

A THEORY OF SOME MULTIPLE DECISION PROBLEMS, I¹

BY E. L. LEHMANN

University of California, Berkeley

Summary. A class of multiple decision procedures is described and its members are shown to possess uniformly minimum risk among all procedures that are unbiased with respect to a certain loss function. This provides a justification for a number of procedures considered by Tukey, Duncan, and others, for certain classes of point estimates, and for some nonparametric decision procedures based on sample cumulative distribution functions and related to tests of the Kolmogoroff-Smirnoff type.

1. Introduction. As has frequently been pointed out, many statistical situations, which it is customary to treat by means of hypothesis testing, really involve a choice between more than two decisions. In such problems, when the hypothesis is rejected, one wants to know in which of a number of possible ways the actual situation differs from the one postulated by the hypothesis. By formulating the problem as one involving only two decisions one not only neglects to differentiate between certain alternative decisions, which may differ considerably in their consequences, but one may also be led to an inappropriate acceptance region for the hypothesis.

As an example suppose that X and Y are independently normally distributed with unit variance and means ξ and η . While there are situations in which one only wishes to determine whether the hypothesis $H: \xi = \eta = 0$ is true or not, it is perhaps more common that in case of rejection one will want to know whether it is ξ or η that is different from 0, or both, and of the nonzero means whether they are positive or negative. Here the first formulation implies complete spherical symmetry between the alternatives, and the appropriate acceptance region is

$$x^2 + y^2 \leq c.$$

On the other hand, when the choice lies between the nine indicated decisions, it seems most natural to accept H when

$$\max(|x|, |y|) \leq k,$$

and in case of rejection to divide the rejection region into the eight subregions shown in Fig. 1 corresponding to the eight possible alternative decisions. This procedure actually will be justified later in terms of a specific loss function.

Received February 8, 1956.

¹ Work done while the author was a Fellow of the John Simon Guggenheim Memorial Foundation.

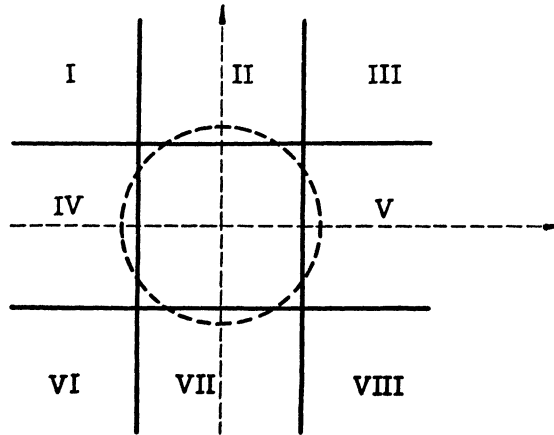


FIG. 1

One of the attractions of formulating statistical problems in terms of hypothesis testing is the resulting structural simplicity. However, at the same time this reduction to a choice between only two decisions frequently causes complications by creating a class of alternatives which combines too many different elements. In many such cases, if one is willing to forego structural simplicity and to divide the class of alternatives into its natural components, one obtains a multiple decision problem, which admits a simpler and more natural solution than the apparently less complex testing problem.

As an example consider the comparison of k variances $\sigma_1^2, \dots, \sigma_k^2$ on the basis of samples $X_{ij} (j = 1, \dots, n; i = 1, \dots, k)$ from normal populations $N(\xi_i, \sigma_i^2)$. There does not seem to exist any really convincing solution to the small sample problem of testing the hypothesis $H: \sigma_1 = \dots = \sigma_k$. On the other hand, there exists a natural multiple decision problem based on the comparisons of the different pairs (σ_i, σ_j) . The associated acceptance region for H is the one discussed by Hartley [6].

In the present paper, a general class of multiple decision problems is described together with procedures that seem appropriate for these problems. The method of constructing the procedures is not new, and is in fact the one used in most cases of multiple comparisons treated in the literature. It was mentioned explicitly in 1950 by Howard Levene in a seminar lecture at Columbia University, and was recently stated, with only a minor difference, by Duncan in [3]. It is also closely related to a principle of test construction proposed by Roy [12], and utilized further by Bose and Roy [2]. As will be shown, the method is applicable not only to the typical multiple comparison problems, but also to problems of point estimation and various nonparametric problems.

The main purpose of the present paper is to prove an optimum property of the above procedures, namely that they are unbiased (in a sense introduced by the author in [9]), and that among all unbiased procedures they uniformly mini-

mize the risk. It should be mentioned that in the application to specific cases a number of distributional problems arise, concerning error control and other properties of the operating characteristic of the procedures. These problems, which require separate treatment for each example, are not considered here but instead the paper is concerned only with general aspects of the method. In particular, the main result involves assumptions concerning the distribution of the observable random variables only indirectly in that the existence of tests of related hypotheses H is assumed, which possess certain optimum properties. Apart from this, the conditions concern only the structure of the multiple decision problems in terms of the hypotheses H .

2. A class of multiple decision procedures. Let X be a random observable, the distribution of which depends on the parameter θ , and suppose that a number of different hypotheses concerning θ are of interest, say $H_\gamma : \theta \in \omega_\gamma, \gamma \in \Gamma$. The class of alternatives K_γ to H_γ is that θ lies in the complement of ω_γ , which will be denoted by ω_γ^{-1} . When considering these hypotheses simultaneously one will wish to determine whether all of them are true, all of them false, or in the intermediate cases, exactly which of the hypotheses hold and which do not.

One is therefore faced with a multiple decision problem in which the different possible decisions correspond to the statements that a certain set of the hypotheses H_γ is correct while the remaining ones are false, or equivalently, that θ lies in a certain set, say $\Omega_i, i \in I$, determined by these conditions. The sets Ω_i , which are the atoms of the field of sets generated by the sets ω_γ , are formally given by

$$(2.1) \quad \Omega_i = \bigcap_{\gamma \in \Gamma} \omega_\gamma^{x_{i\gamma}},$$

where the x 's indicate which of the hypotheses are true ($x_{i\gamma} = 1$) and which are false ($x_{i\gamma} = -1$) for the given Ω_i . If, as is frequently the case, some of the intersections formally defined by (2.1) are empty, we shall restrict the Ω 's to denote the nonempty ones, and shall require that none of the possible decisions should correspond to the empty intersections. Authors dealing with specific multiple comparison problems have frequently not insisted on this restriction, and this is also the point in which Duncan's definition, referred to above, differs from the one given here.

To specify a loss function, suppose that the losses for the individual testing problems are a_γ and b_γ for falsely rejecting and accepting the hypothesis H_γ , and that in the simultaneous consideration of these problems the losses are additive. If then $\theta \in \Omega_i$ and the decision d_k is taken that $\theta \in \Omega_k$, the resulting loss is

$$(2.2) \quad w_{ik} = \sum_{\gamma \in \Gamma} (\epsilon_{ik\gamma} a_\gamma + \epsilon_{ki\gamma} b_\gamma),$$

where

$$(2.3) \quad \epsilon_{ik\gamma} = \begin{cases} 1 & \text{if } x_{i\gamma} = 1, & x_{k\gamma} = -1, \\ 0 & \text{otherwise.} \end{cases}$$

Formula (2.2) expresses the fact that w_{ik} is the sum of all those a_γ for which H_γ is true ($x_{i\gamma} = 1$) but rejected ($x_{k\gamma} = -1$), plus the sum of the b_γ for which H_γ is falsely ($x_{i\gamma} = -1$) accepted ($x_{k\gamma} = 1$). The risk is thus simply a weighted sum of the probabilities of error. Slightly more generally (taking the a 's and b 's to be given) one may wish to put, in the case that Γ is finite,

$$w_{ik} = \sum_{\gamma \in \Gamma} \nu_\gamma (\epsilon_{ik\gamma} a_\gamma + \epsilon_{ki\gamma} b_\gamma),$$

where the ν 's are any positive weights, and in the general case

$$(2.4) \quad w_{ik} = \int_{\Gamma} (\epsilon_{ik\gamma} a_\gamma + \epsilon_{ki\gamma} b_\gamma) d\mu(\gamma).$$

Suppose now that attention is restricted to nonrandomized procedures. This involves no essential loss of generality since it can be achieved for all decision problems with which we shall be concerned, by adjoining to the original random variables a continuous variable which is independent of them. Then a decision procedure for the given multiple decision problem is a partition of the sample space into sets D_i such that, when the observation falls into D_i , the decision $d_i : \theta \in \Omega_i$ is taken. It is natural to try to relate these decision procedures to the tests of the hypotheses H_γ . Let \mathcal{Q} be a family of such tests with acceptance regions A_γ for H_γ and rejection regions A_γ^{-1} , and consider the decision procedure that \mathcal{Q} induces through the relation

$$(2.5) \quad D_i = \bigcap_{\gamma \in \Gamma} A_\gamma^{x_{i\gamma}}.$$

Here it may happen that

$$P(\bigcup_{i \in I} D_i) < 1.$$

This is the case when the union of those intersections $\bigcap_{\gamma} A_\gamma^{x_{i\gamma}}$, for which the corresponding intersection (2.1) is empty, has positive probability. The partition (2.5) is then not a procedure for the given decision problem. When the induced procedure satisfies

$$(2.6) \quad P_\theta(\bigcup_{i \in I} D_i) = 1 \quad \text{for all } \theta,$$

the family \mathcal{Q} is said to be *compatible*. Since the acceptances and rejections making up the intersection (2.5) are consistent with each other if the corresponding intersection (2.1) is nonempty, and are otherwise mutually contradictory, compatibility is equivalent to the condition that the simultaneous application of the tests A_γ not lead to any inconsistencies.

In the case that Γ is noncountable, a further complication is the possibility that a set of measurable A_γ 's through (2.5) may give rise to a nonmeasurable D_i . However, barring such measurability difficulties, which it is usually easy to eliminate in specific problems, one has the following result.

THEOREM 1. *Relation (2.5) defines a 1:1 correspondence between the decision*

procedures for the problem induced by the hypotheses H_γ , $\gamma \in \Gamma$ and the compatible families \mathcal{G} of tests of these hypotheses.

PROOF. Clearly, if \mathcal{G} is compatible, (2.5) does define a valid decision procedure. Conversely there always exists one, and essentially only one, compatible \mathcal{G} leading to a given decision procedure. To see this, suppose first that $\cup D_i$ is the whole sample space X . The desired result is then a consequence of the fact that (2.5) implies and is implied by

$$(2.7) \quad A_\gamma = \bigcup_{\{i: x_{i\gamma}=1\}} D_i,$$

which we shall now prove.

Assume first that (2.5) holds, and let $x \in A_\gamma$. Since x is in some D_i , it must be in some D_i with $x_{i\gamma} = 1$, and we have $A_\gamma \subseteq \bigcup_{\{i: x_{i\gamma}=1\}} D_i$. On the other hand,

$$\bigcup_{\{i: x_{i\gamma}=1\}} D_i = \bigcup_{\{i: x_{i\gamma}=1\}} \bigcap_{\delta \in \Gamma} A_\delta^{x_{i\delta}} \subseteq \bigcup_{\{i: x_{i\gamma}=1\}} A_\gamma^{x_{i\gamma}} = A_\gamma.$$

Thus (2.5) implies (2.6).

Suppose conversely that (2.6) holds. Then $A_\gamma^{-1} = \bigcup_{\{i: x_{i\gamma}=-1\}} D_i$, and hence

$$\bigcap_{\gamma \in \Gamma} A_\gamma^{x_{k\gamma}} = \bigcap_{\gamma \in \Gamma} \bigcup_{\{i: x_{i\gamma}=x_{k\gamma}\}} D_i.$$

Now a point belongs to the set on the right-hand side of this equation if and only if it lies in a set D_i for which $x_{i\gamma} = x_{k\gamma}$ for all γ . But this holds if and only if $i = k$, so that

$$\bigcap_{\gamma \in \Gamma} \bigcup_{\{i: x_{i\gamma}=x_{k\gamma}\}} D_i = D_k,$$

and (2.6) implies (2.5), as was to be proved.

In case

$$\cup D_i = X - N,$$

where N is a null set for all distributions P_θ , relation (2.5) clearly holds also if each A_γ is replaced by $A_\gamma \cap (X - N)$. Applying the correspondence just proved to the space $X - N$, we see that (2.7) is replaced by

$$A_\gamma \cap (X - N) = \cup D_i.$$

Thus the D 's determine the A 's on the set $X - N$ and it is seen that, except on a null set, (2.5) and (2.7) establish the desired 1:1 correspondence quite generally.

For later reference we also give the relationship of the multiple decision procedure to the tests of the individual hypotheses in the case that they may be randomized. Suppose, for this purpose only, that Γ and I are countable. The tests are then described by means of critical functions φ_γ , where $\varphi_\gamma(x)$ denotes the probability with which H_γ is rejected when x is observed. Similarly, a procedure for the given multiple decision problem is a function ψ of the two arguments i and x , the value $\psi_i(x)$ of which is the probability with which decision D_i

is taken when x is observed. If $\Phi = \{\varphi_\gamma, \gamma \in \Gamma\}$ is a family of tests of the hypotheses H_γ , equations (2.5) and (2.7) become

$$\psi_i = \prod_{\gamma \in \Gamma} \varphi_\gamma^{-x_i \gamma},$$

where $\varphi_\gamma^{-1} = 1 - \varphi_\gamma$, and

$$\varphi_\gamma = \sum_{i \in I} \psi_i^{(1-x_i \gamma)/2}$$

3. Some classification problems. We shall now illustrate the concepts of the previous section with some classes of examples. It will be shown later that for many distributions of interest (normal, binomial, Poisson, etc.) the indicated procedures possess the optimum property of uniformly minimizing the risk among all unbiased procedures, provided the levels α_γ of the tests φ_γ of H_γ are related to the losses a_γ and b_γ through the equation

$$(3.1) \quad \alpha_\gamma = b_\gamma / (a_\gamma + b_\gamma).$$

(i) *Three-decision problems.* Perhaps the simplest problem involving more than two decisions, and one which has been previously treated in somewhat similar terms by the author [8] and by Duncan [3], is that of deciding whether a real-valued parameter θ is less than, equal to, or greater than a specified value θ_0 . This may be generated by the hypotheses $H_1: \theta \geq \theta_0$, $H_2: \theta \leq \theta_0$, which according to (2.1) lead to the choice between the three parameter sets $\Omega_0 = \omega_1 \omega_2: \theta = \theta_0$; $\Omega_1 = \omega_1 \omega_2^{-1}: \theta > \theta_0$; $\Omega_2 = \omega_1^{-1} \omega_2: \theta < \theta_0$. If the losses a and b of false rejection and acceptance are taken to be the same in the two component problems, the loss function is given by

	d_2	d_0	d_1
$\theta < \theta_0$	0	b	$a + b$
$\theta = \theta_0$	a	0	a
$\theta > \theta_0$	$a + b$	b	0

Here the greatest weight is attached to the losses resulting from taking decision d_1 when d_2 is correct and vice versa, which are neglected in the usual formulation of testing $H: \theta = \theta_0$.

In several cases of interest, for example those of a binomial or Poisson population, or of a normal population with θ either the mean or the variance where the other parameter may be unknown, the tests φ_1 and φ_2 of H_1 and H_2 depend on a common statistic T , and have rejection regions $T \leq C_1$ for H_1 and $T \geq C_2$ for H_2 , where the constants are determined by

$$(3.2) \quad P_{\theta_0}\{T \leq C_1\} = P_{\theta_0}\{T \geq C_2\} = \alpha.$$

Of the intersections $\omega_1^{x_1} \omega_2^{x_2}$, only $\omega_1^{-1} \omega_2^{-1}$ is empty. The corresponding action, which would consist in the simultaneous rejection of both hypotheses, is impossible when $\alpha < \frac{1}{2}$ since then $C_1 < C_2$. Subject to this restriction, the pair of

tests (φ_1, φ_2) is therefore compatible. The induced decision procedure consists simply in applying the "equal-tails" test of the hypothesis $H: \theta = \theta_0$ at level 2α and drawing the indicated conclusions. It is of interest to note that this two-sided test usually does not coincide with the standard unbiased test to which one is led by combining the two decisions d_1 and d_2 . If the losses are not the same for the two hypotheses, but are a_i, b_i for H_i , equation (3.2) is replaced by

$$P_{\theta_0}\{T \leq C_1\} = \alpha_1, \quad P_{\theta_0}\{T \geq C_2\} = \alpha_2,$$

and the condition for compatibility becomes $\alpha_1 + \alpha_2 < 1$.

Frequently one wishes to determine not whether θ is exactly equal to θ_0 but only whether this equality holds approximately. The corresponding three-decision problem of deciding whether $\theta < \theta_1, \theta_1 \leq \theta \leq \theta_2$ or $\theta > \theta_2$ (where $\theta_1 < \theta_0 < \theta_2$), can be generated by the hypotheses $H_2: \theta \leq \theta_2$ and $H_1: \theta \geq \theta_1$. The tests φ_1 and φ_2 then have the same form as before with the constants C_1, C_2 determined by

$$P_{\theta_1}\{T \leq C_1\} = \alpha_1, \quad P_{\theta_2}\{T \geq C_2\} = \alpha_2.$$

They are compatible with the given decision problem if $\alpha_1 + \alpha_2 < 1$, and if the cumulative distribution function $F_\theta(c) = P_\theta\{T \leq c\}$ is for each c a decreasing function of θ . For we then have

$$C_1 = F_{\theta_1}^{-1}(\alpha_1) < F_{\theta_2}^{-1}(\alpha_1) \leq F_{\theta_2}^{-1}(1 - \alpha_2) = C_2,$$

and hence $C_1 < C_2$, which was seen above to be the condition for compatibility.

In the above two examples the assumptions concerning the form of the tests φ_1 and φ_2 was unnecessarily restrictive. As an illustration of a somewhat more general situation consider the case of two independent binomial variables X_1 and X_2 , and the problem of deciding whether $p_2 < p_1, p_2 = p_1$ or $p_2 > p_1$. The best unbiased tests of the one-sided hypotheses $p_2 \leq p_1$ and $p_2 \geq p_1$ at level α reject the hypotheses when X_2 is too large, respectively too small, on each line segment $X_1 + X_2 = \text{const.}$, where the cutoff points are determined so that the conditional probability in each tail is equal to α . It is seen as before that this pair of tests is compatible provided $\alpha < \frac{1}{2}$. The induced three-decision procedure consists in performing the two-sided test conditionally on each line segment as an "equal tails" test at level 2α , and then making the indicated statement.

(ii) *Classification of two independent parameters.* Let ξ, η be two real-valued parameters, and consider the problem of classifying ξ and η as being \leq or $>$ than ξ_0 and \leq or $>$ than η_0 respectively, so that the choice lies between the four parameter sets $\Omega_1: \xi \leq \xi_0, \eta \leq \eta_0; \Omega_2: \xi \leq \xi_0, \eta > \eta_0; \Omega_3: \xi > \xi_0, \eta \leq \eta_0; \Omega_4: \xi > \xi_0, \eta > \eta_0$. This clearly can be generated by the hypotheses $H_1: \xi \leq \xi_0, H_2: \eta \leq \eta_0$. The problem of compatibility does not arise here since none of the intersections (2.1) is empty. The procedure consists in carrying out the two tests

separately, and combining the results in the obvious fashion.² Examples in which this would be appropriate are that of a normal population $N(\xi, \sigma^2)$ where it is desired to compare ξ and σ^2 with ξ_0 and σ_0^2 respectively, or that of a bivariate normal population where both means are to be compared with certain standards.

The method applies of course equally well when more than two parameters are to be classified. An example is the preference ordering of m objects by n judges. If we assume that the judges constitute a sample from a population with probability p_{ij} of preferring object i to object j , we can test each of the hypotheses $H_{ij} : p_{ij} \leq \frac{1}{2}$ by means of a sign test. The result of applying the set of all of these tests is a judgment concerning each pair (i, j) that either i is preferred to j , or j to i , or that neither is preferred to the other. In general one will of course not obtain a simple ordering but a complex comparison, which may be represented by a preference polygon as shown for example in Fig. 1 of [7].

(iii) *Comparing several populations.* Let samples be given from s populations with distributions depending on the parameters $\theta_1, \dots, \theta_s$ and possibly certain nuisance parameters, which may or may not be common to the different populations, and consider the hypotheses $H_{ij} : \theta_i \leq \theta_j$. To be specific, let the distributions be normal with means θ_i and common variance σ^2 , and let the rejection region for H_{ij} be, in the usual notation, $\bar{X}_i - \bar{X}_j \leq C_{ij}S$ with

$$C_{ij} = C[(1/n_i) + (1/n_j)]^{1/2}.$$

We shall again assume that the level of the tests is less than $\frac{1}{2}$, so that the constants C_{ij} are positive. The procedure, which is essentially the one proposed by Tukey in [14], leads to the decision $\theta_i = \theta_j$ when $|\bar{X}_i - \bar{X}_j| \leq C_{ij}S$, and it is seen that the system is not compatible since with positive probability

$$|\bar{X}_i - \bar{X}_j| \leq C_{ij}S, \quad |\bar{X}_j - \bar{X}_k| \leq C_{jk}S$$

but $|\bar{X}_i - \bar{X}_k| > C_{ik}S$ while the associated parameter sets $\theta_i = \theta_j$, $\theta_j = \theta_k$ and $\theta_i \neq \theta_k$ have an empty intersection. A justification of the resulting inconsistencies may be obtained if one interprets the acceptance of a hypothesis sufficiently loosely. For such inconsistencies occur only if at least one of the hypotheses is accepted. If, for example, H_{ij} and H_{jk} are both rejected, we have

$$\bar{X}_i - \bar{X}_k > (C_{ij} + C_{jk})S > C_{ik}S.$$

Therefore, H_{ik} is then also rejected corresponding to the fact that $\theta_i > \theta_j$, $\theta_j > \theta_k$ implies $\theta_i > \theta_k$.

A perhaps more satisfactory solution is obtained if one replaces the hypotheses H_{ij} by $H'_{ij} : \theta_i \leq \theta_j + \Delta$, ($\Delta > 0$), with rejection regions $\bar{X}_i - \bar{X}_j > C'_{ij}S$, where the constants are determined so that the probability of rejection is α when $\theta_i = \theta_j + \Delta$. It is easily checked that this system is compatible since H'_{ik} may be false when both H'_{ij} and H'_{jk} are true. Each difference $\theta_i - \theta_j$ is

² A different justification of the resulting acceptance region for the combined hypothesis was given by the author in [10].

now classified as being either $< -\Delta$, between $-\Delta$ and Δ , or $> \Delta$. Since the significance statement obtained for the differences $\theta_i - \theta_j$ in this manner are self-consistent, they lead to a classification of the s populations of the kind described by Duncan in Section 3 of [3].

As another example consider samples of equal size from $N(\xi_i, \sigma_i^2)$, and the problem of classifying the populations according to their variances. This can be generated by the hypotheses $H_{ij} : \sigma_i \leq \sigma_j$, for which the rejection regions are $S_i/S_j > C$. In particular the decision that all of the σ 's are equal is taken when $\max(S_i/S_j) \leq C$, so that the present procedure constitutes a refinement of the test for equality of variances discussed by Hartley [6]. For the same reason as in the preceding example the system is incompatible. But for $\alpha < \frac{1}{2}$ inconsistencies can again occur only if at least one acceptance is involved since the rejection of both H_{ij} and H_{jk} implies $S_i/S_k > C^2 > C$, and hence the rejection of H_{ik} . As before the inconsistencies may be avoided altogether by replacing the H_{ij} by the hypotheses $H'_{ij} : \sigma_i \leq \delta\sigma_j$ ($\delta > 1$), and in this way one obtains, as in the case of the means, a satisfactory classification procedure for the variances.

4. Estimation. (i) *Point estimation.* Let θ be a continuous real-valued parameter, and consider the decision problem generated by the set of hypotheses $H(\theta_0) : \theta \leq \theta_0$. If θ^* denotes the true value of the parameter, the hypotheses $H(\theta_0)$ with $\theta^* \leq \theta_0$ are true while those with $\theta^* > \theta_0$ are false. The associated intersection (2.1) is therefore

$$(4.1) \quad \bigcap_{\theta_0 \geq \theta^*} \omega(\theta_0) \cap \bigcap_{\theta_0 < \theta^*} \omega^{-1}(\theta_0).$$

Since $\omega(\theta_0)$ is the interval $\theta \leq \theta_0$, the set (4.1) consists of the single point θ^* , or in the case that nuisance parameters are present, of the totality of points for which $\theta = \theta^*$. In the induced multiple decision problem the possible decisions therefore correspond exactly to the possible true values of θ , that is, the problem is one of point estimation.

Suppose now, as in Section 2, that the tests of $H(\theta_0)$ are nonrandomized. If then $A(\theta_0)$ is the acceptance region for $H(\theta_0)$, a necessary and sufficient condition for compatibility is that, except on a null set,

$$(4.2) \quad A(\theta_0) \subseteq A(\theta_1), \quad \text{whenever } \theta_0 \leq \theta_1,$$

and

$$(4.3) \quad \bigcap_{\theta > \theta_0} A(\theta) = A(\theta_0).$$

That this condition is necessary is obvious since the corresponding relationships do hold for the $\omega(\theta)$'s. To prove sufficiency one must, by the criterion given at the end of Section 2, show that each sample point lies in one of the intersections (4.1) with $\omega(\theta)$ replaced by $A(\theta)$. Consider now the set of θ 's for which the sample point is in $A(\theta)$, and let $\hat{\theta}$ be its greatest lower bound. Then by (4.2) and (4.3), $H(\theta_0)$ is accepted for $\theta_0 \geq \hat{\theta}$ and rejected for $\theta_0 < \hat{\theta}$, and hence the sample point

lies in the intersection

$$\bigcap_{\theta_0 \geq \hat{\theta}} A(\theta_0) \cap \bigcap_{\theta_0 < \hat{\theta}} A^{-1}(\theta_0),$$

as was to be proved. The decision taken in this case is that θ lies in the set

$$\bigcap_{\theta_0 \geq \hat{\theta}} \omega(\theta_0) \cap \bigcap_{\theta_0 < \hat{\theta}} \omega^{-1}(\theta_0) = \{\hat{\theta}\},$$

that is, that θ equals $\hat{\theta}$, which therefore is a point estimate of θ . The relationship

$$\hat{\theta} \leq \theta_0 \Leftrightarrow X \in A(\theta_0)$$

shows furthermore that measurability of the sets $A(\theta_0)$ implies that of the function $\hat{\theta}$, and conversely. One also sees from it that $\hat{\theta}$ is a lower confidence limit for θ with confidence coefficient $1 - \alpha$, if all the tests are carried out at level α .

If $\theta^* < \hat{\theta}$, the hypothesis $H(\theta_0)$ is incorrectly rejected for $\theta^* \leq \theta_0 < \hat{\theta}$ and never falsely accepted. Thus in accordance with (2.4) the loss may be taken as $a(\hat{\theta} - \theta^*)$, where the loss for false rejection is assumed to be a for all of the hypotheses $H(\theta)$. Similarly, when $\hat{\theta} < \theta^*$, the loss is $b(\theta^* - \hat{\theta})$. The loss is therefore the absolute error, multiplied by a or b as θ is an over- or underestimate. If these two kinds of error are considered of equal importance, the loss is simply proportional to the absolute error.

If the losses a and b are the same for the different hypotheses $H(\theta)$, and if (3.1) is assumed to hold, then the different tests must be carried out at a constant level of significance α . Under this assumption, the optimum one-sided tests satisfy (4.2) and (4.3) in many of the standard problems, in particular when one is dealing with an exponential family of distributions. On the other hand, these conditions may also hold in cases in which the losses a and b and hence also the level α at which $H(\theta)$ is tested, vary with θ . As an example suppose that the tests have acceptance regions of the form $T \leq C(\theta)$, where the c.d.f. $F_\theta(c) = P_\theta\{T \leq c\}$ is for each c a continuous and decreasing function of θ . Then $C(\theta) = F_\theta^{-1}[\alpha(\theta)]$, and conditions (4.2) and (4.3) are satisfied provided $C(\theta)$ is an increasing function of θ , which is continuous on the right, or equivalently if $\alpha(\theta)$ is decreasing and continuous on the right.

Slightly more generally, one can take the losses for over- and underestimation to be

$$(4.4) \quad a \int_{\theta^*}^{\hat{\theta}} d\mu(\theta), \quad b \int_{\hat{\theta}}^{\theta^*} d\mu(\theta),$$

with μ not necessarily Lebesgue measure. In case of a scale parameter, for example, an appropriate loss function may be given by

$$a \int_{\theta^*}^{\hat{\theta}} \frac{1}{\theta} d\theta = a \log(\hat{\theta}/\theta^*), \quad b \int_{\hat{\theta}}^{\theta^*} \frac{1}{\theta} d\theta = b \log(\theta^*/\hat{\theta}).$$

(ii) *Point estimation after a preliminary test of significance.* It is frequently of interest to obtain a point estimate of a parameter θ after one has tested, and rejected, some hypothesis concerning it. If for example a new treatment is being

compared with a standard one, the hypothesis may be tested that the new treatment does not represent an improvement, that is, that $\theta = \eta - \xi \leq 0$, where η and ξ denote the means of the new and old treatments. In case the hypothesis is rejected one requires an estimate of $\eta - \xi$.

A procedure for testing the hypothesis $H: \theta \leq \theta_0$, and estimating θ in case of rejection, can be generated by the set of hypotheses $H(\theta_1): \theta \leq \theta_1$, with $\theta_1 \geq \theta_0$. If the tests are carried out at a constant level α , the procedure consists in performing the usual test of $H(\theta_0)$ and in case of rejection estimating θ by the estimate $\hat{\theta}$ of (i), that is, by the lower confidence limit corresponding to confidence coefficient $1 - \alpha$. A drawback of this method is the limitation it imposes on the levels $\alpha(\theta)$. The conservative attitude reflected by the customary choice of a small level α for testing H suggests that also in the estimation part of the problem an overestimate should be considered more serious than the corresponding underestimate. However, one would usually still wish to test H at a lower level than is desirable for the construction of the estimate. Unfortunately such a combination of levels leads to an incompatible procedure.

In some cases a procedure with the desired properties can be obtained by a slight modification of the construction given above. To illustrate the method consider a single normal variable X with mean θ and unit variance. The hypothesis $H: \theta \leq 0$ is to be tested at level $\alpha < \frac{1}{2}$ with the acceptance region $X \leq C$ and in case of rejection θ is to be estimated by X which corresponds to the level $\frac{1}{2}$. This may be generated by the family of hypotheses $H(\theta_1): \theta \leq \theta_1$, with $\theta_1 = 0$ and $\theta_1 \geq C$, at the levels $\alpha(0) = \alpha$, $\alpha(\theta_1) = \frac{1}{2}$ for $\theta_1 \geq C$. In a similar manner one can generate a joint testing and estimation procedure, in which the level for the estimation part of the problem is higher than that of the test, in the case of a binomial or Poisson variable. Another example in which this is possible is that of the ratio τ^2/σ^2 of two variances (for example in components of variance problems), where one wishes to test $H: \tau^2/\sigma^2 \leq k$, and in case of rejection requires a point estimate of the variance ratio. The method, however, does not appear to be applicable without further modification to the case of a sample X_1, \dots, X_n from $N(\xi, \sigma^2)$ on the basis of which one wishes to test $H: \xi \leq 0$ and in case of rejection to use, say \bar{X} , as an estimate of ξ . While this problem may be generated by the class of hypotheses $N(\xi_1): \xi \leq \xi_1$ with $\xi_1 = 0$ and $\xi_1 \geq CS$, this class depends on the random variable S , and can therefore be determined only after the observations have been taken.

The indicated difficulty usually does not exist if the estimation problem arises when H is accepted rather than when it is rejected. An example of this occurs when one wishes to test the hypothesis that a drug has a significant toxic effect ($H: \theta \geq \theta_0$), and in case H is accepted wants to estimate the size of this effect.

5. Some nonparametric problems. (i) *Testing for goodness of fit.* Let

$$X_1, \dots, X_n$$

be independently distributed with cumulative distribution function F , and consider the problem of deciding whether $F = F_0$, or, if this is judged not to be the

case, of determining the sets of points u for which $F(u)$ is $<$, $=$, and $> F_0(u)$. This problem may be generated by the set of hypotheses $H_+(u): F(u) \geq F_0(u)$ and $H_-(u): F(u) \leq F_0(u)$. If in order to be specific we assume F_0 to be continuous and strictly increasing, the set of u for which the true $F(u)$ exceeds $F_0(u)$ is a union of intervals each of which is open on the right. It is necessary for compatibility that the corresponding condition hold for the set of u at which $F(u)$ is judged to exceed $F_0(u)$, that is, for the set of u for which the sample point is in $A_+(u) \cap A_-^{-1}(u)$. A similar condition must be satisfied by the sets $A_+^{-1}(u) \cap A_-(u)$. These conditions are clearly also sufficient for compatibility since given any two such unions of intervals, there exists a cumulative distribution function F which is in the desired relationship to F_0 .

The best unbiased tests of the hypotheses $H_+(u)$ and $H_-(u)$ are the appropriate one-sided sign tests, which reject the hypotheses if the number $X(u)$ of observations $\leq u$ satisfies

$$(5.1) \quad X(u) < a(u) \quad \text{and} \quad X(u) > b(u),$$

respectively. In order to achieve desired levels of significance $\alpha_+(u)$ and $\alpha_-(u)$ it may be necessary to introduce an auxiliary random variable Z , distributed uniformly on $(0, 1)$, and to reject $H_+(u)$ and $H_-(u)$ as

$$(5.2) \quad X(u) + Z < a(u) \quad \text{and} \quad X(u) + Z > b(u).$$

Unfortunately the usual choice of levels, $\alpha_+(u) = \alpha_-(u) = \alpha$, is not satisfactory for the present problem. In fact, with this choice the tests (5.2) will always lead to rejection for u sufficiently large and sufficiently small respectively. The difficulty stems from the circumstance that for sufficiently extreme u , $X(u)$ tends to the sure variable n or 0 and hence contains no information, so that the decisions in the extreme tails depend solely on the value of Z . Since one is usually not even particularly interested in the behavior of F in the extreme tails, it is natural to avoid this difficulty by choosing $\alpha_+(u)$ and $\alpha_-(u)$ in such a way that they tend to 0 as u tends to $\pm \infty$. This will be the case for example if in (5.1) one sets

$$a(u) = F_0(u) - \Delta \quad \text{and} \quad b(u) = F_0(u) + \Delta,$$

so that the acceptance of all of the hypotheses simultaneously reduces to that of Kolmogoroff's test of the hypothesis $F = F_0$.

One obtains a completely analogous problem, only without the complications caused by the behavior in the tails, if the observations are grouped. The procedure will then decide for each interval whether the hypothesis $p_i = p_{i0}$ is to be accepted or whether the observed frequency in the i th interval indicates that p_i exceeds or falls short of its hypothetical value. This is a special case of the classification problems considered in (ii) of Section 3.

(ii) *The two-sample problem.* The problem of deciding whether two unknown cumulative distribution functions F and G are equal, or in the contrary case of determining for each u whether $G(u)$ is $<$, $=$, or $> F(u)$, may be generated by

the hypotheses $H_+(u):G(u) \geq F(u)$ and $H_-(u):G(u) \leq F(u)$. Let X_1, \dots, X_m , and Y_1, \dots, Y_n be samples from F and G , and denote by $X(u)$ and $Y(u)$ the number of observations in these samples that are $\leq u$. The appropriate tests of $H_+(u)$ and $H_-(u)$ are the standard one-sided tests for equality of two binomial distributions, and with $\alpha_+(u) = \alpha_-(u) = \alpha < \frac{1}{2}$ they are clearly compatible. This choice of levels, as in the previous case, puts the weight of the decision in the tails on an irrelevant random experiment. But this is less serious in the present problem since the very small value of α that is required for satisfactory error control implies that one will only rarely reject the hypotheses in the tails, where $X(u) = Y(u) = 0$ or $X(u) = m, Y(u) = n$.

(iii) *Estimating a cumulative distribution function.* Let X_1, \dots, X_n be independently distributed with cumulative distribution function F , and consider the hypotheses $H(u_0, p_0):F(u_0) \leq p_0$. As in Section 4(i), if F^* denotes the true c.d.f., the hypotheses $H(u_0, p_0)$ for which $F^*(u_0) \leq p_0$ are true, and those with $F^*(u_0) > p_0$ are false. The associated intersection (2.1) is therefore

$$(5.3) \quad \bigcap_{\{(u_0, p_0): p_0 \geq F^*(u_0)\}} \omega(u_0, p_0) \quad \bigcap_{\{(u_0, p_0): p_0 < F^*(u_0)\}} \omega^{-1}(u_0, p_0).$$

Since $\omega(u_0, p_0)$ is the set of all F for which $F(u_0) \leq p_0$, the first and second member of (5.3) are the sets of all F satisfying $F(u) \leq F^*(u)$ for all u and $F(u) \geq F^*(u)$ for all u , respectively. The set (5.3) therefore contains as its only element the c.d.f. F^* .

It is seen that for each fixed u_0 , if we set $\theta = F(u_0)$, we are dealing with the problem of Section 4(i), so that in particular the family of sign tests of the hypotheses $H(u_0, p)$ based on the binomial variable $X(u_0)$ leads to the estimate $\hat{\theta} = \hat{F}(u_0)$ derived there. However, for compatibility one must now add the requirement that as u varies the $\hat{F}(u)$ should constitute a c.d.f. This condition is violated in a rather trivial way if one puts $\alpha(u, p) \equiv \alpha$. For when $X(u)$ is 0 or n , the estimate $\hat{F}(u)$ is not 0 and 1 but only close to these values. One can achieve compatibility by putting $\alpha(u, p) = \alpha(u)$, and letting $\alpha(u)$ tend to 0 as u tends to ∞ , and 1 as u tends to $-\infty$. Since it is enough to make this change in the extreme tails, it need not affect the result in practice.

6. Restricted products of decision problems. The method, described in Section 2, of generating a multiple decision problem from a set of hypotheses is a special case of the following process. Consider the definition of a general decision problem in terms of a family of distributions $\mathcal{P} = \{P_\epsilon, \theta \in \Omega\}$, a space of possible decisions $\mathcal{D} = \{d\}$ and a loss function $W(\theta, d)$. Suppose that \mathcal{P} is fixed but that two different decision spaces $\mathcal{D}', \mathcal{D}''$ with the loss functions W', W'' are of interest. From the two associated decision problems one can form a new problem, which consists in the simultaneous consideration of the two given ones, and may be termed their *product*. Its decision space is the Cartesian product $\mathcal{D}' \times \mathcal{D}''$ and the loss resulting from the decision $d = (d', d'')$ is

$$(6.1) \quad W(\theta, d) = W'(\theta, d') + W''(\theta, d''),$$

or slightly more generally $W(\theta, d) = \rho W'(\theta, d') + (1 - \rho)W''(\theta, d'')$. A typical illustration is the first example of Section 1, where the component problems are concerned with the classification of ξ , respectively η , as negative, zero, or positive, and where the product problem is that of simultaneously classifying both parameters.

This concept is however not general enough to cover most of the other problems considered in the previous sections. Consider for example the two hypotheses $H': \theta \leq \theta_0$ and $H'': \theta \geq \theta_0$. If we denote the decisions to accept H' and H'' by d'_0 and d''_0 and the decisions to reject by d'_1 and d''_1 , the product problem offers the choice of the four decisions (d'_0, d''_0) , (d'_0, d''_1) , (d'_1, d''_0) , (d'_1, d''_1) . Of these the first corresponds to the parameter point $\theta = \theta_0$, the second and third to the sets $\theta < \theta_0$ and $\theta > \theta_0$, while the last one combines two inconsistent decisions and hence corresponds to an empty set in the parameter space. In order to obtain the problem of choosing only between the first three of these possibilities, one must eliminate the point (d', d'') from the decision space \mathfrak{D} . In general, we shall speak of a *restricted product* if in a product problem some of the decision pairs (d', d'') are omitted from $\mathfrak{D}' \times \mathfrak{D}''$, so that \mathfrak{D} is a subset of $\mathfrak{D}' \times \mathfrak{D}''$. Given any procedures δ', δ'' for the problems with decision spaces \mathfrak{D}' and \mathfrak{D}'' , let $\delta = (\delta', \delta'')$ be the procedure that takes decision (d', d'') when $\delta' = d'$ and $\delta'' = d''$. In conformance with our earlier terminology we shall say that the pair (δ', δ'') is *compatible* with the given set of restrictions if

$$P_\theta\{(\delta'(X), \delta''(X)) \in \mathfrak{D}\} = 1 \quad \text{for all } \theta,$$

that is, if the probability is zero of the procedure (δ', δ'') leading to one of the forbidden elements of $\mathfrak{D}' \times \mathfrak{D}''$. Under suitable measurability conditions there is then again a 1:1 correspondence between compatible pairs of decision procedures for the component problems and decision procedures for the restricted product problem. The proof is exactly as that of Theorem 1.

$$(4.12) \quad \left[\frac{f(x, \theta_1)}{f(x, \theta_0)} \right]^\rho \left[\frac{f(x, \theta_{-1})}{f(x, \theta_0)} \right]^{1-\rho} = \left[\frac{r(\theta_1)}{r(\theta_0)} \right]^\rho \left[\frac{r(\theta_{-1})}{r(\theta_0)} \right]^{1-\rho} = h.$$

Hence, for each m and all $x^{(m)}$,

$$(4.13) \quad \min \left[\frac{p_{1m}(x^{(m)})}{p_{0m}(x^{(m)})}, \frac{p_{-1m}(x^{(m)})}{p_{0m}(x^{(m)})} \right] \leq h^{\epsilon}$$

where $\epsilon = \min [1/2\rho, 1/2(1 - \rho)]$. Equation (4.10) follows at once from equation (4.13). This completes the proof of Lemma 4.2.

We now prove some consequences of Assumptions A, B, and C. In what follows $\xi^{(0)}$ is a fixed a priori probability measure all of whose components are positive. The reader should recall the italicized statement a few lines above Assumption A. Write $D^{(0)} = \xi_{-1}^{(0)}/\xi_1^{(0)}$, where $\xi^{(0)} = (\xi_{-1}^{(0)}, \xi_0^{(0)}, \xi_1^{(0)})$.

CORRECTION

Page 14, formula (4.12) through page 17, line 23 should be exchanged with page 70, line 8 through page 72, next-to-last line.

LEMMA 4.3. *Under Assumption A there exist positive constants b_m and a_m ($m = 0, 1, 2, \dots$) with $b_m \leq a_m$ and such that*

$$(4.14) \quad \xi^{(m)} \in C_1 \quad \text{if and only if} \quad p_{1m}(x^{(m)}) / p_{-1m}(x^{(m)}) \geq a_m,$$

$$(4.15) \quad \xi^{(m)} \in C_{-1} \quad \text{if and only if} \quad p_{1m}(x^{(m)}) / p_{-1m}(x^{(m)}) \leq b_m,$$

with strict inequality holding if and only if $\xi^{(m)}$ is an interior point of the appropriate C_i .

(Of course, the values a_m, b_m depend on $\xi^{(0)}$.)

PROOF: The method of proof is similar to that of Theorem 3. Let $x^{(m)}$ and $y^{(m)}$ be such that (4.3) holds, and let $\xi^{(m)}(x^{(m)})$, $\xi^{(m)}(y^{(m)})$ be the a posteriori probability measures corresponding to observed values $x^{(m)}, y^{(m)}$. Equation (4.15) will follow if we can show that $\xi^{(m)}(x^{(m)}) \in C_{-1}$ implies $\xi^{(m)}(y^{(m)}) \in C_{-1}$. (The reader will be aided in what follows if he draws a picture.) Now, (4.3) says that the line $V_0\xi^{(m)}(y^{(m)})$ lies toward V_{-1} from (or on) the line $V_0\xi^{(m)}(x^{(m)})$. Moreover, (4.3) implies (4.4) and (4.5), which say that the line $V_{1\xi^{(m)}}(y^{(m)})$ lies toward V_{-1} from (or on) the line $V_{1\xi^{(m)}}(x^{(m)})$ and that the line $V_{-1\xi^{(m)}}(y^{(m)})$ lies toward V_0 from (or on) the line $V_{-1\xi^{(m)}}(x^{(m)})$. Hence, $\xi^{(m)}(y^{(m)})$ lies inside or on the triangle T whose vertices are V_{-1} , $\xi^{(m)}(x^{(m)})$, and the intercept of $V_{1\xi^{(m)}}(x^{(m)})$ with $V_{-1}V_0$. Since T is contained in the triangle $V_0\xi^{(m)}(x^{(m)})V_{-1}$ which (by convexity) is contained in C_{-1} , (4.15) is proved. Moreover, since the last part of Assumption A implies that $\xi^{(m)}(y^{(m)})$ could lie on the line $V_{-1\xi^{(m)}}(x^{(m)})$ only if $\xi^{(m)}(y^{(m)})$ is V_{-1} or $\xi^{(m)}(x^{(m)})$, it is clear that $\xi^{(m)}(y^{(m)})$ is a boundary point of C_{-1} if and only if either $\xi^{(m)}(y^{(m)}) = V_{-1}$ (see the italicized remark a few lines above Assumption A in this case) or else $\xi^{(m)}(x^{(m)})$ is a boundary point of C_{-1} and $\xi^{(m)}(y^{(m)}) = \xi^{(m)}(x^{(m)})$; the latter implies equality in (4.3). Thus (modulo the italicized remark), defining b_m^* to be (for fixed $\xi^{(0)}$) the supremum (over $x^{(m)}$) of those values $p_{1m}(x^{(m)}) / p_{-1m}(x^{(m)})$ for which $\xi^{(m)}(x^{(m)}) \in C_{-1}$, and taking $b_m = b_m^*$ if $\xi^{(m)}(x^{(m)}) \neq V_{-1}$ is on the boundary of C_{-1} for some $x^{(m)}$ and $b_m^* < b_m < \infimum$ of those $p_{1m}(x^{(m)}) / p_{-1m}(x^{(m)})$ for which $\xi^{(m)}(x^{(m)}) \notin C_{-1}$ otherwise, we see from the previous sentence and the fact that $p_{1m}(y^{(m)}) / p_{-1m}(y^{(m)}) < p_{1m}(x^{(m)}) / p_{-1m}(x^{(m)})$ if $\xi^{(m)}(x^{(m)}) \neq \xi^{(m)}(y^{(m)}) \in T$, that the last part of the lemma as it applies to (4.15) is proved. Equation (4.14) (and the corresponding last part) is proved similarly.

LEMMA 4.4. *Under Assumptions A and B, there exist constants $b_m \leq D^{(0)} \leq a_m$ of Lemma 4.3 satisfying $b_m \leq b_{m+1}, a_m \geq a_{m+1}$, for $m = 0, 1, 2, \dots$.*

PROOF: We shall prove the assertion regarding the b_m , a similar proof applying for the a_m . Keeping $\xi^{(0)}$ fixed as before, in order to prove $b_m \leq b_{m+1}$ it clearly suffices to prove that $\xi^{(m)} \in C_{-1}$ and $p_{1,m+1}(x^{(m+1)}) / p_{-1,m+1}(x^{(m+1)}) \leq p_{1m}(x^{(m)}) / p_{-1,m}(x^{(m)})$ imply that $\xi^{(m+1)} \in C_{-1}$ (the case where either ratio is 0/0 or where both are 0 is easily disposed of); i.e., that $\xi^{(m)} \in C_{-1}$ and $f_{1(x_{m+1})} \leq f_{-1(x_{m+1})}$ imply $\xi^{(m+1)} \in C_{-1}$. The last inequality says that the line $V_0\xi^{(m+1)}$ lies toward V_{-1} from (or on) the line $V_0\xi^{(m)}$; by (4.8) and (4.9), it implies that the line $V_{-1\xi^{(m+1)}}$ lies toward V_0 from (or on) the line $V_{-1\xi^{(m)}}$. Thus, $\xi^{(m+1)}$ lies in the triangle

$V_0 \xi^{(m)} V_{-1}$ and hence, by convexity, $\xi^{(m+1)} \in C_{-1}$. The remaining part of the lemma follows at once from the fact that $\xi_1^{(m)} \leq \xi_{-1}^{(m)}$ is equivalent to $p_{1m} / p_{-1m} \leq D^{(0)}$.

If $f_1(x) / f_{-1}(x)$ cannot take on a suitably dense set of values, the a_m and b_m might (for fixed $\xi^{(0)}$) not be unique and might correspond to $\xi^{(m)}$ in the interior of the C_i or the complement of $C_{-1} \cup C_1$. If this is not the case, the previous paragraph and the fact that $V_0 P V_{-1} \in C_{-1}$ show that we can strengthen the weak inequality of Lemma 4.4. One possible formulation of this result is the following:

LEMMA 4.5. *If $b_m < D^{(0)}$ (resp., $a_m > D^{(0)}$) and if for every open interval J containing b_m (resp., a_m) the ratio $p_{1m}(x^{(m)}) / p_{-1,m}(x^{(m)})$ takes on values in $J - \{b_m\}$ (resp., $J - \{a_m\}$) with positive probability under H_1 and H_{-1} (so that b_m, a_m are unique), and if for every open interval J' containing $D^{(0)}$ as a left (resp., right) end-point $f_1(x) / f_{-1}(x)$ takes on values in J' with positive probability under H_1 and H_{-1} , then $b_m < b_{m+1}$ (resp., $a_m > a_{m+1}$). In particular, in case (3) of Lemma 4.2, if μ is equivalent to Lebesgue measure (or if Lebesgue measure is absolutely continuous with respect to μ) on the real line, the a_m and b_m are unique and this last result holds.*

In fact, it remains only to prove the last assertion of the lemma, which follows at once from the fact that $e^{(\theta_1 - \theta_{-1})x}$ takes on values in any interval of positive numbers with positive probability under H_1 and H_{-1} , if Lebesgue measure is absolutely continuous with respect to μ .

LEMMA 4.6. *Under Assumption C, for any $\xi^{(0)}$ all of whose components are (or in fact, for which $\xi_0^{(0)}$ is) positive, there is an integer $N = N(\xi^{(0)})$ such that every Bayes solution with respect to $\xi^{(0)}$ requires fewer than N observations with probability one under f_1, f_{-1} , and f_0 .*

PROOF: Fix $\xi^{(0)}$. Since P is a positive distance from V_0 , there is clearly a positive number c such that every Bayes solution must stop with probability one whenever either $p_{1m}(x^{(m)}) / p_{0m}(x^{(m)}) < c$ or else $p_{-1m}(x^{(m)}) / p_{0m}(x^{(m)}) < c$. The desired result now follows at once from (4.10). (Note again the remark made in italics just before Assumption A).

We may now summarize our results:

THEOREM 4. *Under Assumptions A, B, and C (in particular, under (1), (2), or (3) of Lemma 4.2), any procedure which minimizes $A_0(\delta)$ subject to (4.1) is a GSPRT of H_1 against H_{-1} with $b_m \leq b_{m+1} \leq D^{(0)} \leq a_{m+1} \leq a_m$ for $m = 0, 1, 2, \dots, N$ and some $D^{(0)}$, which stops with probability one under f_i ($i = 0, \pm 1$) after N or fewer observations. Under additional conditions specified in Lemma 4.5 the values a_m, b_m ($m \geq 1$) will be unique and $a_m > a_{m+1}$ or $b_m < b_{m+1}$ unless $a_m = D^{(0)}$ or $b_m = D^{(0)}$, where $D^{(0)}$ corresponds to the a priori distribution with respect to which the optimum procedure is Bayes.*

REMARKS, GENERALIZATIONS, ETC.

1. Of course, a GSPRT of Theorem 4 involves a randomization rule for all $m \leq N$, including a possibly randomized starting rule ($m = 0$) if a_0 or $b_0 = D^{(0)}$. If μ is nonatomic, there will clearly exist an optimum GSPRT involving no randomization, except possibly in the starting rule. The lack of uniqueness of

the a_m and b_m in cases not covered by Lemma 4.5 is of course inessential, reflecting only that certain intervals of values of p_{1m}/p_{-1m} have probability zero under all f_i .

2. In all of the above, the X_i are random elements whose range is immaterial (not necessarily real) as long as the appropriate assumptions are satisfied. To conserve space we have not included statements about the obvious sets of measure zero where various conditions may be permitted to fail.

3. As an example of what can happen when our assumptions are not satisfied, we mention briefly the following example: Suppose $f_j(x) = 1 / \pi[1 + (x - j)^2]$, $j = 0, \pm 1$, $\mu =$ Lebesgue measure. In this Cauchy case it is easy to see that, for $\xi^{(0)}$ with all components positive, the set of possible $\xi^{(1)}$ values is a simple closed curve minus the point $\xi^{(0)}$, and lies entirely in the interior of the triangle $V_0V_1V_{-1}$. Assumption A is not satisfied, and there is no reason why the result of Lemma 4.3. should be valid. Also, since (e.g.) $f_j(\sqrt{3}/2)f_j(-\sqrt{3}/2)$ does not depend on j , there is no reason why the result of Lemma 4.6 should hold here.

4. Remarks analogous to those of Section 2 can be made here: for concave (resp., convex) nondecreasing $c(n)$, the minimization of a linear combination such as (4.2) with $A_0(\delta)$ replaced by $E_0c(n)$ under δ may be compared in an obvious fashion to the minimization when $c(n)$ is replaced by the linear homogeneous cost function $c_m^*(n)$ passing through $(1, c(m + 1) - c(m))$ (i.e., to the solution of the problem we have considered): the stopping region will now change with m , being contained in (resp., containing) that fixed region for the problem concerning $c_m^*(n)$.

To prove the procedures unbiased, we note that unbiasedness of a two-decision procedure φ with losses a and b , by (6.3) is equivalent to

$$(7.4) \quad E_\theta \varphi(X) \begin{matrix} \leq \\ \geq \end{matrix} b/(a + b) \quad \text{for} \quad \begin{matrix} \theta \varepsilon \omega \\ \theta \varepsilon \omega^{-1} \end{matrix},$$

that is, to the Neyman-Pearson condition of unbiasedness at the level

$$(7.5) \quad \alpha = b/(a + b).$$

The result now follows from (iv) of Section 6 since in all of the examples the procedures were obtained as products of unbiased tests.

Unfortunately, as has already been pointed out, it is in general not true that unbiasedness of a product implies the same property for the component problems. Suppose however that for every test of $H: \theta \varepsilon \omega$, the power function $E_\theta \varphi(X)$ is a continuous function of θ . Then unbiasedness of φ entails the somewhat weaker condition of *similarity on the boundary*, namely

$$(7.6) \quad E_\theta \varphi(X) = \alpha \quad \text{for} \quad \theta \varepsilon \Lambda,$$

where Λ is the common boundary of ω and ω^{-1} . For an important class of testing problems, there exists not only among all unbiased tests but also among the larger class of tests satisfying (7.6), one that uniformly maximizes $E_\theta \varphi(X)$ for $\theta \varepsilon \omega^{-1}$ and uniformly minimizes it for $\theta \varepsilon \omega$. This test therefore, among all those that are similar on the boundary, uniformly minimizes the risk (7.1). As was shown in

CORRECTION

Page 14, formula (4.12) through page 17, line 23 should be exchanged with page 70, line 8 through page 72, next-to-last line.

[11], this is the case in particular when $\theta = (\theta_1, \dots, \theta_s)$, the distributions of Ω form an exponential family and ω is of the form $\theta_1 \leq \theta_1^0$ or $\theta_1 \geq \theta_1^0$.

The desired optimum property of the various procedures discussed in the earlier sections, for finite Γ is an easy consequence of the above remarks and the following theorem.

THEOREM 2. *Let $\{H_\gamma: \theta \in \omega_\gamma, \gamma \in \Gamma\}$ be a finite family of hypotheses, and suppose that for each γ the test φ_γ^0 uniformly minimizes the risk among all tests that are similar on the boundary at level $\alpha_\gamma = b_\gamma/(a_\gamma + b_\gamma)$, and that the family $\{\varphi_\gamma^0, \gamma \in \Gamma\}$ is compatible. Suppose further that the following structural assumption is satisfied.*

(*) *For every $\gamma_0 \in \Gamma$ and each common boundary point θ of ω_{γ_0} and $\omega_{\gamma_0}^{-1}$ there exist intersection sets Ω_i and Ω_j of the form (2.1) such that θ is also a common boundary point of Ω_i and Ω_j , and such that $x_{i\gamma} = x_{j\gamma}$ for all $\gamma \neq \gamma_0$ but that $x_{i\gamma_0} \neq x_{j\gamma_0}$. Then if $E_\theta \varphi_\gamma(X)$ is continuous in θ for each γ , the product procedure ψ^0 given by*

$$\psi_k^0 = \prod_{\gamma \in \Gamma} (\varphi_\gamma^0)^{x_{k\gamma}}$$

is unbiased, and uniformly minimizes the risk among all unbiased decision procedures of the restricted product problem, the components of which are the problems of testing H_γ with losses a_γ and b_γ .

PROOF. Let \mathcal{C} be the class of all decision procedures the component tests of which are similar on the boundary at level α_γ . It follows from the assumptions made and from (iii) of Section 6 that the procedure ψ^0 uniformly minimizes the risk within \mathcal{C} . Since the tests φ_γ^0 are unbiased—as is seen by comparison with the tests $\varphi_\gamma(x) \equiv \alpha_\gamma$ —the same is true of ψ^0 . Let \mathcal{C}_0 denote the class of all unbiased procedures of the product problem. We shall now show that $\mathcal{C}_0 \subseteq \mathcal{C}$, which will complete the proof.

Let ψ be any procedure belonging to \mathcal{C}_0 , let γ_0 be any element of Γ and θ_0 any boundary point of ω_{γ_0} and $\omega_{\gamma_0}^{-1}$. Let Ω_i and Ω_j be the sets, the existence of which is guaranteed by (*), and assume without loss of generality that $x_{i\gamma_0} = 1$. Then unbiasedness of ψ implies that for any θ in Ω_i ,

$$\begin{aligned} E_\theta \sum_\gamma \frac{1}{2} [(x_{i\gamma} + 1)a_\gamma \varphi_\gamma(X) - (x_{i\gamma} - 1)b_\gamma \varphi_\gamma^{-1}(X)] \\ \leq E_\theta \sum_\gamma \frac{1}{2} [(x_{j\gamma} + 1)a_\gamma \varphi_\gamma(X) - (x_{j\gamma} - 1)b_\gamma \varphi_\gamma^{-1}(X)], \end{aligned}$$

where the φ_γ are the component tests of ψ . Since $x_{i\gamma} = x_{j\gamma}$ for $\gamma \neq \gamma_0$, $x_{i\gamma_0} = 1$, $x_{j\gamma_0} = -1$, this reduces to

$$a_{\gamma_0} E_\theta \varphi_{\gamma_0}(X) \leq b_{\gamma_0} E_\theta [1 - \varphi_{\gamma_0}(X)].$$

Analogously the opposite inequality is seen to hold for any θ in Ω_j . Because of the continuity of $E_\theta \varphi_{\gamma_0}(X)$ it follows that equality must hold on the boundary of Ω_i and Ω_j , and hence in particular for $\theta = \theta_0$. Thus

$$(7.7) \quad E_\theta \varphi_{\gamma_0}(X) = \frac{b_{\gamma_0}}{a_{\gamma_0} + b_{\gamma_0}} = \alpha_{\gamma_0}$$

for every boundary point of ω_{γ_0} and $\omega_{\gamma_0}^{-1}$, as was to be proved.

The assumptions of this theorem may be weakened slightly at one point, which is important for applications. If Λ_γ is the common boundary of ω_γ and ω_γ^{-1} , it is not necessary for (*) to be satisfied at every point of Λ_γ but enough if it holds for the points of a dense subset Λ_γ^0 . The proof then shows that $E_\theta \varphi_\gamma(X) = \alpha_\gamma$ for all $\theta \in \Lambda_\gamma^0$, and by continuity the equality holds as before for all points of Λ_γ .

The remaining conditions being automatically satisfied for any problem that is generated by one-sided hypotheses concerning one of the parameters in an exponential family, it is necessary only to verify (*) in order to prove the desired optimum property for the various examples of Section 3. This requires no further reference to the possible distributions of the observable random variables, since (*) concerns only the structure of the multiple decision problem, that is, of the sets Ω_i , not the distributions that are represented by the points $\theta \in \Omega$.

In the first example of Section 3, the only boundary point of ω_i and ω_i^{-1} ($i = 1, 2$) is θ_0 . But this is also the boundary point of $\Omega_0 = \omega_i \omega_j$ and $\Omega_i = \omega_i^{-1} \omega_j$ ($i \neq j$), and hence (*) is satisfied. The result here is slightly stronger than the one given by the author in [8] since the condition of unbiasedness can be seen to be less stringent than the restriction imposed on the procedure in [8]. The checking of (*) is exactly analogous in the second version of this example, in which θ_0 is replaced by the interval $\theta_1 \leq \theta \leq \theta_2$.

In example (ii), the common boundary points of ω_1 and ω_1^{-1} for example, are the points with $\xi = \xi_0$. Let (ξ_0, η) be any such point, and suppose without loss of generality that $\eta \leq 0$. Then (ξ_0, η) is also a boundary point of $\Omega_1 = \omega_1 \omega_2$ and $\Omega_3 = \omega_1^{-1} \omega_2$, as was to be proved. The other cases are verified analogously.

In example (iii) where ω_{ij} is the parameter set $\theta_i \leq \theta_j + \Delta$, consider the common boundary of, say, ω_{21} and ω_{21}^{-1} . By the remark following Theorem 2, we may restrict attention to points of this boundary satisfying

$$\theta_{i_1} < \dots < \theta_{i_r} < \theta_2 = \theta_1 + \Delta < \theta_{j_1} < \dots < \theta_{j_{s-r-2}}$$

and we may assume further that all of the differences $\theta_2 - \theta_{i_t}$ ($t = 1, \dots, r$) are $\neq \Delta$. Let

$$\theta^0 = (\theta_1^0, \dots, \theta_s^0) \in \omega_{12} \cap \omega_{21} \cap \bigcap \omega_{i_j}^{x_{i_j}^i}$$

be any specific such point, and consider the points $(\theta_1^0, \theta_2, \dots, \theta_s^0)$ with $0 < \theta_2^0 - \theta_2 < \epsilon$. If ϵ is sufficiently small the relationship between all pairs of coordinates will be the same as before, except that $\theta_2 <$ instead of $= \theta_1 + \Delta$. These points are therefore in the intersection.

$$\omega_{12}^{-1} \cap \omega_{21} \cap \bigcap \omega_{i_j}^{x_{i_j}^i},$$

and since θ^0 is a boundary point of this set as well as of the intersection differing from the present one only in the factor ω_{12} , (*) is verified.

8. Optimality of the procedures of Sections 4 and 5. A basic assumption of Theorem 2 is the finiteness of the set Γ , and the theorem is therefore not applicable to any of the problems of Sections 4 and 5. Since the assumption was used

however only in the proof of the relationship

$$(8.1) \quad \mathfrak{C}_0 \subseteq \mathfrak{C},$$

it will be enough in the following to prove (8.1) in each case. This means showing for these problems that unbiasedness of a decision procedure implies that all of its component tests are similar on the boundary.

(i) *Estimation.* For the problem of estimating a real-valued parameter θ , let the risk function of an estimate Y , in accordance with Section 4, be given by

$$a \int_{\theta}^{\infty} (y - \theta) dP_{\theta}(y) + b \int_{-\infty}^{\theta} (\theta - y) dP_{\theta}(y),$$

where P_{θ} denotes the probability distribution of Y . The condition of unbiasedness then becomes

$$(8.2) \quad \begin{aligned} a \int_{\theta}^{\infty} (y - \theta) dP_{\theta}(y) + b \int_{-\infty}^{\theta} (\theta - y) dP_{\theta}(y) \\ \leq a \int_{\theta'}^{\infty} (y - \theta') dP_{\theta}(y) + b \int_{-\infty}^{\theta'} (\theta' - y) dP_{\theta}(y) \end{aligned}$$

for all θ, θ' . In the case that $a = b$, this states that the estimate, on the average is closer to the true value θ than to any other value θ' .

In the following we shall restrict attention to estimates Y with finite risk. If $\theta < \theta'$, (8.2) then reduces to

$$\begin{aligned} a \left[\int_{\theta}^{\theta'} (y - \theta) dP_{\theta}(y) + (\theta' - \theta) \int_{\theta'}^{\infty} dP_{\theta}(y) \right] \\ \leq b \left[\int_{\theta}^{\theta'} (\theta' - y) dP_{\theta}(y) + (\theta' - \theta) \int_{-\infty}^{\theta} dP_{\theta}(y) \right]. \end{aligned}$$

Dividing both sides by $\theta' - \theta$, and letting θ' tend to θ , we see that

$$0 \leq \frac{1}{\theta' - \theta} \int_{\theta}^{\theta'} (y - \theta) dP_{\theta}(y) \leq P_{\theta}\{\theta < Y \leq \theta'\} \rightarrow 0,$$

and that similarly also the first term on the right-hand side tends to zero. In the limit we therefore get

$$aP_{\theta}\{Y > \theta\} \leq bP_{\theta}\{Y \leq \theta\}$$

or

$$P_{\theta}\{Y > \theta\} \leq \frac{b}{a + b} = \alpha.$$

By letting θ' tend to θ from below, we find analogously that unbiasedness of the estimate Y implies

$$P_{\theta}\{Y \geq \theta\} \geq \alpha$$

Suppose now that the distributions of Ω constitute an exponential family, and that they possess densities with respect to Lebesgue measure. This is the case not only for families of univariate and multivariate normal distributions, of gamma distributions, etc., but also when one is dealing with binomial or Poisson variables to which, as a randomization device, one adds a variable that is uniformly distributed over $(0, 1)$. If then $P_\theta\{Y = c\}$ is positive for some θ , it is positive for all θ . Hence

$$(8.3) \quad P_\theta\{Y > \theta\} = \alpha,$$

except possibly for a countable set of parameter values, and it follows from a theorem of Scheffé [13] that (8.3) must hold for all θ . Since it was shown in Section 4 that $Y > \theta_0$ is the rejection region of the hypothesis $H(\theta_0): \theta \leq \theta_0$, and since θ_0 is the only common boundary point of $\omega(\theta_0)$ and $\omega^{-1}(\theta_0)$, this completes the proof of (8.1).

(ii) *Estimation after a preliminary test of significance.* For the problem generated by the hypotheses $H(\theta_1): \theta \leq \theta_1 (\theta_0 \leq \theta_1)$, and with constant losses a and b , the risk is

$$a \int_{\theta_+}^{\infty} (y - \theta) dP_\theta(y) + b \int_{\theta_+}^{\theta} (\theta - y) dP_\theta(y) + b(\theta - \theta_0) \int_{-\infty}^{\theta_0} dP_\theta(y) \quad \text{if } \theta_0 < \theta,$$

and

$$a \int_{\theta_+}^{\infty} (y - \theta_0) dP_\theta(y) \quad \text{if } \theta \leq \theta_0.$$

It is seen exactly as before that unbiasedness implies (8.3) for all $\theta > \theta_0$, and that this holds also for $\theta = \theta_0$ follows again from Scheffé's theorem.

Consider instead the modified problem of Section 4 (ii) generated by the hypotheses $H(\theta_1): \theta \leq \theta_1$ with $\theta_1 = 0$ or $\theta_1 \geq C$, and with losses a_0, b_0 for $H(0)$ and a, b for the remaining hypotheses. A compatible procedure accepts $H(0)$ when a statistic $Y \leq C$, and otherwise takes Y as an estimate of θ . The associated risk is

$$\begin{aligned} & \int_{c_+}^{\infty} [a(y - C) + \rho a_0] dP_\theta(y) && \text{if } \theta \leq 0, \\ & b_0 \rho \int_{-\infty}^c dP_\theta(y) + a \int_{c_+}^{\infty} (y - C) dP_\theta(y) && \text{if } 0 < \theta \leq C, \\ & a \int_{\theta_+}^{\infty} (y - \theta) dP_\theta(y) + b \int_{c_+}^{\theta} (\theta - y) dP_\theta(y) \\ & \quad + \int_{-\infty}^c [b(\theta - C) + b_0 \rho] dP_\theta(y) && \text{if } C < \theta. \end{aligned}$$

Here we have taken the measure μ of (2.4) to be Lebesgue measure for the hypotheses $H(\theta_1)$ with $\theta_1 \geq C$ and to assign measure ρ to the hypothesis $H(0)$.

For $\theta \geq C$ we see as before that unbiasedness implies (8.3), while for $\theta = 0$ we find

$$a_0 P_0\{Y > C\} = b_0 P_0\{Y \leq C\}$$

and hence

$$P_0\{Y > C\} = \alpha_0$$

by considering the condition of unbiasedness for $\theta = 0$ and $\theta' \downarrow 0$ and vice versa.

(iii) *Testing for goodness of fit.* The problem, as described in Section 5, is generated by the hypotheses $H_-(u): F(u) \leq F_0(u)$ and $H_+(u): F(u) \geq F_0(u)$. Let $d_-(u)$, $d_0(u)$, $d_+(u)$ denote the decisions that the true $F(u)$ is $<$, $=$, $>$ than $F_0(u)$ and let $\psi_-(x, u)$, $\psi_0(x, u)$, $\psi_+(x, u)$ be the probabilities with which these decisions are taken when $x = (x_1, \dots, x_n)$ is observed. Then the corresponding over-all probabilities of these decisions are $P_-(u) = P_-(u, F) = E_F \psi_-(X, u)$, $P_0(u)$ and $P_+(u)$. If the losses resulting from false rejection and acceptance are $a(u)$ and $b(u)$ for both $H_-(u)$ and $H_+(u)$, and if we put

$$R_-(u, F) = R_-(u) = [a(u) + b(u)]P_+(u) + b(u)P_0(u),$$

$$R_0(u, F) = R_0(u) = a(u)[P_-(u) + P_+(u)],$$

$$R_+(u, F) = R_+(u) = [a(u) + b(u)]P_-(u) + b(u)P_0(u),$$

the risk function is

$$R(F) = \int_{S_-} R_-(u) d\mu(u) + \int_{S_0} R_0(u) d\mu(u) + \int_{S_+} R_+(u) d\mu(u),$$

where S_- , S_0 and S_+ are the sets on which the true $F(u)$ is $<$, $=$, and $>$ than $F_0(u)$. In particular

$$R(F_0) = \int_{-\infty}^{\infty} R_0(u, F_0) d\mu(u).$$

We shall assume in the following that $R(F_0)$ is finite, that $a(u)$ and $b(u)$ are bounded in every finite interval, and for the sake of convenience also that F_0 possesses a probability density.

The condition of unbiasedness becomes in the present case

$$\begin{aligned} & \int_{S_-} R_-(u, F) d\mu(u) + \int_{S_0} R_0(u, F) d\mu(u) + \int_{S_+} R_+(u, F) d\mu(u) \\ & \leq \int_{S'_-} R_-(u, F) d\mu(u) + \int_{S'_0} R_0(u, F) d\mu(u) + \int_{S'_+} R_+(u, F) d\mu(u), \end{aligned}$$

where S'_- , S'_0 and S'_+ are the sets on which some alternative c.d.f. F' is $<$, $=$, and $>$ than F_0 . Consider this condition now for some F and F' , both of which agree with F_0 except on a finite interval I on which $F < F_0$ and $F' > F_0$. It

then reduces to

$$\int_I R_-(u, F) d\mu(u) \leq \int_I R_+(u, F) d\mu(u).$$

If, holding I fixed, one considers a sequence of such distributions F , which possess probability densities that tend to the density of F_0 , it follows from Scheffé's theorem that

$$\int_I R_-(u, F_0) d\mu(u) \leq \int_I R_+(u, F_0) d\mu(u),$$

and since I was an arbitrary interval, that

$$R_-(u, F_0) \leq R_+(u, F_0) \quad \text{a.e.}$$

Analogously one sees that the reverse inequality must hold, so that

$$R_-(u, F_0) = R_+(u, F_0) \quad \text{a.e.}$$

If in the above argument F is replaced by F_0 , one finds

$$\int_I R_0(u, F_0) d\mu(u) \leq \int_I R_+(u, F_0) d\mu(u)$$

for all I , and hence

$$R_0(u, F_0) \leq R_+(u, F_0) \quad \text{a.e.}$$

Similarly, on replacing F' by F_0 one gets

$$\int_I R_+(u, F) d\mu(u) \leq \int_I R_0(u, F) d\mu(u)$$

for all I , which by the same argument as before leads to

$$R_+(u, F_0) \leq R_0(u, F_0).$$

Thus, for almost all u ,

$$R_-(u, F_0) = R_0(u, F_0) = R_+(u, F_0),$$

which implies

$$P_-(u, F_0) = P_+(u, F_0) = \frac{b(u)}{a(u) + b(u)} = \alpha(u),$$

as was to be proved.

A very similar proof applies in the two-sample problem discussed in Section 5(ii), and we shall therefore not give the details.

(iv) *Estimating a cumulative distribution function.* The problem of estimating a c.d.f. was treated from a minimax point of view by Agarwal [1] for several loss functions all of which differ from the one below. Following our earlier definitions,

we take the risk function here to be

$$\int_{-\infty}^{\infty} \left\{ \int_{F(u)}^{\infty} a(u)[y - F(u)] dF_u(y) + \int_{-\infty}^{F(u)} b(u)[F(u) - y] dP_u(y) \right\} d\mu(u),$$

where $P_u = P_{u,F}$ is the distribution of the estimate $Y(u)$ of $F(u)$. As in the problem of estimating a single parameter, we shall restrict attention to estimates with finite risk, and the condition of unbiasedness then becomes

$$\begin{aligned} & \int_{-\infty}^{\infty} \left\{ a(u) \int_{F(u)}^{F'(u)} [y - F(u)] dP_u(y) + [F'(u) - F(u)] \int_{F'(u)}^{\infty} dP_u(y) \right\} d\mu(u) \\ & \leq \int_{-\infty}^{\infty} \left\{ b(u) \int_{F(u)}^{F'(u)} [F'(u) - y] dP_u(y) + [F'(u) - F(u)] \int_{-\infty}^{F'(u)} dP_u(y) \right\} d\mu(u), \end{aligned}$$

where the probabilities are computed with respect to F , and where F' is any alternative c.d.f.

We shall consider first the case that F is the uniform distribution on $(0, 1)$, that $I = (u_0, u_1)$ is any subinterval of $(0, 1)$, and that F' is any continuous c.d.f. such that $F'(u) = F(u) + \Delta$ for $u_0 < u < u_1$ and $F'(u) = F(u)$ for $u < u_0 - \Delta$ and $u > u_1 + \Delta$. On dividing by Δ , and letting Δ tend to zero, (8.4) is then seen to reduce to

$$\int_I a(u) P_u \{ Y(u) > F(u) \} d\mu(u) \leq \int_I b(u) P_u \{ Y(u) \leq F(u) \} d\mu(u),$$

and since this holds for all I , to

$$P_u \{ Y(u) > F(u) \} \leq \frac{b(u)}{a(u) + b(u)} = \alpha(u) \quad \text{a.e.}$$

On letting F' tend to F from below one finds similarly

$$P_u \{ Y(u) \geq F(u) \} \geq \alpha(u) \quad \text{a.e.}$$

In exactly the same manner these two inequalities are seen to hold also for any F belonging to the family \mathfrak{F} of mixtures of uniform distribution over nonoverlapping intervals. By considering a countable dense subset of \mathfrak{F} , for example mixtures with rational weights of uniform distributions over intervals with rational endpoints, it is seen that there exists a null set N such that for any $u \notin N$ the two inequalities hold for all $F \in \mathfrak{F}$, and it follows by an argument similar to that given in (i) of this section that for all $u \notin N$, and all $F \in \mathfrak{F}$

$$P_{u,F} \{ Y(u) > F(u) \} = \alpha(u).$$

For $F \in \mathfrak{F}$ and any fixed $u \notin N$, the common boundary points of $\omega(u_0, p_0)$: $F(u_0) \leq p_0$ and $\omega^{-1}(u_0, p_0)$ are exactly the distributions of \mathfrak{F} for which $F(u_0) = p_0$. For these we then have

$$P_{u,F} \{ Y(u_0) > p_0 \} = \alpha(u_0),$$

and since the left-hand side is the probability of rejecting the hypothesis $H(u_0, p_0)$, this completes the proof of (8.1). The desired optimum property of the procedure now follows from the fact, proved by Fraser [4], [5], that for the family of distributions \mathcal{F} , the one-sided sign test uniformly minimizes the probabilities of error among all unbiased tests of the hypothesis $H(u_0, p_0)$.

REFERENCES

- [1] OM P. AGGARWAL, "Some minimax invariant procedures for estimating a cumulative distribution function," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 450-463.
- [2] R. C. BOSE AND S. N. ROY, "Simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 513-536.
- [3] DAVID B. DUNCAN, "Multiple range and multiple F-tests," *Biometrics*, Vol. 11 (1955), pp. 1-42.
- [4] D. A. S. FRASER, "Completeness of order statistics," *Canadian J. Math.*, Vol. 6 (1953), pp. 42-45.
- [5] D. A. S. FRASER, "Non-parametric theory: Scale and location parameters," *Canadian J. Math.*, Vol. 6 (1953), pp. 46-68.
- [6] H. O. HARTLEY, "Maximum F ratio as a short-cut test for heterogeneity of variance," *Biometrika*, Vol. 37 (1950), pp. 308-312.
- [7] M. G. KENDALL, "Further contributions to the theory of paired comparisons," *Biometrics*, Vol. 11 (1955), pp. 43-62.
- [8] E. L. LEHMANN, "Some principles of the theory of testing hypothesis," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1-26.
- [9] E. L. LEHMANN, "A general concept of unbiasedness," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 587-592.
- [10] E. L. LEHMANN, "Testing multiparameter hypotheses," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 541-552.
- [11] E. L. LEHMANN AND HENRY SCHEFFÉ, "Completeness, similar regions, and unbiased estimation, Part II," *Sankhyā*, Vol. 15 (1955), pp. 219-236.
- [12] S. N. ROY, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 220-238.
- [13] HENRY SCHEFFÉ, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434-438.
- [14] J. W. TUKEY, "Quick and dirty methods in statistics, Part II, Simple analyses for standard designs," *Proc. Fifth Annual Convention, Am. Soc. for Quality Control* (1951), pp. 189-197.