Kybernetika

Ivan Kramosil; Jan Šindelář

Statistical deducibility testing with stochastic parameters

Kybernetika, Vol. 14 (1978), No. 6, (385)--396

Persistent URL: http://dml.cz/dmlcz/124270

KYBERNETIKA — VOLUME 14 (1978), NUMBER 6

Statistical Deducibility Testing with Stochastic Parameters

IVAN KRAMOSIL, JAN ŠINDELÁŘ

The deducibility problem, i.e., the problem, how to test whether a formula of a formalized theory is or is not a theorem, is investigated from the statistical point of view. A model is proposed in which the corresponding statistical decision problem can be converted into a parametric test of a simple hypothesis against a simple alternative. Some results following from this possibility are formulated and proved, concerning the probabilities of both types of errors.

1. INTRODUCTION AND MOTIVATION

Since the very beginning of more serious studies in artificial intelligence the automated theorem-proving has been considered to be an important branch of this domain. This situation was probably caused by the fact that also in everyday life of each mathematician theorem-proving is taken for a "highly intelligent", if not "the most intelligent" among his various activities. However, for a long time theorem-proving had been considered as something rather isolated in the sense that the decision whether the investigated formula or statement is or is not a theorem of the theory in question was considered to be the final step of all the preceding effort, i.e., no further decisions depend on it. This way of reasoning led to a certain absolutization when classifying the quality of results achieved during the theorem-proving effort. This means that the positive decision proclaiming a formula to be a theorem had to be always right with no possibility of error admitted; on the other hand, the "third possibility" consisting in saying "I cannot decide" was always possible and it was even preferred to any of the two "decisive" answers supposing they were connected with a (no matter how small) possibility of an error.

This point of view is typical for the so called pure mathematics, but it must be decisively changed when theorem-proving plays its role as a part of a more complex decision making system. Imagine, e.g., a complex stochastic dynamical system developing in time with respect to certain laws which are, in general, of stochastic character, such systems are studied in stochastic control theory. The human subject (user, designer) can intervene, from time to time, into the process by making appropriate decisions with the aim to maximize his expected profit to a certain future instant. The question which among the possible decisions is admissible, good or the best may be solved only if knowing whether an assertion or assertions (e.g., concerning the values of parameters in the present and, maybe, also in some past time instants) is or are valid and this problem can be easily converted to that of theorem-proving.

However, in the new context the classification of the possible outcomes of a theorem-proving procedure will be quite different from that common in pure mathematics and mentioned above. First, the decision must be taken in time, in order to be able to influence the system in the desirable way; a decision, even if optimal in an instant, may be inutile if the corresponding intervention is applied too late. Second, the "third possibility" of abstaining from the decisive answer is avoided. It is possible, of course, to do nothing, i.e., not to intervene into the process, however, non-intervention is also a kind of intervention, and in a particular situation a far not the best one. Hence, in our circumstances, a decision which is not necessarily correct, but which can be taken quickly and the probability of whose error is small enough or acceptable in a sense, may be much more appropriate than the absolutely correct decision connected with an undesirably long decision procedure or than the resignation ("I cannot decide").

The ideas, briefly outlined above, can serve as an intuitive and illustrative justification of various methods for statistical deducibility testing. The first procedure of such a kind was suggested by A. Špaček in 1959 (cf. [7]), his basic idea was developed in a series of papers; some of them can be found in references (cf. [3]–[4]). The common features of all these procedures can be described as follows.

Let $\langle \mathscr{L}, \mathscr{T} \rangle$ be a formalized theory, i.e., $\mathscr{L}$ is the set of all well-formed formulas of a formalized language, $\mathscr{T}$ is the set of all theorems, let the theory be consistent, i.e., $\mathscr{T} \neq \mathscr{L}$. Suppose to have at our disposition a deterministic theorem prover T, formally T is a mapping defined on $\mathscr{L}$ and taking its values in the three-elemented set $\{1, x_0, 0\}$; $T(x) = 0$ is interpreted, for $x \in \mathscr{L}$ as "x is proclaimed to be a theorem" or, briefly, "$x \in \mathscr{T}$", $T(x) = 0$ is interpreted as "$x \in \mathscr{L} - \mathscr{T}$" and $T(x) = x_0$ is interpreted as "we cannot decide about x". Very often this last decision is joined with "$x \in \mathscr{L} - \mathscr{T}$" so that $T : \mathscr{L} \to \{0, 1\}$, but in this case we must keep in mind the different qualitative character of the two remaining decisions, "$T(x) = 1$" and "$T(x) = 0$". The mapping T is supposed to be recursive in order to assure the efectiveness of the corresponding decision procedure. If x contains a free indeterminate, we replace x by its general closure, the same holds for a_i in what follows.

Only very few theories (the most elementary ones) are decidable, i.e., such a mapping T exists that T is recursive, $\{x : T(x) = 1\} = \mathscr{T}$, $\{x : T(x) = 0\} = \mathscr{L} - \mathscr{T}$. In general, $\{x : T(x) = 1\} = \mathscr{T}_0 \subset \mathscr{T}$, $\mathscr{T}_0 \neq \mathscr{T}$ (if a formula is proved by T, it is a theorem, but not every theorem is provable by T). Let $x, y \in \mathscr{L}$, $y \in \mathscr{T}$, then also

$(x \to y) \in \mathcal{T}$, however, if $y \in \mathcal{L} - \mathcal{T}$, then the condition $x \in \mathcal{L} - \mathcal{T}$ is necessary (but not sufficient) for $(x \to y) \in \mathcal{T}$. Written in other way, $\{x : x \in \mathcal{L}, (x \to y) \in \mathcal{T}\} = \mathcal{L}$, if $y \in \mathcal{T}$, $\{x : x \in \mathcal{L}, (x \to y) \in \mathcal{T}\} \subset \mathcal{L} - \mathcal{T}$, if $y \in \mathcal{L} - \mathcal{T}$. Now, the basic idea of statistical deducibility testing as understood here consists in sampling at random some closed formulas a_1 $a_2, \ldots, a_N, \ldots$, testing whether $T(a_i \to x) = 1$ or not (x being the tested formula) and taking the decision that x is or is not a theorem with respect to the relative frequency of those indices i, for which $T(a_i \to x) = 1$, to the total number of indices which have been examined. The testing procedure can be designed in classical way in the extent of random sample and treshold values given a priori or in sequential way when the extent of random sample is a random variable and the decision in taken in the instant when the relative frequency given above leaves an interval depending on the extent of the random sample.

2. THEOREM PROVERS WITH STOCHASTIC PARAMETERS

A more detailed investigation and formalization of the ideas outlined above necessitates to fournish the set $\mathcal{L}$ of well-formed formulas of the formalized language in question by two different structures - logical and probabilistic. The fact that these two structures are hardly compatible causes serious difficulties when forming appropriate theoretical backgrounds for statistical deducibility testing. E.g., the set $\mathcal{T}_0$ of all theorems for which $T(x) = 1$ can be easily described in terms of T (we have just given such a description). Having some more detailed knowledge about the theorem prover T we would be able to describe $\mathcal{T}_0$ in logical terms, e.g., as the set of all logical consequences of such and such axioms derivable from these axiom by an a priori given number of applications of such and such deduction rules. However, it will be very difficult and hardly possible to define, e.g., the probability that a random variable taking its values in $\mathcal{L}$ samples a theorem from $\mathcal{T}_0$. It is why we must be usually satisfied with some very rough and unprecise estimations of the values of interest, i.e., for example, probabilities of errors connected with a statistical deducibility testing procedure, expected extents of random samples in sequentional case, treshold values for decision making etc.

In order to solve at least partially this problem let us turn our attention once more to mathematical statistics. Let a be a random variable, defined on a probability space $\langle \Omega, \mathcal{S}, P \rangle$ and taking its values in $\mathcal{L}$. The probability

$$(1) \qquad P_a(\mathcal{T}_0) = P(\{\omega : \omega \in \Omega,\ a(\omega) \in \mathcal{T}_0\}) = P(\{\omega : \omega \in \Omega,\ T(a(\omega)) = 1\})$$

can be hardly computed on the grounds of a description of $\mathcal{L}$ in logical terms but it can be statistically estimated on the grounds of a finite random sample made by the random variable a from $\mathcal{L}$. Suppose that random samples from $\mathcal{L}$ are made mutually independently and with respect to the same probability distribution as that generated by a and denote by $\bar{p}_a(n, \mathscr{T}_0)$ the relative frequency of samples belonging to $\mathscr{T}_0$ among the first n ones. Laws of statistics give that, for each $\varepsilon > 0$, $\delta > 0$, there exists such an index $n_0 = n_0(\varepsilon, \delta)$ that for all $n > n_0$

$$(2) \qquad P(\{\omega : \omega \in \Omega,\ |\bar{p}_a(n, \mathscr{T}_0) - P_a(\mathscr{T}_0)| < \varepsilon\}) > 1 - \delta\,.$$

At the same time, it is, in a sense, the maximum we are able to obtain about the value $P_a(\mathscr{T}_0)$ on the grounds of a finite sample. The values $\bar{p}_a(n, \mathscr{T}_0)$, ε, δ, and maybe some others introduced below are called *stochastic parameters* of the theorem prover T.

Now, let us investigate what happens when replacing the tested formula x by the implication $a_i \to x$, a_i sampled at random, and testing it by the theorem prover T with stochastic parameters. Because of eliminating the dependence of our results on a particular tested formula x let us suppose that also the tested formula is sampled at random (in the cases of applicational character it has a quite natural justification – the Nature or the environment submit problems, i.e. formulas, to be decided; as the rules according to which they are chosen are unknown, at least a priori, to the theorem prover, they may be assumed to result from a random sample). The following result is a very simple one, but because of the possibility of an easy reference we express them in the form of the following statement.

Theorem 1. Consider a theorem prover T such that, for all $x, y \in \mathscr{L}$, $T(\neg x) = 1$ implies $T(x \to y) = 1$. Let $x, a_1, a_2, \ldots$ be random variables, mutually independent, for $a_1, a_2, \ldots$ equally distributed, defined on a probability space $\langle \Omega, \mathscr{S}, P \rangle$ and taking their values in the set of closed formulas from $\mathscr{L}$. Denote

$$P_a(\mathscr{T}) = P(\{\omega : \omega \in \Omega,\ a_1(\omega) \in \mathscr{T}\}), \text{ similarly for } P_x(\mathscr{T}),$$
$$p = P(\{\omega : T(x(\omega)) = 1\} \mid \{\omega : x(\omega) \in \mathscr{T}\}) = P_x(\mathscr{T}_0 \mid \mathscr{T}),$$
$$p' = P(\{\omega : T(\neg a_1(\omega)) = 1\} \mid \{\omega : a_1(\omega) \in \mathscr{L} - \mathscr{T}\}).$$

Then, for $i = 1, 2, \ldots,$

$$(3) \quad P(\{\omega : T(a_i(\omega) \to x(\omega)) = 1\} \mid \{\omega : x(\omega) \in \mathscr{T}\}) \geqq p + (1 - p)(1 - P_a(\mathscr{T}))\, p'\,,$$

$$(4) \quad P(\{\omega : T(a_i(\omega) \to x(\omega)) = 1\} \mid \{\omega : x(\omega) \in \mathscr{L} - \mathscr{T}\}) \geqq (1 - P_a(\mathscr{T}))\, p'\,.$$

Proof. Assume that $x(\omega) \in \mathscr{T}$, either $x(\omega) \in \mathscr{T}_0$, or $x(\omega) \in \mathscr{T} - \mathscr{T}_0$. The first case implies that $T(x(\omega)) = T(a_i(\omega) \to x(\omega)) = 1$ and its probability equals p. A sufficient condition for $T(a_i(\omega) \to x(\omega)) = 1$ in case when $x(\omega) \in \mathscr{T} - \mathscr{T}_0$ is that $a_i(\omega) \in \in \mathscr{L} - \mathscr{T}$ and $T(\neg a_i(\omega)) = 1$, according to the supposed independence of corresponding random variables the probability of this event equals $(1 - p) . . (1 - P_a(\mathscr{T}))\, p'$. Summing the probabilities of the two mutually disjoint possibilities we obtain (3).

Assume that $x(\omega) \in \mathscr{L} - \mathscr{T}$, then the condition $a_i(\omega) \in \mathscr{L} - \mathscr{T}$ is necessary and the condition $T(\neg a_i(\omega)) = 1$ sufficient for $T(a_i(\omega) \to x(\omega)) = 1$, hence, $(1 - P_a(\mathscr{T})) \geqq P(\{\omega : T(a_i(\omega) \to x(\omega)) = 1\} \mid \{\omega : x(\omega) \in \mathscr{L} - \mathscr{T}\}) \geqq (1 - P_a(\mathscr{T}))\, p'$. Q.E.D.

The two following facts are obvious. First, as

$$P(\{\omega : T(a_i(\omega) \to x(\omega)) = 1\} \mid \{\omega : x(\omega) \in \mathscr{T}\}) > p =$$

$$= P(\{\omega : T(x \mid \omega) = 1\} \mid \{\omega : x(\omega) \in \mathscr{T}\})$$

and, supposing that $P_a(\mathscr{T}) < 1$,

$$P(\{\omega : T(a_i(\omega) \to x(\omega)) = 1\} \mid \{\omega : x(\omega) \in \mathscr{L} - \mathscr{T}\}) > 0 =$$

$$= P(\{\omega : T(x(\omega)) = 1\} \mid \{\omega : x(\omega) \in \mathscr{L} - \mathscr{T}\}),$$

we can see that the intuitive idea that random auxiliary axioms "help us" to prove the tested formula can be formally described and justified. Second, if $P_a(\mathscr{T}) > 1 - p$ then

$$p + (1 - p)(1 - P_a(\mathscr{T}))\, p' > (1 - P_a(\mathscr{T})),$$

i.e., it is more "easy" or probable to prove $a_i(\omega) \to x(\omega)$ under the condition that $x(\omega)$ is a theorem than under the condition that it is not. Both these facts can serve as a theoretical justification of the decision schema outlined in the introductory chapter and they enable to transform the decision problem in question into the following simple and classical form.

Suppose that $P_a(\mathscr{T}) > 1 - p$ and denote $p_1 = p + (1 - p)(1 - P_a(\mathscr{T}))\, p'$, $p_2 = (1 - P_a(\mathscr{T}))$, i.e. $p_1 > p_2$. Sample at random $a_1(\omega), a_2(\omega), \ldots$ and test whether the random events $T(a_1(\omega) \to x(\omega)) = 1$, $T(a_2(\omega) \to x(\omega)) = 1, \ldots$ occured or not. On the ground of these observations we are to decide, whether $P_0 = P(\{\omega : \omega \in \Omega, T(a_i(\omega) \to x(\omega)) = 1\})$ is at least p_1 (hypothesis) or at most p_2 (alternative), abbreviately, we are to test $H : P_0 \geqq p_1$ against $A : P_0 \leqq p_2$. This is a classical parametric statistical testing problem of a composed hypothesis against a composed alternative. As $p_1 > p_2$ this test can be converted to a more simple case of a simple hypothesis $H : P_0 = p_1$ against a simple alternative $A : P_0 = p_2$ (cf. [5], [8], or other source for a more detailed information about statistical hypotheses testing).

In actual cases the hypothesis as well as the alternative can be "improved", i.e., p_1 can be replaced by a $p_1' > p_1$, and p_2 can be replaced by $p_2' < p_2$. For, in case $x(\omega)$ is a theorem not provable by T(i.e., $x(\omega) \in \mathscr{T} - \mathscr{T}_0$), there are always some theorems such that $T(y \to x(\omega)) = 1$, e.g., y can be $x(\omega)$ itself or some formula equivalent to x and such that this equivalence is decidable by T. Having a lower estimate of the probability with which such theorems can be sampled, we can enlarge in an appropriate way the value p_1. Similarly, not every non-theorem y has the property that $T(y \to x(\omega)) = 1$, so we can exclude some formulas from $\mathscr{L} - \mathscr{T}$ when considering this case and, hence, replace p_2 by $p_2' < p_2$. In every case, however, the problem whether $x(\omega) \in \mathscr{T}$ or not will be transformed, again, into a classical parametric test of a simple hypothesis $P_0 = p_1'$ against a simple alternative $P_0 = p_2'$.

Until now, we have proceeded as if p, p' and $P_a(\mathscr{T})$ were known precisely, but we must realize that they are known only by the means of their statistical estimations in the form of (2), i.e., inseparably connected with two stochastic parameters ε and δ. As the values p_1, p_2 are dependent on these parameters, they are also charged by a stochastic indeterminacy and we have to take this fact into account.

Suppose, for the sake of simplicity, that the values p, p', $P_a(\mathscr{T})$ are replaced by their statistical estimates $\bar{p}(n, a)$ $\bar{p}'(n, a)$, $\bar{P}_a(n, \mathscr{T}_0)$ with n chosen in such a way that for a priori given or appropriate stochastic parameters ε, δ

(5)
$$P(\{\omega : \omega \in \Omega,\ |\bar{p}(n, a) - p| < \varepsilon\}) > 1 - \delta\,,$$
$$P(\{\omega : \omega \in \Omega,\ |\bar{p}(n, a) - p'| < \varepsilon\}) > 1 - \delta\,,$$
$$P(\{\omega : \omega \in \Omega,\ |\bar{P}_a(n, \mathscr{T}_0) - P_a(\tau_0)| < \varepsilon\}) > 1 - \delta\,.$$

Instead of $\bar{p}(n, a)$ we shall write $\bar{p}_n$ and similarly for $\bar{p}_n'$ and $\bar{p}_n(\mathscr{T})$ when the dependency on a is not important in the given context. Consider, now, the values p_1 and p_2 as functions of p, p' and $p_a(\mathscr{T})$ and set:

$$\bar{p}_1 = \bar{p}_n + (1 - \bar{p}_n)(1 - \bar{p}_n(\mathscr{T}))\,\bar{p}_n'\,,$$
$$\bar{p}_2 = 1 - \bar{p}_n(\mathscr{T})\,.$$

Given $\delta > 0$, ε can be chosen appropriately small by increasing the value of n (remember the well-known Tchebyshev inequality), so there is, for any $\delta > 0$, $n_0 = n_0(\delta)$ such that, for $n \geqq n_0$, $\bar{p}_1 > \bar{p}_2$. Hence, assuming that all the three inequalities from (5) hold, and this is valid with probability greater than $1 - 3\delta$ we may transform our original problem whether $x(\omega) \in \mathscr{T}$ or not, into a classical parametric test of a simple hypothesis $H : P_0 \geqq \bar{p}_1$ against a simple alternative $A : P_0 \leqq \bar{p}_2$.

Because of the fact that p_1 and p_2 depend on p, p' and $P_a(\mathscr{T})$ in a simple linear form we can find that

$$\bar{p}_1 \geqq (p - \varepsilon) + (1 - p - \varepsilon)(1 - P_a(\mathscr{T}) - \varepsilon)(p' - \varepsilon) > p_1 - 8\varepsilon\,,$$
$$\bar{p}_2 \leqq 1 - P_a(\mathscr{T}) + \varepsilon = p_2 + \varepsilon$$

i.e., ε must be smaller than $1/9$-th of the difference between values $\bar{p}_1$, $\bar{p}_2$ computed on the ground of statistical estimations $\bar{p}_n$, $\bar{p}_n'$ and $\bar{p}_n(\mathscr{T})$.

3. APPLICATIONS OF GENERAL HYPOTHESES TESTING METHODS TO DEDUCIBILITY TESTING

Our main effort in this paper is to show, how the problem of deducibility testing can be converted into a problem common in general parametric hypotheses testing theory, not to investigate or describe in details how this problem is solved at the general level. It is why we limit ourselves to a very short outline of some more simple results which can be said when considering our testing problem.

Theorem 2. Let the conditions and notations of Theorem 1 hold, let N be an integer, let $x, a_1, a_2, \ldots$ be defined as above, let $0 < P_x(\mathscr{T}) < 1$. Let $\mathfrak{f}$ and $\mathfrak{t}$ be two abstract symbols, let $D(N, x, \{a_i\}_{i=1}^N, \cdot)$ be a random variable, defined on the probability space $\langle \Omega, \mathscr{S}, P \rangle$ taking its values in $\{\mathfrak{t}, \mathfrak{f}\}$ and such that

$$(6) \qquad \{\omega : \omega \in \Omega,\ D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{t}\} =$$

$$= \left\{\omega : \omega \in \Omega,\ \left|\frac{1}{N}\sum_{i=1}^N T(a_i(\omega) \to x(\omega)) - \bar{p}_2\right| \geqq \left|\frac{1}{N}\sum_{i=1}^N T(a_i(\omega) \to x(\omega)) - \bar{p}_1\right|\right\}$$

$$D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{f} \quad \text{otherwise}.$$

Then

$$(7) \qquad P(\{\omega : D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{f}\} \mid \{\omega : x(\omega) \in \mathscr{T}\}) \leqq 3\delta + \frac{1}{N|\bar{p}_1 - \bar{p}_2|^2},$$

$$(8) \qquad P(\{\omega : D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{t}\} \mid \{\omega : x(\omega) \in \mathscr{L} - \mathscr{T}\}) \leqq 3\delta + \frac{1}{N|\bar{p}_1 - \bar{p}_2|^2}.$$

Remark. In spite of its rather complicated formalized form the intuitive idea behind this assertion is rather simple. We sample at random formulas $a_1(\omega), a_2(\omega), \ldots \ldots, a_N(\omega)$, and test, for every $i \leqq N$, whether $T(a_i(\omega) \to x(\omega)) = 1$ or whether $T(a_i(\omega) \to x(\omega)) = 0$. Moreover, we compute the relative frequency of the cases when this value equals 1, i.e., when $a_i(\omega) \to x(\omega)$ is provable by T. If this relative frequency is not closer to $\bar{p}_2$ than to $\bar{p}_1$, we accept the hypothesis $P_0 = \bar{p}_1$, i.e., we proclaim $x(\omega)$ to be a theorem (using the decision function D we write this decision formally as $D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{t}$). In other case we proclaim $x(\omega)$ to be a non-theorem (formally, $D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{f}$). There are two possibilities that the decision taken by the decision function D may be wrong; the probabilities of these two possibilities are estimated by (7) and (8).

Proof of Theorem 2. Let $x(\omega) \in \mathscr{T}$, there are two reasons for which the decision made by the decision function D may be wrong. First, the possibility of translating the problem whether $x(\omega) \in \mathscr{T}$ or not into that whether $P_0 = \bar{p}_1$ or $P_0 = \bar{p}_2$ fails, i.e., at least one of the interval estimates for p, p' and $P_a(\mathscr{T})$ given in (5), is not valid.

We have already seen that this probability can be majorized by 3δ. Second, (5) may be valid, but the statistical hypothesis testing of $H : P_0 = \bar{p}_1$ against $A : P_0 = \bar{p}_2$ may lead to an error. Clearly, $T(a_i(\omega) \to x(\omega))$ is a random variable with the expected value $\bar{p}_1$ (as $x(\omega) \in \mathcal{T}$) and with the dispersion at most $1/4$, for different i, j these random variables are mutually independent. Hence, the well-known Tchebyshev inequality gives

$$(9) \qquad P\left(\left\{\omega : \left|\frac{1}{N}\sum_{i=1}^{N} T(a_i(\omega) \to x(\omega)) - \bar{p}_1\right| > \varepsilon_1\right\}\right) \leqq \frac{1}{4N\varepsilon}.$$

In case when $D(N, x, \{a_i\}_{i=1}^{N}, \omega) = \mathfrak{f}$ necessarily $1/N \sum_{i=1}^{N} T(a_i(\omega) \to x(\omega))$ must differ from $\bar{p}_1$ by more than $\frac{1}{2}|\bar{p}_1 - \bar{p}_2|$, replacing ε_1 in (9) by this value we obtain the second expression of the right hand side of (7), hence (7) is proved. (8) can be proved quite analogously. Q.E.D.

Applying once more Tchebyshev inequality, this time to the inequalities in (5), we can see that $\delta < 1/4n\varepsilon^2$, n being the number of random samples on the ground of which the statistical estimates for p, p' and $P_a(\mathcal{T})$ have been obtained. As can be easily seen, a simultaneous increasing of n and N can minimize both the probabilities of error under any a priori given positive bound. An interesting question arises, in which common proportion the values n and N should be enlarged in order to obtain the best (i.e., minimal) estimates for the two probabilities of error? In other words, having given the total sum of N and n, say N_1 of possible random experiments, which proportion of them should be devoted for precising the values p, p' and $P_a(\mathcal{T})$ and which proportion for the statistical hypothesis testing itself? We can offer the following answer.

Lemma 1. Let $c_1 > 0$, $c_2 > 0$ be two reals, let m, n be positive integers such that $N_1 = m + n$ is fixed. Then the expression

$$\frac{c_1}{n} + \frac{c_2}{m}$$

takes its minimal value, iff

$$n = \left[N_1 \cdot \frac{\sqrt{(c_1 c_2)} - c_1}{c_2 - c_1}\right], \quad m = N_1 - n.$$

Proof. Because of its purely analytical character the proof is postponed to the Appendix.

Theorem 3. Let the conditions and notations of Theorem 2 hold with $\delta < 1/4n\varepsilon^2$, let $N_1 = n + N$ be fixed. Then both the probabilities of error in (7) and (8) are minimal, iff

$$(10) \qquad n = \left[\frac{(2\varepsilon)\sqrt{(3)}\,|\bar{p}_1 - \bar{p}_2| - 3|\bar{p}_1 - \bar{p}_2|^2}{(2\varepsilon)^2 - 3|\bar{p}_1 - \bar{p}_2|^2} N_1\right], \quad N = N_1 - n\,.$$

Proof. The assertion follows from Lemma 1 when taking $c_1 = 3/4\varepsilon^2$, $c_2 = 1/|\bar{p}_1 - \bar{p}_2|^2$ and realizing a simple computation. Q.E.D.

The decision function D defined in Theorem 2 can be easily re-defined in such a way that $D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{t}$ if $\sum_{i=1}^{N} T(a_i(\omega) \rightarrow x(\omega)) \geqq M$ with $M \leqq N$ being easily computable from (6). We do not perform this computation for this case, but for a more general one, namely, when the both probabilities of errors are not taken to be comparable. This situation is common in general statistical hypothesis testing theory and it is solved as follows. The "more dangerous" probability of error is strictly requested to be kept below an a priori given treshold value, say α, and the second probability of error is minimized under this condition. In our case, according to the viewpoint accepted in other works dealing with statistical deducibility testing, we take the error consisting in proclaiming a non-theorem to be a theorem for the more dangerous (because this event may cause the set of formulas proclaimed to be theorems to become inconsistent and so useless for a further use). Hence, having N, our aim is to find appropriate $M \leqq N$ as the following theorem precises.

Theorem 4. Let the conditions and notations of Theorem 1 hold, let $0 < M \leqq N$ be two integers, let $\bar{p}_1, \bar{p}_2$ be defined as above, let $\bar{p}_1 > \bar{p}_2$. Let $\mathfrak{t}$ and $\mathfrak{f}$ be two abstract symbols, let $D = D(M, N, x, \{a_i\}_{i=1}^N)$ be a random variable, defined on the probability space $\langle \Omega, \mathscr{S}, P \rangle$, taking its values in $\{\mathfrak{t}, \mathfrak{f}\}$ and such that

$$(11) \qquad D(M, N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{t}\,, \quad \text{if} \quad \sum_{i=1}^{N} T(a_i(\omega) \rightarrow x(\omega)) \geqq M\,,$$

$$D(M, N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{f} \quad \text{otherwise}\,.$$

Let $\alpha > 0$ be given, let

$$(12) \qquad M_1 = [N(u_{1-\alpha} \cdot \sqrt{[N^{-1}\bar{p}_2(1 - \bar{p}_2)]} + \bar{p}_2)] + 1\,,$$

where u_α is the so called α-quantile of the normal distribution $N(0, 1)$. Then

$$(13) \quad P(\{\omega : D(M_1, N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{t}\} \mid \{\omega : x(\omega) \in \mathscr{L} - \mathscr{T}\}) \leqq 3\delta + \alpha\,,$$

$$(14) \quad P(\{\omega : D(M_1, N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{f}\} \mid \{\omega : x(\omega) \in \mathscr{T}\}) =$$

$$= \min_{M=M_1}^{N} \{P(\{\omega : D(M, N, x, \{a_i\}_{i=1}^N, \omega) = \mathfrak{f}\} \mid \{\omega : x(\omega) \in \mathscr{T}\}\}\,.$$

Proof. Consider the classical statistical hypothesis testing problem with $H : p = \bar{p}_1$ against $A : p = \bar{p}_2$. We want to choose $M \leqq N$ such that the probability of at least M events $T(a_i(\omega) \to x(\omega)) = 1$ were majorized by α supposing that $p = \bar{p}_2$. Moreover, we look for the maximal M with this property in order to minimize the other probability of error. Hence, we look for minimal M such that

$$\sum_{i=0}^{M} \binom{N}{i} \bar{p}_2^i (1 - \bar{p}_2)^{N-i} \geqq 1 - \alpha . \tag{15}$$

Denote $\bar{p} = \bar{p}(N) = (1/N) \sum_{i=1}^{N} T(a_i(\omega) \to x(\omega))$, the well-known Central Limit Theorem of probability theory sounds that $\bar{p}$ has, approximately, normal distribution with parameters $\mu = \bar{p}_2$, $\sigma^2 = (1/N)\, \bar{p}_2(1 - \bar{p}_2)$ (under the condition that $p = \bar{p}_2$) i.e., $\bar{p}$ has, approximately, distribution function $\Phi((x - \bar{p}_2)/\sqrt{[N^{-1}\bar{p}_2(1 - \bar{p}_2)]})$, where Φ is the distribution function of the normal distribution $N(0, 1)$. The demand (15) can be transformed into the form

$$\Phi\left(\frac{(M/N) - \bar{p}_2}{\sqrt{[N^{-1}\bar{p}_2(1 - \bar{p}_2)]}}\right) \geqq 1 - \alpha ,$$

hence

$$\frac{(M/N) - \bar{p}_2}{\sqrt{[N^{-1}\bar{p}_2(1 - \bar{p}_2)]}} \geqq u_{1-\alpha} ,$$

and an easy calculation gives the value M_1 as stated above. The values of α-quantiles of the normal distribution $N(0, 1)$ are tabeled and can be found in statistical tables (c.f. [2], e.g.). The problem can be solved also in a non-asymptotical way using the incomplete β-distribution. Q.E.D.

The methods explained until now have one common feature, namely, their length is fixed a priori (i.e., the numbers N or $n + N$ of random experiments which are to be made before a decision is taken are given or limited a priori). As an alternative to these procedures, the general hypotheses testing theory offers the so called sequential tests. In this case the number of random samples necessary to take a decision is a random variable and only its expected value, moments or other statistical characteristics can be computed or estimated. Not wanting to go into details of the theory of sequential tests we limit ourselves by describing a simple variant of sequential test procedure for our case when H is $p = \bar{p}_1$ and A is $p = \bar{p}_2$. The procedure is borrowed from [1], the underlying theoretical results can be found in [5] or [9].

Let $r > 0$ be such a real that we want the sum of both the probabilities of errors not to exceed r. Set

$$k = \frac{\log((1 - \bar{p}_2) \mid (1 - \bar{p}_1))}{\log(\bar{p}_1 \mid \bar{p}_2) + \log((1 - \bar{p}_2) \mid (1 - \bar{p}_1))} , \tag{16}$$

$$q = \frac{\log((1 - r) \mid r)}{\log(\bar{p}_1 \mid \bar{p}_2) + \log((1 - \bar{p}_2) \mid (1 - \bar{p}_1))} . \tag{17}$$

For each $m = 1, 2, \ldots$ set

$$L_1(m) = km + q\,, \quad L_2(m) = km - q\,. \tag{18}$$

Now, sample $a_i(\omega)$ and compute $T(a_i(\omega) \to x(\omega))$. If

$$L_2(m) < \sum_{i=1}^{m} T(a_i(\omega) \to x(\omega)) < L_1(m)\,, \tag{19}$$

sample $a_{m+1}(\omega)$ and continue as above. If $\sum_{i=1}^{m} T(a_i(\omega) \to x(\omega)) \leqq L_2(m)$, stop the sampling and take the decision that $p = \bar{p}_2$, i.e., proclaim $x(\omega)$ to be a non-theorem. If $\sum_{i=1}^{m} T(a_i(\omega) \to x(\omega)) \geqq L_1(m)$, stop the sampling and the take decision that $p = \bar{p}_1$, i.e., proclaim $x(\omega)$ to be a theorem. Under some very general conditions a decision will be eventually taken with the probability one.

APPENDIX: A PROOF OF LEMMA 1

Consider the expression $c_1/n + c_2/m$ with $n + m = z$ fixed. Substituting $m = z - n$ we transform this formula into the form $[c_1 z + (c_2 - c_1)\, n]/(zn - n^2)$.

Suppose the parameter n in this expression to be real-valued and to range from 0 to z. As can be easily seen, the infimum is taken for an $n_0 \in (0, z)$; a simple differentiation gives

$$\frac{\mathrm{d}}{\mathrm{d}n}\left(\frac{c_1 z + (c_2 - c_1)\, n}{zn - n^2}\right) = \frac{(c_2 - c_1)(zn - n^2) - (c_1 z + (c_2 - c_1)\, n)(z - 2n)}{(zn - n^2)^2}\,.$$

Letting the last expression to be zero and solving the resulting quadratic equation for n, we obtain

$$n_{1,2} = \frac{-2c_1 z \pm \sqrt{[4c_1^2 z^2 + 4(c_2 - c_1)\, c_1 z^2]}}{2(c_2 - c_1)} = z\,.\,\frac{-c_1 \pm \sqrt{(c_1 c_2)}}{c_2 - c_1}\,.$$

Choosing the appropriate root we have $n = z(\sqrt{(c_1 c_2)} - c_1/(c_2 - c_1))$ as Lemma 1 claims. Q.E.D.

(Received June 15, 1978.)

REFERENCES

[1] V. Fabian: Základní statistické metody. NČSAV (Publishing House of the Czechoslovak Academy of Sciences), Prague 1963.

[2] J. Janko: Statistické tabulky. NČSAV (Publishing House of the Czechoslovak Academy of Sciences), Prague 1958.

[3] I. Kramosil: Statistical Estimation of Deducibility in Polyadic Algebras. Kybernetika *7* (1971), 3, 181—200.

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

[5] E. L. Lehman: Testing Statistical Hypotheses. John Wiley and Sons, New York, Chapman and Hall, London 1960 (Russian translation: ,,Nauka", Moscow 1964).

[6] J. Šindelář: Některé možnosti použití metody Monte-Carlo ve statistické teorii odvoditelnosti. Institute of Information Theory and Automation, Research Rep. No. 803, 1977.

[7] A. Špaček: Statistical Estimation of Provability in Boolean Logics. In: Transactions of the Second Prague Conference on Information Theory, Prague 1959. NČSAV (Publishing House of the Czechoslovak Academy of Sciences), Prague 1960.

[8] A. Wald: Statistical Decision Functions. John Wiley and Sons, New York 1950.

[9] A. Wald: Sequential Analysis. John Wiley and Sons, New York 1947 (Russian translation: ,,Nauka", Moscow, 1960).

Dr. Ivan Kramosil, CSc., Jan Šindelář, prom. mat., 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.