KYBERNETIKA — VOLUME 18 (1982), NUMBER 5

THREE SEMANTICAL INTERPRETATIONS OF A STATISTICAL THEOREMHOOD TESTING PROCEDURE

IVAN KRAMOSIL

On an informal level, three possibilities are investigated, how the results offered by a statistical deducibility testing procedure may be semantically interpreted. The interpretations differ with respect to the fact whether the logic or meta-logic in question is of classical, many-valued, probabilistic or other nature.

In this paper we concern our attention to the classical propositional calculus formalized by the means of one of its usual formalizations. All the usual propositional connectives are supposed to be at our disposal; the propositional indeterminates are denoted by $p_1, p_2, \ldots, \mathscr{F}$ denotes the set of all well-formed formulas of this calculus, $\mathscr{F}_n \subset \mathscr{F}$ denotes the set of formulas in which only the indeterminates $p_1, p_2, \ldots, p_n$ may occur. $\mathscr{K}_n$ denotes the set of all formulas of the form

$$(1) \qquad \bigwedge_{i=1}^{2^n} \bigvee_{j=1}^{n} a_{ij}, \; a_{ij} \in \{p_1, \neg p_1, p_2, \neg p_2, \ldots, p_n, \neg p_n\},$$

(for the sake of simplicity we write often only $\{a_{ij}\}_{i=1,j=1}^{2^n, n}$), $\neg p_i$ is the negation of p_i, indeterminates together with their negations are called *literals*. Denote by $\mathscr{T} \subset \mathscr{F}$ the set of all theorems (tautologies), set $\mathscr{T}_n = \mathscr{T} \cap \mathscr{F}_n$, $\mathscr{T}\mathscr{K}_n = \mathscr{T} \cap \mathscr{K}_n$, clearly,

$$(2) \qquad \mathscr{K} = \bigcup_{n=1}^{\infty} \mathscr{K}_n, \; \mathscr{T}\mathscr{K} = \mathscr{T} \cap \mathscr{K} = \bigcup_{n=1}^{\infty} \mathscr{T}\mathscr{K}_n, \; \mathscr{K}_n \subset \mathscr{F}_n, \; \mathscr{T}\mathscr{K}_n \subset \mathscr{T}_n, \quad n = 1, 2, \ldots$$

There is a well-known fact that for each formula $A \subset \mathscr{F}_n$ there exists $B \in \mathscr{K}_n$ such that the equivalence $A \leftrightarrow B$ belongs to $\mathscr{T}_n$ (B is a conjunctive normal form for A). It is why we shall limit ourselves, in what follows, to $\mathscr{K}$ as the basic set of formulas.

A pair $\{a_i, a_j\}$ of literals is called *complementary*, if a_i is $\neg a_j$ or a_j is $\neg a_i$. A finite sequence $\{a_1, a_2, \ldots, a_n\}$ of literals is called *closed*, if it contains at least one complementary pair, in the opposite case it is called *open*.

As can be easily seen, a formula $\{a_{ij}\}_{i=1,j=1}^{2^n, n} \in \mathscr{K}$ is a theorem (i.e. a propositional tautology) iff all rows $\langle a_{i_1}, a_{i_2}, \ldots, a_{i_n} \rangle$ are closed. Hence, there is a simple deterministic verification procedure for formulas from $\mathscr{K}$: to check row after row and to check, in each row, all pairs of literals (until an open row is found, in the case of classical logic). The maximal number of pairs which must be, eventually, tested for complementarity is $\frac{1}{2}n(n-1)\,2^n$ for formulas from $\mathscr{K}_n$, as can be easily seen.

Because of the exponential complexity of this deterministic verification procedure we may try to apply the basic idea of the so called *probabilistic algorithms* – i.e., to reduce the computational complexity of the verification procedure by admitting the possibility of error under the condition that the complexity savings are "significant" and the probability of error is "acceptable" or "small", in a sense. For this sake, let us consider a simple statistical verification procedure $\mathscr{S}(k) = \langle B(k), f(k)\}$, $k = 1, 2, \ldots$. Here $B(k) = \{b_1^k, b_2^k, \ldots\}$ is a sequence of mutually independent and equally distributed random variables defined on a probability space $\langle \Omega, \mathscr{S}, \mathsf{P}\rangle$, taking their values in the set of positive integers and such that

$$(3)\qquad \mathsf{P}(\{\omega : \omega \in \Omega,\ b_i^k(\omega) = j\}) = 2^{-k}, \quad j = 1, 2, \ldots, 2^k, \quad i = 1, 2, \ldots, \quad k = 1, 2, \ldots$$

f is a mapping of the set $N = \{1, 2, \ldots\}$ into itself. Let $A \in \mathscr{K}_k (A = \{a_{ij}\}_{i=1,j=1}^{2^k\ \ k})$ be a tested formula, then $\mathscr{S}(k)$ runs as follows.

Sample at random natural numbers $b_i^k(\omega)$, $i = 1, 2, \ldots, f(k)$, set $i_n = b_i^k(\omega)$, $n = 1, 2, \ldots, f(k)$. Denote by $S_1 = S_1(\omega)$ the number of rows among $\{\langle a_{i_n1}, a_{i_n2}, \ldots, a_{i_n,k}\}\}_{n=1}^{f(k)}$ which are closed and set $S_0 = S_0(A, \mathscr{S}, \omega) = (f(k))^{-1}\,S_1(\omega)$. I.e., S_0 is the relative frequency of closed rows among the random sample made from the total collection of 2^k rows. In the classical propositional calculus the only fact about S_0, which is of interest, is whether $S_0 < 1$ (in this case we have discovered at least one open, row, hence, A certainly cannot be a theorem), or whether $S_0 = 1$ (in this case we may proclaim A for theorem with a risk of error). Clearly, in such a case the procedure $\mathscr{S}(k)$ may be modified in such a way that the rows are sampled step by step until $f(k)$-th one and the occurrence of the first open row enables to finish the test sooner. However, the main purpose of this contribution will be to discuss various possible interpretations of the value S_0 in connection with possible interpretations of the propositional languages $\mathscr{F}$ and $\mathscr{K}$.

1. CLASSICAL TWO-VALUED PROPOSITIONAL CALCULUS

In this case the interpretation of formulas from $\mathscr{F}$ is defined as a special mapping which takes $\mathscr{F}$ into a two-element set $T = \{0\ (\mathit{false}),\ 1\ (\mathit{true})\}$ of truth values. This mapping will be denoted by $\mathscr{I}_c$ and is defined in the usual way, i.e., for $A \in \mathscr{F}$, $\mathscr{I}_c(A) = 1$ iff A is a classical propositional tautology. As follows from the considerations above, if $A \in \mathscr{K}$, then $\mathscr{I}_c(A) = 1$ iff all rows in A are closed. Denoting by $r(A)$ the number of closed rows in $A \in \mathscr{K}$ and setting, for $A \in \mathscr{K}_n$, $R(A) = 2^{-n} r(A)$, we may write $\mathscr{I}_c(A) = \mathrm{Int}\,(R(A))$. It is a well-known fact of mathematical logic that in the case of classical two-valued propositional calculus the notions of tautology and theorem coincide, hence, the theoremhood testing problem can be reduced to the decision problem, whether $R(A) = 1$ or $R(A) < 1$ for $A \in \mathscr{K}$. As the number $S_0(A, \omega)$ defined above gives the relative frequency of closed rows in a random sample of rows from A, we may take $S_0(A, \omega)$ when deciding about $R(A)$. Namely, we define a decision function d taking $\mathscr{K} \times \Omega$ into T in this way: $d(A, \omega) = \text{Int}\,(S_0(A, \omega))$, hence

$$(4) \quad d(A, \omega) = 1 \quad \text{iff} \quad S_0(A, \omega) = 1\,, \quad \text{i.e.} \quad d(A, \omega) = 0\,, \quad \text{iff} \quad S_0(A, \omega) < 1\,.$$

The quality of this decision function may be expressed as follows:

Theorem 1. Let $A \in \mathscr{K}_n$, then $\mathsf{P}(\{\omega : \omega \in \Omega, d(A, \omega) = 1\}) = 1$ for $A \in \mathscr{T}\mathscr{K}_n$, $\mathsf{P}(\{\omega : \omega \in \Omega, d(A, \omega) = 1\}) = (R(A))^{f(n)}$ for $A \in \mathscr{K}_n - \mathscr{T}\mathscr{K}_n$.

Proof. If $A \in \mathscr{T}\mathscr{K}_n$, then all rows in A are closed, hence, also all the rows sampled by $\mathscr{S}$ must be closed, so $S_0(A, \omega) = d(A, \omega) = 1$, i.e. each theorem is estimated as theorem. If $A \in \mathscr{K}_n - \mathscr{T}\mathscr{K}_n$, then not all rows are closed, but it is possible that the random sample contains only the closed rows which leads to $S_0(A, \omega) = 1$ and to the wrong decision that $d(A, \omega) = 1$. Because of the conditions imposed to the random variables b_i^n the probability of sampling a closed row by b_1^n equals just $R(A)$, the supposed statistical independence of all the used $f(n)$ samples gives the result, i.e. the probability of error for $A \in \mathscr{K}_n - \mathscr{T}\mathscr{K}_n$ equals $(R(A))^{f(n)}$. □

Let us remark that the decision function d is the best one among all decision functions taking $\mathscr{K}_n$ into T and defined only by the means of the value $S_0(A, \omega)$. More precisely, if the tested formula A is supposed to be sampled at random from $\mathscr{K}_n$ and if the probability of sampling a theorem, i.e. a formula from $\mathscr{T}\mathscr{K}_n$ exceeds $(R(A))^{f(n)}$, then, with respect to all formulas from $\mathscr{K}_n$ with the same $R(A)$, the decision function d has the minimal probability of error. If the probability of sampling a theorem from $\mathscr{K}_n$ does not exceed $(R(A))^{f(n)}$, then the optimal decision function for the same class of formulas consists in a priori setting $d(A) = 0$ without any further investigation. The argumentation of this paragraph can be precised and formalized using the well-known Neyman-Pearson theorem as done, for other statistical deducibility testing method, by Špaček in [1], cf. also the surveyal work [2].

2. MANY-VALUED PROPOSITIONAL CALCULUS

Here we shall limit ourselves to a particular interpretation $\mathscr{I}_m$ taking $\mathscr{F}$ into a countable set $T' = \{m \,.\, 2^{-n} : m = 0, 1, \ldots, 2^n,\ n = 1, 2, \ldots\}$ of truth values, here 0 and 1 belong to T' and have the same interpretation as above, i.e. *false* and *true*. Namely, we define, for $A \in \mathscr{F}$, $\mathscr{I}_m(A) = R(A')$, where A' is a conjunctive normal form of A, so $A' \in \mathscr{K}$. Each open row in A' (if any) can be set in a one-to-one correspondence with a falsifying combination of truth values for A (considered in the classical sense), so $R(A')$ and $\mathscr{I}_m$ do not depend on the particular conjunctive normal form of A. So the truth value of a formula from $\mathscr{F}$ equals, by $\mathscr{I}_m$, to the relative frequency of verifying truth-values combinations among all 2^n possible (for formulas A with $A' \in \mathscr{K}_n$); such an interpretation seems to be quite reasonable.

In this case the value $S_0(A, \omega)$ can be used in two ways: Either as a statistics on the ground of which we decide between an a priori hypothesis that $\mathscr{I}_m(A) = p_1$ against an a priori alternative that $\mathscr{I}_m(A) = p_2 \neq p_1$, or we may use $S_0(A, \omega)$ in order to construct a point estimate $\bar{\mathscr{I}}_m(A)$ of $\mathscr{I}_m(A)$.

Theorem 2. Let $A \in \mathscr{K}_n$, let $\mathscr{H}$ be the hypothesis that $\mathscr{I}_m(A) = p_1$, let $\mathscr{A}$ be the alternative that $\mathscr{I}_m(A) = p_2 < p_1$. Set $N = f(n)$ and define, for $M \leqq N$, a decision function $d_1 = d_1(M, N, A, \mathscr{S}(n))$ in this way:

(5) $\quad d_1(M, N, A, \mathscr{S}(n)) = 1, \quad \text{if} \quad N \,.\, S_0(A, \omega) \geqq M \,, \quad \text{i.e. if} \quad S_1(A, \omega) \geqq M \,,$

$\qquad d_1(M, N, A, \mathscr{S}(n)) = 0 \quad \text{otherwise.}$

Here $d_1 = 1$ is interpreted as the acception of $\mathscr{H}$, $d_1 = 0$ as the acception of $\mathscr{A}$. Let $\alpha > 0$ be given, let u_α be the α-quantile of the normal (Gauss) distribution $N(0, 1)$, let A^* be a random variable defined on $\langle \Omega, \mathscr{S}, \mathsf{P} \rangle$ and taking its values in $\mathscr{K}_n$. Denote

(6) $$M_1 = \operatorname{Int}\left(N(u_{1-\alpha} \sqrt{(N^{-1}\, p_2(1 - p_2))} + p_2)\right) + 1 \,.$$

Then

(7)
$$\mathsf{P}(\{\omega : \omega \in \Omega, d_1(M_1, N, A^*(\omega), \mathscr{S}(n)) = 1\} \mid \{\omega : A^*(\omega) \in \mathscr{K}_n - \mathscr{T}\mathscr{K}_n\}) \leqq \alpha \,,$$

(8)
$$\mathsf{P}(\{\omega : \omega \in \Omega, d_1(M_1, N, A^*(\omega), \mathscr{S}(n)) = 0\} \mid \{\omega : A^*(\omega) \in \mathscr{T}\mathscr{K}_n\}) =$$
$$= \min_{M = M_1}^{N} \{\mathsf{P}(\{\omega : \omega \in \Omega, d_1(M, N, A^*(\omega), \mathscr{S}(n)) = 0\} \mid$$
$$\mid \{\omega : A^*(\omega) \in \mathscr{T}\mathscr{K}_n\}) \,.$$

Proof. Consider the classical statistical hypothesis testing problem with $\mathscr{H}$: $p = p_1$ against $\mathscr{A} : p = p_2$. We want to choose $M \in N$ such that the probability that $S_1(A, \omega) \geqq M$ were majorized by α supposing that $p = p_2$ and we look for the maximal M with this property in order to minimize the other type probability of error. Hence, we look for the minimal M such that $\sum_{i=0}^{M} \binom{N}{i} p_2^i (1 - p_2)^{N-i} \geqq 1 - \alpha$.

The well-known Central Limit Theorem of probability theory sounds that the random variable $S_0(A, \omega) = N^{-1}\, S_1(\omega)$ has, for $p = p_2$, approximately normal distribution with $\mu = p_2$ and $\sigma^2 = N^{-1}\, p_2(1 - p_2)$, i.e. S_0 has, approximately, the distribution function $\Phi((x - p_2)/\sqrt{(N^{-1}\, p_2(1 - p_2))})$, where Φ is the distribution function of the normal distribution $N(0, 1)$. The demand from the end of the last paragraph can be transformed into the form

$$\Phi\left(\frac{(M/N) - p_2}{\sqrt{(N^{-1}\, p_2(1 - p_2))}}\right) \geqq 1 - \alpha\,,$$

hence, $((M/N) - p_2)(\sqrt{(N^{-1}\, p_2(1 - p_2))})^{-1} \geqq u_{1-\alpha}$, and an easy calculation gives the value M, as stated above. The values of α-quantiles of $N(0, 1)$ are tabeled and can be found in statistical tabels. The problem can be solved also in a non-asymptotical way using the incomplete β-distribution. □

Intuitively said, the decision function d_1, saying that $\mathscr{I}_m(A) = p_1 > p_2$ iff $S_1(A, \omega) \geqq M$ (and saying that $\mathscr{I}_m(A) = p_2$ otherwise) has the minimal probability of wrong proclaiming that $\mathscr{I}_m(A) = p_2$ among all decision functions defined by S_1 and assuring that the probability of the other type of error, i.e. the probability of wrong proclaiming that $\mathscr{I}_m(A) = p_1$, is majorized by α. This asymmetric role of the both the types of error is usual in mathematical statistics and is motivated by the aim to majorize, first of all, the probability of the more dangerous (in a sense) type of error. In the case of the classical, two-valued interpretation $\mathscr{I}_c$ it would mean to majorize, by α, the probability that a non-theorem will be, wrongly, proclaimed to be a theorem. This type of error may be considered as the more dangerous as it may cause the set of formulas proclaimed to be theorems to be inconsistent and, hence, useless for a further use at least from the classical point of view. On the other hand, a wrongly refused theorem may be later rejoined with the set of theorems as their logical consequence. This way of argumentation may be applied even in the case of multivalued interpretation $\mathscr{I}_m$ supposing that the formulas with $\mathscr{I}_m(A) \geqq t$ (a given threshold value) are treated as "quasi-theorems". The assumption that the tested formula is sampled at random by A^* is necessary in order to optimize the decision function d_1 in the global sense, not only with respect to a particular formula.

In order to formulate the next assertion let us introduce the notion of *k-binarization* $B(a, k)$ for a real number a. $B(a, k)$ is defined as such a rational number of the form $m\,.\,2^{-k}$, for which $|a - m\,.\,2^{-k}|$ is minimal among all numbers of this form (k is fixed).

Theorem 3. Let $A \in \mathscr{K}_n$, consider the point estimation $\mathscr{I}_m(A)$ of $\mathscr{I}_m(A)$ defined as $\bar{\mathscr{I}}_m(A) = B(S_0(A, \omega), n)$. Then $\bar{\mathscr{I}}_m(A)$ is the optimal n-binarized point estimation of $\mathscr{I}_m(A)$ in the sense of the maximal likelihood principle. More precisely, if Pr (p, m) is the probability that $S_1(A, \omega) = m$ (i.e. that $S_0(A, \omega) = m\,.\,N^{-1}$) under the condition that $R(A)$ (the relative frequency of closed rows in A) is p, and if c_0 maximizes Pr (p, m) taken as a function of p, then $\bar{\mathscr{I}}_m(A) = B(c_0, n)$.

Proof. Let $N = f(n)$, $m = S_1(A, \omega)$, $p = R(A)$, then the probability of sampling at random just m closed among N samples equals $\binom{N}{m} p^m (1 - p)^{N-m}$ because of the supposed statistical independence and equal distribution of the random variables b_i^k. In order to find the maximizing value of p for this probability taken as a function of p we take the derivation (with respect to p) of the logarithm of this probability and set it to be zero. Hence,

$$(10) \qquad \frac{\mathrm{d}}{\mathrm{d}p}\left[\log\left(\binom{N}{m}p^m(1-p)^{N-m}\right)\right] =$$

$$= \frac{\mathrm{d}}{\mathrm{d}p}\left[\log\binom{N}{m} + m\log p + (N-m)\log(1-p)\right] = \frac{m}{p} - \frac{N-m}{1-p} = 0\,,$$

and solving this equation we obtain $p = m\,.\,N^{-1} = S_1(A,\omega)\,(f(n))^{-1} = S_0(A,\omega)$. So $\mathscr{I}_m(A) = B(S_0(A,\omega), n)$ satisfies the assertion of Theorem 3. □

3. TWO-VALUED PROPOSITIONAL CALCULUS WITH MANY-VALUED META-CALCULUS

Let us briefly mention another possibility how to interprete the value $S_0(\omega)$, this time in a meta-level way. Consider the classical two-valued interpretation of formulas from $\mathscr{F}$ as explained above in (1). Moreover, we may associate with $\mathscr{F}$ a meta-language $\mathscr{F}^*$ in which some assertions concerning the formulas from $\mathscr{F}$ or $\mathscr{K}$ can be formalized. Namely, we suppose that assertions of the type "$A \in \mathscr{F}$", i.e. assertions proclaiming the deducibility of a formula A, belong to $\mathscr{F}^*$. Hence, we may consider a many-valued truth evaluation of formulas from $\mathscr{F}^*$, namely such evaluation which ascribes to the formula "$A \in \mathscr{T}$" of $\mathscr{F}^*$ the truth value $S_0(A,\omega)$. Because of the statistical character of the values $S_0(A,\omega)$ this truth evaluation has not necessarily the usual features of probabilistic logics, it may happen, e.g., that $S_0(A,\omega) > S_0(A \vee B,\omega)$ for some $A, B \in \mathscr{K}$, $\omega \in \Omega$. On the other hand, the statistical stability expressed in the so called "laws of large numbers" assures, that these discrepancies will tend to zero, in a sense, with $f(n)$ increasing. More precisely, with probability one, for each $A, B \in \mathscr{K}$,

$$(11) \qquad \mathsf{P}(\{\omega : \omega \in \Omega,\ \lim_{f(n)\to\infty} S_0(A,\omega) \leqq \lim_{f(n)\to\infty} S_0(A \vee B,\omega)\}) = 1\,.$$

This case of statistical meta-level truth value evaluation would deserve a more detailed investigation, which would exceed, however, the limited extent of this contribution.

The results, notions and methods from the domains of mathematical logic, probability theory and mathematical statistics as used here are of very simple nature and can be found in almost all textbooks of undergraduate level of the corresponding branches of mathematics; it is why we do not introduce here special references. Some more investigations concerning the case (1) can be found in [3], namely from the point of view of computational complexity connected with the decision making for various kinds of the function $f(n)$. [2] can serve as a survey of statistical methods in theorem proving and as a source of more detailed references.

(Received November 9, 1981.)

REFERENCES

[1] A. Špaček: Statistical estimation of provability in Boolean logics. In: Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes. NČSAV, Prague 1960, pp. 609—626.

[2] I. Kramosil: Statistical approach to proof theory. Supplement to Kybernetika *15* (1979), 98 pp.

[3] I. Kramosil: Computational complexity of a statistical verification procedure for propositional calculus. In: Third Czechoslovak-Soviet-Hungarian Seminar on Information Theory, Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, 1980, pp. 123—130.

RNDr. Ivan Kramosil, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences). Pod vodárenskou věží 4, 182 00 Praha 8. Czechoslovakia.