KYBERNETIKA — VOLUME 7 (1971), NUMBER 1

The Statistical Interpretation and Modification of GUHA Method

TOMÁŠ HAVRÁNEK

This study is connected with the papers [1], [2], and [3]. The knowledge of [1] and [2] or of [3] is essential for understanding of the following notes. We interpret and modify in a statistical way the concept of almost truthfulness and of the relative almost truthfulness trying to preserve as far as possible the original algorithm of GUHA method as described in [2] or [3].

I.

In the quoted papers the main idea of GUHA method has been formulated, i.e. to obtain automatically all interesting hypotheses from experimental results. In [3] or [2] the formulae of propositional calculus, that are fulfilled at least by a certain given number of objects in the model, are regarded as interesting hypotheses. This idea will be now formulated more exactly. Let there be given a model $\mathcal{M}$ with the number m of objects, let m_A be the number of objects fulfilling a formula A in the model $\mathcal{M}$. On the set of all objects of a corresponding universum (more exactly on some sigma algebra of subsets) we shall define a statistic $X_A(\omega)$ having value 1, if the object ω fulfils the formula A, and value 0 in all other cases. This statistic has an alternative distribution with the parameter p_A, $p_A \in \langle 0, 1\rangle$, which in this case gives directly the probability, that arbitrary object of the universum fulfils the formula A. We call a formula A interesting, if $p_A \geqq p$, where p is a given number, $p \in \langle 0, 1\rangle$. It is convenient to choose the value of p near 1. To select an interesting formula of our model (from the point of view of the universum) we need some decision rule, which we shall try to find. In GUHA [3] or [2] a rule is used, by means of which a formula A is omitted if $m_A/m < p$. The point estimation of parametr p_A is thus used i.e. $\hat{p}_A = m_A/m$. The formula accepted by this rule is called an almost true formula and is used further on. This rule has an unknown probability of error of omitting an interesting formula i.e. unknown $P(\textit{omitting } A/p_A \geqq p)$. Thus we shall try to search for such a rule, for which this probability should be small and limited in advance. We want $P(\textit{omitting } A/p_A \geqq p)$ to be $\leqq \alpha$, where $\alpha \in \langle 0, 1\rangle$

is given in advance (usually $\alpha = 0{,}05$ and similar). We will proceed in an analogical way in the case of relatively almost true implications. The case of the almost truthfulness does not require an alteration of the GUHA algorithm and the explication of this case in this paper may be understood only as a more precise interpretation. On the other hand, in the case of relative almost truthfulness, some alterations are suggested, which are easily linked up with the GUHA algorithm as will be shown in the third part of this paper. The third part contains further reasoning about the applicability of the original GUHA method or of the modification here introduced. The second part contains new definitions of decision rules and some assertions showing that the introduced rules have the required properties and, in addition, that Theorems 6, 7, 8 from [3] (or 1, 2, 3 from [2]) are preserved and consequently the original GUHA algorithm can be easily modified.

The issues discused in this paper have been submitted at the seminar on the aplications of mathematical logics at the department of mathematical logics, Faculty of Mathematics and Physics, Charles University, Prague. I wish to express my gratitude to the head of this seminar Dr. P. Hájek for his valuable comments. I also express my thanks to other members of the seminar. I have based this paper on lectures given by prof. J. Hájek, Dr. Sc. and Ing. J. Machek on mathematical statistics. Moreover, I am obliged to Ing. J. Machek for consultations and for the main idea of the proof of the lemma in the third part of this paper.

II.

First of all we will define in a new manner the notion of almost truthfulness concerning to the universum.

Definition 1. Let p be a rational number, $p \in \langle 0, 1 \rangle$. We say that a formula A is *p-almost-true*, if $p_A \geqq p$.

Analogously we shall define relative p-almost-truthfulness.

Definition 2. We say that an implication $A \to B$ is *relatively p-almost-true*, if it is p-almost-true in the universum of the objects fulfilling A.

These properties of formulae cannot be verified, when only the model $\mathcal{M}$ is given. We must now define properties, which can be verified in this model. The presence or absence of these properties is identical with what we have called a decision rule. We wanted to search for a decision rule, that would guarantee a small and limited propability of error. Such a rule will be construsted in the following way. Let us have a real number α, $\alpha > 0$, arbitrarily small. We construct a one-sided (lower) confidence interval for a parameter p_A with confidence coefficient $1 - \alpha$. Then we have $P(p_A \leqq \bar{p}_A | p_A) = 1 - \alpha$, consequently $P(p_A > \bar{p}_A | p_A) \leqq \alpha$. We will apply this decision rule: if $\bar{p}_A < p$, then we omit A. This rule corresponds to a notion defined in subsequential definition.

Definition 3. Let us have a model $\mathcal{M}$ with m objects, let m_A be the number of objects fulfilling a formula A. We say that the formula A is *suspicious of p-almost-truthfulness on the level* $1 - \alpha$, if $\bar{p}(A, m) \geqq \alpha$, where $\bar{p}(A, m)$ is an upper limit of one-sided (lower) confidence interval with confidence coefficient $1 - \alpha$ for the number of investigations equal to m and for the number of investigations having positive result equal to m_A.

We shall write $\bar{p}_A$ for $\bar{p}(A, m)$ in the sequel.

We have to verify that the defined property so fulfils our requirement as to the probability of error.

Theorem 1. *Under the assumptions of the definition we have* $P(\bar{p}_A < p | p_A \geqq p, p_A) \leqq \alpha$.

Proof. Let us assume $P(p_A \geqq p | p_A) \neq 0$. Then $P(\bar{p}_A < p | p_A \geqq p, p_A) = P(\bar{p}_A < p \,\&\, p \leqq p_A | p_A) : P(p \leqq p_A | p_A)$ and $P(\bar{p}_A < p_A | p \leqq p_A, p_A) = P(\bar{p}_A < p_A \,\&\, p \leqq p_A | p_A) : P(p \leqq p_A | p_A)$. From $\bar{p}_A < p \,\&\, p \leqq p_A \to \bar{p}_A < p_A \,\&\, p \leqq p_A$ follows the following inequality: $P(\bar{p}_A < p \,\&\, p \leqq p_A | p_A) : P(p \leqq p_A | p_A) \leqq P(\bar{p}_A < p_A \,\&\, p \leqq p_A | p_A) : P(p \leqq p_A | p_A)$. Furthermore $\bar{p}_A < p_A$ and $p \leqq p_A$ are independent and therefore $P(\bar{p}_A < p_A | p \leqq p_A, p_A) = P(\bar{p}_A < p_A | p_A)$. The theorem follows from the transitivity of equality and inequality. If is $P(p_A \geqq p | p_A) = 0$ then $P(p_A < p | p_A) = 1$ and no error can occure.

Thus we can see that the probability of omitting an interesting formula is limited by the number α given in advance. Applying a one-sided (upper) confidence interval we can define another notion which will guarantee a great confidence that the formulae selected in accordance with this notion are *p*-almost-true.

Definition 4. Let m and m_A have the same meaning as in Definition 3. We say that a formula A is *probably p-almost-true on the level* $1 - \alpha$, if $p \leqq p_*(A, m)$, where $p_*(A, m)$ is the lower limit of one-sided (upper) confidence interval with confidence coefficient $1 - \alpha$ for the number of investigations equal to m and for the number of investigations with positive result equal to m_A.

A theorem similar to Theorem 1 holds true here. We shall write p_{A*} for $p_*(A, m)$.

Theorem 2. *Under the assumptions of Definition* 4 *we have* $P(p \leqq p_A | p \leqq p_{*A}, p_A) \geqq 1 - \alpha$.

Proof. According to the definition of p_{A*} we have $P(p_{A*} \leqq p_A | p_A) = 1 - \alpha$, analogously to the proof of Theorem 1 we obtain $P(p \leqq p_A | p \leqq p_{A*}, p_A) = P(p \leqq p_A \,\&\, p \leqq p_{A*} | p_A) : P(p \leqq p_{A*} | p_A) \leqq P(p_{A*} \leqq p_A \,\&\, p \leqq p_{A*} | p_A) : P(p \leqq p_{A*} | p_A) = P(p_{A*} \leqq p_A | p_A) \geqq 1 - \alpha$, if $P(p \leqq p_{A*} | p_A) \neq 0$, and the theorem is proved.

Theorem 3. *If a formula is probably p-almost-true on the level* $1 - \alpha$ *then it is suspicious of p-almost-truthfulness on the same level.*

Proof. We have $p_{A*} \leqq \bar{p}_A$.

Let us remark that here p_{A*} and $\bar{p}_A$ do not mean upper and lower limit of the confidence interval (two-sided) with coefficient $1 - \alpha$.

To summarize: The criterion of suspicion of almost-truthfulness gives us an assurance that we select with the probability $1 - \alpha$ all almost-true formulae (for certain p). The criterion of probable p-almost-truthfulness assures that the formulae selected by us are, with the probability $1 - \alpha$, p-almost-true. If we regard the p-almost-true formulae as interesting the criterion of suspicion seams to be more essential for selecting all interesting formulae than the criterion of probable p-almost-truthfulness. Given p, it is possible to determine p_1 such that a formula is suspicious of p-almost-truthfulness if and only if it is probably p_1-almost-true. In this way we can solve the problem a) from the paper [1], p. 45. Let m and m_A have the usual meaning; we choose an α and, given m and m_A, we construct the number p_{A*} such that formula A is p_{A*}-almost true with probability $1 - \alpha$. In particular, $m_A = m$, we obtain a direct solution of the problem mentioned above.

Let us return now to the paper [3] (or [2]). Theorem 6 from [3] (or Theorem 1 from [2]) is valid for formulae suspicious of almost-truthfulness, if in its proof p' is substituted by $\bar{m}$, whose meaning will be explained in the third part of this paper.

If we follow the paper [3] (or [2]), we encounter the problem of implication. In these papers, a submodel $\mathcal{M}'$ of all objects from $\mathcal{M}$ fulfilling A is considered. $\mathcal{M}$ has m objects and $\mathcal{M}'$ has m_A objects. The relative almost-truthfulness of implication $A \to B$ is defined in [3] (or [2]) using the ration $m_{A\&B}/m_A$, where $m_{A\&B}$ is the number of objects fulfilling $A \& B$. This ratio is the point estimation of conditioned probability of fulfilling the formula B under the assumption of validity of formula A (let we call this probability $p_{|B/A|}$ since $p_{|B/A|} = p_{A\&B}/p_A$, $\hat{p}_{A\&B} = m_{A\&B}/m$, $\hat{p}_A = m_A/m$ and therefore $\hat{p}_{|B/A|} = m_{A\&B}/m_A$. We wish again α to be $\geqq P(\text{omitting } A \to B / p_{|B/A|} \geqq p)$. To satisfy this requirement the decision rule must be changed according to the number of objects in $\mathcal{M}'$. In this submodel we determine for a certain α and for the upper limit of one-sided (lower) confidence interval $\bar{p}(A \& B, m_A)$. The decision rule is then as follows: $A \to B$ is omitted if and only if $\bar{p}(A \& B, m_A) < p$. We have $P(\bar{p}(A \& B, m_A) < p / p_{A\&B} \geqq p) \leqq \alpha$ similarly to the preceding case and we can state the following definition.

Definition 5. We say that an implication $A \to B$ is *suspicious of relative p-almost-truthfulness on the level* $1 - \alpha$ in a model $\mathcal{M}$ if the formula B is suspicious of p-almost-truthfulness on the level $1 - \alpha$ in the model $\mathcal{M}'$.

The validity of Theorem 7 from [3] or Theorem 2 from [2] (in which we substitute the terms of relative almost-truthfulness and almost prime disjunction by terms obtained in accordance with Definition 3 and 5) follows from the validity of the following theorem.

Theorem 4. *Every implication suspicious of relative p-almost-truthfulness on the level* $1 - \alpha$ *is suspicious of p-almost-truthfulness on the same level.*

To prove this theorem we shall use the following lemma.

Lemma. *The following statement holds for the F-distribution (Fisher): If* $f_1 \leqq f_2$, *then* $f_1 F_\alpha[f_1, f] \leqq f_2 F_\alpha[f_2, f]$, *where* f, f_1, f_2 *are degrees of freedom and* $F_\alpha[f_i, f]$, $i = 1, 2$ *is* α*-quantil.*

Proof of Theorem 4: The number of objects fulfilling A, $\neg A$, $A \,\&\, B$, $A \,\&\, \neg B$, $\neg A \,\&\, B$, $\neg A \,\&\, \neg B$ in model $\mathscr{M}$ are denoted by m_1, m_0, m_{11}, m_{10}, m_{01}, m_{00}, respectively. We must prove $p \leqq \bar{p}(A \,\&\, B, m_A) \to p \leqq \bar{p}(A \to B, m)$. This implication will be valid if the inequality

$$\bar{p}(A \,\&\, B, m_A) \leqq \bar{p}(A \to B, m) \tag{1}$$

is fulfilled.

We have $m_{A\to B} = m_{11} + m_{01} + m_{00}$, $m = m_1 + m_0$ and $m_0 = m_{01} + m_{00}$, thus $m - m_{A\to B} = m_1 - m_{11} = m_{10}$ for any m_0. $\bar{p}(A \,\&\, B, m_A)$ and $\bar{p}(A \to B, m)$ are solutions of equations

$$\sum_{\nu=0}^{m_{11}} \binom{m_1}{\nu} p_1^\nu (1 - p_1)^{n_1 - \nu} = \alpha \quad \text{and} \quad \sum_{\nu=0}^{m_{A\to B}} \binom{m}{\nu} p_2^\nu (1 - p_2)^{m - \nu} = \alpha\,.$$

Using the non-complete β-function and the distribution function of F-distribution we obtain for the solution of the equation

$$\sum_{\nu=0}^{x} \binom{n}{\nu} p^\nu (1 - p)^{n - \nu} = \alpha$$

the following expression

$$\bar{p}_x(n) = \frac{(x + 1) F_\alpha[2(x + 1), 2(n - x)]}{(n - x) + (x + 1) F_\alpha[2(x + 1), 2(n - x)]}\,. \tag{2}$$

With regard to (2), the inequality (1) can be writetn in the form

$$\frac{(m_{11} + 1) F_\alpha[2(m_{11} + 1), 2(m_1 - m_{11})]}{(m_1 - m_{11}) + (m_{11} + 1) F_\alpha[2(m_{11} + 1), 2(m_1 - m_{11})]} \leqq$$

$$\leqq \frac{(m_{A\to B} + 1) F_\alpha[2(m_{A\to B} + 1), 2(m - m_{A\to B})]}{(m - m_{A\to B}) + (m_{A\to B} + 1) F_\alpha[2(m_{A\to B} + 1), 2(m - m_{A\to B})]}\,.$$

This inequality is equivalent to

$$2(m_{11} + 1) F_\alpha[2(m_{11} + 1), 2(m_1 - m_{11})] \leqq$$

$$\leqq 2(m_{A\to B} + 1) F_\alpha[2(m_{A\to B} + 1), 2(m - m_{A\to B})]\,.$$

Notice that $f = 2(m - m_{A\to B}) = 2(m_1 - m_{11}), f_1 = 2(m_{11} + 1), f_2 = 2(m_{A\to B} + 1)$ and that $m_{11} + 1 \leqq m_{A\to B} + 1$. The last inequality then follows from the lemma.

Proof of the lemma. Let $f_1, f_2, f, f_2 > f_1$ be natural numbers, let $\chi^2(f_1)$, $\chi^2(f_2 - f_1)$ and $\chi^2(f)$ be statistic independent from each other with the χ^2-distribution and with f_1, $f_2 - f_1$ and f degrees of freedom, respectively. Let $\chi^2(f_2)$ be a statistic with the χ^2-distribution and with f_2 degrees of freedom independent from $\chi^2(f)$. Let $W_i = f_i F[f_i, f]$, $i = 1, 2$, be a statistic having F-distribution with f_i, f.degrees of freedom multiplied by f_i. By the definition of F-distribution we have $W_i = \chi^2(f_i)(\chi^2(f)/f)^{-1}$. By properties of χ^2-distribution we obtain $\chi^2(f_2) = \chi^2(f_1) + \chi^2(f_2 - f_1)$. Thus $W_2 = (\chi^2(f_1) + \chi^2(f_2 - f_1))(\chi^2(f)/f)^{-1}$ consequently $W_2 = W_1 + \chi^2(f_2 - f_1)(\chi^2(f)/f)^{-1}$. We have $P(\chi^2(f_2 - f_1)(\chi^2(f)/f)^{-1} > 0) = 1$ hence $P(W_2 > W_1) = 1$ and from this follows that $W_1 \geqq w \Rightarrow W_2 \geqq w$ with probability 1. Hence $P(W_1 \geqq w) \leqq P(W_2 \geqq w)$, which is equaivlent with $P(W_1 < w) \leqq P(W_2 < w)$. By the definition of α-quantil the last inequality implies $W_1(\alpha) \leqq W_2(\alpha)$, where $W_i(\alpha)$ is α-quantil of distribution of the statistic W_i and thus $W_i(\alpha) = f_i F_\alpha[f_i, f]$, $i = 1, 2$. Consequently, $f_1 F_\alpha[f_1, f] \leqq f_2 F_\alpha[f_2, f]$, Q.E.D.

For $f_2 = f_1$, the the assertion of the lemma is evident.

We shall now pay attention to the validity of Theorem 8 from [3] (or Theorem 3 from [2]). We shall interpret the concept of significance as suspiciousness of the antecedent of $\bar{p}_s$-almost-truthfulness on the level $1 - \alpha$, where $\bar{p}_s$ is determined with regard to m, s, α in a usual way. We shall modify the definition of the good antecedent in accordance with Definition 5 and the validity of Theorem 8 from [3] (or Theorem 3 from [2]) will follow from the the following theorem.

Theorem 5. *If a formula $K_1 \,\&\, K_2 \to D$ is suspicious of p-almost-truthfulness on the level $1 - \alpha$, then the formula $K_1 \to (\neg K_2 \vee D)$ is also suspicious of the p-almost-truthfulness on the same level.*

Proof. The theorem follows from the inequality

$$\bar{p}(K_1 \,\&\, K_2 \,\&\, D, m_{K_1 \& K_1}) \leqq \bar{p}(K_1 \,\&\, (\neg K_2 \vee D), m_{K_1}),$$

which can be proved in the same manner as the inequality (1), since

$$m_{K_1} - m_{K_1 \& (\neg K_2 \vee D)} = m_{K_1 \& K_2} + m_{K_1 \& \neg K_2} - m_{K_1 \& \neg K_2} - m_{K_1 \& D} + m_{K_1 \& \neg K_2 \& D} =$$

$$= m_{K_1 \& K_2} - m_{K_1 \& \neg K_2 \& D} - m_{K_1 \& K_2 \& D} + m_{K_1 \& \neg K_2 \& D} = m_{K_1 \& K_2} - m_{K_1 \& K_2 \& D}.$$

Note 1. In the Part II of this paper the expression $P(\omega_1/\omega_2)$ denoted the probability of a random event ω_1 condicioned by a random event ω_2, the expression $P(.../..., p_A)$ denoted the probability under the assumption that p_A is the actual value of parameter.

Note 2. Theorems analogous to Theorems 4 and 5 can be stated and proved also for the probable p-almost-truthfulness.

For processing experimental results on a computer it is unsuitable to determine for numbers m_A upper limits $\bar{p}_A$, since it would be necessary to put in the computer a whole table, which should be always done only for a certain α. For the determination of formulae suspicious of p-almost-truthfulness it would be, however, sufficient – given α, p and m – to determine in the tables the number $\overline{m} = [x]$ such that $p = f(m, x, \alpha)$, where the shape of the function f has been expressed explicitly in the proof of Theorem 4 ($f(m, x, \alpha)$ is a growing continuos function of x and therefore it is possible to invert it). Then the criterion $p \leqq \bar{p}_A$ is replaced by an equivalent criterion $\overline{m} \leqq m_A$. In the case of implication the matter is not so easy, since x depends on the number of objects in the model $\mathcal{M}'$, which is variable, and we should then again put a whole table into the computer. We prefer another way of determining $\overline{m}$ and $\overline{m}_A$. For the model $\mathcal{M}'$, $\overline{m}_A$ plays the same role as m for $\mathcal{M}$. Let p and α be given. The computer computes successively the sums

$$\sum_{v=0}^{n} \binom{m}{v} p^v (1-p)^{m-v} \quad \text{for} \quad n = 1, 2, \ldots$$

until it reaches a number n_1 such that

$$\sum_{v=0}^{n_1} \binom{m}{v} p^v (1-p)^{m-v} \leqq \alpha \quad \text{but} \quad \sum_{v=0}^{n_1+1} \binom{m}{v} p^v (1-p)^{m-v} \geqq \alpha .$$

This n_1 is $\overline{m}$. Analogously, for every m_A the sums

$$\sum_{v=0}^{n} \binom{m_A}{v} p^v (1-p)^{m_A-v}, \quad n = 1, 2, \ldots,$$

are examined. The numbers $\overline{m}_A$ obtained in this way would, of course, have to be inserted into the memory of the computer. However we can modify once more our criteria. For every A, B, the computer computes the sum

$$\sum_{v=0}^{m_{A\&B}} \binom{m_A}{v} p^v (1-p)^{m_A-v} .$$

If that sum is $\geqq$, then $\bar{p}(A \,\&\, B, m_A) \geqq p$ and $A \to B$ is suspicious of relative p-almost-truthfulness on the level $1 - \alpha$. If it is $<$, then $\bar{p}(A \,\&\, B, m_A) < p$ and the implication is omitted. Using this method we can thus verify suspiciousness of relative p-almost-truthfulness directly without putting any table into the computer and its calling up during the computation. This is the reason why this method can easily be included into the GUHA algorithm, except the substitution of the criterion $p \,.\, m \leqq m_A$ by the criterion $\overline{m} \leqq m_A$ and the criterion $m_{A\&B}/m_A \geqq p$ by the criterion

$$\sum_{v=0}^{m_{A\&B}} \binom{m_A}{v} p^v (1-p)^{m_A-v} \geqq \alpha .$$

In the Definition 4 we have defined probable p-almost-truthfulness on the level $1 - \alpha$. We could formulate Theorems 4 and 5 for this concept as well and repeat the reasoning from the beginning of this paragraph. This concept could be useful, if we would need confident assertions.

Let us add some remarks on the applicability of methods from [3] (or [2]) and of method described in this paper. The former method is applicable on models with a big number of objects, for example for $m = 1000$ we have $\bar{p} - p_* = 0{,}02$ and the estimation $\hat{p}$ is sufficiently exact. But for $m = 30$ we have $\bar{p} - p_* = 0{,}30$ i.e. 30% ($\alpha = 0{\cdot}05$, $p \geqq 0{\cdot}90$). Thus the confidence interval is large and the point estimation is nonconfident. The modified method proposed in this paper seams to be useful for applications in three cases. All these cases are similar in that the method is applicable to small models $\mathcal{M}$, but they differ in the way how these models are obtained. The most essential of these cases is that where it is necessary to work with a small model, because we have for different reasons (price of an experiment, infrequented illness etc.) at our disposal only a small number of objects. In such a case we generate on this model hypotheses by the modified method and we study further objects (patients) or we make further experiments with respect to the obtained semantically interesting hypotheses. The second and third cases are similar in that we have at our disposal a large model. In the second case, this model is not representative. We can, however, make a selection from this model by certain methods to obtain a small, more representative submodel. This case is important, because in practice we often meet large but non-representative models. In the third case we have at our disposal a large and sufficiently representative model. We select from it by a random selection a small submodel and process it by the modified method. Then we study the former model only from the point of view of some selected interesting hypothesis. It is necessary to decide, whether the time and charges saved thanks to the fact that we work with a small number of objects outweigh the difficulties with the interpretation of computer results, because the number of generated hypothesis grows with decreasing number of objects in the model.

(Received April 24, 1970.)

REFERENCES

[1] P. Hájek, I. Havel, M. Chytil: GUHA — metoda systematického vyhledávání hypotéz. Kybernetika *2* (1966), 1, 31—47.

[2] P. Hájek, I. Havel, M. Chytil: GUHA — metoda systematického vyhledávání hypotéz II. Kybernetika *3* (1967), 5, 430—437.

[3] P. Hájek, I. Havel, M. Chytil: The GUHA method of automatic hypotheses determination. Computing *1* (1966) fasc. 4, 293—308.

Statistická interpretace a modifikace metody GUHA

TOMÁŠ HAVRÁNEK

Hlavní myšlenka v tomto článku navržené modifikace a interpretace metody GUHA z práce [2] spočívá v definování pojmu skoropravdivosti a relativní skoropravdivosti ne vzhledem k modelu, ale vzhledem k celému příslušnému universu, a v navržení statistických kritérií, z nichž první zaručuje, že podle něj vybereme všechny formule, které jsou skoropravdivé ve smyslu definice 1 tohoto článku (to jest platí pro ně, že $p_A \geqq p$, kde p je předem zadané číslo, $p \in \langle 0, 1\rangle$, a p_A skutečná pravděpodobnost platnosti takové formule A pro libovolný objekt universa) s pravděpodobností $1 - \alpha$, kde α, $\alpha \in \langle 0, 1\rangle$, je libovolně malé. Druhé kritérium zaručuje, že podle něho vybrané formule jsou ve smyslu zde definovaném skoropravdivé s pravděpodobností $1 - \alpha$. Podobná kritéria jsou zavedená i pro případ relativní skoropravdivosti. Všechna kritéria jsou závislá na počtu objektů v modelu $\mathcal{M}$. K jejich konstrukci se používá místo bodového odhadu, který je používán v [2], intervalů spolehlivosti pro parametr binomického rozdělení. Dále je v článku dokázána platnost vět 1 až 3 z [2], upravených ve smyslu zde zavedených kritérií. Na závěr článku jsou probrány způsoby praktického ověřování platnosti kritérií tak, aby byl co nejméně narušen algoritmus metody GUHA z [2], a jsou probírány případy, kdy je vhodné používat modifikované metody zde předložené.

Tomáš Havránek, katedra matematické statistiky MFF UK (Faculty of Mathematics and Physics, Charles University), Sokolovská 83, Praha 8.