Aplikace matematiky

Igor Vajda

A discrete theory of search. I

Aplikace matematiky, Vol. 16 (1971), No. 4, 241–255

Persistent URL: http://dml.cz/dmlcz/103354

A DISCRETE THEORY OF SEARCH I[1])

IGOR VAJDA

(Received September 8, 1967)

1. INTRODUCTION

In the present paper we shall study a statistical problem which in a somewhat simplified version can be described in the following way. Let Θ be a non-empty set (for example a subset of the n-dimensional Euclidean space or a finite set with a large number of elements) and suppose that an object is located at a point $\vartheta \in \Theta$. Suppose that an experimenter cannot observe ϑ directly but only the set Θ is known to him. Suppose moreover that he may choose arbitrary sets $E_{\delta_1}, E_{\delta_2}, \ldots, E_{\delta_N}$ from a given one-parametric class of subsets of Θ and verify whether ϑ belongs to E_{δ_j} or not. In other words, he observes the values $\xi_j = f_{\delta_j}(\vartheta)$, where f_{δ_j} is the characteristic function of E_{δ_j}. The location is identified after N steps iff the intersection

$$\bigcap_{j=1}^{N} f_{\delta_j}^{-1}(\xi_j) \subset \Theta \tag{1.1}$$

contains exactly one element. However, this result can be reasonably expected only if Θ is a discrete set and, in general, it is sufficient to identify ϑ within certain tolerance limits, for example, within the limits of the form

$$\mathit{diameter}\ \left[\bigcap_{j=1}^{N} f_{\delta_j}^{-1}(\xi_j)\right] \leqq \varepsilon\,. \tag{1.2}$$

In the situation described here each observation ξ_j (together with the parameter δ_j of the experiment used) provides the experimenter with a partial information on ϑ, but after making a fairly large number N of such observations, the total information enables him to meet the identification criteria. The optimum rule of experimentation (strategy of search) $\delta = (\delta_1, \delta_2, \ldots)$ is usually defined by the condition that it minimizes an average (over all $\vartheta \in \Theta$) value $\mathsf{E}_\delta N$ of the necessary number of observations N.

In fact, at the first sight this problem does not seem to be of a statistical nature.

---

[1]) Part II of this paper will appear in the following issue of this journal.

But a statistical problem will arise as soon as we suppose that the average number $\mathsf{E}_\delta N$ is evaluated with respect to an a priori probability distribution of ϑ or that ξ_j is a noisy observation of $f_{\delta_j}(\vartheta)$, for example $\xi_j = f_{\delta_j}(\vartheta) + \zeta_j (\mathrm{mod}\ 2)$, where $\zeta_1, \zeta_2, \ldots$ is a binary random sequence (a noise). However, it is to be noted that in the second case usually $\mathsf{E}_\delta N = +\infty$ so that optimality criteria other than $\mathsf{E}_\delta N$ must be used.

It is obvious that the problems of the described form frequently occur in almost every field of human activity. In connection with them we use the word "search" despite of the obvious fact that the activity of the experimenter in the situations where these problems occur need not necessarily be the search for the location in the standard geometric sense (for example, medical diagnosis, classification by means of a questionnaire in sociology etc.). Another terminology could also be used (for example, in [1] the same problems have been interpreted as a "design of experiments").

In the case when Θ is infinite and an identification criterion of the form (1.2) is used, we shall suppose that there exists a finite quantization $\Omega = \{\theta_1, \theta_2, \ldots, \theta_n\}$ of Θ, where $\theta_i \cap \theta_j = \emptyset$ for $i \neq j$ and

$$\bigcup_{i=1}^{n} \theta_i = \Theta .$$

We suppose that ϑ is identified with a satisfactory accuracy when it is known that it belongs to a cell θ_i. This is true, for example, if *diameter* $[\theta_i] \leqq \varepsilon$, $i = 1, 2, \ldots, n$, and Θ is a compact set. In addition to what has been said here, we shall suppose that the one-parameter class of subsets of Θ at the disposal of the experimenter (i.e. the class of the sets $E_{\delta_1}, E_{\delta_2}, \ldots$) contains sets of the form

$$\bigcup_{\text{over some } i} \theta_i$$

only, but the converse need not be true, i.e. all subsets of this form need not necessarily belong to the class at the experimenter's disposal. Therefore, this class is equivalent to a class of subsets of Ω, namely, $E \equiv F$ if

$$E = \bigcup_{\theta_i \in F} \theta_i \quad \text{for} \quad E \subset \Theta , \quad F \subset \Omega .$$

Since the class of all subsets of Ω is finite, the class $\mathscr{E}$ of all possible functions f_{δ_j} at the experimenter's disposal can be enumerated, $\mathscr{E} = \{f_1, f_2, \ldots, f_m\}$. Thus, in this case δ_j assume values from $M = \{1, 2, \ldots, m\}$. Moreover, since we need not distinguish between the elements of the cells θ_i, we can consider Ω and $\theta \in \Omega$ instead of Θ and ϑ.

Hence, under our assumptions, the simplest model of search is described by a pair $(\Omega, \mathscr{E})$, where Ω is a finite set and $\mathscr{E}$ a finite class of binary functions defined on Ω.

This model has been considered in [2]. In [1] a generalized version of this model has been investigated, with $f_l(\theta) \in \mathscr{E}$ replaced by probability distributions $P_l(\,.\mid\theta)$ on a finite set A. The generalized approach presented in [1] is justified from both practical and theoretical points of view, because the noisy observations (for example with the additive noise considered above), which are frequently met in the practice, are well described by the model with $\mathscr{E} = \{P_1(. \mid \theta), P_2(. \mid \theta), \ldots, P_m(. \mid \theta)\}$. On the other hand, in the framework of this model one can very explicitly see a close relation between the theory of search and the Shannon's information theory (coding and choosing $\delta_1, \delta_2, \ldots, \delta_N$ are in a certain sense equivalent).

In the present paper which is rather a review, an introduction to the problems investigated in [1, 2] is given, basic results of these papers are summarized from a unified Bayesian point of view, and some new results are established. By the Bayesian point of view we mean that

(i) an apriori probability distribution on Ω is supposed;

(ii) if the experimentation is carried out in N steps according to a strategy $\delta = (\delta_1, \delta_2, \ldots)$, then the terminal decision on θ is adopted (on the basis of the "information" $(\delta_1, \xi_1), (\delta_2, \xi_2), \ldots, (\delta_N, \xi_N)$) in accordance with the maximum likelihood principle[2]);

(iii) the optimality of δ is numerically measured by the average probability of error $e_\delta N$ corresponding to the maximum likelihood estimator $\hat{\theta}(\delta_1, \xi_1, \ldots, \delta_N, \xi_N)$ of θ, or by the rate of convergence of $e_\delta N \to 0$ for $N \to \infty$.

Papers [1, 2] essentially differ from a large number of papers dealing with discrete models of search by measures of optimality of search strategies used there. The measure defined in (iii) and used also in [1] is very closely related to that used in [2]. It conditions to a great extent a relation between the information theory and the theory of search, in particular, effective applications of coding theory results in search problems. Moreover, it is acceptable from the practical point of view, because its interpretation is clear. On the other hand, it is not so easy to evaluate it, except for the so-called random strategies investigated in [1, 2].

Let us characterize more precisely individual sections of this paper. In the rest of this introductory section some conrete examples of $(\Omega, \mathscr{E})$ will be given, for illustration and for later references. A general model of search and such concepts as a strategy δ and $\mathsf{E}_\delta N$, $e_\delta N$ will be stuudied with more precision in Sec. 2. In Sec. 2–4, properties of $\mathsf{E}_\delta N$ $e_\delta N$ are studied for special classes of strategies, in Sec. 3 the relation between the information and the search theory is described in more detail. Main attention of the following sections will be paid to asymptotic properties of random strategies for which $\delta_1, \delta_2, \ldots$ are mutually independent and equally distributed random variables. It will be seen that for many $(\Omega, \mathscr{E})$ the random strategies are (asymptotically) almost as good as the best systematic strategies, being at the same time much simpler (for example, from the point of view of its programming on a computer).

---

[2]) By ξ_j we denote a random variable conditionally distributed by $P_l(. \mid \theta_i)$ provided $\delta_j = l$, $\theta = \theta_i$.

Example 1.1. Let Ω be arbitrary and let $\mathcal{E} = \{f_1, f_2, \ldots, f_m\}$ be the class of all mappings $\Omega \to \{0, 1\}$, i.e. $m = 2^n$. Let us consider, for example, that Ω is a set of n bulbs, and suppose that one of them fails. Any subset of Ω may be tested by circuiting all the corresponding bulbs in a series. The result of this experiment is 0 (the bulbs are alight) or 1 in the opposite case (the failed bulb belongs to the subset).

Example 1.2 (due to J. Nedoma [3]). Let Ω be the same as above and $\mathcal{E} = \{f_1, f_2, \ldots, f_{n-1}\}$, where

$$f_j(\theta) = \begin{cases} 1 & \text{if} \quad \theta \in \{\theta_1, \theta_2, \ldots, \theta_j\} , \\ 0 & \text{if} \quad \theta \in \{\theta_{j+1}, \theta_{j+2}, \ldots, \theta_n\} . \end{cases}$$

This corresponds to the case when all the bulbs from Ω are already circuited in the series and only subsets E_j of the form $\{\theta_1, \theta_2, \ldots, \theta_j\}$ can be tested (see Fig. 1).

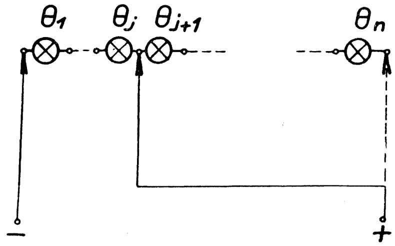

Fig. 1.

Example 1.3. Let Ω be the same as above and let $\mathcal{E} = \{f_1, f_2, \ldots, f_n\}$, where

$$f_j(\theta) = \begin{cases} 1 & \text{if} \quad \theta = \theta_j , \\ 0 & \text{if} \quad \theta \neq \theta_j . \end{cases}$$

This corresponds to the case when only single bulbs can be tested. In this case $f_j(\theta) = 1$ iff $\theta = \theta_j$ i.e. iff the j-th bulb fails.

Example 1.4. Let $\Omega = \{\theta_1, \theta_2, \theta_3\}$, $M = \{1, 2\}$, $A = \{0, 1\}$, i.e. $a, m = 2$, $n = 3$, $\mathcal{E} = \{P_1(. \mid \theta), P_2(. \mid \theta)\}$, where

$$\begin{aligned} P_1(0 \mid \theta_1) &= 1 = 1 - P_1(1 \mid \theta_1) \\ P_1(0 \mid \theta_2) &= p = 1 - P_1(1 \mid \theta_2) \\ P_1(0 \mid \theta_3) &= p = 1 - P_1(1 \mid \theta_3) \end{aligned}$$

$$\begin{aligned} P_2(0 \mid \theta_1) &= 1 - p = 1 - P_2(1 \mid \theta_1) \\ P_2(0 \mid \theta_2) &= 0 \qquad = 1 - P_2(1 \mid \theta_2) \\ P_2(0 \mid \theta_3) &= 1 - p = 1 - P_2(1 \mid \theta_3) . \end{aligned}$$

and $0 < p < 1$. Clearly, on the basis of a subsequent independent observation of θ throughout the channels $P_1(. \mid \theta)$ or $P_2(. \mid \theta)$ we are not able to distinguish θ_2, θ_3 or θ_1, θ_3 respectively, even if the sample size is arbitrarily large. But using an appropriate combination of both the channels, all three values $\theta_1, \theta_2, \theta_3$ are distinguishable.

Remark. In the standard model of mathematical statistics, parameter θ is observed through a unique channel $P(. \mid \theta)$ defined by a list of conditional probability distributions $\{P(. \mid \theta)\}$, $\theta \in \Omega$. However, it is supposed that the channel guarantees $P(. \mid \theta_i) \neq P(. \mid \theta_k)$ for "almost all" $\theta_i \neq \theta_k$. This separability condition makes it possible to find out satisfactory estimators of θ based on a large number of independent observations. In the theory of search this is not the case. Here it is supposed that the separability condition is not satisfied by any of the channels $P_l(. \mid \theta) \in \mathscr{E}$ itself. According to our opinion, this is a characteristic assumption of the theory of search. If this situation occurs, θ must be observed through a fitting combination of channels from $\mathscr{E}$. Within the extent of this paper, this process is considered as a search.

2. GENERAL MODEL

Let $\Omega = \{\theta_1, \theta_2, \ldots, \theta_n\}$ be a non-empty set and let θ be a random variable uniformly distributed on Ω,

$$(2.1) \qquad \mathsf{P}_\theta[\theta = \theta_i] = \frac{1}{n} \quad i = 1, 2, \ldots, n .$$

The uniformity assumption frequently occurs in practice and we accept it for the sake of simplicity, but the theory developed below could be extended to an arbitrary P_θ.

The basic concept of our model of search is a class $\mathscr{E} = \{P_1(. \mid \theta), P_2(. \mid \theta), \ldots, P_m(. \mid \theta)\}$, where $P_l(. \mid \theta)$ is a probability distribution on a set $A = \{0, 1, \ldots, a - 1\}$ for any pair $(l, \theta) \in M \otimes \Omega$, $M = \{1, 2, \ldots, m\}$. The pair $(\Omega, \mathscr{E})$ defines the model.

In the framework of this model we shall consider abstract random variables $\theta, \pi, \eta, \delta, \xi$. θ has been defined above, $\pi = (\pi_1, \pi_2, \ldots)$ is a sequence of mutually (and on θ) independent random variables, $\pi_j \in M$, $\delta = (\delta_1, \delta_2, \ldots)$, $\xi = (\xi_1, \xi_2, \ldots)$ are random sequences (in general depending on θ, π), $\delta_j \in M$, $\xi_j \in A_2$ and $\eta = (\eta_1, \eta_2, \ldots)$ is defined by $\eta_j = (\delta_j, \xi_j) \in M \otimes A = Y$. Now our aim is to define the joint distribution $\mathsf{P}_{\eta\pi\theta} = \mathsf{P}_{\xi\delta\pi\theta}$ (on the standard σ-algebra of subsets of the sample space $Y^\infty \otimes M^\infty \otimes \Omega$ of the random vector (η, π, θ)).

Let us consider a probability distribution $\mu = (\mu_1, \mu_2, \ldots, \mu_m)$ on M and let $d = (d_1, d_2, \ldots)$ be a sequence of mappings, d_1: $M \to M$, d_j: $Y^{j-1} \otimes M \to M$, where $d_1(l) = l$ and $d_j(y_1, y_2, \ldots, y_{j-1}, l) \in M$ is arbitrary, $j = 2, 3, \ldots$ The pair (μ, d) will be called a strategy. Next we shall show in which manner the strategy uniquely defines $\mathsf{P}_{\eta\pi\theta}$.

Let $l_j, l'_j \in M$, $k_j \in A$, $y_j = (l_j, k_j) \in Y$, and define

$$(2.2) \qquad p_N(y_1, y_2, \ldots, y_N, l'_1, l'_2, \ldots, l'_N \mid \theta) = \begin{cases} 0 \\ \prod_{j=1}^{N} \mu_{l_j'} \, P_{l_j}(k_j \mid \theta) \end{cases}$$

depending on whether the condition

$$(2.3) \qquad l_1 = l'_1\,, \quad l_j = d_j(y_1, y_2, \ldots, y_{j-1}, l'_j) \quad j = 2, 3, \ldots, N$$

is satisfied or not. Clearly

$$\sum_{\substack{y_1 \ldots y_N \\ l_1' \ldots l_N'}} p_N(y_1, y_2, \ldots, y_N, l'_1, l'_2, \ldots, l'_N \mid \theta) = 1 \quad \text{for every} \quad \theta \in \Omega,\ N = 1, 2, \ldots$$

If we put

$$(2.4) \qquad \mathsf{P}_{\eta_1 \ldots \eta_N \pi_1 \ldots \pi_N \mid \theta} = p_N(\,.\, \mid \theta) \quad N = 1, 2, \ldots$$

then, using the standard extension argument, we obtain $P_{\eta\pi|\theta}$. Therefore (cf. (2.1))

$$(2.5) \qquad \mathsf{P}_{\eta\pi\theta} = \frac{1}{n} \mathsf{P}_{\eta\pi|\theta}\,.$$

Notice that this definition of $\mathsf{P}_{\eta\pi\theta}$ implies that $\pi_1, \pi_2, \ldots$ are identically distributed, $\mu_l = \mathsf{P}_\pi\,[\pi_j = l]$, $j = 1, 2, \ldots$

An important role in our considerations will be played by the distribution $\mathsf{P}_{\eta\theta}$ on $Y^\infty \otimes \Omega$. We shall not distinguish between two strategies (μ, d), (μ', d') unless the corresponding distributions $\mathsf{P}_{\eta\theta}$, $\mathsf{P}'_{\eta\theta}$ are different. If $\mathsf{P}_{\eta\theta}$, $\mathsf{P}'_{\eta\theta}$ are identical, we shall say that the strategies are equivalent (in symbols, $(\mu, d) \equiv (\mu', d')$). Using this together with the following Lemma (which follows from (2.2), i.e. from the definition of $\mathsf{P}_{\eta\pi\theta}$), we shall be able to identify the concept of $\delta = (\delta_1, \delta_2, \ldots)$ with the concept of the strategy (in symbols, $\delta \equiv (\mu, d)$).

Lemma 2.1. *If $\delta_1, \delta_2, \ldots$ is a realization*[3]) *of the random sequence δ, then*

$$(2.6) \qquad \mathsf{P}_{\xi\theta|\delta} = \frac{1}{n} \bigotimes_{j=1}^{\infty} P_{\delta_j}(\,.\, \mid \theta)\,.$$

The equivalence $\delta \equiv (\mu, d)$ discussed above is justified by the fact that the distribution P_δ of δ together with $\mathsf{P}_{\xi\theta|\delta}$ (which is known if $\mathscr{E}$ is known) determines uniquely $\mathsf{P}_{\eta\theta} = \mathsf{P}_{\delta\xi\theta}$. In other words, to know $\mathsf{P}_{\eta\theta}$ it is not necessary to know $(\Omega, \mathscr{E}, \mu, d)$. It is sufficient to know the triple $(\Omega, \mathscr{E}, \delta) = (\Omega, \mathscr{E}, \mathsf{P}_\delta)$ only.

---

[3]) The random variables δ_j, ξ_j, η_j, π_j, θ as well as their realizations are denoted in this paper by the same symbols.

According to (2.3) it holds

$$(2.7) \qquad \delta_1 = d_1(\pi_1)\,, \quad \delta_j = d_j(\eta_1, \eta_2, \ldots, \eta_{j-1}, \pi_j) \quad j = 2, 3, \ldots$$

for any strategy δ. If $d_1, d_2, \ldots$ are such that δ_j does not depend on π_j (i.e. if (1.8) holds), δ will be called a pure strategy; in the opposite case δ will be called a mixed strategy. A strategy δ will be called sequential or non-sequential depending on whether δ depends on $\eta_1, \eta_2, \ldots, \eta_{j-1}$ for some $j > 1$ or not. If δ is non-sequential, δ_j are mutually (and on θ) independent random variables. If, moreover, δ_j are equally distributed for $j = 1, 2, \ldots$, then δ will be called a random strategy. If δ is a random strategy, then we can put

$$(2.8) \qquad \delta_j = \pi_j\,, \quad j = 1, 2, \ldots . \quad \text{i.e.} \quad \delta \equiv \mu\,.$$

Lemma 2.2. *If δ is a pure strategy, then*

$$(2.9) \qquad \mathsf{P}_{\eta\pi\theta} = \mathsf{P}_{\eta\theta} \otimes \mathsf{P}_\pi \quad \text{where} \quad \mathsf{P}_\pi = \mu \otimes \mu \otimes \ldots$$

If δ is a random strategy, then

$$(2.10) \qquad \mathsf{P}_{\eta\theta} = \bigotimes_{j=1}^{\infty} \mathsf{P}_{\eta_j\theta} \quad \text{where} \quad \mathsf{P}_{\eta_1\theta} = \mathsf{P}_{\eta_2\theta} = \ldots$$

Proof. Cf. (2.2), (2.3).

Notice that this Lemma illustrates the fact that δ uniquely determines $\mathsf{P}_{\eta\theta}$ (provided δ is a pure or random strategy, which are the most interesting cases from both theoretical and practical points of view).

The interpretation of what we have introduced above was already outlined in Sec. 1. Briefly speaking, the experimenter's aim is to establish the value of θ which is not directly observable by him. He has at his disposal only indirect observations through the channels $P_i(. \mid \theta) \in \mathscr{E}$. We suppose that at the moments $j = 1, 2, \ldots$ he decides for a channel $P_{\delta_j}(. \mid \theta)$ and that he observes the value ξ_j. The unknown parameter θ does not depend on j, and the information contained in $\xi_1, \xi_2, \ldots$ concerning θ depends on the channels $P_{\delta_1}(. \mid \theta), P_{\delta_2}(. \mid \theta), \ldots$ (i.e. on $\delta_1, \delta_2, \ldots$). At the moment N experimentation is stopped and on the basis of the information represented by $\eta_1, \eta_2, \ldots, \eta_N$ the terminal decision concerning θ is adopted. We suppose that the distribution $\mathsf{P}_{\eta_1\ldots\eta_N\theta}$ is completely known to the experimenter. Notice that $\mathsf{P}_{\delta_1\ldots\delta_N}$ is defined by the experimenter himself (it is, in fact, the strategy of search) and $\mathsf{P}_{\xi_1\ldots\xi_N\theta\mid\delta_1\ldots\delta_N}$ is known if $\mathscr{E}$ is known (see (2.6)). Thus, in other words, we suppose that $\mathscr{E}$ is completely known to the experimenter.

Example 2.1. Let $(\Omega, \mathscr{E})$ be as in Example 1.4 and let us consider the following sequential pure strategy δ':

$$\delta'_1 = 1\,, \quad \delta'_j = \begin{cases} 1 & \text{if} \quad \sum_{r=1}^{j-1} \xi_r = 0 \\ 2 & \text{if} \quad \sum_{r=1}^{j-1} \xi_r \neq 0 \quad j = 2, 3, \ldots \end{cases}$$

Example 2.2. Let $(\Omega, \mathscr{E})$ be the same as in Example 2.1, $\mu = (\mu_1, 1 - \mu_1)$, $\mu_1 \in \in [0, 1/2]$ and define a strategy δ by (2.7). Clearly, δ is a random strategy.

Suppose now that there exist atoms $f_l(\theta) \in A$ of the distributions from $\mathscr{E}$. More precisely let $P_l(f_l(\theta) \mid \theta) = 1$ for all $l \in M$, $\theta \in \Omega$. Then $\mathscr{E}$ can be defined as a set $\{f_1, f_2, \ldots, f_m\}$ of functions of the form $\Omega \to A$ (cf. Sec. 1).

Example 2.3. Let the situation be the same as in Example 1.2, i.e. let Ω be arbitrary and $\mathscr{E} = \{f_1, f_2, \ldots, f_{n-1}\}$. If $\delta_j^* = j$ for $j = 1, 2, \ldots, n-1$ and $\delta_j^* = n - 1$ for $j \geqq n$, then $\delta^* = (\delta_1^*, \delta_2^*, \ldots)$ is a pure non-sequential strategy.

Example 2.4. Let the situation be the same as in Example 2.3 and suppose that $n = 2^k$ for some integer k. Put $\xi_0 = 0$ and

$$\delta_j' = \sum_{r=0}^{j-1} (-1)^{\xi_r} . 2^{k-r-1}, \quad j = 1, 2, \ldots, k .$$

Then $\delta' = (\delta_1', \delta_2', \ldots)$ is a sequential pure strategy.

Example 2.5. Ω and $\mathscr{E}$ are the same as in Example 2.3 again, and define $\mu = ((n-1)^{-1}, (n-1)^{-1}, \ldots, (n-1)^{-1})$. Then δ defined by (2.7) is a random strategy.

Let us now consider the general model $(\Omega, \mathscr{E})$ again. Except for trivial cases, any such model admits an infinite number of strategies. To be able to appreciate the quality of various strategies we shall try to find a numerical expression of the optimality. Let $\mathsf{P}_{\eta\theta}$ correspond to $(\Omega, \mathscr{E}, \delta)$, and define a random variable $N = N(\eta)$ as the minimum N for which

$$\max_{\theta_i \in \Omega} \mathsf{P}_{\theta|\eta_1 \ldots \eta_N}[\theta = \theta_i] = 1 . \tag{2.11}$$

In other words, N is the minimum number of observations admitting an errorless decision concerning θ. The expectation

$$\mathsf{E}_\delta N = \frac{1}{n} \sum_{i=1}^{n} \int_{Y^\infty} N \, \mathrm{d}\mathsf{P}_{\eta|\theta_i} \tag{2.12}$$

is undoubtedly one of possible numerical measures of optimality of δ. The optimum strategy δ' (if it exists) could be then defined by

$$\mathsf{E}_{\delta'} N = \inf_{\delta \in \Delta} \mathsf{E}_\delta N , \tag{2.13}$$

where Δ denotes the class of all strategies (i.e. the class of all distributions P_δ on M^∞).

However, it is to be noted in advance that, unfortunately, except for the special cases where $\mathscr{E} = \{f_1, f_2, \ldots, f_m\}$, it holds

$$\inf_{\delta \in \Delta} \mathsf{E}_\delta N = +\infty . \tag{2.14}$$

In fact, if $\mathscr{E}$ is not of the special form given above, it holds

$$(2.15) \qquad \liminf_N \mathsf{P}_{\eta_1\ldots\eta_N}\left[\max_{\theta_i\in\Omega} \mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \theta_i] < 1\right] > 0$$

for arbitrarily large N.

Suppose now that $\mathscr{E}$ is of the special form, where f_l: $\Omega \to A$ are arbitrary. We shall say that $\mathscr{E}$ is separating the elements of Ω (for the sake of brevity: $\mathscr{E}$ is separating Ω or $\mathscr{E}$ is separating) if for any pair $\theta_i \neq \theta_k$ there exists $f_l \in \mathscr{E}$ such that $f_l(\theta_i) \neq f_l(\theta_k)$.

Lemma 2.3. *If $\mathscr{E} = \{f_1, f_2, \ldots, f_m\}$, then it holds*

$$(2.16) \qquad \mathsf{E}_\delta N \geqq \log_a n$$

for any strategy δ. The inequality

$$(2.17) \qquad \inf_{\delta\in\Delta} \mathsf{E}_\delta N < +\infty$$

holds iff $\mathscr{E}$ is separating Ω. If this is satisfied, then there exists a pure non-sequential δ such that

$$(2.18) \qquad \mathsf{P}_{\eta\theta}[N(\eta) < n] = 1\,.$$

Proof. Let $\delta_1, \delta_2, \ldots$ be an arbitrary realization of a strategy δ and denote by N_i the least N for which (2.11) holds with $\eta_j = (\delta_j, \varepsilon_{i\delta_j})$, where $\varepsilon_{il} = f_l(\theta_i)$. It is clear that the vectors

$$\begin{aligned} \bar{\varepsilon}_1 &= (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{1N_1}) \\ \bar{\varepsilon}_2 &= (\varepsilon_{21}, \varepsilon_{22}, \ldots, \varepsilon_{2N_2}) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \bar{\varepsilon}_n &= (\varepsilon_{n1}, \varepsilon_{n2}, \ldots, \varepsilon_{nN_n}) \end{aligned}$$

can be interpreted as a variable-length code. The code is defined in such manner that if $N_i \leqq N_k < +\infty$, then $(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iN_i}) \neq (\varepsilon_{k1}, \varepsilon_{k2}, \ldots, \varepsilon_{kN_i})$. It is well known (see § 2.3 in [4]) that if a code possesses this property, it holds

$$\sum_{i=1}^n p_i N_i \geqq \sum_{i=1}^n p_i \log_a p_i$$

for any probability scheme $p = (p_1, p_2, \ldots, p_n)$ and, consequently,

$$(2.19) \qquad \frac{1}{n} \sum_{i=1}^n N_i \geqq \log_a n\,.$$

By (2.12) and Fubini's theorem

$$\mathsf{E}_\delta N = \int_{M^\infty} \left[\frac{1}{n} \sum_{i=1}^n \int_{Y^\infty} N \, \mathrm{d}\mathsf{P}_{\eta|\delta\theta_i}\right] \mathrm{d}\mathsf{P}_\delta\,.$$

Since $\mathsf{P}_{\eta|\delta\theta_i}[\eta \in F_i] = 1$, $i = 1, 2, \ldots, n$ for the cylinder $F_i \subset Y^\infty$ with the base $[(\delta_1, \varepsilon_{i1}), (\delta_2, \varepsilon_{i2}), \ldots, (\delta_{N_i}, \varepsilon_{iN_i})]$ in the first N_i coordinates, we can write

$$\mathsf{E}_\delta N = \int_{M^\infty} N(\delta)\, \mathrm{d}\mathsf{P}_\delta \quad \text{where} \quad N(\delta) = \frac{1}{n} \sum_{i=1}^{n} N_i .$$

This together with (2.19) yields (2.16).

Let us now prove the necessary and sufficient condition for (2.17). If this condition is not satisfied, there exists $\theta_i \neq \theta_k$ for which $f_{\delta_j}(\theta_i) \neq f_{\delta_j}(\theta_k)$, $j = 1, 2, \ldots$ for any $\delta_1, \delta_2, \ldots$ This implies $\mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \theta_k] = \mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \theta_i] \leqq \frac{1}{2}$ for any N, δ. Since

$$\liminf_N \mathsf{P}_{\eta_1\ldots\eta_N}\{\max_{\theta_j \in \Omega} \mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \theta_j] = \mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \theta_k]\} \geqq \frac{1}{n},$$

(2.15) holds as well as (2.14). If, conversely, $\mathscr{E}$ is separating Ω and the pairs $\theta_i \neq \theta_k$ (of the total number $n(n-1)$) are arbitrarily ordered, there exist $\delta_1, \delta_2, \ldots, \delta_{n(n-1)}$ such that f_{δ_j} is separating the j-th pair. However, this implies that

$$\mathsf{P}_{\eta\theta}[N(\eta) \leqq n(n-1)] = 1 \tag{2.20}$$

for example for $\delta = (\delta_1, \delta_2, \ldots, \delta_{n(n-1)}, \ldots)$ and consequently

$$\inf_{\delta \in \Delta} \mathsf{E}_\delta N \leqq n(n-1) .$$

The stronger inequality (2.18) following from Lemma on p. 811 in [2] enables us to establish the following property of the optimum strategies:

Theorem 2.1. *If $\mathscr{E} = \{f_1, f_2, \ldots, f_m\}$ separates Ω, then*

$$\log_a n \leqq \inf_{\delta \in \Delta} \mathsf{E}_\delta N < n \tag{2.21}$$

holds.

Suppose now that $(\Omega, \mathscr{E})$ are the same as in Example 1.2.

Theorem 2.2. *If δ^*, δ', δ are defined in the same way as in Examples* 2.3., 2.4., 2.5., *then*

$$\mathsf{E}_{\delta^*} N = \frac{n}{2} + \frac{1}{2} - \frac{1}{n}, \tag{2.22}$$

$$\mathsf{E}_{\delta'} N = \log_2 n , \tag{2.23}$$

$$\mathsf{E}_{\delta} N = \frac{3n}{2} - \frac{5}{2} + \frac{1}{n}. \tag{2.24}$$

Proof. First we shall prove (2.23). In the definition of δ' it was supposed $n = 2^k$. This definition also implies that the class of all $\theta \in \Omega$ for which the maximum in (2.11) is attained after N steps contains exactly 2^{k-N} elements. Hence the maximum is equal to 2^{N-k}. The equality in (2.11) holds for $N = k = \log_2 n$ independently of $\eta_1, \eta_2, \ldots, \eta_N$. Thus (2.23) holds.

Define a decomposition $\{Y_1, Y_2, \ldots\}$ of Y^∞ by $Y_i = N^{-1}(i)$. According to (2.12),

$$\mathsf{E}_\delta N = \frac{1}{n} \sum_{i=1}^{n} \sum_{N=1}^{\infty} N\, p(N \mid i), \quad \text{where} \quad p(N \mid i) = \mathsf{P}_{\eta|\theta_i}[\eta \in Y_N] \tag{2.25}$$

holds for any δ.

If $\theta = \theta_1$ or θ_n, then $N(\eta) = N$ iff $\delta_N = 1$ or $\delta_N = n - 1$ and $\delta_j \neq 1$ or $\delta_j \neq n - 1$, $j = 1, 2, \ldots, N - 1$ respectively. The probability of every realization $\delta_1, \delta_2, \ldots, \delta_N$ is equal to $(n - 1)^{-N}$ and the number of the realizations for which $N(\eta) = N$ is equal to $(n - 2)^{N-1}$ so that

$$p(N \mid i) = \frac{(n - 2)^{N-1}}{(n - 1)^N}, \quad i = 1, n .$$

Analogously

$$p(N \mid i) = 2(n - 1)^{-N} \left[(n - 2)^{N-1} - (n - 3)^{N-1}\right], \quad 1 < i < n .$$

Now (2.24) holds by (2.25).

If we consider δ^*, then it is easy to see that

$$p(N \mid i) = \begin{cases} 0 & \text{for} \quad N \neq i \\ 1 & \text{for} \quad N = i \end{cases} \qquad i = 1, 2, \ldots, n - 1 ,$$

and

$$p(N \mid n) = \begin{cases} 0 & \text{for} \quad N \neq n - 1 , \\ 1 & \text{for} \quad N = n - 1 . \end{cases}$$

This together with (2.25) implies (2.22) Q.E.D.

Remark. If n is not of the form 2^k, a slightly modified construction described in Example 2.4 yields a strategy δ' which is again optimal in the sense of (2.13). It satisfies the inequality

$$\log_2 n \leqq \mathsf{E}_{\delta'} N < \log_2 n + 1$$

(cf. (2.23)).

As we said above, if $P_l(. \mid \theta) \in \mathscr{E}$ are not monoatomic distributions, $\mathsf{E}_\delta N$ must be replaced by another measure of optimality and for this purpose we adopt the average probability of error corresponding to a Bayesian estimator $\hat{\theta}_N = \hat{\theta}_N(\eta_1, \eta_2, \ldots \ldots, \eta_N)$ of θ. The Bayesian estimator is defined as an arbitrary mapping $Y^N \to \Omega$ such that

$$\mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \hat{\theta}_N] = \max_{\theta_i \in \Omega} \mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \theta_i] . \tag{2.26}$$

It can be defined by (2.26) except for $\eta_1, \eta_2, \ldots, \eta_N$ such that the maximum is attained for more than one θ. For these $\eta_1, \eta_2, \ldots, \eta_N$, $\hat{\theta}_N$ can be defined as a random variable with a uniform distribution on the set of the maximizing θ_i's.

Since the average probability of error corresponding to any estimator $\hat{\theta}_N : Y^N \to \Omega$ can be written in the form

$$\int_{Y^N} (1 - \mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \hat{\theta}_N]) \, d\mathsf{P}_{\eta_1\ldots\eta_N} ,$$

the Bayesian estimator minimizes the average probability of error for any given $(\Omega, \mathscr{E}, \delta, N)$. The minimum average probability of error can be written in the form

$$e_\delta N = \frac{1}{n} \sum_{i=1}^{n} \mathsf{P}_{\eta_1\ldots\eta_N|\theta_i}[\hat{\theta}_N \neq \theta_i] . \tag{2.27}$$

Unless the contrary is explicitely stated, $\hat{\theta}_N$ shall denote the Bayesian estimator.

The separability condition introduced in Lemma 2.3 for the class $\mathscr{E}$ of monoatomic distributions can be extended to any $\mathscr{E}$ in the following way. We shall say that $\mathscr{E} = \{P_1(. \mid \theta), P_2(. \mid \theta), \ldots, P_m(. \mid \theta)\}$ is separating the elements of Ω (separating Ω) if for any pair $\theta_i \neq \theta_k$ there exists $P_l(. \mid \theta) \in \mathscr{E}$ such that $P_l(. \mid \theta_i) \neq P_l(. \mid \theta_k)$.

Theorem 2.3. *For any strategy* δ

$$e_\delta N \geqq e_\delta(N + 1) \quad N = 1, 2, \ldots \tag{2.28}$$

holds, i.e.

$$e_\delta(\infty) = \lim_N e_\delta N \tag{2.29}$$

exists. A strategy $\delta \in \Delta$ *for which*

$$e_\delta(\infty) = 0 \tag{2.30}$$

exists iff $\mathscr{E}$ *separates* Ω. *If this condition is satisfied, then there exists a strategy* δ *and* $\lambda \in [0, 1)$ *such that*

$$e_\delta N \leqq \lambda^N \tag{2.31}$$

for all sufficiently large N.

Proof. Let us define an estimator $\bar{\theta}_{N+1} : Y^{N+1} \to \Omega$ in the following way:

$$\bar{\theta}_{N+1}(\eta_1, \eta_2, \ldots, \eta_{N+1}) = \hat{\theta}_N(\eta_1, \eta_2, \ldots, \eta_N) .$$

Inequality (2.28) is a consequence of the following one:

$$e_\delta(N + 1) \leqq \int_{Y^N} (1 - \mathsf{P}_{\theta|\eta_1\ldots\eta_N}[\theta = \bar{\theta}_N]) \, \mathrm{d}\mathsf{P}_{\eta_1\ldots\eta_N} .$$

Let now $\mathscr{E}$ be not separating Ω. This implies (see (2.6)) $\mathsf{P}_{\xi_1\ldots\xi_N|\delta_1\ldots\delta_N\theta_i} = \mathsf{P}_{\xi_1\ldots\xi_N|\delta_1\ldots\delta_N\theta_k}$ for some $\theta_i \neq \theta_k$ and all realizations $\delta_1, \delta_2, \ldots, \delta_N$ of every strategy δ. Hence also $\mathsf{P}_{\eta_1\ldots\eta_N|\theta_i} = \mathsf{P}_{\eta_1\ldots\eta_N|\theta_k}$ so that

$$\mathsf{P}_{\eta_1\ldots\eta_N|\theta_i}[\hat{\theta}_N \neq \theta_i] + \mathsf{P}_{\eta_1\ldots\eta_N|\theta_k}[\bar{\theta}_N \neq \theta_k] \geqq 1 .$$

This and (2.27) imply $\lim_N e_\delta N \geqq 1/n$. The converse follows from Th. 3.2. proved below, but it can also be deduced by the following argument: If we define a pure non-sequential strategy $\delta_{km+j} = j$, $j = 1, 2, \ldots, m$, $k = 1, 2, \ldots$ and $\mathscr{E}$ separates Ω then clearly $\mathsf{P}_{\xi|\delta\theta_i}$ and $\mathsf{P}_{\xi|\delta\theta_k}$ are singular for every $\theta_i \neq \theta_k$ (see (2.6) and Lemma 1 in [5] or (3.1) in [6]). More precisely, the definition of δ and the assumption of separability imply

$$\liminf_N \frac{1}{N} \sum_{j=1}^{N} \mathrm{Var}\,(\mathsf{P}_{\xi_j|\delta\theta_i}, \mathsf{P}_{\xi_j|\delta\theta_k}) > 0 ,$$

Tab. 1.

Realization of $\xi_1, \xi_2, \ldots$ $\ldots, \xi_N$	The number of such $\xi_1, \xi_2, \ldots$ $\ldots, \xi_N$	Bayesian decision $\hat{\theta}_N$	$P_{\xi_1\ldots\xi_N|\theta_1}$	$P_{\xi_1\ldots\xi_N|\theta_2}$	$P_{\xi_1\ldots\xi_N|\theta_3}$
$0, 0, \ldots, 0$	1	θ_1	1	p^N	p^N
$0, 0, \ldots, 0, 1$	1	θ_2	0	$p^{N-1}(1 - p)$	$p^{N-1}(1 - p)$
$0, \ldots, 0, 1,$ $\xi_{k+2},$ $\xi_{k+3}, \ldots, \xi_N$ where $k < N - 1$	$\binom{N - k - 1}{m}$ where $m = \sum_{i=k+2}^{N} \xi_i$	θ_3 if $m < N - k - 1$ θ_2 if $m = N - k - 1$	0	0 if $m < N - k - 1$ $p^k(1 - p)$ if $m = N - k - 1$	$p^{k+m}(1 - p)^{N-k-m}$

where Var (P, P') denotes the total variation of P and P'. This together with results of [5] (or with the facts presented on p. 459 of [6]) implies not only (2.30), but also (2.31). The applicability of the results of [5, 6] is justified by the assumption that the terms in the random sequence ξ are mutually independent for any fixed realization $\delta_1, \delta_2, \ldots$

In the rest of this paper the asymptotic behaviour of $e_\delta N$ for $N \to \infty$ will be studied. The optimum strategy (if it exists) is defined by

$$e_{\delta'} N = \inf_{\delta \in \Delta} e_\delta N \quad \text{for all } N \text{ greater than some } N_0 \,. \tag{2.32}$$

This asymptotic approach to the optimality is justified by the fact that (2.32) holds for some δ' and for all $N = 1, 2, \ldots$ only in exceptional cases.

Example 2.6. Let $(\Omega, \mathscr{E}, \delta')$ be the same as in Example 2.1. Obviously $\mathscr{E}$ is separating Ω so that $e_{\delta'}(\infty) = 0$. We shall prove that in this case $e_{\delta'} N = p^{N+\sigma(1)}$. It follows from the table that $\mathsf{P}_{\eta_1 \ldots \eta_N | \theta_1}[\hat{\theta}_N \neq \theta_1] = 0$, $\mathsf{P}_{\eta_1 \ldots \eta_N | \theta_2}[\hat{\theta}_N \neq \theta_2] = p^N$, $\mathsf{P}_{\eta_1 \ldots \eta_N | \theta_3} \cdot$ $\cdot [\hat{\theta}_N \neq \theta_3] = N p^N (1 - p)$, so that

$$e_{\delta'} N = \frac{p^N}{3} \left[1 + \frac{N(1-p)}{p} \right] \quad \text{Q.E.D.}$$

Intuitively it is clear that δ' is the optimum in the sense of (2.32), but we shall not go into proving this fact here.

References

[1] *I. Vajda:* On random design of experiments (in Russian). Information Transmission Problems 4 (1968), N. 4, 26—36.

[2] *A. Rényi:* On the theory of random search. Bull. Amer. Math. Soc. 71 (1965), N. 1, 809—828.

[3] *J. Nedoma:* Coding of finite sets of messages. Presented at 2nd Czechoslovak Conf. on Cybernetics, Prague 16.—19. XI. 1965.

[4] *A. Feinstein:* Foundations of Information Theory, McGraw Hill, N. Y. 1958.

[5] *I. Vajda:* Rate of convergence of the information in a sample concerning a parameter. Czechoslovak Math. Journal, 17 (1967), N. 2, 225—231.

[6] *I. Vajda:* On statistical decision problems with finite parameter space. Kybernetika 3 (1967), N. 5, 451—466.

Souhrn

DISKRÉTNÍ TEORIE VYHLEDÁVÁNÍ I

IGOR VAJDA

Tato práce navazuje na dřívější práce [1, 2], věnované otázkám statistického vyhledávání resp. plánování experimentů. Snaží se o široce přístupný a konkrétními příklady hojně ilustrovaný úvod do problematiky obou těchto prací, shrnuje a prohlubuje některé jejich výsledky.

Statistický problém uvažovaný v této práci se liší od klasického v následujícím. V klasickém statistickém modelu se předpokládá, že neznámý parametr θ pozorujeme jediným kanálem $P(. \mid \theta)$, jenž je definován souborem pravděpodobnostních rozložení $\{P(. \mid \theta)\}$, $\theta \in \Omega$, na výběrovém prostoru. Přitom se předpokládá, že $P(. \mid \theta_1) \neq P(. \mid \theta_2)$ pro $\theta_1 \neq \theta_2$. Toto dovoluje odhadnout θ s libovolně malou „chybou", pozorujeme-li ne jednou, ale N-krát, kde N je dostatečně velké. Problém, který jsme zde nazvali problémem vyhledávání, nastává tehdy, když můžeme θ pozorovat ne jedním, ale mnoha kanály, z nichž však žádný sám o sobě nezaručuje rozlišitelnost dvou různých hodnot θ. Abychom i zde dosáhli rozlišení, resp. malé „chyby", musíme parametr postupně pozorovat různými kanály. Pravidlo pro postupný výběr kanálů zde nazýváme strategie vyhledávání. V celé práci uvažujeme pouze model s diskrétním parametrovým i výběrovým prostorem a s konečnou množinou kanálů.

Kvalitu strategií posuzujeme jednak středním počtem pozorování, který je nutný k bezchybnému stanovení θ, anebo tam, kde je tento počet nekonečný, střední pravděpodobností chyby po N krocích. Náš přístup k problému vyhledávání je tedy bayesovský.

V této první části práce jsou dokázány některé věty o obecných a optimálních strategiích, v druhé části převážně věty o tzv. náhodných strategiích, kde se kanál vybírá náhodně a nezávisle na dosavadních výsledcích pozorování. V druhé části se rovněž poukazuje na hlubokou souvislost takto chápané teorie vyhledávání a teorie informace.

Author's address: Igor Vajda, CSc., Ústav teorie informace a automatisace ČSAV, Vyšehradská 49, Praha 2.