KYBERNETIKA — VOLUME 16 (1980), NUMBER 3

Statistical Theory of Logical Derivability

JAN ŠINDELÁŘ

A generalization of notions like "theory" in the sense of mathematical logic, "Markov (normal) algorithm" etc., is described; the new notion is called "formal system". The generalization of notions "validity", and also of notion "provability" of formulas is the notion "derivability". Derivability is studied using statistical methods, analogous to the methods used in the statistical theory of deducibility testing.

1.1. STATISTICAL THEORY OF THEOREM PROVING AND VALIDITY TESTING

Statistical theory of theorem proving studies the possibility of using statistical methods for provability testing (or for non-provability testing) of theorems of some theory $\mathscr{T}$. Usually it is proceeded as follows: we suppose that the provability of some formulas of $\mathscr{T}$ is known. Then we try to decide (or to estimate), with the help of statistical methods, whether some other formulas of $\mathscr{T}$ are provable in $\mathscr{T}$ or not, i.e. we know some relation of "partial provability" and try to approximate (with the help of statistical methods) the relation "provability".

Here two mathematical branches meet – mathematical logic and statistics. It is possible to divide the papers concerning the statistical theory of theorem proving into three groups with respect to use of apparatus of the disciplines mentioned above: the first uses mainly the specific properties of particular theories of mathematical logic and only the basic statistical methods, in the other is on the contrary specially emphasized the statistical aspect of the problem, in the third both disciplines are used in a comparable manner.

Objects investigated in mathematical logic and statistics as well as methods of investigation of these objects are of a substantially different and non-homogeneous nature. Hence, it is quite difficult to combine them and completely use the very developed modern apparatus and possibilities of both these disciplines in the statistical theory of theorem proving.

The effort of the present paper is a relatively general point of view to the questions mentioned above.

As mentioned above, in the statistical theory of theorem proving some formalized theory $\mathscr{T}$ is assumed and the relation of provability of formulas of $\mathscr{T}$ is investigated.

The point of view of this paper pretends to generality, hence there are studied more general objects than formulas and more general relations than provability. Hence, the basic studied structure will not be a theory in the sense of mathematical logic, but it will be a more general structure. We shall call this structure "formal system" and corresponding relation "derivability relation" (it is why the title of this paper is "Statistical Theory of Logical Derivability").

Remark. Sometimes it is easier (or reasonable) to study the validity of formulas of $\mathscr{T}$ instead of their provability; in many papers is chosen this method (and then the concept of "statistical theory of theorem proving" or of "deducibility testing" etc. have not been quite right). The point of view of the present paper generalizes both these approaches to the problem (cf. Examples 1, 2).

Definition D1. The triple $F = \langle B, \{o_p\}_{p=1}^{\infty}, V\rangle$ is called *formal system*, if

(I) $B \neq \emptyset$, B is a finite or countable set;
(II) $V \neq \emptyset$, V is a class;
(III) $\emptyset \neq o_p \subseteq B \times V$ for $p = 1, 2, 3, \ldots$;
(IV) $o_p \subseteq o_{p+1}$ for $p = 1, 2, 3, \ldots$.

Members of B are called *formulas*, members of V *assumptions*, classes o_p are called *partial derivability relations* ($p = 1, 2, \ldots$).

Class

(V) $$o_O = \bigcup_{p=1}^{\infty} o_p$$

is called the *derivability relation*.

For $b \in B$, $v \in V$ we shall denote

$$bo_pv$$

instead of $\langle b, v\rangle \in o_p$.

For $b \in B$, $p = 0, 1, 2, \ldots$ we shall denote

$$b^p = \{v \mid bo_pv, v \in V\}.$$

Formula $b \in B$ is called *derivable* (in formal system F) if

$$b^O = V;$$

otherwise b is called *non-derivable* in F.

The set of derivable formulas will be denoted by T, i.e.

$$T = \{b \in B \mid b^O = V\} .$$

Formula $b \in B$ is called *derivable* (*non-derivable*) *from* $v \in V$, if $bo_O v$ holds (does not hold).

Remark. We distinguish "set" and "class"; for understanding of the studied problems this distinguishing is not essential.

The intuitive sense of notions mentioned in D1 can be like this:
the set B represents formulas investigated for derivability (provability, validity etc.), class V represents objects through which derivability is investigated. Relations o_p are relations of partial derivability; they represent the level of our knowledge about the actual system in which the relation o_O of derivability is investigated. We suppose that (IV) holds, i.e. that with increasing p our knowledge about actual system arises.

Example 1. Let B be the set of closed formulas of some theory $\mathcal{T}$, V be the class of models of $\mathcal{T}$; $bo_O v$ holds iff formula b is satisfied (true) in model V of $\mathcal{T}$. Relations o_p ($p = 1, 2, \ldots$) are chosen to satisfy (III), (IV), (V). Then formula b of $F = \langle B, \{o_p\}, V\rangle$ is derivable in F iff b is true in each model of $\mathcal{T}$.

Example 2. Members of set B and class V are formulas of some theory $\mathcal{T}$; for $p = 1, 2, \ldots$ $bo_p v$ holds iff b is provable in $\mathcal{T}[v]$ at most in p steps (i.e. b can be proved from v in at most p steps). Then $bo_O v$ holds iff b is provable in $\mathcal{T}[v]$.

If b is a theorem of $\mathcal{T}$, then b is a theorem of $\mathcal{T}[v]$ for every $v \in V$, hence $bo_O v$ for every $v \in V$, so $b^O = V$, hence b is derivable in $F = \langle B, \{o_p\}, V\rangle$.

If b is derivable in F, then $bo_p v$ for every $v \in V$; let w be a logical axiom. Then $bo_O w$, hence $bo_q w$ for some $q \in N^+$ (by (V)), so that b is provable in $\mathcal{T}[w]$ (and $\mathcal{T}[w] = \mathcal{T}$ because w is logical axiom), hence b is theorem of $\mathcal{T}$.

We have proved that in Example 2 b is derivable in F iff b is theorem of $\mathcal{T}$.

Example 3. Let B be a set of principles investigated for validity. Class V is some (countable) class of particular (concrete) situations. If principle $b \in B$ is valid in every particular situation (from V), then we called it generally valid principle. But we are not usually able to verify validity of each principle in every particular situation. So we verify the validity of principles $b \in B$ only in finite particular situations; in this way we obtain relations o_p ($p = 1, 2, 3, \ldots$). We suppose that with the time passing by (i.e. with increasing p) our knoweldge about validity of principles in particular situations increases (IV). Moreover we suppose that with the time passing by we shall be able to verify each principle in every situation (V).

It can be easily proved that $b^O = V$ holds iff b is a generally valid principle. Hence, formula $b \in B$ is derivable in the formal system $F = \langle B_1\{o_p\}, V\rangle$ iff b is a generally valid principle.

During this paper two assumptions (A, C) are introduced, which can or need not be fulfilled. In the parts of the paper in which Assumption A or Assumption C will be supposed to be fulfilled this will be explicitly stated; in other parts validity of Assumptions A, C is not supposed.

Assumption A. Let $F = \langle B, \{o_p\}, V\rangle$ be a formal system. Let $V_O \subseteq V$ be such, that: if $b \in B$ is not derivable in F, then b is not derivable from any $v \in V_O$ (i.e. for all $b \in B - T$, $v \in V_O : \neg(bo_Ov)$).

For every $p \in N^+$, $o_p \subseteq o_O$ holds; hence, if Assumption A is in F satisfied, then $\neg(bo_pv)$ holds for every $b \in B - T$, $v \in V_O$, $p \in N$.

Example 4. Let B and V consist of formulas of a formalized theory $\mathscr{T}$. Let bo_Ov holds iff b can be proved from v (i.e. if b is provable in $\mathscr{T}[v]$; cf. Example 2). V_O contains axioms of $\mathscr{T}$ (or V_O contains some theorems of $\mathscr{T}$). In such a case, if $v \in V_O$ and $b \in B$ is non-theorem of $\mathscr{T}$, then b cannot be proved in $\mathscr{T}[v]$; hence, Assumption A is fulfilled.

Asumption A can be fulfilled in every formal system, if we put $V_O = \emptyset$. But the case $V_O \neq \emptyset$ is essential.

Assumptions (members) of V_O have an important role. If $v \in V_O$ and $b \in T$ and $b' \in B - T$ (i.e. b is derivable and b' not derivable), then hold: bo_Ov and $\neg(b'o_Ov)$; moreover, for every $p \in N^+$ holds $\neg(b'o_pv)$.

1.3. RANDOM VARIABLES ON V

First, we shall study the derivability with respect to one random variable on V.

Definition D2. Let $F = \langle B, \{o_p\}, V\rangle$ be a formal system. We define

$$\mathscr{V} = \sigma\{b^p \mid b \in B,\ p = 1, 2, 3, \ldots\} = \sigma\{\{v \mid bo_pv\} \mid b \in B,\ p \in N^+\},$$

where $\sigma\{\ldots\}$ is the smallest σ-algebra over the set $\{\ldots\}$.
Let $\langle \Omega, S, P\rangle$ be a probability space. By v we denote a $\mathscr{V}$-measurable mapping $v\colon \Omega \to V$.

In the following by ω an element of Ω is denoted. It is obvious that (by (VI))

$$P[v(\omega) \in b^p] = P[bo_p\, v(\omega)] \tag{1}$$

holds. The value of $P[bo_p\, v(\omega)]$ is the probability that b is derivable (with respect to o_p and v).

Assumption C. Let the Assumption A be satisfied in formal system $F = \langle B, \{o_p\}, V\rangle$. Let $V_O \in \mathscr{V}$ and

$$0 < y = P[v(\omega) \in V_O] \tag{2}$$

Assumptions from V_O significantly distinguish derivable and not derivable formulas.
If Assumption C holds in formal system F then there is a sufficient number of such "important" assumptions in F.

Lemma L1. Let $F = \langle B, \{o_p\}, V\rangle$ be a formal system, v be as in D2. Then

1°. For $p \to \infty$, $b^p \nearrow b^O$ holds.
2°. For $p \to \infty$, $P[bo_p\, v(\omega)] \nearrow P[bo_O\, v(\omega)]$ holds.
3°. For every derivable $b \in B$, if $p \to \infty$ then $P[bo_p\, v(\omega)] \nearrow 1$ holds.
4°. If Assumption A is satisfied in F and $V_O \in \mathscr{V}$ then for every non-derivable $b \in B$ and $p \in N$ holds $P[bo_p\, v(\omega)] \leqq 1 - P[v(\omega) \in V_O]$.
5°. If Assumption C is satisfied in F, then for every non-derivable $b \in B$ and $p \in N$ $P[bo_p\, v(\omega)] \leqq 1 - y < 1$ hold.

Proof. Assertion 1° follows from (IV), (V) of D1. Assertion 2° follows from 1°, (1) and from the continuity of the probability measure.

If b is derivable (i.e. $b \in T$), then $b^O = V$, hence, by 2°, (1) gives

$$P[bo_p\, v(\omega)] \nearrow_p P[bo_O\, v(\omega)] = P[v(\omega) \in b^O] = P[v(\omega) \in V] = 1\,,$$

which proves 3°.

Let $b \in B - T$. If A is satisfied in F and $V_O \in \mathscr{V}$, then for $v \in b^p$ is $bo_p v$ and by Assumption A is $v \bar{\in} V_O$, hence, $b^p \cap V_O = \emptyset$, so that $b^p \subseteq V - V_O$, hence,

$$P[bo_p\, v(\omega)] = P[v(\omega) \in b^p] \leqq P[v(\omega) \in V - V_0] = 1 - P[v(\omega) \in V_O]\,,$$

which proves 4°.

Let $b \in B - T$. If C is satisfied in F, then by 4° and (2) it can be easily seen that $P[bo_p\, v(\omega)] \leqq 1 - y < 1$ (for every $p \in N$). □

Items 3° and 5° of L1 give a method for distinguishing derivable formulas from non-derivable ones (and vice versa) with the help of statistical methods.

1.4. LOGICAL DERIVABILITY AND STATISTICAL METHODS

The procedure indicated by L1 has been used many times. A formula $b \in B$ is declared derivable, if b is derivable from a sufficient number of assumptions from V; more precisely: we choose some relation $o_p (p \in N^+)$ and naturals $0 \leqq m < n$. We sample assumptions $v_1, v_2, \ldots, v_n \in V$ indepedently, using a random mechanism. If the formula $b \in B$ is derivable from more than m assumptions among $v_1, \ldots, v_n$, we proclaim b to be derivable; otherwise we declare b to be non-derivable.

Definition D3. Let $F = \langle B, \{o_p\}, V\rangle$ be a formal system, $\langle \Omega, S, P\rangle$ a probability space. Let $v_1, v_2, v_3, \ldots$ be independent random variables from $\langle \Omega, S, P\rangle$ to $\langle V, \mathscr{V}\rangle$ with the same probability distribution. Let $0 \leqq m < n$ be naturals.

A formula $b \in B$ is declared to be derivable (with respect to o_p and $\omega \in \Omega$) if

$$\sum_{i=1}^{n} X(bo_p\, v_i(\omega)) > m .$$

A formula $b \in B$ is declared to be non-derivable (with respect to o_p and $\omega \in \Omega$), if

$$\sum_{i=1}^{n} X(bo_p\, v_i(\omega)) \leqq m .$$

Let b be a random variable from $\langle \Omega, S, P \rangle$ to $\langle B, exp.\, B \rangle$, let $b, v_1, v_2, v_3, \ldots$ be independent random variables. The first and second type of erorr probabilities are defined by the following expressions:

$$PE_1 = P[\bigcup_{b \in T} \{\sum_{i=1}^{n} X(bo_p\, v_i(\omega)) \leqq m \,\&\, b(\omega) = b\}]$$

$$PE_2 = P[\bigcup_{b \in B-T} \{\sum_{i=1}^{n} X(bo_p\, v_i(\omega)) > m \,\&\, b(\omega) = b\}] .$$

Remark. For every $n \in N^+$, $p \in N$ and $0 \leqq m < n$ we should define statistical decision function $D : B \times V^n \to \{0, 1\}$ as follows:

$$D(b, v_1, \ldots, v_n) = 0 \quad \text{if} \quad \sum_{i=1}^{n} X(bo_p v_i) \leqq m$$

$$D(b, v_1, \ldots, v_n) = 1 \quad \text{if} \quad \sum_{i=1}^{n} X(bo_p v_i) > m .$$

In such case

$$PE_1 = P[D(b(\omega), v_1(\omega), \ldots, v_n(\omega)) = 0 \,\&\, b(\omega) \in T] ,$$

$$PE_2 = P[D(b(\omega), v_1(\omega), \ldots, v_n(\omega)) = 1 \,\&\, b(\omega) \in B - T] .$$

hold (and could be define).

1.5. BASIC PARAMETERS AND RELATIONS

Three parameters occur in the presented procedure of statistical derivability: an index p of relation o_p, characterizing the extent of our knowledge about a (concrete) formal system studied for derivability; parameter n determines the number of assumptions (from (V)) for which the partial derivability is considered when the derivability of formulas from B is tested; parameter m (or the ratio m/n) – a formula $b \in B$ is declared to be derivable iff the relative frequency of assumptions from which is b partially derivable (with respect to o_p) is greater than m/n.

In the following we introduce estimations for the probability of declaring a formula $b \in B$ to be derivable (or non-derivable) with respect to o_p ($p = 0, 1, 2, \ldots$) and estimations of probability errors of both types. We also introduce asymptotical characteristics of these four values when some of parameters p, n, m turns to infinity.

In the following we suppose that $m, p, n \in N$ and $0 \leqq m < n$. When the specification (of values) of parameters p, n, m is necessary in order to avoid misunderstanding or when it is suitable, we prescribe these parameters as indexed. For example we denote the first type of error probability by PE_1, or PE_1^p, $PE_1^{p,n,m}$ etc.

We denote the probability of partial derivability of $b \in B$ (with respect to o_p) as P_b, i.e.

$$(3) \qquad P_b^{p,n,m} = P_b = P[bo_p\, v_1(\omega)]$$

the probability of declaring $b \in B$ to be non-derivable and derivable (with respect to o_p) as P_{bnd} and P_{bd}, i.e.

$$(4) \qquad P_{bnd}^{p,n,m} = P_{bnd} = P[\sum_{i=1}^{n} X[bo_p\, v_i(\omega)] \leqq m],$$

$$(5) \qquad P_{bd}^{p,n,m} = P_{bd} = P[\sum_{i=1}^{n} X[bo_p\, v_i(\omega)] > m] = 1 - P_{bnd},$$

the probability of sampling a formula $b \in B$ for testing to derivability as α_b, i.e.

$$(6) \qquad \alpha_b = P[b(\omega) = b].$$

Remark. With the help of statistical decision function D P_{bd}, P_{bnd} can be defined as

$$P_{bd} = P[D(b, v_1(\omega), \ldots, v_n(\omega)) = 0]$$
$$P_{bnd} = P[D(b, v_1(\omega), \ldots, v_n(\omega)) = 1].$$

Lemma L2. For $p, n, m \in N^+$, $0 \leqq m < n$ the following hold:

$$(7) \qquad P_{bnd} = \sum_{j=0}^{m} \binom{n}{j} P_b^j (1 - P_b)^{n-j},$$

$$PE_1 = \sum_{b\in T} \alpha_b P_{bnd} = \sum_{b\in T} \alpha_b \sum_{j=0}^{m} \binom{n}{j} P_b^j (1 - P_b)^{n-j},$$

$$PE_2 = \sum_{b\in B-T} \alpha_b \cdot (1 - P_{bnd}) = \sum_{b\in T} \alpha_b \sum_{j=m+1}^{n} \binom{n}{j} P_b^j (1 - P_b)^{n-j}.$$

Lemma L3. If $0 \leqq m < n$ are naturals, then the function

$$(8) \qquad f(x) = \sum_{j=0}^{m} \binom{n}{j} x^j (1 - x)^{n-j}$$

is continuous and decreasing in $\langle 0, 1 \rangle$, $f(0) = 1$, $f(1) = 0$.

2.1. RESULTS

In this section a fixed formal system $F = \langle B, \{o_p\}, V \rangle$ is considered, the probability space $\langle \Omega, S, P \rangle$ and random variables $b, v_1, v_2, \ldots$ satisfying conditions of D1, D2.

Theorem T1. Let parameters n, p be fixed. If parameter m increases from 0 to $n - 1$, then for every $b \in B$ the probability P_{bd} of declaring b to be derivable is non-increasing, the probability P_{bnd} of declaring b to be non-derivable is nondecreasing; the probability PE_1 is nondecreasing, PE_2 is nonincreasing.

Remark. From T1 it can be easily seen, that if the declaration of a derivable formula to be non-derivable is the more important error then the contrary error (or if we want to do PE_1 as small as possible), we choose the value of m/n close to 0; in the opposite case we take ratio m/n close to 1.

Proof. The proof of T1 easily turns from L2, L3 and (5). □

Theorem T2.a. Let parameters m, n be fixed. If parameter p increases, then the probabilities P_{bd}, PE_2 are nondecreasing, the probabilities P_{bnd}, PE_1 are non-increasing.

Remark. Probability of the error PE_1 (the error caused by declaring derivable formulas to be non-derivable) and the probability of declaring $b \in B$ as non-derivable can be, as T2 shows, made smaller by increasing of parameter p, i.e. by increasing of our knowledge about concrete system investigated for derivability. The increasing of p is not sufficient to make the probability of error PE_2 and the probability of declaring $b \in B$ as derivable smaller. We will shown (in T6 and in what follows) that for decreasing of PE_2 and P_{bnd} also the increasing of parameter n is necessary (n determines the number of assumptions in V from which the partial derivability is considered when the derivability of formulas in B is tested.)

Proof of T2.a. We shall show that if m, n are fixed then P_{bnd} is nonincreasing in p. The proof of the remaining relations of T2.a. is similar.

Let m, n be fixed. From (3) and assertion 2° of L1 it can be seen, that P_b is non-decreasing in p. Hence, by L3 and (7) P_{bnd} is nonincreasing in p. □

Theorem T2.a. is precised by T2.b, c.

Theorem T2.b. Let m, n be fixed and $b \in T$. Then

$$P_{bd} \to_p 1 \quad \text{and} \quad P_{bnd} \to_p 0$$

(i.e. probability P_{bd} of declaring b to be derivable tends to 1 and probability P_{bnd} of declaring b to be non-derivable tends to 0 when p tends to ∞).

Theorem T2.c. Let m, n be fixed. Then $PE_1 \to_p 0$ (i.e. the first type of error probability PE_1 tends to 0, when p tends to ∞).

Proof of Theorem T2.b, c.

b. Let b be derivable in F (i.e. $b \in T$), $p \to \infty$. Then by (3) and 3° of L1 P_b tends to 1. (7), (8) gives

$$P_{bnd} = f(P_b),$$

hence, by L3

$$\lim_{p\to\infty} P_{bnd} = \lim_{p\to\infty} f(P_b) = \lim_{x\to 1^-} f(x) = f(1) = 0\,,$$

hold, i.e. P_{bnd} tends to 0. By (5) $P_{bd} = 1 - P_{bnd}$, hence, P_{bd} tends to 1 (when $p \to \infty$).

c. The set $T \subseteq B$ of theorems is finite or countable. By L2 is

$$PE_1 = \sum_{b\in T} \alpha_b P_{bnd}$$

hence,

$$|PE_1| \leqq \sum_{b\in T} |\alpha_b P_{bnd}| \leqq \sum_{b\in T} |\alpha_b| \leqq \sum_{b\in B} |\alpha_b| = \sum_{b\in B} \alpha_b = 1 < +\infty$$

hence, because $P_{bnd} \to_p 0$ for $b \in T$,

$$PE_1 = \sum_{b\in T} \alpha_b P_{bnd} \to_p 0\,. \quad \square$$

It can be proved, that the following holds.

Lemma L4. Let m/n tend to β when n increases to ∞ and let $e \in \langle 0, 1\rangle$, $\beta \neq e$. Then the limit value

$$L = \lim_{n\to\infty} \sum_{j=0}^{m} \binom{n}{j} e^j (1 - e)^{n-j}$$

exists and

$$L = 0 \quad \text{iff} \quad e > \beta\,,$$

$$L = 1 \quad \text{iff} \quad e < \beta\,.$$

Theorem T3.a. Let p be fixed. Let us denote

$$D = \sup\{P_b \mid b \in B - T,\ \alpha_b \neq 0\}$$

$$d = \inf\{P_b \mid b \in T,\ \alpha_b \neq 0\}\,.$$

Let $\beta \in \langle 0, 1\rangle$ be fixed, let $m/n \to_n \beta$. Then

1.a) for $\beta \neq d$: $PE_1 \to_n 0$ iff $\beta < d$;

b) if $\beta > d$, then $\liminf_{n\to\infty} PE_1 > 0$;

2.a) for $\beta \neq D$: $PE_2 \to_n 0$ iff $\beta > D$;

b) if $\beta < D$ then $\liminf_{n\to\infty} PE_2 > 0$.

T3.a. tolds us the following: if we cannot increase the parameter p (i.e. enlarge our knowledge about the particular formal system tested to derivability) or if we have only one (or only finite number of) particular derivability relation and we want to make probabilities of both errors small, we have to choose large n and m/n close to β satisfying conditions

$$\sup\{P_b \mid b \in B - T,\ \alpha_b \neq 0\} < \beta < \inf\{P_b \mid b \in B - T,\ \alpha_b \neq 0\}\,.$$

 If

$$\sup\{P_b \mid b \in B - T,\ \alpha_b \neq 0\} > \inf\{P_b \mid b \in B - T,\ \alpha_b \neq 0\}\,,$$

then, having the fixed relation of partial derivability and increasing n, at least one of probabilities PE_1, PE_2 will be greater than some positive (real) number independent of n.

Proof T3.1.a) By L2 is

$$PE_1 = \sum_{b \in T} \alpha_b \sum_{j=0}^{m} \binom{n}{j} P_b^j (1 - P_b)^{n-j}\,,$$

hence

$$PE_1 = \sum_{\substack{b \in T \\ \alpha_b \neq 0}} \alpha_b \sum_{j=0}^{m} \binom{n}{j} P_b^j (1 - P_b)^{n-j}\,,$$

further, if for $b \in T$, $\alpha_b \neq 0$, then $\beta < d < P_b$, hence, by L3

$$\lim_{n \to \infty} \binom{n}{j} P_b^j (1 - P_b)^{n-j} = 0\,,$$

from which can be easily seen, that $PE_1 \to_n 0$. The proof of 2.a) is similar.

1.b) If $\beta > d$, then there exists $b_O \in T$ such that $\alpha_{b_O} \neq 0$ (then $\alpha_{b_O} > 0$) and $\beta > P_{b_O}$. Hence,

$$PE_1 = \sum_{\substack{b \in T \\ \alpha_b \neq 0}} \alpha_b \sum_{j=0}^{m} \binom{n}{j} P_b^j (1 - P_b)^{n-j} \geqq$$

$$\geqq \alpha_{b_O} \sum_{j=0}^{m} \binom{n}{j} P_{b_O}^j (1 - P_{b_O})^{n-j}\,;$$

by L4 the value of the last mentioned expression tends to α_{b_O} (when $n \to \infty$), hence, $\liminf_{n \to \infty} PE_1 > 0$. The proof of 2.b) is similar. □

Theorem T3.b. Let p be fixed. If the Assumption C is satisfied in a formal system F, then

$$D = \sup\{P_b \mid b \in B - T,\ \alpha_b \neq 0\} \leqq 1 - y < 1\,.$$

Proof. T.3.b. immediately follows from 5° of L1 and (3). □

From T3.b. and 2.a) of T3.a. can be easily, seen, that if Assumption C is satisfied in F and m/n tends to a number from $(1 - y, 1)$ then (with increasing n) the probability of error PE_2 tends to 0 (when p is fixed, i.e. when our possibilities of studying of derivability are limited by some relation of partial derivability).

From T2 and T3 follows, that by increasing only some of parameters m, n, p is it possible to reach the convergence of PE_1, PE_2 to 0 only in some special cases. Hence, in the following we shall study probabilities of both errors and probabilities P_{bd}, P_{bnd} when all parameters m, n, p tend to ∞; in addition we suppose, that $m/n \to_n \beta$, where $\beta \in \langle 0, 1 \rangle$.

Theorem T4.a. Let $\beta > \sup\{P_b^{p_0} \mid b \in B - T,\ \alpha_b \neq 0\}$ for $p_0 = 0$. Then

1. There are numbers $f \in (0, 1)$, $n_0 \in \{1, 2, 3, \ldots\}$ such that, for every $n \geqq n_0$, $p = 0, 1, 2, \ldots$
$$PE_2^{p,n} \leqq nf^n .$$
2. If parameters $p, n \to \infty$, then $PE_2 \to 0$.

Proof. Part 2 is a consequence of part 1. Part 1 can be proved from L2, L4. □

Theorem T4.b.

1. If, for some $p_0 \in N^+$, $\beta < \sup\{P_b^{p_0} \mid b \in B - T,\ \alpha_b \neq 0\}$, holds then
   a) there is $H > 0$ such that $\liminf_{n\to\infty} PE_2^{p_0} > H$.
   b) there is $n_0 > 0$ such that, for every
$$p \geqq p_0\,,\quad n \geqq n_0\,,\quad PE_2^{p,n} > H\,.$$
2. If, for some $p_0 \in N^+$, $\beta > \sup\{P_b^{p_0} \mid b \in B - T,\ \alpha_b \neq 0\}$ holds then, for every $p = 1, 2, \ldots, p_0$, $PE_2^{p,n} \to_n 0$.
3. If $\beta > \sup\{P_b^{p_0} \mid b \in B - T,\ \alpha_b \neq 0\}$ for $p_0 = 0$ then $PE_2^{p,n} \to_n 0$ uniformly in p (i.e. with increasing n tends $PE_2^{p,n}$ to 0 uniformly in p).

In T4 are presented detailed conditions under which the second type of error probability tends and does not tend to 0, when our knowledge about the actual relation of (partial) derivability increases so as the number of assumptions from V considered when derivability is investigated (i.e. parameter n) increases too.

Proof T4.b. Proof of 1.a) is similar to proof of 1.b) in T3.a. 1.b). 1.a) give that $\liminf_{n\to\infty} PE_2^{p_0} > H$, hence, there is n_0 such that, for $n \geqq n_0$, $PE_2^{p_0,n} > H$. By T2.a. PE_2 is nondecreasing in p, hence, for every $p \geqq p_0$, $n \geqq n_0$, $PE_2^{p,n} \geqq PE_2^{p,n_0} > H$. The proof of part 2 is similar to that of 1a) in T3.a. The proof of part 3 follows from 1 of T4.a. (as the expression $n \,.\, f^n$ is indepdent on p). □

Theorem T5. If $\beta \in \langle 0, 1)$, then $p, n \to \infty$ implies $PE_1 \to 0$.

Proof. 1. Let be $\varepsilon > 0$. We search n_0 and p_0 such that for $p \geqq p_0$, $n \geqq n_0$, $PE_1 < \varepsilon$.

2. By the definition of α, $\sum_{b\in T} \alpha_b \leqq 1$ and $\alpha_b \geqq 0$ (for $b \in T$), hence, there is a finite $C \subseteq T$ such that $\sum_{b\in T-C} \alpha_b < \varepsilon/2$.

3. By L2, $PE_1 = \sum_{b\in T} \alpha_b \sum_{j=0}^{m} \binom{n}{j} \,.\, P_b^j(1 - P_b)^{n-j}$, (and $0 \leqq P_b \leqq 1$ for $b \in T$), hence,

$$PE_1 \leqq \sum_{b\in C} \alpha_b \sum_{j=0}^{m} \binom{n}{j} \cdot P_b^j (1 - P_b)^{n-j} + \sum_{b\in T-C} \alpha_b \cdot 1 \leqq$$
$$\leqq \sum_{b\in C} \alpha_b \sum_{j=0}^{m} \binom{n}{j} \cdot P_b^j (1 - P_b)^{n-j} + \varepsilon/2 \, .$$

4. By 3° in L1 and (3), for every $b \in T$, $P_b \to_p 1$ hence, for every b in finite set C, $P_b \to_p 1$, hence, there is $p_0 \in N^+$ such that, for every $b \in C$, $P_b^{p_0} > \beta$.
5. $P_b^{p_0} > \beta$ holds for every $b \in C$, hence, by L4, for every $b \in C$,

$$\lim_{n\to\infty} \sum_{j=0}^{m} \binom{n}{j} (P_b^{p_0})^j (1 - P^{p_0})^{n-j} = 0 \, ,$$

hence, (because C is finite), there is n_0 such that for every $n \geqq n_0$ and $b \in C$

$$\sum_{j=0}^{m} \binom{n}{j} \cdot (P_b^{p_0})^j \cdot (1 - P_b^{p_0})^{n-j} < \frac{\varepsilon}{2} \, ,$$

hence, for every $n \geqq n_0$,

$$PE_1^{p_0,n} < \sum_{b\in C} \alpha_b \cdot \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$$

and by 2. also

$$PE_1^{p_0,n} < 1 \cdot \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \, .$$

6. By T2.a. PE_1 is nonincreasing in p, hence, for every $p \geqq p_0$, $n \geqq n_0$

$$PE_1^{p,n} \leqq PE_1^{p_0,n} < \varepsilon \, . \quad \square$$

Summarizing the results of T4, T5 we can see, that, to be able to make probabilities of both errors smaller than $\varepsilon > 0$ by increasing p and n, we have to choose m/n close to some number from the interval

$$(\sup \{P_b^{p_0=0} \mid b \in B - T, \ \alpha_b \neq 0\}, 1) \, ;$$

it can be proved that it is sufficient to find $\delta > 0$ such that

$$\sup \{P_b^{p_0=0} \mid b \in B - T, \ \alpha_b \neq 0\} + \delta < 1 - \delta \tag{9}$$

and choose m/n between the both sides of (9).

Let us return to Assumption C. If C is satisfied in a formal system then for every $p = 0, 1, 2, \ldots$ (by T3.b.)

$$\sup \{P_b^p \mid b \in B - T, \ \alpha_b \neq 0\} \leqq 1 - y < 1 \, ;$$

from these inequalities and results above can be easily seen that the following statement holds.

Theorem T6. If the Assumption C is satisfied in a formal system F and $1 - y < < \beta < 1$ holds, then

a) if n is fixed, then $PE_1 \to_p 0$;

b) if p is fixed, then $PE_2 \to_n 0$;
c) if $p, n \to \infty$ then $PE_1 \to 0$ and $PE_2 \to 0$;
moreover, $PE_2 \to 0$ uniformly in p.

Proof. a) By 3° of L1 for $b \in T$, $P_b \to_p 1$ hence, by L3, $\sum_{j=0}^{m} \binom{n}{j} P_b^j (1 - P_b)^{n-j} \to_p 0$ (when m, n are fixed), hence, by L2, $PE_1 \to_p 0$.

b) Let p be fixed. Let $b \in B - T$; from Assumption C follows, that $P_b \leqq 1 - y$ (by 5° of L1 and (3)), hence, $P_b < \beta$. Hence, for every $b \in B - T$, $P_{bnd} \to_n 0$ (by L4 and (7)), hence, by L2 $PE_2 \to_n 0$.

The proof of c) follows from T5 and 2 of T4.a. (with the help of 5° of L1 and (3)). □

As a conclusion let us introduce two nontraditional examples.

Continuation of Example 4 (*shows the importance of one fundamental formula of mathematical logic*).

Let B, V be sets of formulas of some theory $\mathscr{T}$. Let the formula

(10) $$b \to b$$

be true for every formula b of $\mathscr{T}$ (or, let every formula in $\mathscr{T}$ follow from itself, i.e. let $b \to b$ be "easily" provable in $\mathscr{T}$). Hence, let for every $b \in B$, $p = 1, 2, 3, \ldots$

(11) $$bo_p b$$

hold. In addition we assume that for every $b \in B$

$$P[b(\omega) = b] \neq 0 \quad \text{and} \quad P[v_1(\omega) = b] \neq 0 .$$

Let us denote

$$\alpha_b = P[b(\omega) = b] \quad (>0)$$
$$\pi_b = P[v_1(\omega) = b] \quad (>0) .$$

Then for every $b \in B$, $p = 0, 1, 2, \ldots$ by (11)

(12) $$P_b = P[bo_p\, v_1(\omega)] \geqq P[b = v_1(\omega)] = \pi_b > 0 ,$$

hence, by L3 and (12)

(13) $$P_{bnd} \leqq f(\pi_b) < 1 ,$$

(14) $$P_{bd} = 1 - P_{bnd} \geqq 1 - f(\pi_b) > 0 ,$$

where f is defined by (8). From L2, (13), (14) can be easily seen that

(15) $$PE_1 = \sum_{b \in T} \alpha_b \,.\, P_{bnd} \leqq \sum_{b \in T} \alpha_b \,.\, f(\pi_b) < 1$$

i.e.

(16) $$PE_1 < 1$$

and

(17) $$PE_2 = \sum_{b \in B - T} \alpha_b \,.\, P_{bd} \geqq \sum_{b \in B - T} \alpha_b \,.\, (1 - f(\pi_b)) .$$

 If there is at least one non-derivable formula in B, then by (17)

(18) $$PE_2 > 0 .$$

Consequently, denoting by b, c, d, the following statements

b) the formula $b \to b$ is true (or "easily" provable) for every formula b of $\mathscr{T}$ (i.e. $bo_p b$ holds for every $p \in N^+$, $b \in B$);

c) every formula is choosen to be tested to validity (provability) with positive probability (i.e. $P[b(\omega) = b] > 0$ for every $b \in B$);

d) the probability of choosing a formula v of $\mathscr{T}$ as an assumption (when validity (probability) of formulas of $\mathscr{T}$ is tested) is positive for every v of $\mathscr{T}$ (i.e. $P[v(\omega) = v] > 0$ for every $v \in V$);

then (b), (c) and (d) imply that,

by (14), every formula b of $\mathscr{T}$ is, with a positive probability, proclaimed to be true (provable),

by (13) every formula b of $\mathscr{T}$ is, with a probability less than 1 proclaimed to be not true (false, non-provable),

by (18) is the second type of error probability positive.

Hence, the relation (10) is quite strong for theories of mathematical logic (cf [2]) and the relation (11) is quite strong for formal systems.

Example 5. If we suppose, that the basis of our reasoning is formed by a certain class of principles, whose validity is tested by consistence investigating of every principle with other principles (by logical considerations and by confrontation with experience and with the environment).

If we suppose, in addition, that

a) our conviction (opinion) about validity of every principle increases with the number of principles consistent with it,

b) every principle is consistent with itself,

c) every principle can be submitted for validity testing (or every principle will be eventually taken to be tested),

d) every pair of principles can be (or will be) tested for consistence,

e) principle is proclaimed to be valid (true) if it is consistent with the greater part of principles;

then we obtain formal system and (statistical decision) procedure of validity testing of principles, described in Example 4.

Hence, we can say, that under assumptions described above the following hold: probability of proclaiming a principle to be valid is positive for every principle; probability of declaring principle not to be valid does not reach 1 for every principle; the second type of error probability is positive.

Investigated formal system and procedure of statistical derivability are rather general and, at the same time, relatively simple. Hence, the domain of application of the presented results may be quite wide.

An investigation of parameter "the extent of knowledge about concrete system tested for derivability", i.e. the study of properties of derivability assuming different levels of knowledges about concrete systems, have given some new results. Further examples, information and results are in [5].

(Received October 3, 1979.)

REFERENCES

[1] J. R. Schoenfield: Mathematical Logic. Addison-Wesley, 1967.

[2] G. Getzen: Untersuchungen über das logische Schliessen I, II. Mathematische Zeitschrift *39* (1935), 176—210 (I), 404—431 (II).

[3] I. Kramosil: A Method for Random Sampling of Well-Formed Formulas. Kybernetika *8* (1972), 2, 135—148.

[4] I. Kramosil: A Method for Statistical Testing of an at Random Sampled Formula. Kybernetika *9* (1973), 3, 162—173.

[5] J. Šindelář: Některé otázky statistické teorie dokazatelnosti. Research Report UTIA No. 760, 1976.

Jan Šindelář, Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.