and 390 tails ($n_{\overline{h}} = 390$). Thus $n_h + n_{\overline{h}} = 531 + 390 = 921 = n = n_{SEC}$. According to the completeness interpretation we get the lower probability constraint $p_{hmin} = n_h/n_{SEC} = 531/921 = 0.5765$ and the higher constraint $p_{hmax} = 1 - n_{\overline{h}}/n_{SEC} = 531/921 = 0.5765$. Thus $p_{hmin} = p_{hmax} = p_h$. Uncertainty of the probability decreased to the minimum in terms of Chernoff bound (the possible error $\epsilon = 0.01$). If we apply the FFr-interpretation, formula (3) we get $p_h = n_h/n = 531/921 =$

0.5765. In this case results delivered by both the completeness- and by the FFr-interpretation are the same. They are shown in Fig. 4.

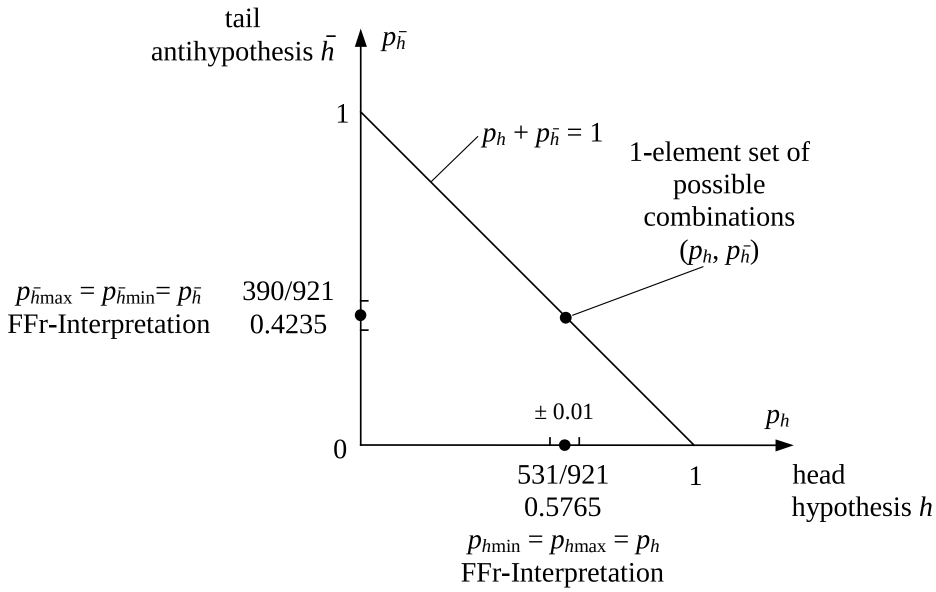

Figure 4: Illustration of results of determining the p_h-probability at the number n of trials equal to the number $n_{SEC} = 921$ required by the satisfactory evidential completeness. The possible error of $p_h = 0.5765$, $\epsilon = 0.01$.

On the basis of the completeness interpretation the following first conclusion can be formulated: If we have n_h confirmations of the hypothesis h and $n_{\overline{h}}$ confirmations of the anti-hypothesis $\overline{h}$ and $n_h + n_{\overline{h}} < n_{SEC}$ then the probability p_h lies in interval (7).

$$p_h = \left[\frac{n_h}{n_{SEC}}, 1 - \frac{n_{\overline{h}}}{n_{SEC}}\right] \tag{7}$$

In real decision-making problems often the simplified singleton-representation is necessary (one, single number is easily understandable for non-specialists). Therefore a question arises: "Which probability value from interval (7) could in the best way fulfill this task?". To answer this question an optimality criterion has to be chosen. One of possible criteria is given by (8). 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 by p_{hR} the best representation of the probability interval (7) among all possible representations p^*_{hR} contained in this interval. The best representation is determined by criterion (8).

$$p_{hR} = \min_{p^*_{hR} \in [p_{hmin}, p_{hmax}]} \left[\max\left(|p^*_{hR} - p_{hmin}|, |p_{hmax} - p^*_{hR}|\right)\right] \qquad (8)$$

This optimal value is simply the mean of the constraints p_{hmin} and p_{hmax} (9).

$$p_{hR} = 0.5(p_{hmin} + p_{hmax}) \qquad (9)$$

A very important remark: the optimal representation p_{hR} given by (9) **is not a statement** that it is the true and precise value of probability p_h, because this value is not known. We only know that it lies in the interval $[p_{hmin}, p_{hmax}]$. The optimal representation is only a single value chosen from this interval as our best guess, chosen to help us in decision-making. Use of this representation prevents very large and maximal errors of problem solutions. An interesting question is how do the lower and higher limits and the optimal representation change with increasing number n of trials, at $n \leq n_{SEC}$. Because of the limitations of the paper volume a table with detailed results can't be shown here. However, the results of investigation are shown in Fig. 5.

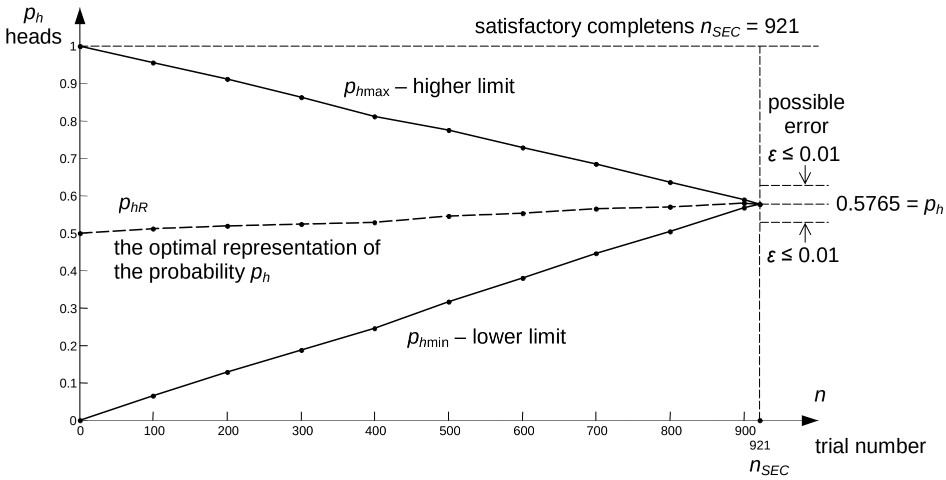

Figure 5: Illustration of the process of decreasing uncertainty of probability with increasing number n of trials.

As Fig. 5 shows, at small number n of trials (small amount of evidence) the uncertainty $(p_{hmax} - p_{hmin})$ of probability p_h is very large but with increasing number of trials it gradually decreases to the minimum (to the possible error in terms of Chernoff bound $\epsilon = 0.01$) at $n = n_{SEC}$. The optimal representation p_{hR} of the uncertainty interval slowly and gradually, without oscillations converges into the highly precise (in the sense of Chernoff bound accuracy 0.99)) value p_h of probability. Fig. 5 presented results of investigations for $n \in \{0, \ldots, 921\}$. An investigation was also made for smaller numbers of trials $n \in \{0, \ldots, 10\}$. Table with detailed results cant be presented here because of the paper volume limitation, but the results achieved are shown in Fig. 6.

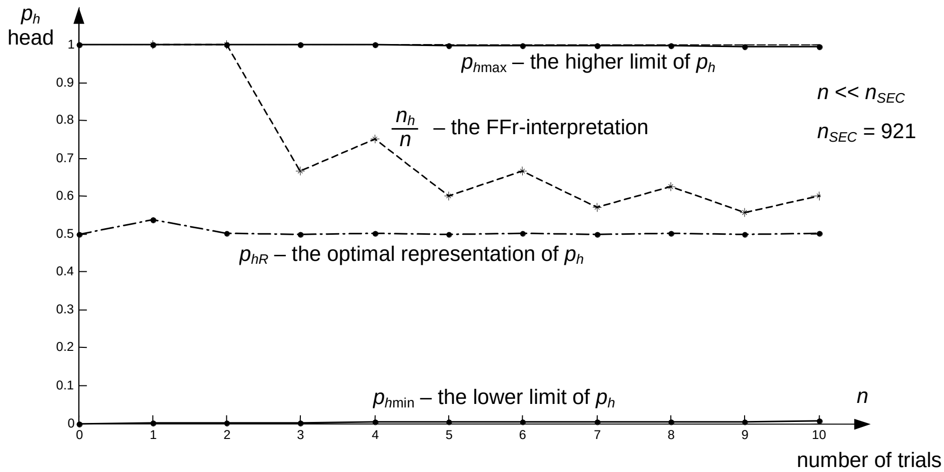

Figure 6: Results of determining the head probability p_h for number of trials $n \in \{0, \ldots, 10\}$.

6 Conclusions

In situation of limited number of data pieces it is not possible to precisely determine probability. It is uncertain and the uncertainty can be very large. The completeness interpretation allows for determining this uncertainty. As Fig. 6 shows, the values n_h/n calculated by the FFr-interpretation are some kind of representation of the uncertainty interval $[p_{hmin}, p_{hmax}]$ of probability. But such representation seems not optimal and has many faults. Its **first fault** is lack of optimality criterion. We don't know in which sense the FFr- representation would be optimal. The **second fault** is that it creates considerable variations and oscillations of the n_h/n-value with increasing number of evidence pieces (see Fig. 6). Single pieces strongly change 'the opinion' of the FFr-interpretation about probability estimation. Thus, this representation resembles an undecided and hesitant person. From this point of view the completeness interpretation looks better. The optimal representation p_{hR} is not 'hesitant'. It is stable. It gradually changes its value after adding next pieces of evidence because this representation 'knows' that weight of a single piece of evidence is not large (1/921 in the example) and that this weight depends proportionally on the number n_{SEC} of evidence pieces required by SEC. The **third fault** of the FFr-representation is that it is not able to account for the zero-evidence case ($n = 0$) whereas the completeness interpretation is (see Fig. 6 and formula (9)). The **fourth fault** of the n_h/n representation is that it calculates incredible values (0 or 1) for the single case (only 1 piece of evidence, $n = 1$). The **fifth fault** of this representation and next faults can be found e.g. in [1]. All the above faults of the n_h/n representation of probability do not mean that this representation is completely incorrect. It is strongly incorrect at small numbers of evidence pieces $n \ll n_{SEC}$. If the number n approaches the number n_{SEC} required by satisfactory completeness, then its accuracy improves.

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

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

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

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

[8] http//en.wikipedia.org/wiki/Chernoff_bound

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

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

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

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

[13] Popper K.R. (1957) The propensity interpretation of the calculus of probability and the quantum theory. In S. Koerner (ed), The Colston Papers, 9, pp. 65-70.