SOME DECISION MAKING METHODS APPLICABLE TO THE MEDICAL SCIENCES

J. T. CHU
UNIVERSITY OF PENNSYLVANIA

1. Introduction and summary

Decision making problems are often encountered in the medical sciences. In this paper, two types of problems are investigated, namely, medical diagnosis and planning of medical research. For medical diagnosis, we consider the use of statistical decision functions in minimizing the risks and error probabilities, and methods and advantages in some semiautomated diagnoses. The existence of certain types of optimal decision functions are proved, and methods for obtaining some of these decision functions are discussed. For planning of medical research, an optimal method is given for the selection of research projects under budget limitation. A certain estimation problem related to that optimal method is investigated. Both parametric and nonparametric methods of estimation are given.

2. Medical diagnosis

2.1. *A description of the problem.* By medical diagnosis we mean the act of recognizing the disease of a patient, and of classifying his state of health. In terms of the standard statistical decision theory, the problem may be analyzed and formulated as follows.

2.1.1. *The state space.* Let θ denote the disease that a given patient has (or the state of health of a given person). The state space Ω is defined as the set of all diseases having similar symptoms to those which the patient shows. We assume that Ω contains a finite number of elements $\theta_1, \cdots, \theta_m$, where the θ_i correspond to either no disease, or one or several diseases. The probability that $\theta = \theta_i$, that is, the patient has disease θ_i, is denoted by p_i, which may be known or unknown.

2.1.2. *The action space.* The objective of medical diagnosis is, of course, to identify the disease of the patient. However, in some cases, a clean cut decision

Research supported in whole by U. S. Public Health Service Grant GM 12869-01 from the National Institute of General Medical Sciences.

may be very difficult to reach. In other cases, an inconclusive type of decision, such as the patient has either disease θ_i or θ_j, may also serve some useful purposes. For example, as a guide for further diagnosis or initial treatment. Using the terminology of decision theory, any decision is called an action. The action space A is defined as the set of all the actions that are allowed to be taken. Depending on the preference of the decision maker, the action space may or may not contain inconclusive types of actions. An element of A is denoted by a_j.

2.1.3. *Experiments of diagnosis.* As is well known, conventional methods for diagnosis include questions, laboratory tests, X-rays, and so forth. The combination of some or all of these is called an experiment. The results of such an experiment are, in general, subject to random variations, and will be denoted by a random vector X. The choice of an experiment, or equivalently, that of X, is an important and often difficult problem. In this paper, we assume that X has already been chosen. We also assume, for mathematical generality, that a measure μ is defined on a σ-field of sets in the space S of all possible values x of X, and that with respect to μ, $f(x|\theta_i)$ is the probability density function of X given that $\theta = \theta_i$.

2.1.4. *Randomized decision functions.* A randomized decision function $\phi = \phi(j|x)$ is a function defined for every $x \in S$ and $a_j \in A$ such that $\phi(j|x) \geqq 0$ and $\sum_{a_j \in A} \phi(j|x) = 1$. If x is the observed value of X, then $\phi(j|x)$ is the probability that action a_j will be taken. If for all j and x, $\phi(j|x) = 0$ or 1 only, then ϕ is called a nonrandomized decision function. Both kinds of decision functions are of interest to us.

2.1.5. *Optimal decision functions.* An important problem for decision making in general is the selection of decision functions that in some senses are optimal. For medical diagnosis, we shall consider decision functions which minimize the average risk and error probabilities of various kinds, subject to some additional requirements. Details will be given in the next section.

2.1.6. *Automation and subjective judgment.* In recent years electronic computers in medical diagnosis have been used in different ways. One type of semiautomation of medical diagnosis is the following. Based on the outcome of the experiment X, the computer makes the necessary computation and tries to reach a decision concerning the diseases of the patient. If no conclusion can be made, then the medical specialist takes over the responsibility of diagnosis. However, the process by which a specialist reaches his decision may involve some subjective judgments which are not statistical in nature. In this paper, we investigate this type of semiautomated medical diagnosis from the viewpoints of both cost and error probability.

2.2. *Average risks.* In this section, we investigate the use of Bayes decision functions which minimize the average risk, and that of inconclusive type of actions. As was described before, the state space Ω contains $\theta_1, \cdots, \theta_m$; and an element in the action space A is denoted by a_j. If $\theta = \theta_i$ and action a_j is taken, we assume that a loss $w_{ij} \geqq 0$ is incurred. Let ϕ be a randomized decision function given in section 2.1.4. The risk in using ϕ when $\theta = \theta_i$ is defined as

$$r(\theta_i, \phi) = \sum_{a_j \in A} \int w_{ij}\phi(j|x)f(x|\theta_i)\, d\mu. \tag{2.1}$$

The average risk in using ϕ is then defined as

$$\rho(p, \phi) = \sum_{i=1}^{m} r(\theta_i, \phi)p_i, \tag{2.2}$$

where $p = (p_1, \cdots, p_m)$. One type of optimal decision functions, known as Bayes decision functions, is any ϕ^* that satisfies

$$\rho(p, \phi^*) \leqq \rho(p, \phi), \tag{2.3}$$

for all $\phi = \phi(j|x)$ defined on A and S. Bayes decision functions can be obtained without difficulty and need not be randomized ([1], p. 152 and p. 157). The corresponding average risk is given by

$$\rho^* = \int \min_{j \in A} \left\{ \sum_{i=1}^{m} w_{ij}g_i(x) \right\} d\mu, \tag{2.4}$$

where $g_i(x) = f(x|\theta_i)p_i$, $i = 1, \cdots, m$, and $\min_{j \in A}$ denotes the minimum taken over all $a_j \in A$. Therefore, when a Bayes decision function is used, the average risk is minimized, and the average risk taken is ρ^*.

Obviously, the quantity ρ^* depends, among other things, on A. Assuming that all other factors are fixed, we shall use the notation ρ_A^* to emphasize its dependence on A, and investigate how ρ_A^* may be reduced by expanding A. The main objective is to find conditions under which the inconclusive type of actions described in section 2.1.2 may be profitably used.

THEOREM 2.1. *Let B be an action space and $A \subseteq B$. Then $\rho_B^* \leqq \rho_A^*$ and inequality holds if and only if $P\{E\} > 0$, where*

$$E = \left\{ x \in S : \min_{j \in BA'} \sum_{i=1}^{m} w_{ij}g_i(x) < \min_{j \in A} \sum_{i=1}^{m} w_{ij}g_i(x) \right\}, \tag{2.5}$$

and

$$P\{E\} = \int_E \sum_{i=1}^{m} g_i(x)\, d\mu, \tag{2.6}$$

and A' is the complement of A.

PROOF. Since $A \subseteq B$, it follows from (2.4) that

$$\rho_A^* - \rho_B^* = \int_E \left[\min_{j \in A} \sum_{i=1}^{m} w_{ij}g_i(x) - \min_{j \in BA'} \sum_{i=1}^{m} w_{ij}g_i(x) \right] d\mu. \tag{2.7}$$

Hence $\rho_A^* > \rho_B^*$ if and only if $\mu(E) > 0$. From the definition of E and theorems on integration in measure theory, ([5], p. 104) we see that $\mu(E) > 0$ is equivalent to $P\{E\} > 0$.

Theorem 2.1 states an advantage of using Bayes decision functions. Namely, if the inclusion of an additional action can reduce the average risk, then Bayes decision functions will take that action with positive probability.

EXAMPLE 1. Suppose that $A = \{a_1, \cdots, a_m\}$, where a_i is the action of deciding that $\theta = \theta_i$. Let a_0 denote the action of making no decisions, and

$A_0 = \{a_0, a_1, \cdots, a_m\}$. Assume that if $\theta = \theta_i$ and the action a_0 is taken, the corresponding loss $w_{i0} \equiv w_0$ for all $i = 1, \cdots, m$; and that $w_{ij} = 0$, if $i = j$, and $w_{ij} = w_i$, if $i \neq j$, where $i, j = 1, \cdots, m$. In the type of semiautomated diagnosis mentioned in section 2.1.6, w_0 may be viewed as specialist's fee and w_i, the loss due to incorrect diagnosis by the computer, when $\theta = \theta_i$. From (2.5), the corresponding

$$E = \{x \in S: \sum_{i=1}^{m} (w_i - w_0)g_i(x) > w_j g_j(x), j = 1, \cdots, m\}. \tag{2.8}$$

In order that $P\{E\} > 0$, a necessary condition is that $E \neq 0$, that is, nonempty. By summing with respect to j, the two sides of the inequality in E, we see that $E \neq 0$ only if $w_0 < (1 - 1/m) \max_i w_i$. This result indicates that in a semiautomated diagnosis system, specialist's service may often be needed if losses due to incorrect decisions are great. On the other hand, if, compared with w_0, all the w_i are small, then E is likely to be empty and diagnosis by computer alone may be sufficient.

EXAMPLE 2. Following the previous example, we may also consider actions such as a_u, where $u = (k, j)$, and for fixed $k = 1, \cdots, m$, and $j = 1, \cdots, \binom{m}{k}$; thus a_u denotes the action of identifying k specific elements of Ω: $\theta_{u_1}, \cdots, \theta_{u_k}$, as possible values of θ. For a given k, we use A_k to denote the set of all a_u, where $u = (k, j)$. In this new and more general notation, a_i, for $i = 0, 1, \cdots, m$, of example 1 become $a_{m1}, a_{11}, \cdots$, and a_{1m}, respectively, and $A = \{a_1, \cdots, a_m\}$ now becomes A_1. Let $w(\theta_i, a_u)$ denote the loss due to action a_u when $\theta = \theta_i$. It seems reasonable to assume for every $\theta_i \in \Omega$, that $w(\theta_i, a_u) \leqq w(\theta_i, a_v)$, where both a_u and $a_v \in A_k$, but $\theta_i \in \{\theta_{u_1}, \cdots, \theta_{u_k}\}$, and $\theta_i \notin \{\theta_{v_1}, \cdots, \theta_{v_k}\}$. Now consider the case where for all $\theta_i \in \Omega$ and a_u, $w(\theta_i, a_u) \equiv w$ if $\theta_i \notin \{\theta_{u_1}, \cdots, \theta_{u_k}\}$ and $w(\theta_i, a_u) = w_k$, if $a_u \in A_k$ and $\theta_i \in \{\theta_{u_1}, \cdots, \theta_{u_k}\}$. For example, if $a_u = \{\theta_1, \theta_2\}$ and $a_v = \{\theta_2, \theta_3, \theta_4\}$, then $w(\theta_1, a_v) = w$, $w(\theta_2, a_u) = w_2$ and $w(\theta_2, a_v) = w_3$. In general, for every fixed $a_u \in A_k$ and $t > k$, there exists an $a_v \in A_t$ such that $\{\theta_{u_1}, \cdots, \theta_{u_k}\} \subsetneqq \{\theta_{v_1}, \cdots, \theta_{v_t}\}$. For simplicity, let us assume that the two sets are $\{\theta_1, \cdots, \theta_k\}$ and $\{\theta_1, \cdots, \theta_t\}$, respectively. If $w_k \geqq w_t$, then

$$\begin{aligned} \sum_{i=1}^{m} w(\theta_i, a_u)g_i(x) &= w_k \sum_{i=1}^{k} g_i(x) + w \sum_{i=k+1}^{m} g_i(x) \\ &\geqq w_t \sum_{i=1}^{k} g_i(x) + w_t \sum_{i=k+1}^{t} g_i(x) + w \sum_{i=t+1}^{m} g_i(x) \\ &= \sum_{i=1}^{m} w(\theta_i, a_v)g_i(x) \geqq \min_{a_v \in A_t} \sum_{i=1}^{m} w(\theta_i, a_v)g_i(x). \end{aligned} \tag{2.9}$$

By theorem 2.1, we see that if all elements a_v in A_t are already contained in the action space, then the minimum average risk ρ^* will not be further reduced by introducing into consideration any a_u in A_k where $k < t$, but $w_k \geqq w_t$.

The results obtained in the above examples may be interpreted as follows. To avoid an incorrect diagnosis, inconclusive type of decisions may be considered so that further investigation may be planned and conducted. On the other hand, more definitive types of diagnosis should be attempted, only if there is possibly something to gain. We also note that the results obtained for the inconclusive types of actions are preliminary in nature. An extensive investigation is needed in order to explore their full potentiality. Finally, for previously done work concerning inconclusive types of actions, we cite references [2], [6], [8].

2.3. *Error probabilities.* An often raised objection to the use of Bayes decision functions is that the losses w_{ij} due to incorrect actions are often difficult, if not impossible, to estimate. As an alternative, we shall consider certain optimal decision functions based on minimizing the error probabilities. An existence theorem will be proved and some applications discussed. We point out that even though the w_{ij} still appear in the theorem, they are no longer necessarily losses. In most cases, they are constants to be chosen for specific purposes. Also, to simplify the proof and notations, we shall assume that S is countable and use i and j to denote θ_i and a_j.

THEOREM 2.2. *Let $B_1, \cdots, B_s$ and B_{s+1} be subsets of $\Omega \times A = \{(i, j) | i \in \Omega, j \in A\}$ such that for certain $j_0 \in A$, $(i, j_0) \notin B_k$, for all $i \in \Omega$ and $k = 1, \cdots, s$, unless $w_{ij_0} = 0$. Let $c_k \geqq 0$, $k = 1, \cdots, s$, be given constants. Then the set T of all randomized functions φ satisfying the following is nonempty:*

$$B_k(\varphi) \leqq c_k, \qquad k = 1, \cdots, s, \tag{2.10}$$

where $w_{ij} \geqq 0$ and

$$B_k(\varphi) = \sum_{B_k} \int w_{ij}\varphi(j|x)g_i(x)\, d\mu, \qquad k = 1, \cdots, s+1. \tag{2.11}$$

There also exists a $\phi^* \in T$ which minimizes $B_{s+1}(\phi)$.

PROOF. Let φ_0 be a randomized decision function such that for all $x \in S$, $\varphi_0(j_0|x) \equiv 1$ and $\varphi_0(j|x) \equiv 0$ for all $j \neq j_0$. Then $B_k(\varphi_0) = 0$ for all $k = 1, \cdots, s$. Hence, $\varphi_0 \in T$ and T is nonempty. Next we show that there is a $\phi^* \in T$ satisfying

$$B_{s+1}(\varphi^*) = \inf_{\varphi \in T} B_{s+1}(\varphi). \tag{2.12}$$

Let φ^k be a sequence of elements in T such that $B_{s+1}(\varphi^k) \to \inf_{\varphi \in T} B_{s+1}(\varphi)$ as $k \to \infty$. For all fixed x and j, $\varphi^k(j|x)$ is a bounded sequence of k, therefore by the Bolzano-Weierstrass compactness theorem, it contains a convergent subsequence. Since S is assumed to be countable, it follows from the well known diagonal method that there exists a subsequence $\varphi^{(k)}$ such that for every x and j, $\varphi^{(k)}(j|x) \to \varphi^*(j|x)$ as $k \to \infty$. Obviously, φ^* is a randomized decision function. By the dominated convergence theorem ([7], p. 125), $\varphi^* \in T$ and satisfies (2.12).

REMARK. Under certain conditions, any φ^* in the above theorem satisfies the equalities of (2.10). One set of such conditions is: for all $i \in \Omega$, $(i, j_0) \in B_{s+1}$; if $(i, j_1) \in B_{s+1}$, then $w_{ij_1} < w_{ij_0}$; and for at least one i_0,

$$\int w_{i_0 j_{i_0}} \varphi^*(j_0|x) g_{i_0}(x)\, d\mu > 0. \tag{2.13}$$

The following is a proof. There exists a $\delta > 0$ such that $\mu(H) > 0$, where

$$H = \{x \in S: \varphi^*(j_0|x) \geqq \delta,\ g_{i_0}(x) > 0\}. \tag{2.14}$$

Now suppose that $B_1(\varphi^*) < c_1$ and $(i, j_1) \in B_1$ for some i and $(i, j_1) \in B_{s+1}$ for $i = 1, \cdots, k$. Let $\varphi'(j_0|x) = \varphi^*(j_0|x) - \delta$ and $\varphi'(j_1|x) = \phi^*(j_1|x) + \delta$, for all $x \in H$; and $\varphi'(j|x) \equiv \varphi^*(j|x)$ for all $j \neq j_0$ and j_1, or $x \notin H$. Obviously, φ' is a randomized decision function and $B_1(\varphi') \leqq c_1$, if δ is sufficiently small. Now

$$B_{s+1}(\varphi^*) - B_{s+1}(\varphi') = \delta \left[\sum_{i=1}^{m} \int_H w_{ij_0} g_i(x) \, d\mu - \sum_{i=1}^{k} \int_H w_{ij_1} g_i(x) \, d\mu \right] \tag{2.15}$$

If the second summation in the bracket is 0, then

$$B_{s+1}(\varphi^*) - B_{s+1}(\varphi') \geqq \delta \int_H w_{i_0 j_0} g_{i_0}(x) \, d\mu > 0. \tag{2.16}$$

If the second summation is positive, then

$$B_{s+1}(\varphi^*) - B_{s+1}(\varphi') \geqq \delta [\min_{i \leqq k} (w_{ij_0} - w_{ij_1})] \int_H \sum_{i=1}^{k} g_i(x) \, d\mu > 0. \tag{2.17}$$

It follows that φ^* does not satisfy (2.12). Contradiction! Hence $B_1(\varphi^*) = c_1$, and similarly $B_k(\varphi^*) = c_k$ for all $k = 1, \cdots, s$.

The following are some applications of theorem 2.2.

EXAMPLE 3. (Neyman-Pearson lemma). Suppose that $\Omega = \{\theta_1, \theta_2\}$, $A = \{a_1, a_2\}$, and a_j is the action of deciding that $\theta = \theta_j$. Let $w_{11} = w_{22} = 0$, $w_{12} = 1/p_1$, $w_{21} = 1/p_2$, and $c_1 = \alpha$, where $0 < \alpha < 1$ is given. Let $B_1 = \{(i, 2): i = 1, 2\}$, $s = 1$, and $B_{s+1} = \{(i, 1): i = 1, 2\}$. Then $j_0 = 1$, and (2.10) becomes

$$\begin{aligned} B_1(\varphi) &= \int \varphi(2|x) f(x|\theta_1) \, d\mu \leqq \alpha, \\ B_{s+1}(\phi) &= \int \varphi(1|x) f(x|\theta_2) \, d\mu. \end{aligned} \tag{2.18}$$

Hence there exists a φ^* such that $B_1(\varphi^*) \leqq \alpha$ and $B_{s+1}(\varphi^*)$ is minimum. This is a part of the Neyman-Pearson lemma.

EXAMPLE 4. *Controlling the average error probability.* Consider the action space A_0 and the loss structure of example 1, with $w_0 = w_i = 1$, for all $i = 1, \cdots, m$. Let $s = 1$, $B_1 = \{(i, j): i, j = 1, \cdots, m\}$, and $B_{s+1} = \{(i, 0): i = 1, \cdots, m\}$. Then

$$B_1(\varphi) = \sum_{i=1}^{m} \int [1 - \varphi(0|x) - \varphi(i|x)] g_i(x) \, d\mu, \tag{2.19}$$

and

$$B_{s+1}(\varphi) = \sum_{i=1}^{m} \int \varphi(0|x) g_i(x) \, d\mu. \tag{2.20}$$

Hence for a given α, $0 < \alpha < 1$, there exists a φ^* for which $B_{s+1}(\varphi^*)$ is minimum. However, such a φ^* may not satisfy $B_1(\varphi^*) = \alpha$. An example is the case where decisions concering θ can be made without any error. Then $B_1(\varphi^*) = B_{s+1}(\varphi^*) = 0$. In the kind of semiautomated medical diagnosis discussed previously, $B_1(\varphi)$ is the average probability of error in computer diagnosis and $B_{s+1}(\varphi)$ is the average probability of calling upon a specialist for further diagnosis. We also note that decision functions of the above kind were previously considered in connection with problems of computer character recognition [2].

EXAMPLE 5. *Controlling the error probabilities.* Consider again the action space A_0 of example 1. Let $s = m$, $B_k = \{(k, j): j = 1, \cdots, m\}$, $k = 1, \cdots, s$, and $B_{s+1} = \{(i, 0): i = 1, \cdots, m\}$. Let $w_{ii} = 0$, $w_{ij} = 1/p_i$, with $i \neq j$, for $i, j = 1, \cdots, m$, and $w_{i0} = 1$, for $i = 1, \cdots, m$, and let $c_k = \alpha_k$, where $0 < \alpha_k < 1$, for $k = 1, \cdots, s$. Then

$$(2.21) \qquad \begin{aligned} B_k(\varphi) &= \int [1 - \varphi(0|x) - \varphi(k|x)] f(x|\theta_k)\, d\mu, \qquad k = 1, \cdots, s, \\ B_{s+1}(\varphi) &= \sum_{i=1}^{m} \int \phi(0|x) g_i(x)\, d\mu. \end{aligned}$$

Hence there exists a ϕ^* such that $B_k(\phi^*) \leqq \alpha_k$, $k = 1, \cdots, s$, and $B_{s+1}(\phi^*)$ is minimum. In this way, all error probabilities are controlled under specified levels.

EXAMPLE 6. Consider the action space $\sum_{k=1}^{m} A_k$, where A_k is defined in example 2. For each $i = 1, \cdots, m$, we may choose B_i to be the set of all (θ_i, a_u), where a_u denotes the action of identifying k specific elements of Ω, $\theta_{u_1}, \cdots, \theta_{u_k}$, as possible values of θ, but θ_i is not one of them. We may also choose B_{m+1} to be the complement of $\sum_{k=1}^{m} B_k$ in the space of $\Omega \times \sum A_k$. Using theorem 2.2, we obtain decision functions for which the error probabilities are under control while the probability of further diagnosis is minimized.

To conclude this section, we note that we have only proved the existence of certain types of optimal decision functions. To actually construct such decision functions is another matter. For the case described in example 3, the Neyman-Pearson lemma provides the answer. For the case described in example 4, a method of constructing the decision functions is given in [2]. However, we do not have a method of construction in general.

3. Planning of medical research

Planning of research, medical or otherwise, is often a very difficult and complicated job. On the one hand, the planner is usually faced with limitations imposed by the availability of the various kinds of resources, such as manpower, funds, time, and so forth. On the other hand, there are different and possibly conflicting objectives that the planner may wish to reach. The objectives may, in general, be difficult to measure numerically; the number of research projects that the planner may select are often large, and furthermore, there may be some not known to the planner when selection must begin. Hence, the problem of optimal planning is certainly not simple. In this paper, we shall consider a very simplified version of the problem where there is only one requirement to meet and there is only one objective to reach, where both can be measured numerically. Whether this type of model is suitable for the planning of medical research is, of course, open to question. However, in medical research, the most important objective by far is probably the discovery of new drugs and technology and the severest limitation is probably the available fund. Hence, the one requirement and one objective assumption may be applicable.

For this simplified problem, an optimal selection procedure is given. For the case where not all research projects are known to the planner, a related method of estimation is investigated. We note that the results are modifications of some of those obtained previously for automatic abstracting in [3].

3.1. *An optimal selection procedure.* Mathematically, the problem of optimal selection described previously may be stated as follows. Suppose that a finite population is given, whose elements are denoted by (x_i, y_i), where $x_i, y_i > 0$, for $i = 1, \cdots, N$. Let $u_i = 0$ or 1. Find u_i, $i = 1, \cdots, N$, such that $\sum_{i=1}^N u_i y_i$ is maximized subject to the condition that $\sum_{i=1}^N u_i x_i \leqq \alpha \sum_{i=1}^N x_i$, where $0 < \alpha < 1$ is given. In other words, x_i and y_i are, respectively, the cost and return of the ith research project. Due to budget limitation, the planner wishes to select, in terms of cost, a given fraction of all the proposed research projects in such a way that the total return is maximized. When all x_i and y_i are known, the problem can be solved by either integer or dynamic programming. A more general problem for the case where the given population may be infinite is the following. Let $F(x, y)$ be a given cumulative distribution function defined on a given sample space S of elements (x, y), where $x, y > 0$. By a randomized selection function we mean a function $\varphi(x, y)$ defined on S such that $0 \leqq \varphi(x, y) \leqq 1$. Find $\varphi(x, y)$ which maximizes

$$\int_S \varphi(x, y) y \, dF(x, y), \tag{3.1}$$

subject to

$$\int_S \varphi(x, y) x \, dF(x, y) \leqq \alpha \int x \, dF(x, y), \qquad 0 < \alpha < 1. \tag{3.2}$$

The problem is very similar in form to a selection problem considered in [4], and may be solved in a similar way. For simplicity, we assume that both $\int x \, dF$ and $\int y \, dF$ are finite.

THEOREM 3.1. *There exists a randomized selection function $\varphi^*(x, y)$ of the following form which maximizes* (3.1) *subject to* (3.2). *Furthermore, $\varphi^*(x, y)$ satisfies the equality in* (3.2). *And*

$$\varphi^*(x, y) = \begin{cases} 1 & \text{if} \quad y > cx, \\ p & \text{if} \quad y = cx, \\ 0 & \text{if} \quad y < cx, \end{cases} \tag{3.3}$$

where $0 \leqq p \leqq 1$ *and* $c \geqq 0$.

PROOF. Let T_1, T_2, and T_3 denote, respectively, the sets $\{(x, y) \in S | y > cx\}$, $\{(x, y) \in S | y = cx\}$, and $\{(x, y) \in S: y < cx\}$. Let $\varphi(x, y)$ be any randomized selection function for which (3.2) holds. Assume, for the time being, that it is possible to determine c and p so that the corresponding $\varphi^*(x, y)$ of (3.3) satisfies the equality in (3.2). Then

$$\begin{aligned} \int [\varphi^* - \varphi] y \, dF &= \sum_{i=1}^{3} \int_{T_i} [\varphi^* - \varphi] y \, dF \\ &\geqq \sum_{i=1}^{3} \int_{T_i} [\varphi^* - \varphi] cx \, dF = c \left[\alpha \int x \, dF - \int \varphi x \, dF \right] \geqq 0. \end{aligned} \tag{3.4}$$

Hence, φ^* maximizes (3.1) subject to the condition (3.2).

Let $I(\varphi^*)$ and μ_x denote $\int \varphi^* x \, dF$ and $\int x \, dF$, respectively. We now show that c and p may be determined so that the corresponding φ^* satisfies the equality in (3.2), that is, $I(\varphi^*) = \alpha\mu_x$. It is easy to see that $I(\varphi^*)$ is a nonincreasing function of c and tends to 0 as $c \rightarrow \infty$. For a given c, let φ_c^1 and φ_c^2 denote functions of the type (3.3) for which $p = 1$ and 0, respectively. For a given α, with $0 < \alpha < 1$, let c_α be the greatest lower bound of all real $c \geqq 0$ such that $I(\varphi_c^1) \leqq \alpha\mu_x$. Then $I(\varphi_c^1) \leqq \alpha\mu_x$, if $c > c_\alpha$ and $I(\varphi_c^1) \geqq \alpha\mu_x$, if $c < c_\alpha$. Using known results concerning the indicators of monotonic sequences of sets and their limits ([7], p. 59) and theorems on monotone convergence ([7], p. 124), we see that $I(\varphi_{c_\alpha}^1) \geqq \alpha\mu_x$ and $I(\varphi_{c_\alpha}^2) \leqq \alpha\mu_x$. Hence $I(\varphi^*) = \alpha\mu_x$, if $c = c_\alpha$; and

$$p = \frac{\alpha\mu_x - I(\varphi_{c_\alpha}^2)}{I(\varphi_{c_\alpha}^1) - I(\varphi_{c_\alpha}^2)}, \tag{3.5}$$

if the denominator is not 0, and $p = 0$, if it is 0.

3.2. *A problem of estimation.* The optimal selection procedure obtained in the previous section can be determined only if we know the distribution function of the population π from which selections are to be made. If this is not the case, we will have to estimate the distribution function and use some approximate methods. In the following we shall see two such methods of approximation. One is for the case where the population π is finite, and the other for the case where the distribution function may be approximated by a bivariate normal distribution.

3.2.1. *Finite population.* Suppose that the probability function of the given population π is $p(x, y)$. Then the optimal selection function $\varphi^*(x, y)$ satisfies the relation $\sum \varphi^*(x, y)xp(x, y) = \alpha \sum xp(x, y)$. Let a random sample of size n be taken from π, and $p_n(x, y)$, the sample probability function. Using $p_n(x, y)$, we may determine $\varphi_n^*(x, y)$ such that $\sum \varphi_n^*(x, y)xp_n(x, y) = \alpha \sum xp_n(x, y)$. Now apply $\varphi_n^*(x, y)$ to the entire population π. Since it has the form (3.3), it is also an optimal selection procedure. The question which needs to be investigated is whether $I(\varphi_n^*) = \alpha\mu_x$. We now show that $I(\varphi_n^*) \rightarrow \alpha\mu_x$ in probability, as $n \rightarrow \infty$.

$$\begin{aligned} I(\varphi_n^*) - \alpha\mu_x = \sum \varphi_n^*(x, y)x[p(x, y) - p_n(x, y)] \\ + \alpha \sum x[p_n(x, y) - p(x, y)]. \end{aligned} \tag{3.6}$$

By the law of Large Numbers, for every fixed x and y, $p_n(x, y) \rightarrow p(x, y)$ in probability as $n \rightarrow \infty$. Since π is finite and φ_n^* is bounded, it is clear that both terms on the right side of (3.6) tend to 0 in probability, as $n \rightarrow \infty$.

3.2.2. *Bivariate normal distribution.* Consider the case where the given population has a bivariate normal distribution $f(x, y)$. For simplicity, we shall assume that the correlation of x and y is 0. (If not, see [9], p. XXVIII). Let the means and variances of x and y be $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, respectively. Let $u = (x - \mu_1)/\sigma_1$ and $v = (y - \mu_2)/\sigma_2$. Then the constant c in the corresponding φ^* of (3.3) satisfies the following equation

$$\sigma_1 \iint ug(u)g(v) \, du \, dv + \mu_1 \iint g(u)g(v) \, du \, dv = \alpha\mu_1, \tag{3.7}$$

where $g(u) = (1/2\pi)^{1/2}e^{-u^2/2}$ and the integrals are to be evaluated over the domain: $c\sigma_1 u - \sigma_2 v \leqq \mu_2 - c\mu_1$, and $-\infty < u, v < \infty$. The first integral may be evaluated by first integrating with respect to u. The second integral may be evaluated by the fact that $c\sigma_1 u - \sigma_2 v$ is normally distributed with mean 0 and variance equal to $c^2\sigma_1^2 + \sigma_2^2$. Hence (3.7) becomes

$$[c\sigma_1^2/(c^2\sigma_1^2 + \sigma_2^2)^{1/2}]g[(\mu_2 - c\mu_1)/(c^2\sigma_1^2 + \sigma_2^2)^{1/2}] + \mu_1 G[(\mu_2 - c\mu_1)/(c^2\sigma_1^2 + \sigma_2^2)^{1/2}] = (1 - \alpha)\mu_1, \tag{3.8}$$

where $G(x) = \int_x^\infty g(y)\,dy$. The parameters μ_1, μ_2, σ_1^2, and σ_2^2 may be estimated by drawing a random sample, and an approximate solution of c may then be obtained by numerical methods.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, New York, Wiley, 1954.
[2] J. T. Chu, "Optimal decision functions for computer character recognition," *J. Assoc. Comput. Mach.*, Vol. 12 (1965), pp. 213–226.
[3] ———, "Optimal procedures for automatic abstracting," *Technical Preconditions for Retrieval Center Operations* (edited by B. F. Cheydleur), Washington, Spartan Books 1965, pp. 103–116.
[4] W. G. Cochran, "Improvement by means of selection," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 449–470.
[5] P. R. Halmos, *Measure Theory*, New York, Van Nostrand, 1955.
[6] D. V. Lindley, "Statistical inference," *J. Roy. Statist. Soc. Ser. B*, Vol. 15 (1953), pp. 30–76.
[7] M. Loève, *Probability Theory*, New York, Van Nostrand, 1955.
[8] C. R. Rao, "The utilization of multiple measurements in problems of biological classification," *J. Roy. Statist. Soc. Ser. B*, Vol. 10 (1948), pp. 159–203.
[9] National Bureau of Standards, *Tables of the Bivariate Normal Distribution Function and Related Functions*, Applied Mathematics Series 50, Washington, U. S. Government Printing Office, 1959.