STATISTICAL ESTIMATION OF SEMANTIC PROVABILITY

ANTONÍN ŠPAČEK
CZECHOSLOVAK ACADEMY OF SCIENCES

1. Introduction

Let us point out that there is nothing unexpected in this paper. The sole element of novelty is the formal description of a simple relation between a chapter of mathematical logic and mathematical statistics. The word semantic occurring in the title indicates that, roughly speaking, provability or nonprovability is to be estimated on the basis of truth and falsehood in interpretations in models. The logical formalism used in this paper is monadic logic introduced by P. R. Halmos in [2]. In principle it is possible to replace the monadic logic by a more developed formalism, for instance, by polyadic logic [3]. The elements, the provability or nonprovability of which is to be estimated, as well as the interpretations, are chosen at random by appropriate chance mechanisms, hence the whole problem is probabilistic in nature. The estimation procedures established in this paper possess a natural optimum property. The study of the behavior of these procedures at infinity shows that the statistical decision functions of finite size, which estimate provability are, in fact, asymptotically good approximate proofs. One may hope that the questions treated in this paper reflect at least the most elementary features of heuristic reasoning which is so perfectly realized by the human brain.

All that is necessary for an easy understanding is developed in the paper in full detail and with intuitive justification. The main reason is that one cannot expect that, in general, specialists in mathematical logic are familiar with concepts, methods and results of statistical decision theory or that statisticians are familiar with formalisms of mathematical logic.

The basic concepts and results of statistical decision theory on an appropriate level of generality are summarized in section 2. These results are then applied in section 3 to the problem of statistical estimation of belonging relations. The passage from the considerations of section 3 to the solution of our main problem of statistical estimation of provability is completely transparent and forms the contents of section 4.

The present paper, which is closely connected with [8], does not furnish more than may be intuitively expected and, therefore, its practical value is very limited. Further developments in this direction, however, will probably throw some light into the mechanism of human behavior in problem solving.

2. The Neyman-Pearson theorem

A wide variety of problems of mathematical statistics can be reduced to a simple application of a classical theorem due to J. Neyman and E. S. Pearson [5], [6]. It is not surprising that this famous theorem plays a decisive role in our considerations. Its original version, however, does not fulfil our requirements. The main reason is that it does not allow the discussion of cases in which more general sample spaces occur. We shall see later that an adequate generalization of the Neyman-Pearson theorem can be easily obtained.

Our basic probability space will be denoted, as usual, by $(\Omega, \mathfrak{S}, \mu)$, where Ω is the set of elementary events, $\mathfrak{S}$ the sigma-algebra of random events and μ the probability measure on $\mathfrak{S}$. The symbol ω will always mean an element of Ω. Throughout this paper the notation just introduced will be preserved.

A statistical decision problem is defined to be a pair (φ, ξ) of random variables, where φ takes its values in the parameter space and ξ ranges over the sample space. The parameter space is assumed to consist of exactly two elements, namely 0 and 1, hence, the measurability of φ is assured by the requirement that

$$\{\omega : \varphi(\omega) = 1\} \in \mathfrak{S}.$$

On the other hand, no restriction will be imposed on the range X of ξ except that it is supplied with a fixed sigma-algebra $\mathfrak{X}$ of subsets of X. The measurable space $(X, \mathfrak{X})$ is said to be the sample space. The transformation ξ of Ω into X will be called a random sample if

$$\{\omega : \xi(\omega) \in E\} \in \mathfrak{S}$$

for every set E from $\mathfrak{X}$.

Roughly speaking, a statistical decision is an action determined by the value of the random sample. This action can be formally described using the concept of decision function. The domain of a decision function is the sample space and its range is usually called the space of decisions. In our case, however, the space of decisions is assumed to coincide with the parameter space, hence, a decision function δ is a function defined on X and taking the values 0 or 1. But this is not enough. In order to ensure that the compound transformation $\delta[\xi(\cdot)]$ becomes a random variable, it is reasonable to impose on δ an additional condition of measurability, namely,

$$\{x : \delta(x) = 1\} \in \mathfrak{X}.$$

A natural manner of how to evaluate statistical decisions with respect to the random occurrence of parameters is the convention that

$$(\{\omega : \varphi(\omega) = 1\} \cap \{\omega : \delta[\xi(\omega)] = 0\}) \cup (\{\omega : \varphi(\omega) = 0\} \cap \{\omega : \delta[\xi(\omega)] = 1\})$$

means the random event of incorrect decisions.

Our main question is how to choose the decision function δ in order to make the probability of the random event of incorrect decisions as small as possible. The answer is quite satisfactory.

THEOREM 1. *There always exists a statistical decision function which minimizes the probability of the random event of incorrect decisions.*

The proof is a simple application of the Hahn decomposition theorem [4]. Let us write

$$\nu(E) = \mu[\{\omega:\varphi(\omega) = 1\} \cap \xi^{-1}(E)] - \mu[\{\omega:\varphi(\omega) = 0\} \cap \xi^{-1}(E)]$$

for every E from $\mathfrak{X}$. Clearly, ν is a signed measure on $\mathfrak{X}$. It is well known that there exists a set H from $\mathfrak{X}$ such that $\nu(H \cap E) \geqq 0$ and $\nu(H' \cap E) \leqq 0$ for every E from $\mathfrak{X}$, where $H' = X - H$. Since

$$\nu(E) = \nu(H \cap E) + \nu(H' \cap E) \leqq \nu(H) + \nu(H' \cap E) \leqq \nu(H)$$

for every E from $\mathfrak{X}$, hence the number $\nu(H)$ is the maximum of ν on $\mathfrak{X}$. Now let us define the decision function β by the requirement that

$$\{x:\beta(x) = 1\} = H.$$

Since for every decision function δ the probability of the random event of incorrect decisions is equal to

$$\mu\{\omega:\varphi(\omega) = 1\} - \nu\{x:\delta(x) = 1\}$$

hence, using the fact that β is determined by the Hahn decomposition (H, H') of ν, we can write

$$\mu\{\omega:\varphi(\omega) = 1\} - \nu\{x:\beta(x) = 1\} \leqq \mu\{\omega:\varphi(\omega) = 1\} - \nu\{x:\delta(x) = 1\}$$

for every decision function δ, Q.E.D.

The decision function β, whose existence is assured by theorem 1, is said to be the Bayes solution of the statistical decision problem (φ, ξ).

It is easy to verify that the signed measure ν is absolutely continuous with respect to the probability measure $\mu\xi^{-1}$ in $\mathfrak{X}$, hence, using the Radon-Nikodym theorem [4], we can state that there exists a real valued measurable function h on X such that

$$\nu(E) = \int_E h(x)\, d\mu\xi^{-1}$$

for every set E from $\mathfrak{X}$. We see at once that the set

$$\{x:h(x) > 0\}$$

and its complement determine a Hahn decomposition of ν and this is in fact the content of the Neyman-Pearson theorem. It is, however, more appropriate to formulate this theorem in terms of the measurable functions h^+ and h^- defined for every element x of X and every set E from $\mathfrak{X}$ by the equations

$$\mu\Big(\{\omega:\varphi(\omega) = 1\} \cap \xi^{-1}(E)\Big) = \alpha \int_E h^+(x)\, d\mu\xi^{-1},$$

$$\mu\Big(\{\omega:\varphi(\omega) = 0\} \cap \xi^{-1}(E)\Big) = (1 - \alpha) \int_E h^-(x)\, d\mu\xi^{-1},$$

where

$$\alpha = \mu\{\omega:\varphi(\omega) = 1\}.$$

The number α is said to be the a priori probability in the parameter space. Clearly, if $\alpha > 0$ then h^+ is a conditional probability density and if $\alpha < 1$ then h^- is a conditional probability density. Since

$$\mu\xi^{-1}\{x:h(x) = \alpha h^+(x) - (1 - \alpha)h^-(x)\} = 1,$$

the Neyman-Pearson theorem can be formulated as follows:

THEOREM 2. *The statistical decision function β determined by the relation*

$$\{x:\beta(x) = 1\} = \{x:\alpha h^+(x) > (1 - \alpha)h^-(x)\}$$

minimizes the probability of the random event of incorrect decisions.

In applications of this theorem the densities h^+ and h^- are always assumed to be known, hence the Bayes solution β of the statistical decision problem (φ, ξ) depends only on the a priori probability α in the parameter space.

Now we shall introduce the abstract substitute of the concept of sample size. The classical model shows that one of the most important consequences of the reduction of sample size is a restriction imposed on the measurability of the decision functions. This fact motivates our definition of the size of a decision function.

Let $\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3, \cdots$ be a nondecreasing sequence of sigma-algebras of subsets of X and suppose that the union

$$\bigcup_{n=1}^{\infty} \mathfrak{X}_n$$

is a base of the sigma-algebra $\mathfrak{X}$. This sequence will serve as a scale of the sizes of decision functions.

The decision function δ is said to be of size n if it is measurable with respect to the sigma-algebra $\mathfrak{X}_n$, that is, if

$$\{x:\delta(x) = 1\} \in \mathfrak{X}_n$$

but it is not measurable with respect to the sigma-algebra $\mathfrak{X}_m$ for $m = 1, 2, \cdots, n - 1$. We shall say that δ is of finite size if there exists a positive integer n such that δ is of size n. The decision function δ, which is by definition measurable with respect to the whole sigma-algebra $\mathfrak{X}$, is said to be of infinite size if it is not of finite size. Clearly, if there exists a decision function of infinite size then, roughly speaking, the scale $\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3, \cdots$ must have effectively an infinite number of divisions.

Denoting by Δ the set of all decision functions in X and by Δ_n that of all decision functions in X at most of size n for $n = 1, 2, 3, \cdots$, we see at once that

$$\Delta_1 \subset \Delta_2 \subset \Delta_3 \subset \cdots \subset \Delta,$$

hence, if ϵ is the probability of the random event of incorrect decisions associated with the Bayes decision function β from Δ and ϵ_n that associated with the Bayes decision function β_n from Δ_n for $n = 1, 2, 3, \cdots$ then

$$\epsilon_1 \geqq \epsilon_2 \geqq \epsilon_3 \geqq \cdots \geqq \epsilon$$

that is, as may be intuitively expected, the least probabilities of making incorrect decisions do not increase whenever the sizes of the decision functions admitted to the concurrence increase to infinity.

By theorem 2 a Bayes decision function β_n of size n is determined by the relation

$$\{x:\beta_n(x) = 1\} = \{x:\alpha h_n^+(x) > (1 - \alpha)h_n^-(x)\}$$

for $n = 1, 2, 3, \cdots$, where h_n^+ and h_n^- are defined using the sigma-algebra $\mathfrak{X}_n$ in the same way as h^+ and h^- were defined using the whole sigma-algebra $\mathfrak{X}$.

The main effect of increasing the sample size can be expressed as

THEOREM 3. *The sequence of random variables $h_1^+[\xi(\cdot)], h_2^+[\xi(\cdot)], h_3^+[\xi(\cdot)], \cdots$ converges to the random variable $h^+[\xi(\cdot)]$ with probability one, the sequence of random variables $h_1^-[\xi(\cdot)], h_2^-[\xi(\cdot)], h_3^-[\xi(\cdot)], \cdots$ converges to the random variable $h^-[\xi(\cdot)]$ with probability one, and the sequence $\epsilon_1, \epsilon_2, \epsilon_3, \cdots$ of probabilities of the random event of incorrect decisions, associated successively with the Bayes decision functions $\beta_1, \beta_2, \beta_3, \cdots$ converges to the probability ϵ of the random event of incorrect decisions associated with the Bayes decision function β.*

The first two assertions of theorem 3 are immediate consequences of a well-known martingale theorem [1] and the last assertion is contained in [7] as a particular case.

Let us note that if $\epsilon = 0$ then the last assertion of theorem 3 expresses the well-known property of consistency of the Bayes decision functions $\beta_1, \beta_2, \beta_3, \cdots$.

3. Statistical estimation of belonging relations

A wide variety of questions concerning statistical estimation of provability possesses a common statistical structure of very elementary nature and this fact enables us to treat the basic statistical problem separately and independently of any consideration belonging purely to the domain of mathematical logic. After establishing the general results it remains only to interpret them appropriately in order to obtain the desired final answer to various questions of statistical estimation of provability. The realization of this last step is, however, rather only a routine matter.

Suppose that one wants to decide whether an element chosen at random by an appropriate chance mechanism from a fixed set A belongs or does not belong to a fixed nonempty proper subset M of A.

The random variable η taking values in A is assumed to be a formal substitute of our basic chance mechanism. One of the most natural requirements concerning measurability is

$$\{\omega:\eta(\omega) \in M\} \in \mathfrak{S}.$$

The direct observation on M is replaced by observations on the subsets $Q(m)$ of A for $m = 1, 2, 3, \cdots$, hence, it is also natural to impose on η an additional condition, namely,

$$\{\omega:\eta(\omega) \in Q(m)\} \in \mathfrak{S}$$

for $m = 1, 2, 3, \cdots$ and this completes the definition of the random variable η.

Now let τ be an ordinary random variable taking on values of positive integers. The compound transformation $Q[\tau(\cdot)]$ is a random variable in the sense that

$$\{\omega : p \in Q[\tau(\omega)]\} \in \mathfrak{S} \tag{1}$$

for every element p of A. This follows from the obvious identity

$$\{\omega : p \in Q[\tau(\omega)]\} = \bigcup_{j=1}^{\infty} \{\omega : \tau(\omega) = m_j\},$$

where m_j is the jth positive integer for which $p \in Q(m_j)$.

Clearly,

$$\{\omega : \eta(\omega) \in Q[\tau(\omega)]\} = \bigcup_{m=1}^{\infty} [\{\omega : \eta(\omega) \in Q(m)\} \cap \{\omega : \tau(\omega) = m\}];$$

hence we can state that

$$\{\omega : \eta(\omega) \in Q[\tau(\omega)]\} \in \mathfrak{S} \tag{2}$$

and this is the most important fact concerning the relation between the two kinds of random variables.

In elementary set theory the relation $p \in M$ is often expressed in terms of the characteristic function c of M by the equivalent statement that $c(p) = 1$. A slightly more complicated concept is that of the characteristic function of a random set. If, for each $m = 1, 2, 3, \cdots$, $c(m)$ denotes the characteristic function of the set $Q(m)$ then by (1) the compound function $c[\tau(\cdot)]$ is an ordinary random variable taking the values 1 or 0. The random variable $c[\tau(\cdot)]$ is said to be the characteristic function of the random set $Q[\tau(\cdot)]$. The element p of A belongs to $Q[\tau(\omega)]$ or to its complement $A - Q[\tau(\omega)]$ according as $c[\tau(\omega)](p) = 1$ or $c[\tau(\omega)](p) = 0$.

Clearly, the compound transformation $c[\eta(\cdot)]$ is an ordinary random variable taking the values 1 or 0. The value of η at ω belongs to M or to its complement $A - M$ according as $c[\eta(\omega)] = 1$ or $c[\eta(\omega)] = 0$. By (2) the compound transformation $c[\tau(\cdot)][\eta(\cdot)]$ is an ordinary random variable taking the values 1 or 0. The value of η at ω belongs to $Q[\tau(\omega)]$ or to its complement $A - Q[\tau(\omega)]$ according as $c[\tau(\omega)][\eta(\omega)] = 1$ or $c[\tau(\omega)][\eta(\omega)] = 0$. We have thus defined a probabilistic extension of belonging relations.

In order to simplify the notation we shall write $\varphi(\cdot)$ instead of $c[\eta(\cdot)]$ and $\chi(\cdot)$ instead of $c[\tau(\cdot)][\eta(\cdot)]$.

Let X be the set of all sequences $x = (x_1, x_2, x_3, \cdots)$ every term of which is either equal to 1 or 0. Coincidence in the first n terms of sequences from X is an equivalence relation in X. The class $\mathfrak{X}_n$ of all unions of equivalence sets induced by this equivalence relation is a complete algebra of subsets of X for every $n = 1, 2, 3, \cdots$. The sets from $\mathfrak{X}_n$ are called n-dimensional cylinders. Our basic sigma-algebra $\mathfrak{X}$ of subsets of X is that induced by the union

$$\bigcup_{n=1}^{\infty} \mathfrak{X}_n.$$

Let $\tau_1, \tau_2, \tau_3, \cdots$ be a sequence of integral-valued random variables. Then the sequence

$$\xi = (\chi_1, \chi_2, \chi_3, \cdots),$$

where $\chi_n(\cdot) = c[\tau_n(\cdot)][\eta(\cdot)]$ for $n = 1, 2, 3, \cdots$, is a random vector with values in X. Clearly, $\mathfrak{X}$ is the smallest sigma-algebra of subsets of X for which the vector ξ is measurable.

Now the ground is prepared to put the traditional machinery of statistical decision functions into action. The passage from the general scheme of statistical decision to our particular case is very simple because the notation of section 2 is preserved. As has already been pointed out in section 2, the Bayes solution of a statistical decision problem depends on the a priori probability in the parameter space. We shall see, however, that, as compared with the general case, our particular version of the statistical decision problem is, roughly speaking, less sensitive to the exact knowledge of the a priori probability, provided that a very simple and natural condition, namely,

$$M \subset Q(m) \tag{3}$$

for $m = 1, 2, 3, \cdots$, is satisfied. We shall see that under this condition either the decision function β_n^* which associates with every sample point x of X the decision 0, or the decision function β_n which associates with the sample point x of X the decision 1 or 0 according as the first n coordinates of x are equal to 1 or at least one of these coordinates is equal to 0, can occur. More precisely

THEOREM 4. *Under (3) the Bayes solution of size n of the statistical decision problem (φ, ξ) is determined by the decision function β_n or β_n^* and the probability of the random event of incorrect decisions is equal to*

$$(1 - \alpha)h_n^-(1, 1, \cdots, 1, x_{n+1}, x_{n+2}, \cdots) \tag{4}$$

or to α according as

$$h_n^-(1, 1, \cdots, 1, x_{n+1}, x_{n+2}, \cdots) < \frac{\alpha}{1 - \alpha} \tag{5}$$

or the opposite inequality holds.

The details of the proof can be omitted because theorem 4 is nothing else but a particular version of theorem 2. It suffices to note that, as compared with theorem 2, the main simplification arises from (3) and from the definition of X, $\mathfrak{X}_n$ and ξ. Under these conditions $h_n^+(x) = 1$ or 0, according as the first n coordinates of x are equal to 1 or one at least of these coordinates is equal to 0, and $0 \leqq h_n^-(x) \leqq 1$ for every x from X, hence theorem 2 is immediately applicable.

In order to make the intuitive content of the theorem just established more transparent we shall give the informal description of an experimental procedure of how to estimate that an element of A chosen by η belongs to M or to its complement $A - M$ using the Bayes decision procedure of size n, that is, that determined by the random variables $\tau_1, \tau_2, \cdots, \tau_n$. Whenever the inequality (5) does not hold then the value of η is always estimated to belong to $A - M$.

If (5) holds then the decision procedure runs as follows: At the first step we choose the set $Q(m)$ determined by the value of τ_1. If the value of η does not belong to this set, the procedure is stopped and the value of η is estimated to belong to $A - M$. In the opposite case we continue the inspection choosing the set $Q(m)$ determined by the value of τ_2. If the value of η does not belong to this set, the procedure is stopped and the value of η is estimated to belong to $A - M$. In the opposite case we continue the inspection choosing the set $Q(m)$ determined by the value of τ_3 and so on. Exhausting all the sets $Q(m)$ determined successively by the values of $\tau_1, \tau_2, \cdots, \tau_n$ without reaching the decision that the value of η belongs to $A - M$ we accept the decision that the value of η belongs to M. We see that the final decision that the value of η belongs to $A - M$ can be reached at every step of the decision procedure. On the other hand, the opposite decision that the value of η belongs to M can be reached only at the last step.

Now we shall show that under the two additional conditions

$$\bigcap_{m=1}^{\infty} Q(m) = M, \tag{6}$$

$$\mu\left[\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \{\omega : \tau_n(\omega) = m\}\right] = 1, \tag{7}$$

the Bayes solution of the statistical decision problem (φ, ξ) becomes asymptotically independent of the a priori probability α.

Roughly speaking, condition (6) together with (3) express the natural requirement that the approximation of M by the successive intersections of the sets $Q(m)$ can be arbitrarily close and the condition (7) means that the sequence $\tau_1, \tau_2, \tau_3, \cdots$ exhausts with probability one the whole set of positive integers.

For instance, condition (7) is satisfied whenever the integer valued random variables $\tau_1, \tau_2, \tau_3, \cdots$ are mutually independent, identically distributed, and such that

$$\mu\{\omega : \tau_1(\omega) = m\} > 0$$

for $m = 1, 2, 3, \cdots$.

Clearly, under the last condition,

$$\mu\left[\bigcap_{n=1}^{k} \{\omega : \tau_n(\omega) \neq m\}\right] = \left[\mu\{\omega : \tau_1(\omega) \neq m\}\right]^k,$$

$$\mu\{\omega : \tau_1(\omega) \neq m\} < 1,$$

for $k = 1, 2, 3, \cdots, m = 1, 2, 3, \cdots$; hence,

$$\mu\left[\bigcap_{n=1}^{\infty} \{\omega : \tau_n(\omega) \neq m\}\right] = 0$$

for $m = 1, 2, 3, \cdots$, that is,

$$\mu\left[\bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} \{\omega : \tau_n(\omega) \neq m\}\right] = 0$$

or, equivalently, (7) holds.

Our theorem 4 can be completed as

THEOREM 5. *If $\alpha > 0$ and the conditions* (3), (6), (7) *are satisfied then there exists a positive integer k such that β_n is the Bayes solution of size n of the statistical decision problem (φ, ξ) whenever $n > k$.*

Since by theorem 3 $h_n^-[\xi(\cdot)] \to h^-[\xi(\cdot)]$ with probability one as $n \to \infty$, hence, by theorem 4 and by the assumption $\alpha > 0$ of theorem 5 it suffices to show that $h^-(1, 1, 1, \cdots) = 0$, that is, that

$$\mu\left(\{\omega:\eta(\omega) \in A - M\} \cap \left\{\omega:\eta(\omega) \in \bigcap_{n=1}^{\infty} Q[\tau_n(\omega)]\right\}\right) = 0. \tag{8}$$

In order to simplify the notation we shall write

$$\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \{\omega:\tau_n(\omega) = m\} = G.$$

It follows from (6) that

$$G \cap \{\omega:\eta(\omega) \in A - M\} \cap \left\{\omega:\eta(\omega) \in \bigcap_{n=1}^{\infty} Q[\tau_n(\omega)]\right\} = 0,$$

hence

$$\mu\left(G \cap \{\omega:\eta(\omega) \in A - M\} \cap \left\{\omega:\eta(\omega) \in \bigcap_{n=1}^{\infty} Q[\tau_n(\omega)]\right\}\right) = 0,$$

and since by (7) $\mu(G) = 1$, we obtain (8), Q.E.D.

Let us denote by β the decision function which associates with every sample point x from X the decision 1 or 0, according as $x = (1, 1, 1, \cdots)$ or $x \neq (1, 1, 1, \cdots)$. By theorem 2, β is the Bayes solution of the statistical decision problem (φ, ξ) with respect to the whole sigma-algebra $\mathfrak{X}$, hence it is of infinite size. Since the probability of the random event of incorrect decisions is equal to zero, the decision function β, in fact, becomes a proof that the value of η belongs to M or to its complement $A - M$.

Now we shall introduce a function l on Ω whose values are positive integers or ∞ as follows:

$$\{\omega:l(\omega) = 1\} = \{\omega:\chi_1(\omega) = 0\},$$

$$\{\omega:l(\omega) = n\} = \{\omega:\chi_n(\omega) = 0\} \cap \bigcap_{j=1}^{n-1} \{\omega:\chi_j(\omega) = 1\}$$

for $n = 2, 3, 4, \cdots$ and

$$\{\omega:l(\omega) = \infty\} = \bigcap_{n=1}^{\infty} \{\omega:\chi_n(\omega) = 1\}.$$

We see at once that l is an ordinary random variable, provided that the definition is modified in such a way that the possibility

$$\mu\{\omega:l(\omega) = \infty\} > 0$$

is not excluded.

The random variable l is said to be the length of the decision function β.

THEOREM 6. *If conditions* (3), (6) *and* (7) *are satisfied then the length of the decision function β is infinite with conditional probability one under the condition* *that the value of η belongs to M and it is finite with conditional probability one under the condition that the value of η belongs to $A - M$.*

By the definition of l,

$$\mu[\{\omega:l(\omega) = \infty\}|\{\omega:\eta(\omega) \in M\}] = \frac{1}{\alpha}\mu\left(\{\omega:\eta(\omega) \in M\} \cap \left\{\omega:\eta(\omega) \in \bigcap_{n=1}^{\infty} Q[\tau_n(\omega)]\right\}\right).$$

Using the conditions (3), (6), and (7), we see that (8) holds, hence the first assertion of theorem 6 is an immediate consequence of (8). Since, in addition,

$$\mu[\{\omega:l(\omega) = \infty\}|\{\omega:\eta(\omega) \in A - M\}] = \frac{1}{1-\alpha}\mu\left(\{\omega:\eta(\omega) \in A - M\} \cap \left\{\omega:\eta(\omega) \in \bigcap_{n=1}^{\infty} Q[\tau_n(\omega)]\right\}\right),$$

the second assertion of theorem 6 follows at once from (8).

Let us note that under the assumptions of the theorem just proved it is not true that the conditional moments of l under the condition that the value of η belongs to $A - M$ are finite, that is, the analogue of theorem 2 in [8] does not hold. This disadvantage, however, can be removed by adding further restrictive conditions.

4. Semantic concepts

The statistical decision problem of section 3 is based on observations on the sets $Q(1)$, $Q(2)$, $Q(3)$, $\cdots$ which replace the direct observation on M. The most natural way of how to get the sets $Q(1)$, $Q(2)$, $Q(3)$, $\cdots$ is the effect of reduction of resolving power in A which can be formally described by an appropriate application of equivalence relations.

A binary relation R in the set A is said to be an equivalence relation if it is reflexive, symmetric and transitive. Every equivalence relation in A induces a partition of A into equivalence sets and vice versa. Two elements p, q of A belong to the same equivalence set if and only if pRq. The equivalence relation S in A is said to be finer than R and we shall write $S < R$ if pSq implies pRq or, in other words, if every equivalence set induced by S is included in at least one, hence in exactly one, equivalence set induced by R. Clearly, the set of all equivalence relations in A is partially ordered by $<$ and the identity I is the finest equivalence relation.

The formal description of reduction of resolving power by equivalence relations is intuitively justified by the convention that two elements of A which belong to the same equivalence set cannot be distinguished. Under this convention it is reasonable to introduce the concept of closure M^R of M induced by the equivalence relation R, requiring that

$$M^R = \bigcup_{q \in M} \{p:pRq\}.$$

Clearly, $M^I = M$, that is, the application of the identity I on M has no effect, and $M^R \subset M^S$, whenever $R < S$.

Let $R_1, R_2, R_3, \cdots$ be a sequence of equivalence relations in A. Putting

$$M^{R_n} = Q(n)$$

for $n = 1, 2, 3, \cdots$, we see that, in fact, the decision problem of section 3 is based on observations at reduced resolving power. This artificial reduction of resolving power is justified by the fact that, in general, M^R has a simpler structure than M^S, whenever $S < R$.

The application of the elementary facts summarized in section 3 to our main question of statistical estimation of provability by interpretations in models requires a number of restrictions which must be imposed on the sets A and M and on the equivalence relations $R_1, R_2, R_3, \cdots$.

First of all we shall suppose that A is a Boolean algebra. As usual, we shall denote by 0 and 1 the zero and unity of A, by p' the complement of the element p of A, by $\wedge$ the operation of forming the greatest lower bound, and by $\vee$ that of forming the least upper bound in A.

In order to eliminate misunderstandings we recall that the subset M of A is said to be a Boolean ideal in A if it contains the greatest lower bound of any two of its elements as well as the least upper bound of any two elements of A one at least of which belongs to M. The algebraic structure just defined is usually called dual Boolean ideal. We shall, however, omit the suffix dual because no other ideals will occur in this paper.

The relation R defined in A is said to be a Boolean congruence relation if it is an equivalence relation which, in addition, satisfies the condition

$$pRq \quad \text{implies} \quad p' \vee rRq' \vee r. \tag{9}$$

The simplest algebraic structure, which enables the treatment of propositional functions of mathematical logic and for which the concept of interpretation is natural, is that of monadic algebra introduced by P. R. Halmos [2].

A monadic algebra is a Boolean algebra A together with an operator $\forall$ which assigns to every element p of A an element $\forall p$ of A in such a way that

$$\forall 1 = 1,$$

$$\forall p \leqq p$$

for every element p of A, and

$$\forall(p \vee \forall q) = \forall p \vee \forall q$$

for every p and q in A. The operator $\forall$ is said to be the universal quantifier in A.

A subset M of a monadic algebra A is said to be a monadic ideal in A whenever M is a Boolean ideal in A and

$$p \in M \quad \text{implies} \quad \forall p \in M.$$

A monadic ideal M in the monadic algebra A is called maximal if it is proper, that is, if $M \neq A$ and M is not a proper subset of any other proper ideal in A.

The relation R defined in the monadic algebra A is said to be a monadic congruence relation if it is a Boolean congruence relation and if, in addition,

$$pRq \quad \text{implies} \quad \forall pR\forall q. \tag{10}$$

A monadic congruence relation R in the monadic algebra A is called simple whenever the monadic residual class algebra $A(R)$ of A modulo R is simple, that is, when there is no proper monadic ideal in A other than that containing the sole element 1.

The relevant properties of closures of monadic ideals will be expressed by the following lemma:

If M is a monadic ideal and R a monadic congruence relation in the monadic algebra A then the closure M^R of M induced by R is a monadic ideal in A. If, in addition, R is simple then either $M^R = A$ or M^R is maximal.

We shall first show that M^R is a Boolean ideal in A. Let $p \in M^R$ and $q \in A$. By the definition of the closure M^R of M, there exists an element r_p of M such that r_pRp. By (9) $r_p' \vee 0Rp' \vee 0$, that is, $r_p'Rp'$. Hence $r_p \vee qRp \vee q$. Since M is a Boolean ideal in A, we have $r_p \vee q \in M$, hence $p \vee q \in M^R$. Now let us suppose, in addition, that $q \in M^R$. Then there exists an element r_q of M such that r_qRq. By (9) we have

$$r_p' \vee r_q'Rp' \vee r_q', \qquad p' \vee r_q'Rp' \vee q',$$

and hence

$$(r_p' \vee r_q')' \vee 0R(p' \vee r_q')' \vee 0,$$
$$(p' \vee r_q')' \vee 0R(p' \vee q')' \vee 0$$

or, equivalently,

$$r_p \wedge r_qRp \wedge r_q,$$
$$p \wedge r_qRp \wedge q,$$

and, using the transitivity of R, we obtain

$$r_p \wedge r_qRp \wedge q.$$

Since M is a Boolean ideal in A we have $r_p \wedge r_q \in M$, hence, $p \wedge q \in M^R$. We see that M^R is a Boolean ideal in A. Now it remains to show that M^R is a monadic ideal in A, that is, that $\forall p \in M^R$ whenever $p \in M^R$. Since r_pRp, it follows from (10) that $\forall r_pR\forall p$, hence, using the assumption that M is a monadic ideal in A, we have $\forall r_p \in M$ and, consequently, $\forall p \in M^R$. This completes the proof of the first part of our lemma. If R is simple then, by the definition of simplicity, the class of all congruence sets which have a nonempty intersection with M is equal either to $A(R)$ or to the monadic ideal $\{1\}$ in $A(R)$, hence, either $M^R = A$ or $M^R = \{p:pR1\}$, Q.E.D.

A monadic logic is a pair (A, M), where A is a monadic algebra and M is a monadic ideal in A. The monadic logic (A, M) represents a deductive theory in A. The elements of A which belong to M are called provable. If R is a simple congruence relation in A then the closure M^R of M induced by R is said to be an interpretation of M in the model $A(R)$. If an element p of A belongs to the interpretation M^R of M we shall say that p is true in that interpretation and otherwise that it is false.

The monadic logic (A, M) is said to be semantically consistent if there exists at least one interpretation of M in a model.

Since $M \subset M^R$, we can state that a provable element of A is true in every interpretation. Whenever the opposite conclusion is possible then the monadic logic (A, M), is called semantically complete. More precisely, the monadic logic (A, M) is said to be semantically complete if M is equal to the intersection of all its interpretations.

For our purposes, however, a restricted version of semantic completeness is more appropriate. Let $\mathfrak{Q}$ be a class of interpretations of M. The monadic logic (A, M) is said to be semantically $\mathfrak{Q}$-complete, whenever

$$M = \bigcap_{Q \in \mathfrak{Q}} Q.$$

In order to eliminate degenerate cases it is natural to assume that the monadic logic (A, M) is semantically consistent and semantically $\mathfrak{Q}$-complete. Clearly, the assumption of semantic consistency can be replaced by $M \neq A$ and, by our lemma, there is no restriction of generality if we assume that the interpretations from $\mathfrak{Q}$ are maximal monadic ideals.

The estimation of provability or nonprovability of elements of a monadic logic is based upon the inspection of its truth or falsehood in interpretations in models. Since to each interpretation Q from $\mathfrak{Q}$ there corresponds a simple monadic congruence relation R_Q such that $Q = M^{R_Q}$, the idea of artificial reduction of resolving power by simple monadic congruence relations is justified by the fact that, by the lemma, the induced closures are maximal monadic ideals which evidently have an extremely simple algebraic structure.

The application of the results established in section 3 to the question of statistical estimation of provability in monadic logic requires a further restriction, namely, that $\mathfrak{Q}$ is denumerable. In this case we can write

$$\mathfrak{Q} = \{Q(1), Q(2), Q(3), \cdots\}.$$

The random variable η chooses an element of the monadic algebra A, the provability or nonprovability of which is to be estimated on the basis of interpretations of M chosen from $\mathfrak{Q}$ by the random variables $\tau_1, \tau_2, \cdots, \tau_n$.

One may intuitively expect that the following decision procedure is the most favorable one. At the first step we choose the interpretation $Q(m)$ determined by the value of τ_1. If the value of η is false in this interpretation, the procedure is stopped and the value of η is estimated to be nonprovable. In the opposite case we continue the inspection choosing the interpretation $Q(m)$ determined by the value of τ_2. If the value of η is false in this interpretation, the procedure is stopped and the value of η is estimated to be nonprovable. In the opposite case we continue the inspection choosing the interpretation $Q(m)$ determined by τ_3 and so on. Exhausting all the interpretations $Q(m)$ determined successively by the values of $\tau_1, \tau_2, \cdots, \tau_n$ without reaching the decision that the value of η is nonprovable we accept the decision that the value of η is provable.

In fact, the decision procedure just described minimizes the probability of making an incorrect decision only if the a priori probability α that η chooses a provable element of A is sufficiently large. Whenever α is small then the degenerate decision procedure which always estimates the value of η to be nonprovable is better. The exact discrimination between these two decision procedures is contained in theorem 4.

If the a priori probability α is positive, if the values of the random variables $\tau_1, \tau_2, \tau_3, \cdots$ exhaust with probability one the whole set of positive integers, and if the monadic logic (A, M) is $\mathfrak{Q}$-complete then, by theorem 5, for a sufficiently large number of interpretations to be inspected, the nondegenerate estimation procedure is the most favorable one in the sense that the probability of making an incorrect estimate becomes a minimum. Let us note that the condition of semantic consistency is in this case always fulfilled automatically whenever $\alpha > 0$ and (A, M) is semantically $\mathfrak{Q}$-complete.

In the language of monadic logic the decision function β of infinite size occurring in theorem 6 is said to be the heuristic reasoning about the element of A chosen by η and the random variable l is called the length of the heuristic reasoning β.

The content of theorem 6 can be expressed as follows: If $\alpha > 0$, if the values of the random variables $\tau_1, \tau_2, \tau_3, \cdots$ exhaust with probability one the whole set of positive integers and if the monadic logic (A, M) is semantically $\mathfrak{Q}$-complete, then the length of the heuristic reasoning about the value of η is infinite with conditional probability one under the condition that a provable element of A has been chosen by η and it is finite with conditional probability one under the condition that the element of A chosen by η was nonprovable.

Clearly, only the last assertion is practically effective because only nonprovability can be discovered after a finite number of steps. On the other hand, this pessimistic opinion concerning heuristic reasoning is weakened by the fact that if provability is estimated then this result is asymptotically good.

REFERENCES

[1] J. L. DOOB, *Stochastic Processes*, New York, Wiley, 1953.

[2] P. R. HALMOS, "Algebraic logic. I. Monadic algebras," *Compositio Math.*, Vol. 12 (1955), pp. 217–249.

[3] ———, "Algebraic logic. II. Homogeneous locally finite polyadic Boolean algebras of infinite degree," *Fund. Math.*, Vol. 43 (1956), pp. 255–325.

[4] ———, *Measure Theory*, New York, Van Nostrand, 1950.

[5] J. NEYMAN and E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses. I. Unbiassed critical regions of type A and type A₁," *Statist. Res. Mem. London*, Vol. 1 (1936), pp. 1–37.

[6] ———, "Contributions to the theory of testing statistical hypotheses. II. Certain theorems on unbiassed critical regions of type A. III. Unbiassed tests of simple statistical hypotheses specifying the values of more than one unknown parameter," *Statist. Res. Mem. London*, Vol. 2 (1938), pp. 25–57.

[7] A. PEREZ, "Transformation ou sigma-algèbre suffisante et minimum de la probilité d'erreur," *Czechoslovak Math. J.*, Vol. 7 (1957), pp. 115–123.

[8] A. ŠPAČEK, "Statistical estimation of provability in Boolean logic," *Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, 1960, pp. 609–626.