Journal of Theoretical and Applied Computer Science
ISSN 2299-2634 (printed), 2300-5653 (online)

Vol. 7, No. 3, 2013, pp. 3–20
http://www.jtacs.org

Mean square error optimal completeness estimator Ep_{h2} of probability

Andrzej Piegat[1], Marek Landowski[2]

[1] Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Szczecin, Poland
[2] Institute of Quantitative Methods, Maritime University of Szczecin, Poland

apiegat@wi.zut.edu.pl, m.landowski@am.szczecin.pl

Abstract: *The paper presents the optimal estimator of probability for the binomial and multinomial case that was called "completeness estimator Ep_{h2}" and theoretical proof of its optimality. The estimator accuracy was compared with accuracy of the universally used frequency estimator. The comparison was realized both theoretically and experimentally. Both comparison ways show superiority of the completeness estimator Ep_{h2} over the frequency estimator $fr_h = n_h/n$. A prooved solution of the single case problem is given.*

Keywords: *probability, probability estimation, completeness interpretation of probability, probability interpretations, single-case problem*

1. Introduction

The uncertainty of future events, of unknown states of matter etc. is a subject that has fascinated scientists and not scientists for centuries. There are a few approaches to uncertainty description: probability theory [1, 4, 7, 8, 9, 11, 12, 16, 17], fuzzy system theory [10, 20], belief and plausibility theory [18], possibility theory [6], info-gap theory [19], and others. Probability theory is the oldest and frequently used in practice. It is lectured in all technical and mathematical oriented universities. There are at least 5 of probability interpretations described in [1, 4, 7, 8, 9, 11, 13, 14, 15, 16, 17], but the best known ones are the classic and the frequency definitions. The classic interpretation has been presented [8] as follows: If a random experiment can result in N mutually exclusive and equally likely outcomes and if N_A of the outcomes results in the occurrence of an event A, the probability of event A is defined by (1).

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

This interpretation was proposed by Laplace [11] in 1814. Its application is limited only to problems with a finite number of possible outcomes, e.g. dice experiment. It cannot be applied to problems with an infinite number of possible outcomes as e.g. unfair coin tossing. Therefore "frequentists" with their main representative von Mises [13] proposed the frequency interpretation of probability. According to this interpretation [8] "probability of an event is its relative frequency of occurrence after repeating a process a large number of times under similar conditions. Let us 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 $P(A) = p$", [8].

This interpretation was called the long-run frequency interpretation. However, in real life an infinite number of experiments cannot be realized. Thus, we have to use the finite frequency interpretation. Its sense is as follows [8]: "the probability of an attribute A in a finite reference class B is the relative frequency of actual occurrence of A within B". The probability estimate is calculated by (3),

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

where n is a finite number. However, the frequency interpretation also has many weak points described e.g. in [1, 3, 8, 14, 15]. They were the reason for scientists to create other interpretations of probability as the subjective interpretation of de Finetti [7], the logical interpretation of Carnap [4], the propensity interpretation of Popper [16], the interpretation of Khrennikov [9], and Rocchi [17]. These new interpretations are also discussed. Professor K. Burdzy in his book "Search for certainty – On the clash of philosophy of probability" [1, 2] evaluated the present situation of probability very critically. The discussion about the probability state caused one of the authors, A. Piegat, to elaborate and propose the completeness interpretation of probability that was described in [14, 15]. According to this interpretation, to determine probability of a hypothesis h concerning an event, first, the complete evidential set *evidential completeness (EC)* should be determined. It contains such a set of evidence pieces which would fully prove the truth / validity of hypothesis. In practice we often possess only a partial and incomplete evidential set and can determine only the minimal, lower limit of probability $p_{h\min}$ of the hypothesis h and the minimal probability $p_{\bar{h}\min}$ of the anti-hypothesis $\bar{h} = NOTh$. It allows for the calculation of the upper limits of the probability (4).

$$\begin{aligned} p_{h\max} &= 1 - p_{\bar{h}\min} \\ p_{\bar{h}\max} &= 1 - p_{h\min} \end{aligned} \tag{4}$$

In problems where we do not have the full evidence set but only a part of it we are not able to precisely determine the probability p_h of the hypothesis. Only an interval of the probability can be determined (5).

$$p_h \in [p_{h\min}, p_{h\max}] \tag{5}$$

We are also interested in the estimate Ep_h of the probability p_h. In [14] and [15] the first and simple completeness estimator p_{hR} was proposed. This estimator represents the uncertainty interval (5). It minimizes the maximal possible, absolute error of the estimate and is expressed in (6).

$$p_{hR} = 0.5(p_{h\min} + p_{h\max}) \tag{6}$$

In case of the binomial event as the occurrence of a certain event or not, n_h means a number of confirmations of the hypothesis h concerning the event and $n_{\bar{h}}$ a number of confirmations of the anti-hypothesis, the estimator formula takes the form in (7).

$$p_{hR} = \frac{1}{2} + \frac{n_h - n_{\bar{h}}}{2n_{SEC}} \tag{7}$$

Denotation n_{SEC} means a number of evidence pieces necessary for a satisfactorily precise (the required precision can be e.g. 99%) proof of the hypothesis [14, 15]. Though estimates

Ep_h calculated by (7) correctly converge with an increasing number n of evidence pieces to the precise value of probability, the convergence speed is small, because the estimator is not an informed one. Based on theoretical analysis the idea of a new completeness estimator was found which will be demonstrated in the next section.

2. The idea of the completeness estimator Ep_{ha} of probability

The universally used frequency estimator $fr_h = n_h/n$ has many disadvantages that are described e.g. in [1, 14, 15]. It has great estimation errors for a small number n of sample items with whicfh we often have to do in practical problems. It also gives difficult acceptable results in case of a single sample item, where it suggests probability values 0 or 1 that means certainty. This phenomenon was called "single case problem", [1, 8]. In case, when we have at disposal only homogeneous data as e.g. $\{H, H, H, H, H\}$, where H means confirmation of a hypothesis, the frequency estimator suggests the hypothesis probability $p_h = 1$, which means full certainty and is not acceptable. The frequency estimator produces oscillating probability values also at large number of sample items when estimation should stabilize [1, 12, 14, 15]. The proposed completeness estimator Ep_{ha} (8) is free of at least a part of these faults.

$$Ep_{ha} = \frac{1}{2} + \frac{n_h - n_{\bar{h}}}{2(n+a)} \tag{8}$$

In formula (8) n_h means a number of confirmations of the hypothesis h and $n_{\bar{h}}$ a number of confirmations of the anti-hypothesis $\bar{h} = NOTh$, and n means the entire number of sample items ($n = n_h + n_{\bar{h}}$). Three examples of binomial problems are given below:
h - smoking increases the cancer danger, $\bar{h}$ - smoking does not increase the cancer danger,

h - Justyna Kowalczyk will win the competition, $\bar{h}$ - Justyna Kowalczyk will not win the competition,

h - today the driving time from A to B will be longer than 23 minutes, $\bar{h}$ - today the driving time from A to B will be shorter than 23 minutes.

It can be easily proved that the completeness estimator has some features presented below.

1. Estimates of probability calculated by the completeness estimator Ep_{ha} converge for large number of sample items $n \to \infty$ to the true and precise probability value p_h defined by formula (2). The proof is given below.

$$\begin{aligned}\lim_{n\to\infty} Ep_{ha} &= \lim_{n\to\infty}\left(\tfrac{1}{2} + \tfrac{n_h - n_{\bar{h}}}{2(n+a)}\right) = \lim_{n\to\infty} \tfrac{n+a+n_h-n_{\bar{h}}}{2(n+a)} \\ &= \lim_{n\to\infty} \tfrac{n_h+n_{\bar{h}}+a+n_h-n_{\bar{h}}}{2(n+a)} = \lim_{n\to\infty} \tfrac{2n_h+a}{2(n+a)} \\ &= \lim_{n\to\infty} \tfrac{2n_h}{2(n+a)} + \lim_{n\to\infty} \tfrac{a}{2(n+a)} = \lim_{n\to\infty} \tfrac{n_h}{n+a} \\ &= \lim_{n\to\infty} \tfrac{n_h/n}{1+a/n} = \lim_{n\to\infty} \tfrac{n_h}{n} = p_h\end{aligned} \tag{9}$$

2. The probability estimate $Ep_{ha}(1)$ calculated by the completeness estimator Ep_{ha} from one single sample item that confirms hypothesis h is not equal to 1 as in case of the frequency estimator ($fr_h(1_h) = 1$), but is less or equal to 1 (10) depending on the assumed value of "a".

$$Ep_{ha}(1_h) = \frac{1}{2}\left(\frac{2+a}{1+a}\right) \tag{10}$$

For $a > 0$ the estimate $Ep_{ha}(1_h)$ satisfies the condition $0.5 < Ep_{ha}(1_h) \leq 1$. If e.g. $a = 1$ then $Ep_{ha}(1_h) = 0.75$. The frequency estimator fr_h realizes from one confirming sample item 1_h an infering that can be called not cautious (it infers the probability $p_h = 1$). Instead, the infering realized by the completeness estimator can be called cautious. If the single sample item is a negation ($1_{\bar{h}}$) of the hypothesis h then the estimate Ep_{ha} is given by (11).

$$Ep_{ha}(1_{\bar{h}}) = \frac{1}{2}\left(\frac{a}{1+a}\right) \tag{11}$$

If $a > 0$ then $0 \leq Ep_{ha}(1_{\bar{h}}) < 0.5$. If $a = 1$ then the estimate value is given by (12).

$$Ep_{ha}(1_{\bar{h}}) = 0.25 \tag{12}$$

And in case of the frequency estimator $fr_h(1_{\bar{h}})$ the estimate from one sample item negating the hypothesis equals 0 (13).

$$fr_h(1_{\bar{h}}) = \frac{n_h}{n} = \frac{0}{1} = 0 \tag{13}$$

The above means that the frequency estimator realizes drastic infering from one sample item, because from only one and single negation it concludes zero-probability of the hypothesis. Instead, the infering made by the completeness cannot be called drastic but rather moderate and cautious one. The infering caution can be controlled by the coefficient ”a”. The value $a = 0$ means ”no caution” or ”maximal radicalism” of infering. Increasing a increases the cautious infering that becomes maximal for $a \to \infty$.

3. The estimate value $Ep_{ha}(0)$ calculated by the completeness estimator at lack of sample items.
This value is given by (14), $n = n_h = n_{\bar{h}} = 0$.

$$Ep_{ha}(0) = \frac{1}{2} + \frac{n_h - n_{\bar{h}}}{2(n+a)} = \frac{1}{2} \tag{14}$$

This estimate value is reasonable because it minimizes the maximal possible absolute error of the estimate to 0.5. Any other estimate value different from 0.5 would increase the maximal possible error over 0.5. For comparison, the frequency estimator $fr_h = n_h/n$ is not able to infer the hypothesis probability at lack of sample items ($n = 0$).

3. Derivation of the optimal value of the coefficient a representing the cautious inferring

Increasing the value of the coefficient a in formula (8) of the completeness estimator Ep_{ha} increases the caution of probability inferring from one sample item. Let us remember that the hypothesis probability inferred by the frequency estimator from one sample item 1_h confirming the hypothesis h equals 1 ($fr_h = n_h/n = 1/1 = 1$). Such a conclusion can be characterized as not cautious and maximally exaggerated one. Let us try another estimator which would be less risky and which would conclude from one confirming sample item an estimate $Ep_h(1_h)$ less than one ($Ep_h(1_h) \leq 1$) and from one sample item negating the hypothesis an estimate $Ep_h(1\bar{h}) \geq 0$. The both estimates should satisfy the condition (15).

$$Ep_h(1_h) + Ep_h(1_{\bar{h}}) = 1 \tag{15}$$

The condition (15) is satisfied not only by the completeness estimator but also by the frequency one. Let us assume now that the true probability of the hypothesis h equals p_h and the conclusion concerning this probability, calculated by the completeness estimator on the basis of only one hypothesis confirmation (1_h) equals $Ep_h(1_h)$. In the general case $Ep_h(1_h) \neq p_h$. Formula (16) determines the square error of the completeness estimate.

$$\Delta^{sqr}(1_h) = [p_h - Ep_h(1_h)]^2 \tag{16}$$

If an evidence piece negates the hypothesis (denotation - $1_{\bar{h}}$) then the square error of the complete estimate is given by (17).

$$\Delta^{sqr}(1_{\bar{h}}) = [p_h - Ep_h(1_{\bar{h}})]^2 = [p_h - [1 - Ep_h(1_h)]]^2 \tag{17}$$

If we have at disposal N sample items and the number N approaches infinity then the number N_h of sample items confirming the hypothesis h equals $N \cdot p_h$ and the number $N_{\bar{h}}$ of sample items negating the hypothesis equals $N \cdot (1 - p_h)$. The sum of squared errors $\Delta^{sqr}(N_h)$ of individual conclusions from all sample items 1_h confirming the hypothesis is given by formula (18).

$$\Delta^{sqr}(N_h) = [p_h - Ep_h(1_h)]^2 \cdot N_h = [p_h - Ep_h(1_h)]^2 \cdot p_h N \tag{18}$$

The sum of squared errors $\Delta^{sqr}(N_{\bar{h}})$ of all sample items $1_{\bar{h}}$ negating the hypothesis is given by formula (19).

$$\Delta^{sqr}(N_{\bar{h}}) = [p_h - [1 - Ep_h(1_h)]]^2 \cdot N_{\bar{h}} = [p_h - [1 - Ep_h(1_h)]]^2 \cdot (1 - p_h)N \tag{19}$$

The global sum of squared errors $\Delta^{sqr}(N)$ of all single estimates, both from the confirming and from negating sample items is determined by formula (20).

$$\begin{aligned} \Delta^{sqr}(N) &= \Delta^{sqr}(N_h) + \Delta^{sqr}(N_{\bar{h}}) = \\ &= [p_h - Ep_h(1_h)]^2 \cdot p_h N + [p_h - [1 - Ep_h(1_h)]]^2 \cdot (1 - p_h)N \end{aligned} \tag{20}$$

If the global sum of squared errors $\Delta^{sqr}(N)$ will be divided by N then the average error $\Delta^{sqr}_{aver}(1)$ will be achieved. It is the average error of probability estimation independently of whether the sample item confirms (1_h) or negates $(1_{\bar{h}})$ the hypothesis, formula (21).

$$\begin{aligned} \Delta^{sqr}_{aver}(1) &= \Delta^{sqr}_{aver}(1_h) + \Delta^{sqr}_{aver}(1_{\bar{h}}) = \\ &= [p_h - Ep_h(1_h)]^2 \cdot p_h + [p_h - [1 - Ep_h(1_h)]]^2 \cdot (1 - p_h) \end{aligned} \tag{21}$$

Fig. 1 presents the functional surface ($y = f(x_1, x_2)$) of the first component $\Delta^{sqr}_{aver}(1_h)$ of the average error (21) of one sample item (1) and Fig. 2 presents the surface of the second component $\Delta^{sqr}_{aver}(1_{\bar{h}})$. Fig. 3 demonstrates the full functional surface of the error $\Delta^{sqr}_{aver}(1)$.

Fig. 3 allows for interesting analysis. The true value p_h of the probability is not known. It is to be estimated. But we can choose such value $Ep_h(1_h)$ of the probability estimate from one confirming sample item (1_h) that will be optimal in the sense of the mostly used criterion of the sum of squared errors. The optimal estimate value $Ep_h^{opt}(1_h)$ will minimize the risk of committing large errors of the probability estimation. Fig. 4 presents the section of the functional surface of the square error $\Delta^{sqr}_{aver}(1)$ for the estimate value $Ep_h(1_h) = 1$ of the probability resulting from one confirming sample item. This estimate value corresponds to the conclusion made from one sample item 1_h by the frequency estimator fr_h.

The visual analysis of Fig. 3 allows for the statement that there are other values of the estimate $Ep_h(1_h)$ which generate smaller values of the square error than $Ep_h(1_h) = 1$. It

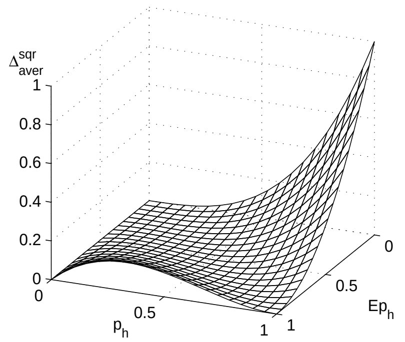


Figure 1. The functional surface of the first component $\Delta^{sqr}_{aver}(1_h) = [p_h - Ep_h(1_h)]^2 \cdot p_h$ of the average error $\Delta^{sqr}_{aver}(1)$ caused by sample items confirming the hypothesis h.

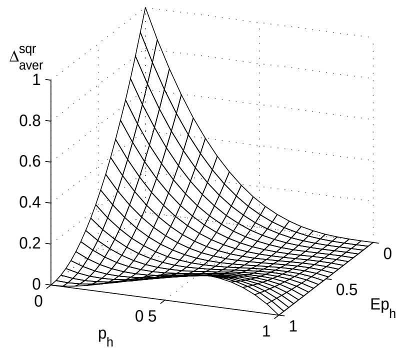


Figure 2. The functional surface of the second component $\Delta^{sqr}_{aver}(1_{\bar{h}}) = [p_h - [1 - Ep_h(1_h)]]^2 \cdot (1 - p_h)$ of the average error $\Delta^{sqr}_{aver}(1)$ caused by sample items negating the hypothesis h.

means that assigning very strong confirmation equal to 1 by the universally used frequency estimator $fr_h = n_h/n$ to the single sample item 1_h is not the best idea. In the first step of investigation the optimal value of the one sample item estimate $Ep_h(1_h)$ that minimizes the cross section area A of the one sample item square error function $\Delta^{sqr}_{aver}(1)$ will be determined. This function is expressed by formula (22).

$$\Delta^{sqr}_{aver}(1) = [p_h - Ep_h(1_h)]^2 \cdot p_h + [p_h - [1 - Ep_h(1_h)]]^2 \cdot (1 - p_h) \tag{22}$$

The square error area A that should be minimized is expressed by (23).

$$A = \int_0^1 \Delta^{sqr}_{aver}(1) dp_h \tag{23}$$

After integrating (23) the formula (24) for the error-area A is achieved.

$$A = Ep_h^2(1_h) - \frac{4}{3}Ep_h(1_h) + \frac{1}{2} \tag{24}$$

Derivative of A equated to 0 is given by (25).

$$\frac{dA}{dEp_h(1_h)} = 2Ep_h(1_h) - \frac{4}{3} = 0 \tag{25}$$

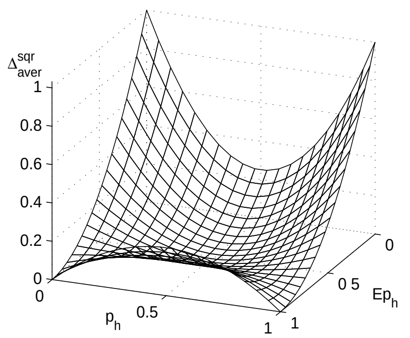


Figure 3. The functional surface $\Delta^{sqr}_{aver}(1) = \Delta^{sqr}_{aver}(1_h) + \Delta^{sqr}_{aver}(1_{\bar{h}})$ of the average error $\Delta^{sqr}_{aver}(1)$ caused by both sample items confirming and negating the hypothesis h, formula (21).

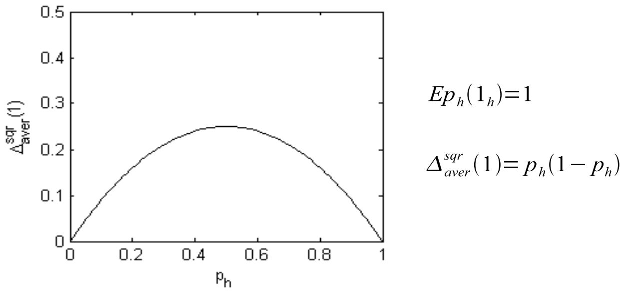


Figure 4. Cross section of the average single sample item, square error $\Delta^{sqr}_{aver}(1)$ from Fig. 3 for the probability estimate $Ep_h(1_h) = 1$ that corresponds to the estimate calculated by the frequency estimator $fr_h = n_h/n$.

Solution of equation (25) delivers the optimal value of the one sample item estimate $Ep^{opt}_h(1_h) = 2/3$. After inserting this value in formula (24) the minimal value of the square error area $A = 1/18$ is achieved. Fig. 5 presents the minimal cross-section with this area.

For comparison Fig. 6 presents the cross-section of the error function for the value $Ep_h(1_h) = 1$ corresponding to the frequency estimator fr_h ($fr_n = n_h/n = 1/1 = 1$).

As the comparison of Fig. 5 and Fig. 6 shows, the error area $A = 1/6$ of the frequency estimator fr_h is 3 times larger than the error area of the optimal completeness estimator with $A = 1/18$. Now, on the basis of the optimal, confirming sample item estimate $Ep_h(1_h)$ the corresponding value of the caution coefficient a can be determined. The general formula of the completeness estimator is given by (8).

As determined by solving equation (25) the optimal value of conclusion from one sample item (1_h) confirming the hypothesis h is $Ep_{ha}(1_h) = 2/3$. The one, alone confirming sample item 1_h is corresponded by $n_h = 1$, $n_{\bar{h}} = 0$, and $n = 1$. Taking this into account and inserting these values in (8), equation (26) is achieved.

$$Ep_{ha}(1_h) = \frac{1}{2} + \frac{1}{2(1+a)} = \frac{1}{2}\left(\frac{2+a}{1+a}\right) = 2/3 \tag{26}$$

Solving equation (26) gives the optimal value of the caution coefficient $a^{opt} = 2$. Thus, the formula of the optimal completeness estimator that further on will be denoted as Ep_{h2} has the form of (27).

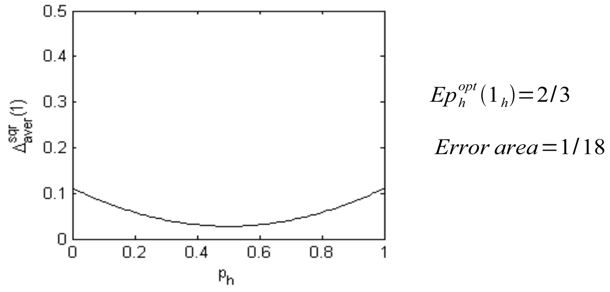


Figure 5. The minimal cross-section of the function $\Delta^{sqr}_{aver}(1) = f(p_h, Ep_h(1_h))$ of the average, square, one sample item error for the optimal one sample item estimate value $Ep_h^{opt}(1_h) = 2/3$. The area $A = 1/18$.

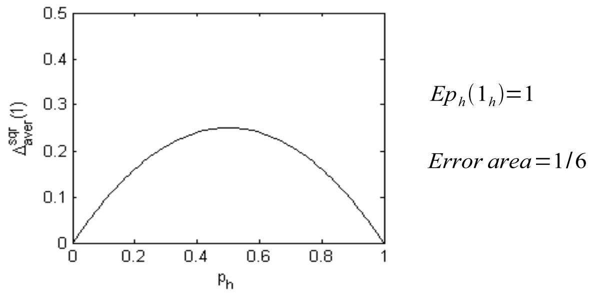


Figure 6. Cross-section of the square-error function $\Delta^{sqr}_{aver}(1) = f(p_h, Ep_h(1_h))$ for the value $Ep_h(1) = 1$ which corresponds to the universally used frequency estimator of probability $fr_h = n_h/n$. The error area $A = 1/6$.

$$Ep_{ha} = \frac{1}{2} + \frac{n_h - n_{\bar{h}}}{2(n+2)} \tag{27}$$

Let us remember that in formula (27) n_h means the number of sample items confirming the binomial hypotheses h concerning an event, $n_{\bar{h}}$ means the number of sample items negating the hypothesis h, and n means the entire number of all sample items ($n = n_h + n_{\bar{h}}$). Formula (27) of the optimal probability estimator was derived theoretically. Therefore one could doubt its practical correctness. To check these doubts, in the next section the results of test experiments will be presented. In these experiments the universally used frequency estimator will be compared with the new, completeness estimator of probability.

4. Results of comparative experiments of probability estimation by the completeness estimator Ep_{h2} and by the frequency estimator $fr_h = n_h/n$

To test and to compare accuracy of the completeness and the frequency estimator, experiments were realized in which 1000 series with 10000 of 1s and 0s were generated with following p_h-probabilities: 0.01, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.99. Thus, the number of the experiment-probabilities was equal to 11. Each "1" generated by a random-number generator was understood as confirmation of the hypothesis h and "0" as

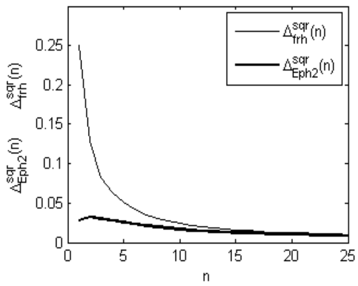


Figure 7. Diagram of average square estimation errors $\Delta^{sqr}_{Ep_{h2}}(n)$ and $\Delta^{sqr}_{fr_h}(n)$ for numbers of sample items $n \leq 25$ for the completeness estimator Ep_{h2} (thick line) and frequency estimator fr_h (thin line). The average errors were calculated on the basis of 1000 experiments with 10000 numbers (1 and 0) generated in each single experiment. The estimated probability was equal to 0.5.

negation of it. Computer generators are used by many scientists in their investigations, e.g. by Larose [12]. Due to the fact that in each experiment the p_h-probability of generated 1s was known with high accuracy after calculation of the estimates Ep_{h2} and fr_h for the series comparison of their values with the probability p_h and calculation of the square error was made possible. The diagram in Fig. 7 demonstrates a square errors of both estimators for computing of probability $p_h = 0.5$ on the basis of a small number of sample items $n \leq 25$. These 25 sample items are only the beginning of the full generated series of 10000 numbers.

Fig. 7 clearly shows a considerable superiority of the completeness estimator over the frequency one for a small number of sample items $n \leq 25$. The error sum for all sample item numbers $n \leq 25$ equals 0.9491 for the frequency estimator and is over 2 times higher than for the completeness estimator where it is equal to 0.4131. In particular, for $n = 1$ (single case problem) the average square error of the frequency estimator equals 0.25 whereas the error of the completeness estimator equals 0.0278. The difference is greater than 800%. For sample item numbers greater than 25 sample items differences between both estimators disappear, because both estimates converge and for n approaching infinity they calculate the precise probability value.

Diagram in Fig. 8 shows the average square errors for estimation of probability $p_h = 0.4$ and $p_h = 0.6$. These probabilities are antonym and the results achieved for them were very similar, which is compatible with the theory, see formula (21) and Fig. 3. Fig. 9 shows a corresponding diagram for antonym probabilities $p_h = 0.3$ and $p_h = 0.7$.

Fig. 8 and Fig. 9 also, similarly as Fig. 7, show considerable superiority of the completeness estimator over the frequency one. The smaller the sample item number n the greater the superiority is. For probabilities 0.4 and 0.6 the sum of mean, square-errors equals 0.8864 for the frequency estimator fr_h and 0.3981 for the completeness estimator Ep_{h2}. Thus, the difference in accuracy is larger than 100%. In case of the estimated probabilities 0.3 and 0.7 the sum of mean, square errors is equal to 0.8329 for the frequency estimator fr_h and 0.4343 for the completeness estimator Ep_{h2}. Thus, the difference in accuracy of both estimators is also equal to about 100%. Fig. 10 presents diagram of square errors of both competing estimators for estimation of antonym probabilities 0.2 and 0.8.

Estimation results of probabilities 0.2 and 0.8 also show considerable superiority of the completeness estimator over the frequency one. The mean error for the first 25 sample item numbers is equal for the completeness estimator 0.4141 and for the frequency estimator 0.6151

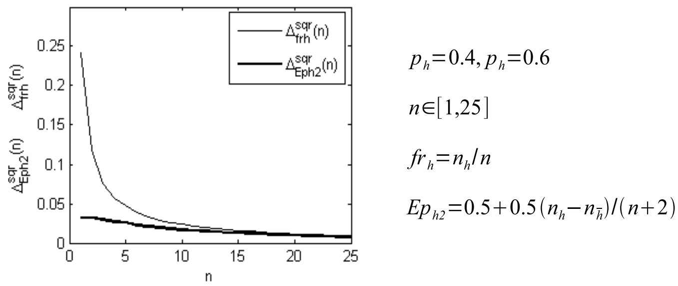


Figure 8. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ and $\Delta^{sqr}_{Ep_{h2}}(n)$ of the frequency estimator fr_h and the completeness estimator Ep_{h2} for estimation of the antonym probabilities $p_h = 0.4$ and $p_h = 0.6$ for sample item numbers $n \in [1, 25]$. The diagram shows mean results of 1000 experiments.

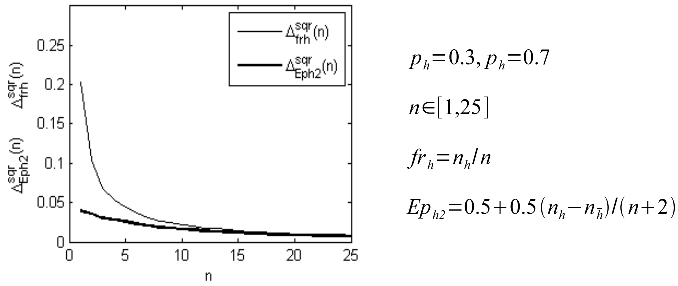


Figure 9. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ and $\Delta^{sqr}_{Ep_{h2}}(n)$ of the frequency estimator fr_h and the completeness estimator Ep_{h2} for estimation of the antonym probabilities $p_h = 0.3$ and $p_h = 0.7$ for sample item numbers $n \in [1, 25]$. The diagram shows mean results of 1000 experiments.

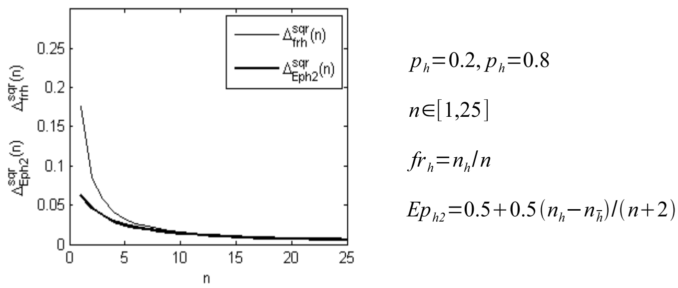


Figure 10. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ of the frequency estimator fr_h and $\Delta^{sqr}_{Ep_{h2}}(n)$ of the completeness estimator Ep_{h2} for estimation of the antonym probabilities $p_h = 0.2$ and $p_h = 0.8$ for sample item numbers $n \in [1, 25]$. The diagram shows mean results of 1000 experiments.

and is larger at about 50% than the mean error of the completeness estimator. Fig. 11 presents estimation results of antonym probabilities 0.1 and 0.9.

For the frequencies 0.1 and 0.9 the frequency estimator has a small superiority over the completeness one. The square error sum for all sample item numbers $n \in [1, 25]$ equals 0.3478 for the frequency estimator and 0.3859 for the completeness estimator.

Fig. 12 presents estimation results for frequencies 0.01 and 0.99. These frequencies mean linguistic probabilities ”almost zero” and ”almost certainty”.

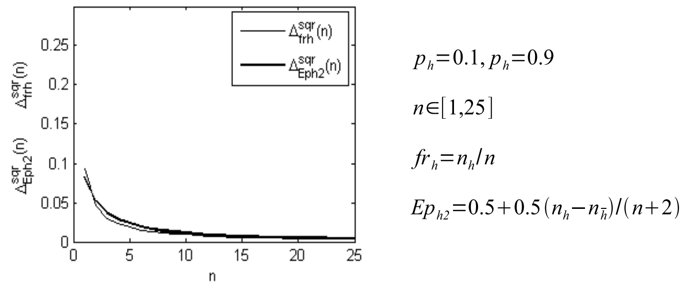


Figure 11. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ and $\Delta^{sqr}_{Ep_{h2}}(n)$ of the frequency estimator fr_h and the completeness estimator Ep_{h2} for estimation of the antonym probabilities $p_h = 0.1$ and $p_h = 0.9$ for sample item numbers $n \in [1, 25]$. The diagram shows mean results of 1000 experiments.

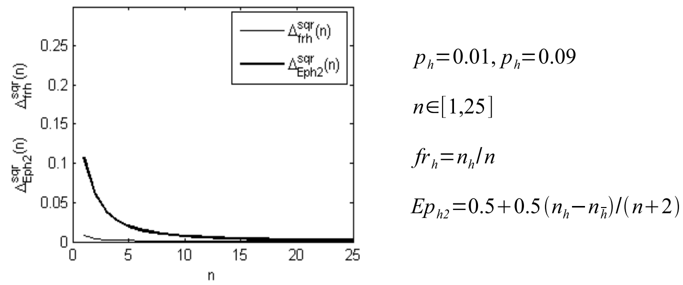
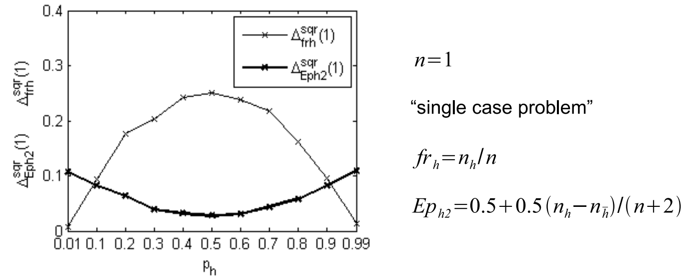


Figure 12. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ and $\Delta^{sqr}_{Ep_{h2}}(n)$ of probabilities 0.01 (almost zero) and 0.99 (almost certainty) made by the frequency estimator fr_h and completeness estimator Ep_{h2}. The results are mean of 1000 experiments.

As it can be seen in Fig. 12, this time the superiority of the frequency estimator is considerable. The error sum for all 25 sample item numbers $n \in [1, 25]$ equals 0.0442 for the frequency and 0.3672 for the completeness estimator.

Fig. 13 demonstrates collected results of estimation errors made by both competitive estimators for all 11 estimated probabilities.

The results presented in Fig. 13 clearly show that the completeness estimator Ep_{h2} allows for considerable, general decreasing of average errors of one sample item estimation in comparison to the frequency estimator $fr_h = n_h/n$. However, it should be reminded here that the remark concerns the average single sample item error resulting from many experiments (here from 1000 experiments). In a single experiment the maximal absolute single sample item error can take a value from interval $[0, 1]$. Thus, de Finetti [7] claiming that it has no sense to speak about single-case probability was partially right, because the maximal, absolute estimation-error can reach the value of almost 1.00 in case of the frequency estimator. Therefore concluding probability from one experiment is in case of the frequency estimator very dangerous. In case of the completeness estimator the situation is better because the maximal, absolute error can in a single experiment reach the value of $2/3$, which is smaller. In everyday situations we are frequently forced to conclude probability from a single experiment, from a single fact. In such situations it will be better to use estimation results of the completeness estimator. Fig. 13 also shows that the frequency estimator is a "specialist" of easy frequencies which are close to 1 (full confirmative certainty) or close to zero (full



p_h	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
$\Delta^{sqr}_{frh}(1)$	0.0079	0.0932	0.1756	0.2040	0.2420	0.2500	0.2372	0.2180	0.1612	0.0956	0.0128
$\Delta^{sqr}_{Eph2}(1)$	0.1072	0.0822	0.0630	0.0391	0.0318	0.0278	0.0302	0.0438	0.0582	0.0830	0.1088

Figure 13. Diagram of the average square one sample item error$\Delta^{sqr}_{fr_h}(1)$ and $\Delta^{sqr}_{Ep_{h2}}(1)$ of the frequency estimator fr_h and completeness estimator Ep_{h2} representing estimation of different 11 probabilities p_h for $n = 1$ (single-case problem). The results are mean of 11 times 1000 experiments.

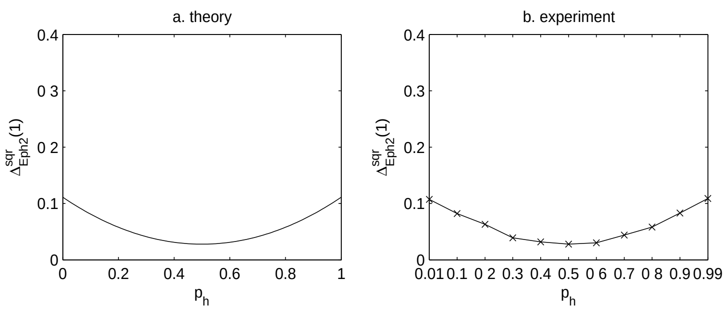


Figure 14. The minimal, theoretically derived square error function $\Delta^{sqr}_{Ep_{h2}}(1)$ of the completeness estimator Ep_{h2}, Fig. 14.a, and the same function achieved experimentally from 1000 experiments, Fig. 14.b, (the single-case problem).

negating certainty). Probabilities of almost certain events are easy to identify by people. However, probabilities close to 0.5 are very difficult identifiable for the frequency estimator and its average, square-error obtained from estimation of this probability is the greatest one in relation to other probability values and equals 0.25, Fig. 13. In case of the completeness estimator the situation is completely inverse. This estimator is estimation "specialist" of just the probability 0.5 and close to it. When the average square error of the frequency estimator equals 0.25 for this probability, the error of the completeness estimator equals 0.0278 and is 9 times smaller (in the single-case problem). Theoretical and experimental results display astonishing compatibility. Fig. 14 demonstrates a comparison of the minimal, theoretically derived error function for the value $Ep_h^{opt}(1) = 2/3$ and the same function achieved from experiments and shown in Fig. 14.

The same similarity show also error functions of the frequency estimator fr_h, Fig. 15.

Fig. 16 presents collected square errors of probability estimation from a very small number of sample items $n \in [1, 5]$.

The diagram shown in Fig. 16 unambiguously proves that the completeness estimator Ep_{h2} for most estimated probabilities is superior over the frequency estimator fr_h. In case of the frequency estimator the collected square error equals 0.1150 and is 4 times larger than

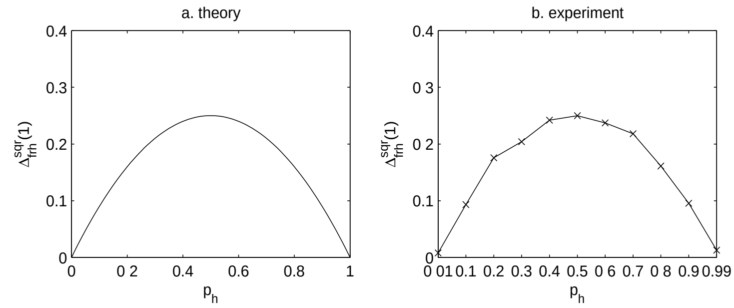
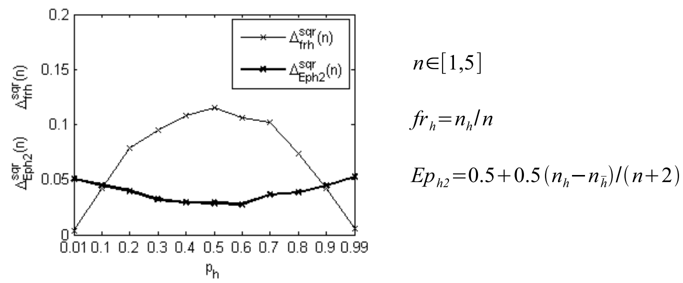


Figure 15. The square error function $\Delta^{sqr}_{fr_h}(1)$ of the frequency estimator fr_h derived theoretically (Fig. 15.a) and obtained experimentally from 1000 experiments (Fig. 15.b).



p_h	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
$\Delta^{sqr}_{frh}(n)$	0.0036	0.0423	0.0784	0.0947	0.1080	0.1150	0.1058	0.1017	0.0736	0.0426	0.0058
$\Delta^{sqr}_{Eph2}(n)$	0.0507	0.0447	0.0399	0.0318	0.0293	0.0287	0.0276	0.0360	0.0381	0.0445	0.0523

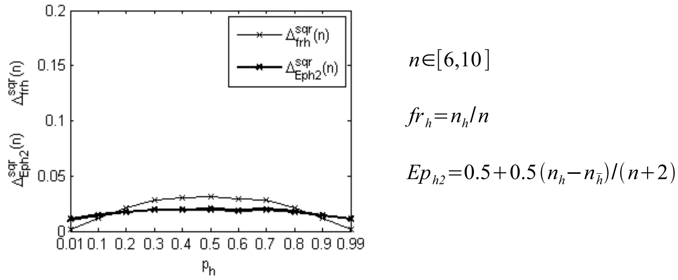
Figure 16. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ and $\Delta^{sqr}_{Ep_{h2}}(n)$ of the frequency estimator fr_h and the completeness estimator Ep_{h2} for estimation of 11 different probabilities p_h for sample item numbers $n \in [1, 5]$. The results are mean of 1000 experiments for each p_h-value, $\Delta^{sqr}_{\dots}(n) = \frac{1}{5}\sum_{i=1}^{5} \Delta^{sqr}_{\dots}(i)$

the error of the completeness estimator that equals 0.0287. Fig. 17 shows collected, average square errors of probability estimation for small sample item number $n \in [6, 10]$.

As Fig. 17 shows, also for the sample item numbers in interval $[6, 10]$ the completeness estimator is superior over the frequency one, apart from probabilities lying near 0 or 1. The frequency estimator has the collected, mean error equal to 0.0317 for frequency 0.5 and the completeness estimator 0.0198. Fig. 18 shows mean collected errors of both estimators for sample item numbers $n \in [11, 15]$.

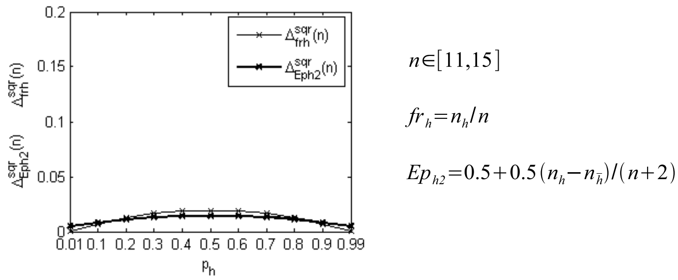
Fig. 18 also shows small superiority of the completeness estimator Ep_{h2} over the frequency one. For the probability 0.5 the square error equals 0.0185 for the frequency estimator and 0.0138 for the completeness one. Fig. 19 presents collected results for 10 sample item numbers from interval $[16, 25]$.

The differences between both estimators for the sample item number interval $[16, 25]$ are very small, because for larger sample item numbers the estimators converge. Summing up results shown in Fig. 13-Fig. 19 one can say without any doubt that the completeness estimator is for the most probability values in respect of accuracy superior over the frequency estimator



p_h	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
$\Delta^{sqr}_{frh}(n)$	0.0011	0.0117	0.0205	0.0281	0.0300	0.0317	0.0293	0.0281	0.0206	0.0117	0.0013
$\Delta^{sqr}_{Eph2}(n)$	0.0107	0.0145	0.0173	0.0192	0.0194	0.0198	0.0186	0.0198	0.0176	0.0143	0.0112

Figure 17. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ and $\Delta^{sqr}_{Ep_{h2}}(n)$ of the frequency estimator fr_h and the completeness estimator Ep_{h2} for estimation of 11 different probabilities p_h for sample item numbers $n \in [6, 10]$. The results are mean of 1000 experiments for each p_h-value, $\Delta^{sqr}_{...}(n) = \frac{1}{5}\sum_{i=6}^{10} \Delta^{sqr}_{...}(i)$



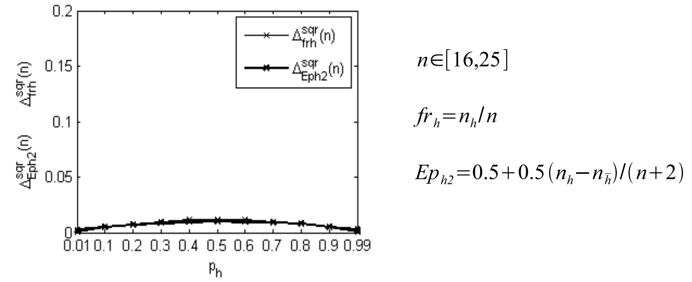
p_h	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
$\Delta^{sqr}_{frh}(n)$	0.0007	0.0068	0.0122	0.0167	0.0186	0.0185	0.0185	0.0163	0.0126	0.0066	0.0008
$\Delta^{sqr}_{Eph2}(n)$	0.0048	0.0081	0.0111	0.0132	0.0141	0.0138	0.0139	0.0132	0.0116	0.0078	0.0051

Figure 18. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ and $\Delta^{sqr}_{Ep_{h2}}(n)$ of the frequency estimator fr_h and the completeness estimator Ep_{h2} for estimation of 11 different probabilities p_h for sample item numbers $n \in [11, 15]$. The results are mean of 1000 experiments for each p_h-value, $\Delta^{sqr}_{...}(n) = \frac{1}{5}\sum_{i=11}^{15} \Delta^{sqr}_{...}(i)$

in respect of accuracy. It can estimate with much smaller errors a probability range (near 0.5) which is difficult to estimate for the frequency estimator.

Fig. 20 and Fig. 21 show diagrams of the average, minimal number of sample items $n^{fr_h}_{0.05}$, $n^{Ep_{h2}}_{0.05}$, $n^{fr_h}_{0.01}$ and $n^{Ep_{h2}}_{0.01}$ that are necessary for estimation of various probability values with the absolute error less than 0.05 and 0.01. Obtained values are results from experiments.

The comparison of Fig. 20 and Fig. 21 demonstrates that achieving of higher estimation accuracy is connected with a non-proportional, strong increase in the required sample item number. The results shown in these figures differ from the corresponding results calculated from Chernoff-bound [5]. Chernoff-bound does not take into account the calculation accu-



p_h	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
$\Delta^{sqr}_{frh}(n)$	0.0005	0.0043	0.0076	0.0102	0.0116	0.0123	0.0118	0.0102	0.0081	0.0045	0.0005
$\Delta^{sqr}_{Eph2}(n)$	0.0024	0.0049	0.0071	0.0087	0.0096	0.0101	0.0097	0.0089	0.0077	0.0050	0.0024

Figure 19. Diagram of the average square error $\Delta^{sqr}_{fr_h}(n)$ and $\Delta^{sqr}_{Ep_{h2}}(n)$ of the frequency estimator fr_h and the completeness estimator Ep_{h2} for sample item numbers $n \in [16, 25]$. Results are the mean of 1000 experiments for each p_h-value, $\Delta^{sqr}_{...}(n) = \frac{1}{10} \sum_{i=15}^{25} \Delta^{sqr}_{...}(i)$

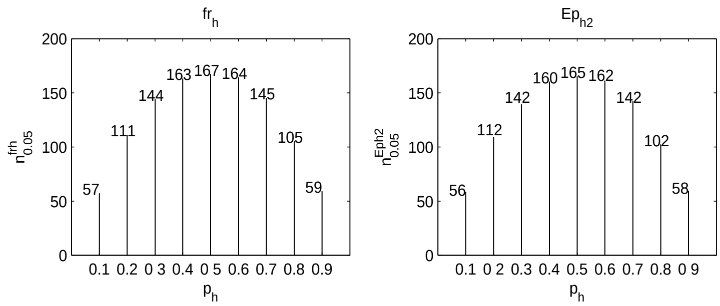


Figure 20. Approximate minimal sample item numbers $n^{fr_h}_{0.05}$ and $n^{Ep_{h2}}_{0.05}$ that are necessary for satisfactory estimation of various probability values with the use of the frequency estimator fr_h and the completeness estimator Ep_{h2} with the absolute error less than 0.05. Results are the mean of of 1000 experiments for each value of p_h.

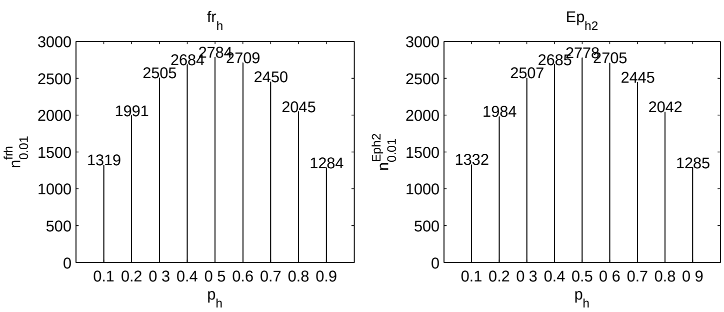


Figure 21. Approximate minimal sample item numbers $n^{fr_h}_{0.01}$ and $n^{Ep_{h2}}_{0.01}$ that are necessary for identification of various probability values with the use of the frequency estimator fr_h and the completeness estimator Ep_{h2} with the absolute error less than 0.01. Results are the mean of 1000 experiments for each probability value.

racy and gives smaller sample item numbers for probabilities near 0 and near 1. However, it requires much higher sample item number for probabilities near 0.5.

Further on, a simple example of probability calculation according to the completeness interpretation is given. Let us assume that after 5 coin tosses ($n = 5$) the results are as follows: 3 heads (3 confirmations of the hypothesis h about the head superiority, $n_h = 3$) and 2 tails (2 confirmations of tail superiority in the coin, $n_{\bar{h}} = 2$). The frequency estimator fr_h gives the following estimation of the p_h - probability:

$$fr_h = n_h/n = 3/5$$

An application of the completeness interpretation [14, 15] and of the completeness estimator Ep_{h2} gives different results. Let us assume according to Chernoff-bound the satisfactory evidence set SEC with $n_{SEC} = 23026$ sample items. Such an evidence set allows, on the basis of Chernoff bound [5, 14, 15], for estimation of probability with error ε less than ± 0.01 with credibility 0.99. With the use of formula (28) the minimal probability $p_{h\min}$ of the hypothesis about the head superiority can be calculated as shown:

$$p_{h\min} = n_h(1-2\varepsilon)/n_{SEC} = 3(1-0.02)/23026 = 0.000128 \quad (28)$$

Similarly the minimal probability of the anti-hypothesis $\bar{h}$ (NOT head superiority = tail superiority) can be calculated (29).

$$p_{\bar{h}\min} = n_{\bar{h}}(1-2\varepsilon)/n_{SEC} = 2(1-0.02)/23026 = 0.000085 \quad (29)$$

With formula (30) the maximal probability of the hypothesis h can be calculated.

$$p_{h\max} = 1 - p_{\bar{h}\min} = 1 - 0.000128 = 0.999872 \quad (30)$$

From formula (31) the estimation of the hypothesis about the head superiority can be calculated as below.

$$Ep_{h2} = 0.5[1 + (n_h - n_{\bar{h}})/(n+2)] = 0.5[1 + (3-2)/(5+2)] = 4/7 = 0.571429 \quad (31)$$

Interpretation of the achieved results is given below.

It is not possible to determine the probability p_h of the hypothesis h precisely because of the largely insufficient number $n = 5$ sample items (evidence pieces) we have at disposal. This number is much smaller than the required number $n_{SEC} = 23026$ sample items. Therefore the true probability remains unknown. The only thing we can say is that this probability lies somewhere between two limits $p_{h\min}$ and $p_{h\max}$.

$$p_{h\min} \le p_h \le p_{h\max} : 0.000128 \le p_h \le 0.999872$$

The calculated completeness estimate of the hypothesis probability $Ep_h = 4/7 = 0.571429$ is not the true value of the probability p_h but only a very uncertain estimate. However, it is the best estimate we can calculate on the basis of such small number of sample items ($n = 5$), the best in the sense of the square-error optimality criterion.

The value $4/7$ estimated by the completeness estimator may seem incompatible with our intuition and common sense. The frequency estimate $fr_h = 3/5$ intuitively seems more appropriate. However, as the theoretical proof and the experimental verification presented in the paper show the value proposed by the completeness estimator is more credible than the value proposed by the frequency estimator.

5. Completeness estimator Ep_{h2} of probability for the multinomial case

If the number k of possible hypotheses in a problem is larger than 2 ($k > 2$) then this k-nomial problem can be decomposed in k binomial subproblems of type: hypothesis h_i and negation $\bar{h}_i$ of the hypothesis. For each of the k binomial subproblems the binomial estimate $Ep^*_{h_i2}$ of probability can be determined from formula (32).

$$Ep^*_{h_i2} = \frac{1}{2} + \frac{n_{h_i} - n_{\bar{h}_i}}{2(n+2)} = \frac{1}{2} + \frac{n_{h_i} - (n - n_{h_i})}{2(n+2)} = \frac{1}{2} \cdot \frac{2n_{h_i} + 2}{n+2} = \frac{n_{h_i} + 1}{n+2} \tag{32}$$

where: $n = \sum\limits_{j=1}^{k} n_{h_j}$.

However, the sum $n = \sum\limits_{j=1}^{k} Ep^*_{h_j2}$ of the binomial estimates calculated with (32) is not normalized in the sense of Kolmogorov's axioms [9] and is larger than 1. Therefore the normalization of the binomial estimates $Ep^*_{h_i2}$ has to be done according to formula (33). The normalization gives normalized estimates Ep_{h_i2}.

$$Ep_{h_i2} = \frac{Ep^*_{h_i2}}{\sum\limits_{j=1}^{k} Ep^*_{h_j2}} = \frac{\frac{n_{h_i}+1}{n+2}}{\sum\limits_{j=1}^{k} \frac{n_{h_j}+1}{n+2}} = \frac{n_{h_i} + 1}{\sum\limits_{j=1}^{k} (n_{h_i} + 1)} = \frac{n_{h_i} + 1}{n+k} \tag{33}$$

Now, let us compare the frequency estimate $fr_h = n_h/n$ with the multinomial completeness estimate Ep_{h_i2} of probability. In case of the frequency estimate $fr_h = n_h/n$ the formula is independent of the estimation type. The formula $fr_h = n_h/n$ is the same for the binomial case ($k = 2$), trinomial case ($k = 3$), ..., m-nomial case ($k = m$). It looks rather strange. In case of the completeness estimator Ep_{h_i2} the estimation formula (33) changes if k varies and thus it is different for the binomial, trinomial, ..., m-nomial case. This fact is of a large importance for probability estimation, e.g. histograms.

6. Conclusions

The paper presents a new probability estimator Ep_{h2} that minimizes the estimate error. In the paper were derived formulas (32) and (33) of the completeness estimator for binary and k-ary case. This estimator estimates with high accuracy, in sense of square-error-sum, probability not only from a great but also from a small sample item. It was shown that even from a single sample item probability can be estimated ($Ep_{h2} = 2/3$), though uncertainty of this estimate is very great. It is a very important fact in the discussion, whether probability estimation from one sample item is sensible or not. The probability estimators determined by formulas (32) for binary case and by (33) for k-ary case are very useful especially for data mining problems where frequently only a small evidence-sample item is at disposal. The author of the Ep_{h2} estimator idea and of its optimality proof is Andrzej Piegat. Computer programs for the experimental verification and experiment were realized by Marek Landowski. The aim of the experiment was to determine the accuracy of the completeness estimator and to compare it with the widely used frequency estimator $fr_h = n_h/n$. Both the theoretical proof and experimental verification show a considerably larger accuracy of the completeness estimator Ep_{h2} in comparison with the frequency one in case of small sample item $n = 25$ and the same accuracy for larger numbers of $n > 25$.

References

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

[2] Burdzy, K.: Philosophy of probability. *http://www.math.washington.edu/~burdzy/Philosophy/*, 2013, online.

[3] Burdzy, K.: Blog related to the book "The Search for Certainty. On the Clash of Science and Philosophy of Probability" by K. Burdzy. *http://search4certainty.blogspot.com/*, 2013, online.

[4] Carnap, R.: Logical foundations of probability. University Press, Chicago, 1952.

[5] Chernoff, H.: A measure of asymptotic efficiency for test of a hypothesis based on the sum of observations. Annals of Mathematical Statistics, 23 (4), pp. 493–507, 1952.

[6] Dubois, D., Prade, H.: Possibility theory. Plenum Press, New York and London, 1988.

[7] De Finetti, B.: Theory of probability, a critical introductory treatment. Willey, London, 1975.

[8] Hajek, A.: Interpretations of probability. The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), (ed. E.N. Zalta), *http://plato.stanford.edu/entries/probability-interpret/*, 2013, online.

[9] Khrennikov, A.: Interpretations of probability. Brill Academic Pub. Utrecht, Boston, 1999.

[10] Klirr, G. J., Yuan, B. (editors): Fuzzy sets, fuzzy logic, and fuzzy systems. Selected papers by Lotfi Zadeh, World Scientific, Singapore, New Jersey, London, Hong Kong, 1996.

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

[12] Larose, D. T.: Discovering statistics. W.H. Freeman and Company, New York, 2010.

[13] von Mises, R.: Probability, statistics and the truth. Macmillan, second revised English edition, Dover, New York, 1957.

[14] Piegat, A.: Uncertainty of probability. In Recent Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics, Volume 1: Foundations, Systems Research Institute, Polish Academy of Sciences, Warsaw, pp. 159–173, 2011.

[15] Piegat, A.: Basic lecture on completeness interpretation of probability. *http://kmsiims.wi.zut.edu.pl/pobierz-pliki/cat_view/47-publikacje*, 2011, online.

[16] Popper, K. R.: The propensity interpretation of the calculus of probability and the quantum theory, In Observation and Interpretation: A Symposium of Philosophers and Physicists (eds S. Korner), London: Butterworth Scientific Publications, pp. 65–70, 1957.

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

[18] Shafer, G.: A mathematical theory of evidence. Princetown University Press, Princetown and London, 1976.

[19] Yakov, B. H.: Info-gap decision theory. Second edition, Elsevier, Oxford, Amsterdam, 2006.

[20] Zadeh, L. A.: Fuzzysets. Information and Control, Vol.8, pp. 338–353, 1965.