SOME MODEL I PROBLEMS OF SELECTION

By E. L. Lehmann[1]

University of California, Berkeley

1. Summary. There are given a populations $\Pi_1, \cdots, \Pi_a$, of which we wish to select a subset. The quality of the ith population is characterized by a real-valued parameter θ_i, and a population is said to be

(1) *positive* (or *good*) if $\theta_i \geqq \theta_0 + \Delta$,

(2) *negative* (or *bad*) if $\theta_i \leqq \theta_0$,

where Δ is a given positive constant and θ_0 is either a given number or a parameter that may be estimated. A number of optimum properties of selection procedures are defined (Section 3) and it is shown that for some of these, the optimum procedure selects Π_i when

(3) $$T_i \geqq C_i,$$

where T_i is a suitable statistic, the distribution of which depends only on θ_i, and where C is a suitable constant. (Sections 4 and 6.) Applications are given to distributions with monotone likelihood ratio in the case that θ_0 is known (Sections 5 and 6), and to normal distributions when instead observations on θ_0 are included in the experiment (Sections 10 and 11).

2. Introduction. An important class of classification problems is concerned with *selection*, that is with the classification of items into a superior category (the selected items) and an inferior one. We shall not be concerned here with more general classification procedures which would divide the items into possibly more than two categories. Selection problems have been treated in many different formulations. A basic distinction is that corresponding to Models I and II in the analysis of variance. In Model I, the items being classified are considered fixed; only the observations made on each item are random. In Model II, on the other hand, the items themselves are drawn at random from some population and would therefore change under a replication of the experiment. Model II problems have been treated recently, among others by Z. W. Birnbaum [1], Birnbaum and Chapman [2], T. W. Anderson [3], Cochran [4], Finney [5, 6], Davies [7], Curnow [8] and Dunnett [9]. For the related problem of the rejection of outliers, see for example [10]. We shall in the present paper be concerned only with Model I.

We shall assume therefore that a number of varieties, treatments, production

Received February 18, 1960, revised July 13, 1961.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43 and Nonr-2842(00)). This paper in whole or in part may be reproduced for any purpose of the United States Government.

methods, etc. (in general we shall speak of *populations*) are at our disposal. Their quality is characterized by a measure which we shall assume to be scalar. From the available set we wish to make a selection, selecting as far as possible the best ones.

It is useful to consider two cases according to the size of the group to be selected.

PROBLEM 1. A first possibility is that we wish to select only a single population (if possible the best one): the variety to be planted, the production method we are going to adopt, etc. As a slight generalization we may wish to select a fixed number, say two or three. We may have a fixed number of prizes or fellowships to award, or we may not wish to put all eggs into one basket.

PROBLEM 2. In the second case the group size is variable and is determined by the observations. This arises for example when we wish to select all worthwhile treatments or if we want to be reasonably sure that the selected group contains the best treatment.

PROBLEM 3. There is finally the intermediate possibility that the group size is variable but has a fixed upper limit. It may for example be desirable to investigate all treatments that appear promising but budget restrictions may limit the research program to the investigation of at most three treatments.

The traditional formal treatment of the class of problems described here, which has always been recognized as inadequate, is through tests of homogeneity (as for example in the analysis of variance). The only question answered by such a test is whether there is any difference at all among the available populations.

The first step toward a more realistic formulation is due to Mosteller [11] who gave a procedure for testing the hypothesis of homogeneity against the *slippage alternatives* that exactly one of the populations has slipped to the right and for deciding, in case of rejection of the hypothesis, just which of the populations slipped. Mosteller's paper was at least a partial answer to such an urgent need that, in spite of his warnings regarding certain inadequacies in the formulation, it inspired a large literature on slippage tests. At the same time, it led to further clarification of the issues. The first completely satisfactory proposal for dealing with a problem of type 2 above was made by Paulson [12], while problem 1 was formulated and essentially solved[2] by Bahadur [13].

Most of the literature on selection problems so far has been concerned with the definition of suitable procedures, an evaluation of their performance characteristics and the determination of the sample size. An optimum theory was developed for problem 1 by Bahadur in [13] and by Bahadur and Goodman in [14]. An optimum property of a slippage test, with reference only to slippage alternatives, was first proved by Paulson [15]. His proof was applied to other problems, was generalized and simplified in papers by Doornbos and Prins [16], Kudo [17], Pfanzagl [18], Ramachandran and Khatri [19], Truax [20], and Karlin and Truax [21]. Finally, contributions toward optimum properties of

[2] For the nonsequential case which is the only one considered here.

procedures for problem 2 were made by Gupta [22], Robbins [23] and Seal [24], [25], [26]. In the present paper, we shall be concerned with optimum procedures for certain cases of problem 2.

It is useful to introduce here another distinction according to the definition of the quality of a population.

A. In the simplest case, the quality is defined in absolute terms. If a number of new treatments are being compared with a standard treatment, it may happen that the latter has been observed so extensively that its effect can be taken as known. A treatment is then "good" if it is better (or sufficiently much better) than the standard. Another example is furnished by the selection of binomial populations, where success probabilities are compared with the "pure chance" value $p = \frac{1}{2}$, a probability p_i being considered as good if it exceeds this value or exceeds it by at least a given amount.

B. Usually, in the comparison of new treatments with a standard, it is of course better not to treat the standard as known but to let it participate in the experiment as a control. A new treatment is then "good" if it compares favorably with the control, the effect of which is also determined by the experiment.

C. Comparisons are not always relative to a standard or control. If a new product is being developed, it may be a question of selecting the most promising of a number of variants or a number of production methods. In such a case, each population must be compared with the totality of the remaining populations. A population may then be considered as "good" if it is (sufficiently much) better than the average of the remaining populations or if it does not fall too much below the best one. In the present paper, only problem A and B will be considered.

We mention in conclusion that the applications of selection theory are even wider than may appear at first: The emphasis instead of on selection may be on elimination. Thus we may wish to eliminate those regression coefficients or interactions, which can safely be neglected or those observations that represent gross errors. In the latter context slippage procedures, that is, procedures derived under the assumption of at most one "outlier" were proposed and their disadvantages discussed quite early by Pearson and Chandrasekhar [27].

3. Formulation of the problem. As in the Neyman-Pearson theory of hypothesis testing, there are two possible sources of error in any set of selections. There is the possibility of *false positives*, that is, populations which are selected although they are negative ($\leqq \theta_0$), and of *false negatives*, that is, populations which are not selected although they are positive ($\geqq \theta_0 + \Delta$). Instead of on false negatives we shall focus attention on *true positives*, that is, on those positive populations which are included in the selected group. This is analogous to the replacement of the consideration of an error of the second kind by that of power in the Neyman-Pearson theory.

Roughly speaking, it is the aim of a selection procedure to seek out the true positives while holding false positives to a minimum.

For measuring how well a procedure carries out its task of identifying the positive populations, a number of criteria are available.

(a) The expected number of true positives.

(b) The expected proportion of true positives, that is, the quantity (a) divided by the total number of positives.

These criteria are appropriate if it is desired to include in the selected group as many of the positive populations as possible.

(c) The probability of at least one true positive.

(d) The probability of including in the selected group the best population (that is, the population with the largest θ-value), provided it is positive.

These two criteria may be appropriate if the selection is only a step in a scheme, of which the eventual aim is the selection of a single population.[3]

(e) The probability of including all good populations.

This criterion implies that one would prefer the selection with probability γ of all good populations and with probability $1 - \gamma$ of none of them to the selection with probability $\gamma - \epsilon$ of all and with probability $1 - \gamma + \epsilon$ of all but one of the good populations. The criterion would thus seem to be appropriate only in rare cases.

As a measure of the performance of a procedure with respect to false positives we shall take either

(i) the expected number of false positives

or

(ii) the expected proportion of false positives, that is, the quantity (i) divided by the total number of negatives.

As a generic notation for any one of the quantities (a)–(e), all of which depend on the parameter point θ and on the particular selection procedure δ under investigation, we shall use $S(\theta, \delta)$. Here it is to be understood that S is defined only for the set Ω' of those parameter-points for which at least one of the populations is positive.

Similarly, we shall let $R(\theta, \delta)$ denote the quantity (i) or (ii). With these definitions of R and S, it is desirable to have $S(\theta, \delta)$ as large and $R(\theta, \delta)$ as small as possible. Specifically, we shall consider the problem of determining a procedure for which, subject to

$$\inf_{\theta \epsilon \Omega'} S(\theta, \delta) \geqq \gamma \tag{4}$$

we have

$$\sup_{\theta \epsilon \Omega} R(\theta, \delta) = \min, \tag{5}$$

(where Ω denotes the whole parameter space), or the dual problem in which $\inf S(\theta, \delta)$ is maximized subject to an upper bound on $\sup R(\theta, \delta)$.

Which of the various formulations is most appropriate, depends of course on the particular circumstances of each problem. In the absence of such more specific

[3] A complete sequential procedure for dealing with this problem was proposed by Stein [28]. For more recent work on such sequential procedures, see [29].

considerations, it seems perhaps most reasonable to control the minimum value of either (b) or (d). Subject to this condition one might, since each false positive provides a nuisance disturbance, wish to minimize the maximum value of (i).

It is of interest to note that condition (4) with S given by (b), that is,

$$\text{(6)} \qquad \inf_{\theta \varepsilon \Omega'} [\text{expected proportion of true positives}] \geqq \gamma,$$

implies

$$\text{(7)} \qquad \inf_{\theta \varepsilon \Omega'} P\{\text{at least one true positive}\} \geqq \gamma.$$

This follows from the fact that the left-hand side of (7) always exceeds that of (6). To see this, denote by A_i the event of including the ith population in the selected group, let I denote the set of indices i for which the ith population is positive, and let k be the number of elements of I. Then

$$P(\bigcup_{i \varepsilon I} A_i) \geqq \max_{i \varepsilon I} P(A_i) \geqq \sum_{i \varepsilon I} P(A_i)/k$$

as was to be proved.

The other proposed condition, (4) with S given by (d), that is,

$$\text{(8)} \qquad P\{\text{best population is in selected group}\} \geqq \gamma \qquad \text{for all } \theta \varepsilon \Omega',$$

can be given the following interpretation. Let Σ denote the set of selected indices. Then Σ constitutes a confidence set for the index i corresponding to the best population, provided attention is restricted to the parameter set Ω'.

We shall prove in the next sections, for certain families of distributions, that the solution with any of the formulations (a)–(d) combined with either (i) or (ii) is given by (3) of section 1 but that this is not true for formulation (e).

4. A minimax solution. If attention is restricted to nonrandomized procedures, as can always be done by enlarging the sample space, a selection procedure is a partition of the sample space into the sets $D_{i_1, \cdots, i_k}$ of those sample points for which the selected group consists of the populations with subscripts $i_1, \cdots, i_k$ and no others. To these must be added the set D_0 for which none of the populations is selected. If the number of available populations is a, the number of sets D is 2^a since each subscript may or may not occur with all combinations of the remaining subscripts.

Fortunately, selection procedures possess an equivalent, and for most purposes much simpler, representation. Let E_i be the set of sample points for which the ith population is included in the selected group. Then each of the two systems of sets $\{D\}$ and $\{E\}$ is uniquely determined by the other. In fact, E_i is the union of all those sets D which have i as one of their subscripts. Conversely,

$$D_{i_1, \cdots, i_k} = E_{i_1} \cap \cdots \cap E_{i_k} \cap \bar{E}_{j_1} \cap \cdots \cap \bar{E}_{j_{a-k}}$$

where $j_1, \cdots, j_{a-k}$ are the subscripts different from $i_1, \cdots, i_k$ and $\bar{E}$ denotes the complement of E. Instead of working with the sets E_i in the enlarged sample space, it is now more convenient to return to the possibility of randomization.

Each E_i is then represented by a function ψ_i defined over the sample space and taking on values between 0 and 1, where $\psi_i(x)$ denotes the probability with which the ith population is included in the selected group. A selection procedure is characterized by the vector $\psi = (\psi_1, \cdots, \psi_a)$.[4]

According to the formulation of the preceding section we are concerned with a minimax problem subject to side conditions. This type of problem was investigated by Blyth [31] but his conditions do not apply to the cases to be considered here. The following lemma is an immediate extension of the standard method of characterizing minimax solutions as Bayes solutions corresponding to a least favorable *a priori* distribution. As in its more usual form, it is essentially an application of the Lagrange method of undetermined multipliers.

LEMMA 1. *Let* $\mathcal{B}$ *be a* σ*-field of subsets of the parameter space* Ω *and let* λ *and* μ *be probability distributions over* $(\Omega, \mathcal{B})$. *Let* A, B *be two positive constants and let* δ_0 *maximize the integral*

$$B \int S(\theta, \delta)\, d\mu(\theta) - A \int R(\theta, \delta)\, d\lambda(\theta). \tag{9}$$

Then δ_0 *minimizes* $\sup R(\theta, \delta)$ *subject to*

$$\inf S(\theta, \delta) \geqq \gamma \tag{10}$$

provided

$$\int R(\theta, \delta_0)\, d\lambda(\theta) = \sup R(\theta, \delta_0) \tag{11}$$

and

$$\int S(\theta, \delta_0)\, d\mu(\theta) = \inf S(\theta, \delta_0) = \gamma. \tag{12}$$

If δ_0 *is the unique procedure maximizing* (9), *it is also the unique solution of the restricted minimax problem.*

PROOF. Let δ be any procedure satisfying (10). Then

$$B \int S(\theta, \delta)\, d\mu(\theta) - A \int R(\theta, \delta)\, d\lambda(\theta)$$
$$\leqq B \int S(\theta, \delta_0)\, d\mu(\theta) - A \int R(\theta, \delta_0)\, d\lambda(\theta) = B\gamma - A \sup R(\theta, \delta_0).$$

[4] This representation was first utilized in a slightly more special form by Robbins [23]. A generalization was given by the author in Theorem 1 of [30]. I am grateful to Professor L. LeCam for pointing out an error in the generalization of Theorem 1 to randomized procedures. The displayed equivalence formulae at the top of p. 6 of [30] are not correct. However, the equivalence theorem itself remains correct even when randomization is permitted. This can be seen as above, by representing a randomized procedure as a nonrandomized procedure in an enlarged sample space and applying Theorem 1.

Since δ satisfies (10), it follows that

$$\sup R(\theta, \delta_0) \leqq \int R(\theta, \delta)\, d\lambda(\theta) \leqq \sup R(\theta, \delta)$$

as was to be proved.

If condition (10) is to hold only when the sup is taken over a subset Ω' of Ω, λ must be a distribution over Ω' but no other changes are necessary.

As is usually the case with minimax problems, the more difficult part of the solution is not the maximization of (9) but the determination of an appropriate λ and μ. In this connection, the following standard devices are helpful.

1. Condition (11) implies that λ assigns probability one to the set ω of parameter points θ for which

$$R(\theta, \delta_0) = \sup_{\theta'} R(\theta', \delta_0). \tag{13}$$

Similarly, μ must assign probability one to the set for which

$$S(\theta, \delta_0) = \inf_{\theta'} S(\theta', \delta_0).$$

2. The pair (λ, μ) is least favorable in the sense that it minimizes the maximum (with respect to δ) value of (9).

3. If the problem exhibits any symmetries, it pays to look for distributions λ, μ possessing the corresponding symmetries.

We shall in the following consider procedures, which determine the selection or nonselection of the ith population on the basis of real-valued statistics T_i, and in particular we shall prove certain minimax properties for procedures of the type

$$\psi_i = 1, \lambda_i, 0 \qquad \text{as } T_i >, =, < C_i. \tag{14}$$

For this purpose it is convenient first to state the following lemma, the proof of which is immediate.

LEMMA 2. *Suppose that the distribution of T_i depends only on θ_i and is stochastically increasing in θ_i. Let $\delta = (\psi_1, \cdots, \psi_a)$ be any procedure satisfying* (14) *and let I be the set of subscript i for which*

$$E_{\theta_0+\Delta}\, \psi_i = \min_{j=1,\cdots,a} E_{\theta_0+\Delta}\, \psi_j .$$

Then, if S is given by one of the quantities (a), (c) *or* (d) *of Section* 3, $\inf_{\Omega'} S(\theta, \delta)$ *is attained at all points θ such that for some $i \in I$*

$$\theta_i = \theta_0 + \Delta \qquad \text{and} \qquad \theta_j < \theta_0 + \Delta \qquad \text{for all } j \neq i.$$

If S is given by (b), $\inf_{\Omega'} S(\theta, \delta)$ *is attained at all points θ such that for some subset $\{i_1, \cdots, i_k\}$ of I*

$$\theta_{i_1} = \cdots = \theta_{i_k} = \theta_0 + \Delta \qquad \text{and} \qquad \theta_j < \theta_0 + \Delta \quad \text{for the remaining } \theta\text{'s}.$$

If R is given by (i) *or* (ii), *then* $\sup_\Omega R(\theta, \delta)$ *is attained at the point*

$$\theta^{(0)} = (\theta_0, \cdots, \theta_0).$$

We note also that if in addition to the assumptions made above, the joint distribution of $(T_1, \cdots, T_a)$ is stochastically increasing in $(\theta_1, \cdots, \theta_a)$, and if S is given by (e) of Section 3, then $\inf_{\Omega'} S(\theta, \delta)$ is attained at the point $(\theta_0 + \Delta, \cdots, \theta_0 + \Delta)$.

The minimax solution can now be obtained under the following conditions.

THEOREM. *Let the probability density of X be denoted by p_0 when $\theta_1 = \cdots = \theta_a = \theta_0$, and by p_i when $\theta_i = \theta_0 + \Delta$ and the parameters θ_j for $j \neq i$ have a common value $\theta' \leqq \theta_0 + \Delta$ determined so that the conditions below are satisfied. Suppose that $p_i(x)/p_0(x)$ is a nondecreasing function of a real-valued statistic T_i, that the distribution of T_i depends only on θ_i, is stochastically increasing in θ_i, and is independent of i. Then the procedure δ_0 satisfying* (4) *and* (5) *with S equal to any one of the quantities* (a)–(d) *and R defined by* (i) *or* (ii) *of the preceding section, is given by*

$$\psi_i = 1, \lambda_0, 0 \qquad \text{as} \quad T_i >, =, < C, \tag{15}$$

where λ_0 and C are determined by

$$E_{\theta_0+\Delta}\psi_i = \gamma. \tag{16}$$

The solution of the dual problem in which (4) *is replaced by*

$$\sup R(\theta, \delta) \leqq \gamma' \tag{17}$$

is also given by (15), *with λ and C now determined by*

$$R(\theta^0, \delta_0) = \gamma' \tag{18}$$

where $\theta^0 = (\theta_0, \cdots, \theta_0)$.

PROOF.

1. Let μ be the distribution which assigns probability one to the point $(\theta_0, \cdots, \theta_0)$ and λ the distribution which assigns probability $1/a$ to the points $\theta^{(i)}$ given by

$$\theta_i = \theta_0 + \Delta, \qquad \theta_j = \theta' < \theta_0 + \Delta \qquad \text{for } j \neq i. \tag{19}$$

Then it follows from Lemma 2 that δ_0 satisfies conditions (11) and (12).

2. For the distributions λ and μ specified in 1, and with p_i denoting the probability density of X when $\theta = \theta^{(i)}$, we have

$$R(\theta^0, \delta) = \int (\psi_1 + \cdots + \psi_a)p_0; \qquad S(\theta^{(i)}, \delta) = \int \psi_i p_i \tag{20}$$

and hence (9) reduces to

$$\frac{B}{a}\sum_{i=1}^{a}\int \psi_i p_i - A\sum_{i=1}^{a}\int \psi_i p_0 = \int \sum \psi_i\left(\frac{B}{a}p_i - Ap_0\right). \tag{21}$$

Since $0 \leqq \psi_i \leqq 1$, (21) is maximized by putting $\psi_i = 0$ or 1 as

$$(B/a)p_i < \text{or} > Ap_0,$$

and hence as T_i is $<$ or $>$ C, as was to be proved.

We note that it is actually not necessary for $p_i(x)/p_0(x)$ to be an increasing function of T_i, but only that there exist a constant k such that the regions $T_i > C$ and $T_i < C$ (for the particular value C determined by the side conditions) are equivalent to the regions $p_i(x)/p_0(x) > k$ and $< k$ respectively.

This theorem provides the basis for determining the sample size necessary to control the risks R and S at any desired levels. For suppose that we wish the selection procedure to satisfy

$$R(\theta, \delta) \leqq \gamma' \qquad \text{for all } \theta \in \Omega$$

and

$$S(\theta, \delta) \geqq \gamma \qquad \text{for all } \theta \in \Omega'.$$

Then for the smallest sample size (possibly randomized) which constitutes a solution to this problem, the associated procedure δ_0 minimizes sup $R(\theta, \delta)$ subject to (4). If the conditions of the theorem are satisfied, δ_0 is therefore given by (15) and hence satisfies (if we assume for simplicity that it is nonrandomized)

$$\sup R(\theta, \delta_0) = R(\theta_0, \delta_0) = \begin{cases} aP_{\theta_0}(T_i \geqq C) & \text{if } R \text{ is given by (i)} \\ P_{\theta_0}(T_i \geqq C) & \text{if } R \text{ is given by (ii)}. \end{cases}$$

It further satisfies the condition

$$\inf S(\theta, \delta_0) = P_{\theta_0+\Delta}(T_i \geqq C).$$

If we let

$$\gamma^* = \begin{cases} \gamma'/a & \text{when } R \text{ is given by (i)} \\ \gamma' & \text{when } R \text{ is given by (ii)}, \end{cases}$$

the sample size is therefore determined by the conditions

$$P_{\theta_0}(T_i \geqq C) \leqq \gamma^*, \qquad P_{\theta_0+\Delta}(T_i \geqq C) \geqq \gamma.$$

These are exactly the conditions appropriate for testing the hypothesis $\theta_i = \theta_0$ against the alternative $\theta_i = \theta_0 + \Delta$ if we wish to have significance level γ^* and power at least γ. In all particular cases considered in the following section, the sample size determination therefore reduces to a problem whose solution is known from the corresponding problem of hypothesis testing.

5. Families with monotone likelihood ratio. The theorem of the preceding section applies directly to the case of independent samples $X_{i1}, \cdots, X_{in}$ from populations with probability density f_{θ_i} depending only on the real-valued parameter θ_i with respect to which we wish to select, if there exists a sufficient statistic T_i for $(X_{i1}, \cdots, X_{in})$ with monotone likelihood ratio. Let the probability density of T_i be g_{θ_i}, and take θ_0 for the value θ' of the theorem. Then

$$\frac{p_i(x)}{p_0(x)} = \frac{g_{\theta_0+\Delta}(t_i)}{g_{\theta_0}(t_i)} \prod_{j \neq i} \frac{g_{\theta_0}(t_j)}{g_{\theta_0}(t_j)} = \frac{g_{\theta_0+\Delta}(t_i)}{g_{\theta_0}(t_i)},$$

which is nondecreasing in t_i, and it follows that the minimax procedure is given by (15).

In particular, if $n = 1$ and the probability densities $f_{\theta_i}(x_i)$ have monotone likelihood ratio in x_i, the result holds with $T_i = X_i$ so that the procedure is given by

$$\psi_i = 1, \lambda_0, 0 \qquad \text{as } X_i >, =, < C. \tag{22}$$

Examples of this are the case in which the random variables X_i are independently distributed with binomial distributions $b(p_i, m)$, with Poisson distributions $P(\tau_i)$, or more generally with distributions having densities of the form

$$C(\theta_i)e^{\theta_i x_i}h(x_i),$$

that is, belonging to an exponential family.

As another example, consider samples $(X_{i1}, \cdots, X_{in})$ from normal distributions $N(\xi_i, \sigma_i^2)$ and suppose we wish to select the populations with small variances. Attention may be restricted to the sufficient statistic $\bar{X}_1, \cdots, \bar{X}_a$ and $S_1^2, \cdots, S_a^2$ where

$$\bar{X}_i = \sum_{j=1}^n X_{ij}/n; \qquad S_i^2 = \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2.$$

Since the problem remains invariant[5] under addition of arbitrary constants c_i to $\bar{X}_i$, it follows from a trivial extension of the Hunt-Stein theorem ([32], p. 336) that there exists a minimax solution depending only on the variables $S_1^2, \cdots, S_a^2$. To these variables the theorem is now applicable with $n = 1$, and shows that the minimax procedure consists in selecting the populations for which $S_i^2 \leqq C$.

Suppose that instead it is desired to select the populations for which the parameters $\theta_i = \xi_i/\sigma_i$ are sufficiently large. This time the problem remains invariant under multiplication of $\bar{X}_i$ by any positive constant c_i and of S_i^2 by c_i^2. There exists therefore a minimax solution depending only on the variables $T_i = \bar{X}_i/S_i$. Since the T_i have noncentral t-distributions which possess monotone likelihood ratio, it follows that the minimax procedure consists in selecting the populations for which $\bar{X}_i/S_i \geqq C$.

If in this last problem the variances σ_i^2 are assumed to be independent of i, with the common variance σ^2 still being unknown, the problem (of selecting for ξ_i/σ or ξ_i) surprisingly is much less simple since the T_i are then dependent. We shall consider this case in Section 7.

We conclude the present section by showing that in all the problems considered above, if criterion (e) of Section 1 is used instead of one of the criteria (a)–(d), the minimax solution is no longer given by (15). The argument is sufficiently clearly indicated by considering the case $a = 2$. Suppose therefore that X_1, X_2 are independently distributed with densities $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ which have monotone likelihood ratio. We shall also assume that the associated cumulative distribution functions are continuous; if the region of positive density is

[5] In the same strong sense as in the theory of hypothesis testing that is without performing any transformations of the decision space.

independent of θ, this can always be achieved by adjoining a uniformly distributed random variable.

Subject to

$$\text{sup [Expected number of false positives]} \leqq \gamma' \tag{23}$$

we wish to maximize the minimum probability of including in the selected group all good populations (if there are any). The procedure of the preceding section includes the ith population when $X_i \geqq C$ where $P_{\theta_0}\{X_i \geqq C\} = \gamma'/2$. Then the maximum expected number of false positives is exactly γ' and, if

$$\beta = P_{\theta_0+\Delta}(X_i \geqq C),$$

the minimum probability of including all good populations is β^2.

Consider now the following alternative procedure:

Include θ_1 in the selected group
if $X_1 \geqq C$ and if $X_2 < C - \epsilon$ or $X_2 \geqq C$.

Include θ_2 in the selected group
if $X_2 \geqq C$ and if $X_1 < C - \epsilon$ or $X_1 \geqq C$.

In addition include both θ_1 and θ_2 if $C - \epsilon \leqq X_1, X_2 < C$.

For ϵ sufficiently small, it is then easily checked that the probability of including θ_i in the selected group when $\theta_i \leqq \theta_0$ is still $\leqq \gamma'/2$ so that the procedure continues to satisfy (23). The probability of including in the selected group all good populations if only one of the θ's is $\geqq \theta_0 + \Delta$ is now less than its previous value β but by continuity is $> \beta^2$ for ϵ sufficiently small. On the other hand, the probability of including all good populations when both θ_1 and θ_2 are $\geqq \theta_0 + \Delta$ is now clearly $> \beta^2$ since the set of sample points for which both populations are included has been increased by the set $C - \epsilon \leqq X_1, X_2 < C$. Hence for ϵ sufficiently small, the minimum probability of including all good populations is now $> \beta^2$ as was to be proved.

6. Unequal sample sizes. The case of unequal sample sizes requires a slight generalization of the theorem of Section 4.

THEOREM. *Let T_i $(i = 1, \cdots, a)$ be a real valued statistic whose distribution depends only on θ_i and is stochastically increasing in θ_i. Let $\delta_0 = (\psi_1, \cdots, \psi_a)$ be the selection procedure defined by*

$$\psi_i = 1, \lambda_i, 0 \qquad \text{as } T_i >, =, < C_i \tag{24}$$

where λ_i, C_i are determined by (16). *Let $p_i(x)$ and $p_0(x)$ be defined as in Section* 4 *and suppose that there exist constants k_i such that*

$$p_i(x)/p_0(x) >, =, < k_i \qquad \text{as } T_i >, =, < C_i .$$

Then, if S is equal to one of the quantities (a)–(d) *and R is defined by* (i) *or* (ii), *δ_0 minimizes $\sup_{\theta\epsilon\Omega} R(\theta, \delta)$ subject to* (4). *If instead of by* (16), *the constants λ_i, C_i are determined by* (18) *and*

$$E_{\theta_0+\Delta}\psi_i \text{ is independent of } i, \tag{25}$$

δ_0 maximizes $\inf_{\theta\epsilon\Omega'} S(\theta, \delta)$ subject to (17).

Thus in both cases, the critical functions ψ_i, which can be interpreted as tests of the hypotheses $H_i : \theta_i \leqq \theta_0$ against the alternatives $K_i : \theta_i \geqq \theta_0 + \Delta$, are not determined to give a constant (independent of i) significance level but instead so that the minimum power against K_i is constant.

The reason for this is clear: the probability of the ith population giving rise to a false positive takes on its maximum value at the same point $\theta^{(0)}$ for all i; hence the contributions of the various populations to the expected number or proportion of false positives are combined and do not have to be controlled individually. On the other hand, the probability of the ith population resulting in a true positive takes on its minimum value at a different point $\theta^{(i)}$ for each i. For the minimum of these minima to be a maximum, they have to be equal.

The lack of symmetry shown by the solution is of course a consequence of the asymmetric formulation of Section 3, where the consideration was shifted from false negatives to true positives. However, this asymmetry is not artificial but only reflects a corresponding asymmetry of the problem.

The proof of the result for unequal sample sizes parallels that of the special case when the sample sizes are equal. Let δ_0 be the procedure determined by (24), (25) and (16) or (18). Let the distribution μ be defined as before but let λ assign to the point $\theta^{(i)}$, instead of $1/a$, a probability π_i to be determined later. The quantity (21) then becomes

$$\int \sum \psi_i(B\pi_i p_i - Ap_0) \tag{26}$$

which is maximized by putting $\psi_i = 0$ or 1 as

$$p_i/p_0 < \text{ or } > A/B\pi_i .$$

Hence if B/A and π_i are determined so that $A/B\pi_i$ is equal for each i to the constant k_i defining δ_0, it follows that δ_0 has the desired minimax property.

7. Normal populations with common unknown variance. Let X_{ij} ($j = 1, \cdots, n_i$; $i = 1, \cdots, a$) be normally distributed as $N(\xi_i, \sigma^2)$, and suppose we wish to select the populations with large values of ξ_i/σ. More specifically, we shall consider a population as negative if $\xi_i/\sigma \leqq 0$ and as positive if $\xi_i/\sigma \geqq \Delta$. A set of sufficient statistics is given by the means $\bar{X}_i$ together with

$$S^2 = \sum \sum (X_{ij} - \bar{X}_i)^2.$$

Since the problem remains invariant under multiplication of each $\bar{X}_i$ by the same positive constant c and of S^2 by c^2, there exists a minimax procedure depending only on the variables $Y_i = \bar{X}_i/S$. If $\theta_i = \xi_i/\sigma$, the joint density of the Y's, is up to a constant,

$$\begin{aligned} p(y_1, \cdots, y_a) &= \left(1 + \textstyle\sum n_i y_i^2\right)^{-a-f/2} \\ &\quad \cdot \int_0^\infty w^{2a+f-1} \exp\left[-w^2/2 + w\textstyle\sum n_i\theta_i y_i/\sqrt{1 + \sum n_i y_i^2}\right] dw, \end{aligned} \tag{27}$$

where

$$f = \sum_{i=1}^{a} (n_i - 1).$$

From a consideration of the individual hypotheses $H_i : \theta_i \leqq 0$, it appears natural to include the ith population in the selected group if

$$\bar{X}_i/S \geqq C_i . \tag{28}$$

For this procedure it follows as before that the expected number of proportion of false positives takes on its maximum value at $\theta^{(0)} = (0, \cdots, 0)$ while the quantities (a)–(d) take on their minimum value at (among other points) the points $\theta^{(i)} = (0, \cdots, 0, \Delta, 0, \cdots, 0)$.

If the joint density of the Y's at $\theta^{(0)}$ and $\theta^{(i)}$ is denoted by p_0 and p_i, it is seen from (27) that p_i/p_0 is an increasing function of

$$\frac{\Delta Y_i}{\sqrt{1 + \sum n_j Y_j^2}} = \frac{\Delta \bar{X}_i}{\sqrt{S^2 + \sum n_j \bar{X}_j^2}}.$$

The selection of the ith population when p_i/p_0 is sufficiently large thus leads to selecting the populations for which

$$\frac{\bar{X}_i}{\sqrt{S^2 + \sum n_j \bar{X}_j^2}} > C_i . \tag{29}$$

This corresponds to the solution proposed by Paulson [15] and Pfanzagl [18] for the associated slippage problem in which the standard is replaced by a control.[6] It is however not a solution to the present problem. For as $\theta_j \to -\infty$ for some $j \neq i$, the probability of the inequality (29) tends to zero, and so therefore does the minimum value of each of the quantities (a)–(d).

Sometimes it is not unreasonable to assume *a priori* that

$$\theta_i \geqq 0 \qquad \text{for all } i. \tag{30}$$

If for example we wish to select among a number of possible enrichments of a substandard diet, we may be willing to assume that the effect of each, if any, is beneficial. While under this assumption, for most significance levels and sample sizes, the performance of the procedure (29) is no longer as drastic as before, the procedure is nevertheless still not a minimax solution. This is easily seen from the fact that the probability of the inequality (29), subject to $\theta_i = \Delta$ and $0 \leqq \theta_j \leqq \Delta$ for all $j \neq i$ takes on its minimum value not at $\theta^{(i)}$ but instead at the point $\theta_1 = \cdots = \theta_a$. The proof is similar (but simpler) to the one given below.

The question arises whether the intuitive procedure (28) is, as one might expect, a minimax solution. We shall prove in the next section that this is not the case, under the *a priori* assumption (30). It seems likely that the situation is the

[6] A slippage procedure, corresponding to (28) was proposed by Paulson in [9]; the complete procedure (28), with the standard replaced by a control, is discussed by Dunnett [33].

same even without this assumption. However, as we shall now show, (28) is then at least approximately minimax in the sense that subject to (4), the maximum expected number of false positives can at least be improved only slightly over its value for (28). We shall show this with the constant C_i determined by (25).

Condition (4) implies that for all i,

$$E\psi_i \geqq \gamma \quad \text{when} \quad \theta_i \geqq \Delta \quad \text{and} \quad \theta_j < \Delta \qquad \text{for } j \neq i. \tag{31}$$

Since the procedure (28), which for the moment we shall denote by $\psi^* = (\psi_1^*, \cdots, \psi_a^*)$, attains the maximum expected number of false positives when $\theta_1 = \cdots = \theta_a = 0$, the minimax value of this quantity can be no higher than

$$E_{0,\cdots,0}(\psi_1^* + \cdots + \psi_a^*) = \alpha_1 + \cdots + \alpha_a$$

where

$$\alpha_i = P_0\{\bar{X}_i/S > C_i\},$$

C_i being determined by $E_\Delta\psi_i^* = \gamma$.

On the other hand, this minimax value cannot be much lower than $\sum \alpha_i$. For consider any procedure $\psi = (\psi_1, \cdots, \psi_a)$ satisfying (31). Its maximum expected number of false positives is greater than or equal to $E_{0,\cdots,0}(\psi_1 + \cdots + \psi_a)$. Consider now the problem of minimizing $E_{0,\cdots,0}(\psi_1 + \cdots + \psi_a)$ subject to (31). If we restrict (31) to the parameter values $\theta_i \geqq \Delta$ and $\theta_j = 0$ for $j \neq i$, the solution becomes

$$\psi_i = 1 \quad \text{when} \quad \bar{X}_i \Big/ \sqrt{S^2 + \sum_{j \neq i} n_j \bar{X}_j^2} > C_i'.$$

Thus $\sum \alpha_i'$ where $\alpha_i' = P_{0,\cdots,0}\{\bar{X}_i/\sqrt{S^2 + \sum_{j\neq i} n_j \bar{X}_j^2} > C_i'\}$ constitutes a lower bound for the sought for minimax value. However, for typical values of the sample sizes, α_i' will be only slightly lower than α_i, the only difference being the $a - 1$ added degrees of freedom in the denominator of the t-statistic, which now has $\sum n_i - 1$ degrees of freedom instead of the $\sum (n_i - 1)$ in the case of (28).

The same argument shows that (28) approximately minimizes the maximum expected proportion of false positives.

8. A counterexample. We shall now construct, for the case of equal sample sizes, a procedure satisfying (23) and with larger minimum value of (a)–(d) than (28). To this end we shall first prove the existence of points $(y_1, \cdots, y_a)$ and $(y_1', \cdots, y_a')$ with $y_1 < C < y_1'$ and such that

$$\frac{p_{\Delta,\theta_2,\cdots,\theta_a}(y_1, \cdots, y_a)}{p_{0,0,\cdots,0}(y_1, \cdots, y_a)} > \frac{p_{\Delta,\theta_2,\cdots,\theta_a}(y_1', \cdots, y_a')}{p_{0,0,\cdots,0}(y_1', \cdots, y_a')}$$

for all $0 \leqq \theta_2, \cdots, \theta_a \leqq \Delta$. It is seen from (27), that the probability ratio $p_{\Delta,\theta_2,\cdots,\theta_a}/p_{0,\cdots,0}$ is an increasing function of

$$(\Delta y_1 + \sum_{j=2}^{a} \theta_j y_j) \Big/ \sqrt{1 + n \sum_{j=1}^{a} y_j^2}. \tag{32}$$

Hence it is enough to construct the points y and y' in such a way that the expression (32) (where without loss of generality we can put $n = 1$) exceeds the corresponding expression when y is replaced by y'. Putting $y_2 = \cdots y_a$, $y_2' = \cdots = y_a'$, letting $\rho = (\theta_2 + \cdots + \theta_a)/\Delta$ and writing k for $a - 1$, we must find pairs (y_1, y_2), (y_1', y_2') with $y_1 < C < y_1'$ and such that

$$\frac{y_1 + \rho y_2}{\sqrt{1 + y_1^2 + (a-1)y_2^2}} > \frac{y_1' + \rho y_2'}{\sqrt{1 + y_1'^2 + (a-1)y_2'^2}} \qquad \text{for all } 0 \leqq \rho \leqq k.$$

Since all coordinates will be chosen to be nonnegative, the inequality can be squared and on collection of powers of ρ becomes

$$f(\rho) = a_0 \rho^2 + 2a_1\rho + a_2 > 0$$

with

$$a_0 = y_2^2(1 + y_1'^2) - y_2'^2(1 + y_1^2)$$

$$a_1 = y_1y_2[1 + y_1'^2 + (a-1)y_2'^2] - y_1'y_2'[1 + y_1^2 + (a-1)y_2^2]$$

$$a_2 = y_1^2[1 + (a-1)y_2'^2] - y_1'^2[1 + (a-1)y_2^2].$$

A sufficient set of conditions for $f(\rho)$ to be positive for all $0 \leqq \rho \leqq k$ is $a_0 < 0$, $f(0) > 0$, $f(k) > 0$, and hence

$$a_0 < 0, \qquad a_2 > 0, \qquad a_0k^2 + 2a_1k + a_2 > 0.$$

For any fixed y_1, y_1' and y_2, the first two of these conditions are satisfied if y_2' is sufficiently large. The coefficient of $y_2'^2$ in $a_0k^2 + 2a_1k + a_2$ is $-k^2(1 + y_1^2) + 2ky_1y_2(a-1) + (a-1)y_1^2$. If y_2 is chosen large enough so that this coefficient is positive, the third condition is also satisfied for y_2' sufficiently large, and this completes the proof. The two points constructed in this way will be denoted by $y^0 = (y_1^0, \cdots, y_a^0)$ and $y^{0'} = (y_1^{0'}, \cdots, y_a^{0'})$. We note for later use that the points can be chosen in such a way that $C < y_j, y_j'$ for all $j > 1$.

Let R and R' denote two spheres with centers at y^0 and $y^{0'}$, and radii determined so that

$$P(R) = P(R') \qquad \text{when } \theta_1 = \cdots = \theta_a = 0. \tag{33}$$

In addition, the spheres are to be sufficiently small so that

$$\frac{y_1 + \sum_{j=2}^{a} \rho_j y_j}{\sqrt{1 + \sum_{j=1}^{a} y_j^2}} > \frac{y_1' + \sum_{j=2}^{a} \rho_j y_j'}{\sqrt{1 + \sum_{j=1}^{a} y_j'^2}} \qquad \text{for all } 0 \leqq \rho_j \leqq 1, \tag{34}$$

that $y_1 < C < y_1'$ for all $y \varepsilon R$ and $y' \varepsilon R'$ and that further conditions are satisfied which will be specified later.

Consider now the following modification of (28):

the 1st population is selected if the sample point satisfies

$$(y \varepsilon R) \quad \text{or} \quad (y_1 > C \quad \text{and} \quad y \not\varepsilon R'), \tag{35}$$

and the rule for selecting the other populations is defined by symmetry.

By (35), the expected number of false positives, when $\theta_1 = \cdots = \theta_a = 0$, is the same for the modified procedure as for (28). Since under assumption (30) the ith population can give rise to a false positive only when $\theta_i = 0$, it follows that the modified procedure satisfies (23) if R and R' are sufficiently small.

Consider on the other hand one of the criteria (a)–(d), for example the probability of including the best population in the selected group when its θ-value is $\geqq \Delta$. Under (28) this attains its minimum value at the points whose coordinates for some i satisfy $\theta_i = \Delta$ and $0 \leqq \theta_j < \Delta$ for $j \neq i$. By (34), the probability of including the best population is larger at all these points under the modified procedure than under (28).

In order to prove that also the minimum probability of including the best population has been increased by the modification, it is sufficient to show that with the modified procedure this minimum probability is still attained at points satisfying $\theta_i = \Delta$, $0 \leqq \theta_j < \Delta$ for $j \neq i$. This follows easily if we can show that for the modified procedure the probability of including the ith population is an increasing function of θ_i for fixed values of the other θ's. This result finally is an immediate consequence of the following two facts.

1. The partial derivative

$$\frac{\partial}{\partial \theta_i} P_{\theta_i}\{\bar{X}_i/S > C\}$$

is positive for all θ_i and is bounded away from 0 in any finite interval $a \leqq \theta_i \leqq b$.

2. As the radius of the sphere R (and hence also of R') tends to zero, the derivatives

$$\frac{\partial}{\partial \theta_i} P_{\theta_1,\cdots,\theta_a}(R) \quad \text{and} \quad \frac{\partial}{\partial \theta_i} P_{\theta_1,\cdots,\theta_a}(R')$$

tend to zero.

PROOF.

1. Putting $\sigma = 1$ so that $\xi_i = \theta_i$, the derivative is equal to

$$\frac{\partial}{\partial \theta_i} \int_0^\infty P_{\theta_i}\{\bar{X}_i > Cs\}\, dP(s).$$

We can differentiate under the integral sign and the derivative of the integrand is known to be positive. (See for example [32], p. 114. Problem 18.) Since this derivative is a continuous function of θ_i, it is bounded away from zero in any finite interval.

2. Writing $P(R) = \int_R P_{\theta_1,\cdots,\theta_a}(y)\, dy$, the differentiation can be carried out under the integral sign. Since for $(\theta_1, \cdots, \theta_a)$ in any finite interval, the integrand is uniformly bounded, the result follows.

Exactly the same argument applies if criterion (d) is replaced by (a) or (c). However, with (b) the difficulty arises that the expected proportion of true positives takes on its minimum value at all points which for some $1 \leqq i_1 < \cdots < i_k \leqq a$, $1 \leqq k \leqq a$, satisfy

$$\theta_{i_1} = \cdots = \theta_{i_k} = \Delta; \quad 0 \leqq \theta_j < \Delta \quad \text{for all } j \neq i_1, \cdots, i_k.$$

To obtain an increase at all these points, we note that at the beginning of the section we proved the existence of points $y = (y_1, \cdots, y_a)$ and $y' = (y_1', \cdots, y_a')$ with

$$y_1 < C < y_1' \quad \text{and} \quad C < y_j, y_j' \qquad \text{for all } j \neq 1$$

and such that (34) holds.

Let R_1 and R_1' denote the spheres previously denoted by R and R', let R_i and R_i' be defined by symmetry. The modified procedure as before consists in including θ_i in the selected group if

$$(y \,\varepsilon\, R_i) \quad \text{or if} \quad (y_i > C \quad \text{and} \quad y \,\varepsilon\, R_i').$$

For this procedure it was shown previously that if $\theta_i = \Delta, 0 \leqq \theta_j < \Delta$ for $j \neq i$, the expected proportion of true positives has been increased by the modification. Suppose now that two of the θ's are equal to Δ, say $\theta_1 = \theta_2 = \Delta, 0 \leqq \theta_j < \Delta$ for $j > 2$. Then twice the expected proportion of true positives equals

$$P\{\text{selecting } \theta_1\} + P\{\text{selecting } \theta_2\}.$$

Since for the points $(\Delta, \Delta, \theta_3, \cdots, \theta_a)$ with $0 \leqq \theta_j < \Delta$ for $j > 2$ we have both $P(R_1) > P(R_1')$ and $P(R_2) > P(R_2')$, it is seen that the expected proportion is increased also in this case, and in the same way that it is increased at all points at which it takes on its minimum under (28). The remainder of the argument requires no change.

In conclusion we mention, without going into details, that even without the restriction (30) the procedure (28) is not the solution of the problem of minimizing the maximum expected number of false negatives subject to

$$\sup \text{[Expected number of false positives]} \leqq \gamma'.$$

This follows more simply but by the same method as before from the fact that the expected number of false negatives under (28) takes on its maximum at the single point $\theta_1 = \cdots = \theta_a = \Delta$.

9. Decision theoretic approach. Although the formulations of Section 3 appear to the author to be more useful for most applications, the problems can also be treated from a purely decision theoretic point of view, with general loss functions replacing the consideration of true and false positives. In the present section such a treatment of the problems of Sections 5 and 6 will be sketched very briefly.

Suppose that $X_{i1}, \cdots, X_{in_i}$ $(i = 1, \cdots, a)$ are independent samples, that the distribution of the ith sample depends only on the parameter θ_i and that we wish to select the populations with high θ-values. Let the loss resulting from the selection or nonselection of the ith population depend only on θ_i and be denoted by $L_i(\theta_i)$ and $L_i'(\theta_i)$ respectively. Finally, let the over-all loss be the sum of the individual losses.

Consider now the ith component problem, a two-decision problem for the parameter θ_i with losses L_i and L_i'. Suppose that the minimax solution ψ_i for this problem is a Bayes solution with respect to a least favorable *a priori* distribution λ_i of θ_i. Then ψ_i is also the Bayes solution for this same two-decision problem on the basis of all $\sum n_i$ observations with respect to the *a priori* distribution $\lambda_1(\theta_1) \times \lambda_2(\theta_2) \times \cdots \times \lambda_a(\theta_a)$ for the combined parameter $\theta = (\theta_1, \cdots, \theta_a)$. This is an immediate consequence of the fact that the *a posteriori* (marginal) distribution of θ_i given all the $\sum n_i$ x's depends only on $x_{i1}, \cdots x_{in_i}$. It then follows from result (ii) (on p. 15) of [30] that the selection procedure $(\psi_1^0, \cdots, \psi_a^0)$ is a minimax solution of the over-all problem.

Conditions under which the minimax solutions for the component problems are of the form (24) are given for example in [34] and [35]. The minimax property of procedure (24) in these cases is a slight generalization of a result of Robbins [23] and Hannan and Robbins [36], which was established there by quite different methods. It is suggested by these papers (see also Johns [37]) that for the problem under consideration there exist asymptotic subminimax procedures so that for large a, certain improvements over the above minimax procedure may be possible.

10. Comparison of normal means with a control. In the remaining two sections we shall be concerned with problems in which the quality of the standard is not assumed known but where instead a control group X_{0j} $(j = 1, \cdots, m)$ is observed in addition to the observations X_{ij} $(j = 1, \cdots, n_i; i = 1, \cdots, a)$ on the a treatments.

We consider first the case that the X_{ij} are independently distributed with normal distributions $N(\xi_i, \sigma_1^2)$ for $i = 1, \cdots, a$ and $N(\xi_0, \sigma_0^2)$ for $i = 0$, and assume to begin with that σ_0^2, σ_1^2 are known and that $n_i = n$ for $i = 1, \cdots, a$. The averages $\bar{X}_0, \bar{X}_1, \cdots, \bar{X}_a$ are then sufficient statistics, independently distributed with normal distributions $N(\xi_0, \tau_0^2)$ for $\bar{X}_0$ and $N(\xi_i, \tau_1^2)$ for $\bar{X}_i$ $(i = 1, \cdots, a)$ where

$$\tau_0^2 = \sigma_0^2/m \quad \text{and} \quad \tau_1^2 = \sigma_1^2/n.$$

As in Section 5, a slight generalization of the Hunt-Stein theorem permits a reduction of the data. We may restrict attention to the variables $Y_i = \bar{X}_i - \bar{X}_0$ since by this theorem there exists a minimax solution which is invariant under a common translation of all variables and since $Y_1, \cdots, Y_a$ constitute a maximal set of invariants with respect to these transformations.

Putting $\theta_i = \xi_i - \xi_0$, the conditional joint density of the Y's given $\bar{X}_0 = x_0$ is (up to a constant factor)

$$\exp\left\{-\frac{1}{2\tau_1^2}\sum [(y_j - \theta_j) + (x_0 - \xi_0)]^2\right\}.$$

The joint density of the Y's is therefore

$$C \int \exp\left\{-\frac{1}{2\tau_1^2}\sum [(y_j - \theta_j) + y_0]^2 - \frac{1}{2\tau_0^2} y_0^2\right\} dy_0,$$

which after some simplification becomes

$$p(y) = C \exp\left\{-\frac{1}{2\tau_1^2}\left[\sum (y_j - \theta_j)^2 - \frac{a^2\tau_0^2}{a\tau_0^2 + \tau_1^2}(\bar{y} - \bar{\theta})^2\right]\right\}. \tag{36}$$

We shall now apply the theorem of Section 4 with $\theta_0 = 0$ and $\theta' = \rho\Delta$. Then $p_i(y)/p_0(y)$ is the ratio of the two densities (36) for $\theta_i = \Delta$, $\theta_j = \rho\Delta (j \neq i)$ and $\theta_1 = \cdots = \theta_a = 0$. The quadratic terms in the exponent cancel and the linear terms are, up to a factor Δ/τ_1^2,

$$\rho \sum_{j \neq i} y_j + y_i - \frac{a^2\tau_0^2}{a\tau_0^2 + \tau_1^2} \frac{1 + (a-1)\rho}{a} \bar{y}$$

$$= (1 - \rho)y_i + a\bar{y}\left\{\rho - \frac{\tau_0^2}{a\tau_0^2 + \tau_1^2}[1 + (a-1)\rho]\right\}.$$

For

$$\rho = \frac{\tau_0^2}{\tau_0^2 + \tau_1^2} \tag{37}$$

the coefficient of $a\bar{y}$ vanishes, so that

$$p_i(y)/p_0(y) = C \exp\left\{\frac{\Delta(1-\rho)}{\tau_1^2} y_i\right\}$$

is an increasing function of y_i. For this value of ρ, the conditions of the theorem of Section 4 are satisfied and the minimax procedure (15) thus reduces to

$$\psi_i = 1 \quad \text{when} \quad y_i = \bar{X}_i - \bar{X}_0 > C, \tag{38}$$

where C is defined by

$$P_\Delta(\bar{X}_i - \bar{X}_0 > C) = 1 - \Phi\left(\frac{C}{\sqrt{\frac{\sigma_0^2}{m} + \frac{\sigma_1^2}{n}}} - \Delta\right) = \gamma. \tag{39}$$

This solution is easily extended to the case of unequal sample sizes and unequal variances. If the variance of $\bar{X}_j$ is τ_j^2 we find for the joint distribution of the Y's,

$$p(y) = C \exp\left\{-\frac{1}{2}\sum_{j=1}^{a} \frac{1}{\tau_j^2}(y_j - \theta_j)^2 + \frac{1}{2}\frac{[\sum (y_j - \theta_j)/\tau_j^2]^2}{\frac{1}{\tau_0^2} + \sum \frac{1}{\tau_j^2}}\right\}.$$

We now apply the theorem given in Section 6, this time with $\theta' = \theta_i' = \rho_i\Delta$, so that $p_i(y)$ is the density of the Y's when

$$\theta_i = \Delta, \qquad \theta_j = \rho_i\Delta \qquad \text{for } j \neq i.$$

Then the quadratic terms in the exponent again cancel in the ratio $p_i(y)/p_0(y)$, and the linear term is

$$\sum \theta_j y_j/\tau_j^2 - \left[\frac{1}{\tau_0^2} + \sum \frac{1}{\tau_k^2}\right]^{-1} \sum \frac{y_j}{\tau_j^2} \sum \frac{\theta_j}{\tau_j^2}$$

all sums extending from 1 to a. For $j \neq i$, the coefficient of y_j is (up to a factor $\Delta[(1/\tau_0^2) + \sum (1/\tau_k^2)]^{-1}$)

$$\frac{\rho_i}{\tau_j^2}\left[\frac{1}{\tau_0^2} + \sum \frac{1}{\tau_k^2}\right] - \frac{1}{\tau_j^2}\left[\frac{1}{\tau_i^2} + \rho_i \sum_{k\neq i} \frac{1}{\tau_k^2}\right].$$

This will be zero if $\rho_i = \tau_0^2/(\tau_0^2 + \tau_i^2)$ and the coefficient of y_j then becomes

$$\frac{1}{\tau_i^2}\left[\frac{1}{\tau_0^2} + \sum_{j\neq i} \frac{1}{\tau_j^2}(1 - \rho_i)\right].$$

Since this is positive, $p_i(y)/p_0(y)$ is then an increasing function of $Y_i = \bar{X}_i - \bar{X}_0$ and the minimax procedure is therefore given by (24) where C_i is given by (39) with n_i in place of n, or by (25).

If the variables X_{ij} ($j = 1, \cdots, n_i$; $i = 0, \cdots, a$) all have common but unknown variances, it follows as in Section 7 that the procedure given by

$$\psi_i = 1 \quad \text{when} \quad \frac{(\bar{X}_i - \bar{X}_0)/\sqrt{\dfrac{1}{n_i} + \dfrac{1}{n_0}}}{\sqrt{\sum_{j=0}^{a} \sum_{k=1}^{n_j} (X_{jk} - \bar{X}_j)^2}} > C \tag{40}$$

is approximately minimax, where n_0 replaces the earlier m.

11. Comparison of normal variances with a control. Let X_{ij} ($j = 1, \cdots, n_i$; $i = 0, \cdots, a$) be independently distributed with normal distributions $N(\xi_i, \sigma_i^2)$ (ξ_i, σ_i unknown), and consider the problem of selecting the populations for which $\sigma_i^2/\sigma_0^2 \leqq \delta$. Application of the generalized Hunt-Stein theorem proves the existence of a minimax procedure depending only on the statistics

$$S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2.$$

We may therefore restrict attention to $S_0^2, \cdots, S_a^2$ where the distribution of S_i^2/σ_i^2 is $\chi_{f_i}^2$ with $f_i = n_i - 1$, and by another application of the same theorem to the variables $V_i = S_i^2/S_0^2$ ($i = 1, \cdots, a$). The joint densities of the V's, is up to a constant factor

$$\left[\left(1 + \sum_{j=1}^{a} \frac{v_j}{\sigma_j^2/\sigma_0^2}\right)^{(f_0+f_1+\cdots+f_a-1)/2}\right]^{-1} \prod_{i=1}^{a} \frac{v_i^{(f_i-2)/2}}{(2\sigma_i^2/\sigma_0^2)^{f_i/2}\Gamma(f_i/2)}.$$

Let us now apply the theorem of Section 6 with $\theta_i = \sigma_0^2/\sigma_i^2$, with $\theta_0 = 1$ (so that the conditions $\theta_1 = \cdots = \theta_a = \theta_0$ are equivalent to $\sigma_1^2 = \cdots = \sigma_a^2 = \sigma_0^2$); with $1 + \Delta = 1/\delta$ (so that the conditions $\theta_i \geqq \theta_0 + \Delta = 1 + \Delta$ is equivalent to $\sigma_i^2/\sigma_0^2 \leqq \delta$); and $\theta' = \rho_i(1 + \Delta)$, where $\rho_i < 1$ is to be determined later. With these values, the probability ratio p_i/p_0 is an increasing function of $(1 + \sum v_j)/(1 + \sum \theta_j v_j)$.

Let us therefore consider the region

$$1 + \sum v_j \geqq k_i(1 + \sum \theta_j v_j)$$

which is equivalent to

$$k_i(1 - \rho_i)(1 + \Delta)v_i \leqq (1 - k_i) + [1 - \rho_i k_i(1 + \Delta)] \sum_{j=1}^{a} v_j .$$

If we put $\rho_i = 1/k_i(1 + \Delta)$, this reduces to

$$v_i \leqq \frac{1 - k_i}{k_i(1 - \rho_i)(1 + \Delta)}$$

or equivalently to $v_i \leqq C_i$ with

$$C_i = \frac{1 - k_i}{k_i(1 + \Delta) - 1}.$$

As k_i goes from $1/(1 + \Delta)$ to 1, C_i goes from ∞ to 0, and for these values of k_i, it is seen that $\rho_i < 1$. The theorem of Section 6 therefore shows that the procedure $\psi_i = 1$ if $v_i \leqq C_i$ has the desired minimax property.

As a last problem, consider a set of Poisson populations with Poisson parameter $\lambda_0, \lambda_1, \cdots, \lambda_a$. The problem of selecting the populations for which $\lambda_i/\lambda_0 \geqq 1 + \Delta$ is not meaningful in the formulation given here since the parameter pairs $(\lambda_0, \lambda_i = \lambda_0)$ and $(\lambda_0, \lambda_i = (1 + \Delta)\lambda_0)$ become indistinguishable as $\lambda_0 \to \infty$. The problem could be treated with the minimax principle replaced by a suitable unbiasedness principle. Alternatively, if one is concerned with Poisson processes, one may instead of observing the number of occurrences in fixed intervals, take as observations the times required to get a specified number of occurrences. These times then follow gamma distributions, and the solution of the present section is directly applicable.

REFERENCES

[1] Z. W. Birnbaum, "Effect of linear truncation on a multinormal population," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 272–279.

[2] Z. W. Birnbaum and D. G. Chapman, "On optimum selections from multinormal populations," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 443–447.

[3] T. W. Anderson, "Classification by multivariate analysis," *Psychometrika*, Vol. 16 (1951), pp. 31–50.

[4] W. G. Cochran, "Improvement by means of selection," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951, pp. 449–470.

[5] D. J. Finney, "The consequences of selection for a variate subject to errors of measurement," *Rev. Inst. Internat. Stat.*, Vol. 24 (1956), pp. 1–10.

[6] D. J. Finney, "Statistical problems in plant selection," *Bull. Internat. Stat. Inst.*, Vol. 36 (1958), pp. 242–268.

[7] O. L. Davies, "The design of screening tests in the pharmaceutical industry," *Bull. Internat. Stat. Inst.*, Vol. 36 (1958), pp. 226–241.

[8] R. N. Curnow, "The consequences of errors of measurement for selection from certain non-normal distributions," *Bull. Internat. Stat. Inst.*, Vol. 37 (1960), pp. 291–308.

[9] C. W. Dunnett, "On selecting the largest of k normal population means," *J. Roy. Stat. Soc.*, Ser. B, Vol. 22 (1960), pp. 1–40.

[10] F. J. Anscombe, "Rejection of outliers," *Technometrics*, Vol. 2 (1960), pp. 123–147.

[11] Frederick Mosteller, "A k-sample slippage test for an extreme population," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 58–65.

[12] Edward Paulson, "A multiple decision procedure for certain problems in analysis of variance," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 95–98.

[13] Raghu Raj Bahadur, "On a problem in the theory of k populations," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 362–375.

[14] Raghu Raj Bahadur and Leo A. Goodman, "Impartial decision rules and sufficient statistics," *Ann. Math. Stat*, Vol. 23 (1952), pp. 553–562.

[15] Edward Paulson, "An optimum solution to the k-sample slippage problem for the normal distribution," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 610–616.

[16] R. Doornbos and H. J. Prins, "On slippage tests," *Indagationes Mathematicae*, Vol. 20 (1958), pp. 38–55 and pp. 438–447.

[17] Akio Kudô, "On the invariant multiple decision procedures," *Bull. Math. Stat.*, *Res. Assn. Stat. Sci.*, Vol. 6 (1956), pp. 57–68.

[18] J. Pfanzagl, "Ein kombiniertes Test und Klassifikations-Problem," *Metrika*, Vol. 2 (1959), pp. 11–45.

[19] K. V. Ramachandran and C. G. Khatri, "On a decision procedure based on the Tukey statistic," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 802–806.

[20] Donald R. Traux, "An optimum slippage test for the variances of k normal distributions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 669–674.

[21] Samuel Karlin and Donald Truax, "Slippage problems," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 296–324.

[22] S. S. Gupta, "On a decision rule for a problem in ranking measures," unpublished Ph.D. Thesis, University of North Carolina.

[23] Herbert Robbins, "Asymptotically subminimax solutions of compound statistical decision problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951. pp. 131–148.

[24] K. C. Seal, "On a class of decision procedures for ranking means of normal populations," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 387–398.

[25] K. C. Seal, "An optimum decision rule for ranking means of normal populations," *Calcutta Stat. Assn. Bull.*, Vol. 7 (1957), pp. 131–150.

[26] K. C. Seal, "On ranking parameters of scale in type III populations," *J. Amer. Stat. Assn.*, Vol. 53 (1958), pp. 164–175.

[27] E. S. Pearson and C. Chandrasekhar, "The efficiency of statistical tools and a criterion for the rejection of outlying observations," *Biometrika*, Vol. 28 (1936), pp. 308–320.

[28] Charles M. Stein, "On the selection of the largest of a set of means," (Abstract), *Ann. Math. Stat.*, Vol. 19 (1948), p. 429.

[29] Robert E. Bechofer, "A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs," *Biometrics*, Vol. 14 (1958), pp. 408–429.

[30] E. L. Lehmann, "A theory of some multiple decision problems, I," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1–25.

[31] Colin R. Blyth, "On minimax statistical decision procedures and their admissibility," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 22–42.

[32] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1959.

[33] Charles W. Dunnett, "A multiple comparison procedure for comparing several treatments with a control," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1096–1121.

[34] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[35] SAMUEL KARLIN AND HERMAN RUBIN, "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.

[36] JAMES F. HANNAN AND HERBERT ROBBINS, "Asymptotic solutions of the compound decision problem for two completely specified distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 37–51.

[37] M. V. JOHNS, JR., "Non-parametric empirical Bayes Procedures," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 649–669.