

A THEORY OF SOME MULTIPLE DECISION PROBLEMS*. II

BY E. L. LEHMANN¹

University of California, Berkeley

Summary. The theory of Part I is extended to problems in which it is permitted not to come to a definite conclusion regarding one or more of the questions under consideration. Some problems are also investigated in which, from a single set of observations, one wishes to answer a number of questions in sequence. Here the nature of the question at a later stage will depend on the answers obtained at the earlier stages.

9. Decision procedures permitting partial conclusions. It is a common feature of all the problems treated in Part I, that a fixed partition of the parameter space Ω into sets Ω_i is given, and that one wishes to determine which of these sets contains the true parameter point. There are however many statistical problems, such as the estimation by confidence sets, in which the possible decisions do not correspond to the sets of a fixed partition. In particular, this is the case in the field of statistical inference, when the statistician is free to decide how sharp a statement he can reliably make on the basis of the observations. We shall show in the present section how such problems may be generated by the simultaneous consideration of a number of two-decision problems as in Part I, if one suitably modifies the interpretation of the decisions involved.

Previously we were concerned with testing a set of hypotheses $H_\gamma: \theta \in \omega_\gamma$, so that in each component problem the choice lay between the two decisions $\theta \in \omega_\gamma$ (acceptance of H_γ) and $\theta \in \omega_\gamma^{-1}$ (rejection of H_γ). Suppose now instead that the statistician is asked only whether the data reject the hypothesis, and that in case they do not, no alternative positive statement is required. The choice may then be said to lie between the statements $\theta \in \omega_\gamma^{-1}$ and $\theta \in \omega_\gamma^0 = \Omega$.

This actually appears to be the point of view taken by Duncan ("Multiple range and multiple F tests", *Biometrics*, Vol. 11 (1955), pp. 1-42) in his formulation of this class of multiple decision problems to which reference is made in section 1 of the present paper.

If one considers simultaneously a number of such problems, one is faced with a multiple decision problem in which the different possible decisions correspond to the statements that a certain number of the hypotheses H_γ are false, but where nothing is said regarding the remaining hypotheses. This is equivalent to the statement that the parameter point θ lies in the set

$$(9.1) \quad \Omega_i = \bigcap_\gamma \omega_\gamma^{y_i}$$

Received October 12, 1956.

* To rectify a printing error in Part I, in the March, 1957 issue of these *Annals* replace page 14, formula (4.12) through page 17, line 23 by page 70, line 8 through page 72, next-to-last line. Ed.

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

where the y 's are -1 for each rejected H_γ and 0 for the others. In the particular case that all of the y 's are zero, we have $\Omega_i = \Omega$, and thus make no statement whatever about the position of θ . As before, it may of course happen that some of the formal intersections (9.1) are empty, and we shall then restrict the Ω 's to denote the nonempty ones and require that none of the possible decisions should correspond to these empty intersections. If we assume that in the simultaneous consideration of a number of problems the losses are additive so that the total loss is the sum (or a weighted sum) of the losses of the component problems, and if we suppose the losses to be a_γ for rejecting H_γ when $\theta \in \omega_\gamma$, b_γ for not rejecting it when $\theta \in \omega_\gamma^{-1}$, and zero in the other two cases, the total loss is again given by (2.2) and (2.3) with

$$(9.2) \quad x_{i\gamma} = 2y_{i\gamma} + 1.$$

Since in particular the loss function of the basic two-decision problem is unchanged, the associated optimum unbiased procedure of Part I will retain its optimum property in spite of the reinterpretation of one of the decisions. It now leads to the statements $\theta \in \omega_\gamma^{-1}$ and $\theta \in \omega_\gamma^0$ as $X \in A_\gamma^{-1}$ or $X \in A_\gamma$ where A_γ is the acceptance region of the best unbiased test of H_γ . The simultaneous carrying out of a number of these tests then still leads to a procedure in which the decision $d_i: \theta \in \Omega_i$ is taken when X falls in the set

$$(9.3) \quad D_i = \bigcap_\gamma A^{x_i}$$

but where Ω_i is now defined by (9.1) and (9.2) instead of (2.1).

As has already been pointed out the sets Ω_i , which define the possible decisions, no longer constitute a partition of the parameter space. Instead, they are generated through intersections from the class $\{\omega_\gamma^{-1}, \gamma \in \Gamma\}$, that is, they constitute the smallest class that is closed under intersections and contains the sets ω_γ^{-1} . It may happen that two of these Ω 's are equal, $\Omega_i = \Omega_j$, say, and one would then wish to identify the associated decisions. On the other hand, viewing the problem as a product one must consider all of the formal intersections (9.1) as distinct. Otherwise the definition of the loss function, for example, would become ambiguous since the losses resulting from decisions d_i and d_j when θ is in some Ω_h would usually not be the same even though $\Omega_i = \Omega_j$. Fortunately, the difficulty arises in the applications we wish to make only in cases in which it can be overcome by a natural further restriction of the decision space. Suppose namely that H_γ and H_δ are two hypotheses with

$$\omega_\gamma^{-1} \subset \omega_\delta^{-1}$$

so that the two intersections $\omega_\gamma^{-1} \cap \omega_\delta^{-1}$ and $\omega_\gamma^{-1} \cap \omega_\delta^0$ are identical. It then seems reasonable that whenever the data lead to the rejection of H_γ one would also wish to reject the more restrictive hypothesis H_δ . (In part I this was actually part of the compatibility requirement.) With such tests the decisions $\theta \in \omega_\gamma^{-1} \cap \omega_\delta^0$ would never be reached, and the conflict would thus be avoided. For this reason we shall eliminate from the list of permissible decisions not only the formal intersections

(9.1) that are empty, but also those for which the intersection Ω_i equals some other Ω_j satisfying $y_{j\gamma} \leq y_{i\gamma}$ for all γ . It follows from the general discussion of Sec. 7 that (9.3) defines a 1:1 correspondence between families of tests which are compatible with this set of restrictions, and decision procedures for the restricted product problem. It is useful to note further that for these restrictions the formation of restricted products of decision problems also retains the property of being commutative and associative, which it obviously had in the problems of Part I.

Let $\mathcal{P}_+(\theta_0)$ denote the two-decision problem discussed above, in which the choice lies between the statements $\theta > \theta_0$ and $-\infty < \theta < \infty$, and let $\mathcal{P}_-(\theta_0)$ denote the dual problem in which the first of these possibilities is replaced by the statement $\theta < \theta_0$. By considering these two problems simultaneously one obtains a three-decision problem with loss function

(9.4)

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

As an example suppose that θ measures the difference in quality of two products which are being compared by an impartial research organization. The decisions d_1 and d_2 claim superiority for one or the other of the products, while d_0 states that the data are inconclusive and that neither of the two products can be ascertained to be better than the other. It is an advantage of such a formulation over the more conventional one in which d_0 is replaced by the statement $\theta = \theta_0$, that it enables the statistician to control the probability of error. In the standard situation with D_2 and D_1 given by

(9.5) $D_2:T \leq C_1$ and $D_1:T \geq C'_1$

where $P_\theta\{T \leq C_1\}$ and $P_\theta\{T \geq C'_2\}$ are $< \alpha$ for $\theta > \theta_0$ and $\theta < \theta_0$ respectively, and where $P_{\theta_0}\{T \leq C_1\} = P_{\theta_0}\{T \geq C'_2\} = \alpha$, the maximum probability of error occurs when $\theta = \theta_0$, and is 2α . (A very similar formulation was discussed by Bahadur, "A property of the t -statistic," *Sankhya*, Vol. 12 (1952), pp. 79-88.)

The loss function (9.4) is not appropriate in situations in which a definite decision is preferred to d_0 even when $\theta = \theta_0$. A formulation which is more suitable for this case is obtained if in the one-sided problems $\mathcal{P}_-(\theta_0)$ and $\mathcal{P}_+(\theta_0)$ one replaces the decisions $\theta < \theta_0$ and $\theta > \theta_0$ by $\theta \leq \theta_0$ and $\theta \geq \theta_0$ respectively, so that these two component problems are given by

(9.6a)

	$-\infty < \theta < \infty$	$\theta \leq \theta_0$
$\theta > \theta_0$	0	a
$\theta \leq \theta_0$	b	0

and

$(b < a)$

(9.6b)

	$-\infty < \theta < \infty$	$\theta \geq \theta_0$
$\theta < \theta_0$	0	a
$\theta \geq \theta_0$	b	0

The simultaneous consideration of these two problems leads to a four-decision problem with loss table

(9.7)

	d_2 $\theta \leq \theta_0$	d_0 $-\infty < \theta < \infty$	d_1 $\theta \geq \theta_0$	d_3 $\theta = \theta_0$
$\theta < \theta_0$	0	b	a + b	a
$\theta = \theta_0$	b	2b	b	0
$\theta > \theta_0$	a + b	b	0	a

It turns out that this formulation leads to essentially the same solution as the previous one, with D_2 and D_1 given by (9.5), decision d_0 being taken when $C_1 < T < C'_1$, and decision d_3 not occurring at all.

We mention finally that still another problem leads to the same solution, namely that given by

(9.8)

	d_2 $\theta \leq \theta_0$	d_0 $-\infty < \theta < \infty$	d_1 $\theta \geq \theta_0$	$\left(\begin{matrix} b < a' \\ b < b' \end{matrix} \right)$
$\theta < \theta_0$	0	b	a'	
$\theta = \theta_0$	0	b'	0	
$\theta > \theta_0$	a'	b	0	

The level α of (9.5) is in this case given by

$$\alpha = \frac{b' - b}{a' + 2(b' - b)}$$

10. Partial classification of one or more parameters. (i) Let θ be a real parameter and suppose that we wish to determine, as far as possible, its position relative to two given values $\theta_1 < \theta_2$. A procedure may be generated by considering simultaneously the four problems $\mathcal{P}_\pm(\theta_i)$, $i = 1, 2$. The resulting problem offers the choice between the decisions

(10.1)

$$\begin{aligned} d_1: \{\theta < \theta_1\} &= \{\theta < \theta_1\} \cap \{\theta < \theta_2\} \\ d_2: \{\theta < \theta_2\} &= \{-\infty < \theta < \infty\} \cap \{\theta < \theta_2\} \\ d_3: \{\theta_1 < \theta < \theta_2\} &= \{\theta_1 < \theta\} \cap \{\theta < \theta_2\} \\ d_4: \{\theta > \theta_1\} &= \{\theta > \theta_1\} \cap \{-\infty < \theta < \infty\} \\ d_5: \{\theta > \theta_2\} &= \{\theta > \theta_2\} \cap \{\theta > \theta_1\} \\ d_6: \{-\infty < \theta < \infty\} &= \{-\infty < \theta < \infty\} \cap \{-\infty < -\theta < \infty\}. \end{aligned}$$

Here the sets $\{\theta < \theta_1\}$ and $\{\theta > \theta_2\}$ may also be represented as the intersections $\{\theta < \theta_1\} \cap \{-\infty < \theta < \infty\}$ and $\{\theta > \theta_2\} \cap \{-\infty < \theta < \infty\}$. However these decisions are ruled out by the convention of the previous section.

Suppose now that the tests of the four generating hypotheses have rejection regions

$$T \leq C_1, \quad T \leq C'_1, \quad T \geq C_2, \quad T \geq C'_2$$

for $H: \theta \leq \theta_1, \theta \geq \theta_1, \theta \leq \theta_2, \theta \geq \theta_2$ respectively, where the constants are determined by

$$P_{\theta_1}\{T \leq C_1\} = P_{\theta_1}\{T \geq C'_1\} = P_{\theta_2}\{T \leq C_2\} = P_{\theta_2}\{T \geq C'_2\} = \alpha.$$

Compatibility requires that the intersections

$$\begin{aligned} &\{T \leq C_i\} \cap \{T \geq C'_i\}, \quad \{T \leq C_1\} \cap \{C_2 < T < C'_2\} \\ &\text{and } \{C_1 < T < C'_1\} \cap \{T \geq C'_2\} \end{aligned}$$

should be empty, and hence that

$$C_1 < C'_1, \quad C_2 < C'_2, \quad C_1 < C_2 \text{ and } C'_1 < C'_2.$$

These conditions are satisfied if $\alpha < \frac{1}{2}$, and if for each fixed C

$$(10.2) \quad P_{\theta}\{T \leq C\} > P_{\theta'}\{T \leq C\} \quad \text{when } \theta < \theta'.$$

According to (9.3) the resulting procedure is given by

$$(10.3) \quad \begin{array}{ll} D_1: T \leq C_1 & D_4: \max(C'_1, C_2) \leq T < C'_2 \\ D_2: C_1 \leq T < \min(C'_1, C_2) & D_5: T \geq C'_2 \\ D_3: C'_1 \leq T \leq C_2 & D_6: C_2 < T < C'_1 \end{array}$$

where \leq is $<$ when $C'_1 < C_2$ and \leq when $C'_1 \geq C_2$. Depending on the sign of the difference $C_2 - C'_1$ and hence on the distance between θ_1 and θ_2 , only one of the decisions d_3 and d_6 will occur. For intermediate values of T the positive statement $\theta_1 < \theta < \theta_2$ will be made only if θ_1 and θ_2 are not too close. Otherwise such T -values will leave the position of θ in doubt.

If (10.3) holds, the probability of the procedure leading to a false statement never exceeds 2α . For the probability of error is equal to

$$\begin{aligned} P_{\theta}\{T \geq C'_1\} &\leq P_{\theta_1}\{T \geq C'_1\} = \alpha && \text{for } \theta < \theta_1 \\ P_{\theta_1}\{T \leq C_1\} + P_{\theta_1}\{T \geq C'_1\} &= 2\alpha && \text{for } \theta = \theta_1 \\ P_{\theta}\{T \leq C_1\} + P_{\theta}\{T \geq C_2\} &\leq P_{\theta_1}\{T \leq C_1\} + P_{\theta_2}\{T \geq C_2\} = 2\alpha && \text{for } \theta_1 < \theta < \theta_2, \end{aligned}$$

and similarly for $\theta \geq \theta_2$.

In the usual applications T is a function of a sample, and as the sample size increases T tends to θ in probability. The procedure is then consistent in the

sense that

$$P(D_i) \rightarrow 1 \quad \text{for } \theta \in \Omega_i \quad (i = 1, 2, 3).$$

Except when θ is exactly equal to θ_1 or θ_2 , the probability is therefore 1 that one of the three sharp statements d_1 , d_3 , or d_5 will be made, and that this statement will be correct. Since

$$P_{\theta_1}(D_2) = P_{\theta_2}(D_4) = 1 - 2\alpha,$$

it also follows that

$$P_{\theta_1}(D_2) = P_{\theta_2}(D_4) \rightarrow 1$$

if one lets α tend to zero as n tends to infinity.

A slightly different procedure results for problem (10.1) when one is concerned with a sample X_1, \dots, X_n from $N(\theta, \sigma^2)$. Here the tests of the four generating hypotheses have rejection regions

$$\begin{aligned} (\bar{X} - \theta_1)/S \leq -C, \quad (\bar{X} - \theta_1)/S \geq C, \\ (\bar{X} - \theta_2)/S \leq -C, \quad (\bar{X} - \theta_2)/S \geq C, \end{aligned}$$

and the induced multiple decision procedure is given by (10.3) with

$$(10.4) \quad T = \bar{X}/S, \quad C_i = -C + \frac{\theta_i}{S}, \quad C'_i = C + \frac{\theta_i}{S}.$$

Both of the decisions d_3 and d_6 occur in this procedure with positive probability, d_3 only in cases in which $S \leq (\theta_2 - \theta_1)/C$ and d_6 only when the opposite inequality holds. The remarks concerning error control and consistency require no change.

As an application consider the comparison of two normal populations $N(\xi, \sigma^2)$ and $N(\eta, \sigma^2)$ on the basis of samples X_1, \dots, X_m and Y_1, \dots, Y_n . With $\theta = \eta - \xi$, $\theta_1 = -\Delta$, $\theta_2 = \Delta$, the possible decisions become

$$\begin{aligned} d_1: \eta - \xi < -\Delta, \quad d_3: -\Delta < \eta - \xi < \Delta, \quad d_5: \eta - \xi > \Delta, \\ d_2: \eta - \xi < \Delta, \quad d_4: \eta - \xi > -\Delta, \quad d_6: -\infty < \eta - \xi < \infty. \end{aligned}$$

Here d_1 states that ξ is significantly larger than η , d_2 that η is not significantly larger than ξ , d_3 that the two means do not differ significantly, etc. The procedure is given by (10.3) with $T = (\bar{Y} - \bar{X})/S$ and

$$C_1 = -C - \frac{\Delta}{S}, \quad C_2 = -C + \frac{\Delta}{S}, \quad C'_1 = C - \frac{\Delta}{S}, \quad C'_2 = C + \frac{\Delta}{S}.$$

Problem (10.1) leads to still another type of procedure in the case of two independent Poisson variables, say X and Y , where one wishes to classify the ratio $\rho = \lambda/\mu$ of the parameters of the two distributions with respect to two values $\rho_1 < \rho_2$. Here the tests of the four generating hypotheses $\rho \leq \rho_i$, $\rho \geq \rho_i$ ($i = 1, 2$) are carried out conditionally, given the value of $X + Y$. The

conditional distribution of Y given $X + Y = m$ is the binomial distribution $b(p, m)$ corresponding to m trials and probability $p = \mu/(\lambda + \mu)$ of success. The conditional situation is therefore of the kind discussed at the beginning of this section, and leads to the procedure (10.3) with $T = Y$ and $\theta = p$.

Whether d_3 or d_6 occurs depends here on m , d_3 being associated with large values of m . To see this, let $F_{p,m}(t)$ denote the conditional cumulative distribution function, given $X + Y = m$, of $Y + U$ where U is independent of Y and is uniformly distributed on $(0, 1)$. Then $C'_1 = C'(m)$ and $C_2 = C_2(m)$ are determined by

$$(10.5) \quad F_{p_2,m}(C_2) = 1 - F_{p_1,m}(C'_1) = \alpha.$$

We shall show that $C'_1(m_1) < C_2(m_1)$ implies that also $C'_1(m_2) < C_2(m_2)$ for all $m_2 > m_1$. Since $F_{p_2,m}(C_2) = \alpha$, $F_{p_1,m}(C_2)$ is the power of the most powerful level α test for testing p_2 against p_1 in a binomial distribution $b(p, m)$. If $C'_1(m_1) < C_2(m_1)$ it follows from (10.9) that in the case of m trials this power is greater than $1 - \alpha$; but then it must also exceed $1 - \alpha$ for $m_2 > m_1$ trials so that

$$F_{p_1,m_2}(C'_1) = 1 - \alpha < F_{p_1,m_2}(C_2),$$

and hence $C'_1(m_2) < C_2(m_2)$.

It is interesting to note that in all of these problems the choice between decision d_3 and d_6 depends on the distance between θ_1 and θ_2 relative to the amount of information the data contain for the problem.

While the classification of θ with respect to a single value θ_0 , or two values $\theta_1 < \theta_2$, are the most interesting cases, let us consider briefly also the problem of classifying θ with respect to a countable set of values $\dots < \theta_{-2} < \theta_{-1} < \theta_0 < \theta_1 < \dots$. This is generated by the problems $\mathcal{P}_{\pm}(\theta_i)$, $i = 0, \pm 1, \pm 2, \dots$. Supposing that $\theta_i \rightarrow \pm \infty$ as $i \rightarrow \pm \infty$ and letting $\theta_{-\infty} = -\infty$, $\theta_{\infty} = \infty$, the possible decisions consist of the totality of statements $\theta_i < \theta < \theta_j$. The decision $\theta_i < \theta < \theta_j$ corresponds to the individual decisions that $\theta < \theta_k$ for $k \geq j$ and $\theta > \theta_k$ for $k \leq i$, and that the position of θ is left in doubt with respect to the points θ_k with $i < k < j$.

The limiting case of this problem, in which one wishes to classify θ with respect to all possible values of θ_0 , is obtained by considering simultaneously the problems $\mathcal{P}_{\pm}(\theta_0)$ for all θ_0 . The possible decisions then consist of the totality of statements $\underline{\theta} < \theta < \bar{\theta}$, and if $a_{\gamma} = a$, $b_{\gamma} = b$ for all γ , (9.3) yields precisely the standard confidence intervals for θ with confidence coefficient $1 - 2\alpha$. The loss function resulting from the additivity assumption is in the simplest case²

$$(10.6) \quad \begin{aligned} &(a + b)(\underline{\theta} - \theta) + b(\bar{\theta} - \underline{\theta}) && \text{if } \theta < \underline{\theta} \\ &b(\bar{\theta} - \underline{\theta}) && \text{if } \underline{\theta} < \theta < \bar{\theta} \\ &(a + b)(\theta - \bar{\theta}) + b(\bar{\theta} - \underline{\theta}) && \text{if } \theta > \bar{\theta}. \end{aligned}$$

² A loss function of this type was suggested by Wolfowitz [4].

More generally one may replace $\bar{\theta} - \underline{\theta}$ by $\int_{\underline{\theta}}^{\bar{\theta}} d\mu(\theta)$, and similarly for the lengths of the other intervals. Unfortunately the condition of unbiasedness becomes very difficult to interpret in the present problem, and it is doubtful that for the standard distributions the procedure is unbiased with uniformly minimum risk, as is the case with the other problems of this section.

It should be pointed out that all these procedures concerning a single real-valued parameter θ can be obtained from the standard confidence intervals concerning θ . When classifying θ with respect to θ_1 and θ_2 for example, one can state $\theta < \theta_1$ if $\bar{\theta} < \theta_1$, and $\theta_1 < \theta < \theta_2$ if $\theta_1 < \underline{\theta} < \bar{\theta} < \theta_2$. On the other hand, if $\underline{\theta} < \theta_1 < \bar{\theta} < \theta_2$, only the conclusion $\theta < \theta_2$ is possible, the relation of θ to θ_1 being left in doubt. This approach, however, does not yield any optimum properties for these procedures, and does in fact not carry over to problems involving more than one parameter.

In Example (iii) of Sec. 3 we considered the comparison of a number of normal populations $N(\theta_i, \sigma^2)$. The consistency difficulties that occurred in combining the decisions $\theta_i \leq \theta_j$ for different pairs (i, j) disappear if one treats the problem from the present point of view, which is exactly that from which the problem was treated by Duncan, "Multiple range and multiple F -tests," *Biometrics*, Vol. 11 (1955), pp. 1-42. For each pair (θ_i, θ_j) the possible decisions are now $\theta_i < \theta_j$, $\theta_i > \theta_j$, and $-\infty < \theta_j - \theta_i < \infty$ instead of the earlier $\theta_i = \theta_j$. Since there is no loss in omitting the vacuous statements, the total decisions consist in the ordering of some but not necessarily all of the pairs (θ_i, θ_j) . In the case of three populations, for example, the following decision types will occur:

- (a) $\theta_i < \theta_j < \theta_k$, (c) $\theta_i < \theta_j, \theta_i < \theta_k$
 (b) $\theta_i < \theta_k, \theta_j < \theta_k$, (d) $\theta_i < \theta_j$
 (e) no statement.

We shall now show that for this procedure the probability of error can be controlled through the choice of α , and that its maximum is in fact attained when all of the θ 's are equal and is then given by

$$(10.7) \quad P \left\{ \frac{|\bar{X}_j - \bar{X}_i|}{S} > C_{ij} \text{ for some } i, j \right\}.$$

In the particular case of equal sample sizes this becomes

$$(10.8) \quad P \left\{ \frac{\max |\bar{X}_j - \bar{X}_i|}{S} > C \right\}$$

where C is the cut-off point of the one-sided t -test at level α .³ The probability of error is the probability measure of the set

$$(10.9) \quad \cup \left\{ \frac{\bar{X}_j - \bar{X}_i}{S} > C_{ij} \right\} + \cup \left\{ \frac{\bar{X}_\ell - \bar{X}_k}{S} < C_{k\ell} \right\}$$

³ For a table of the values C for which this probability is 1 per cent or 5 per cent see [2], where also a number of related tables are discussed.

where the first union is extended over all pairs (i, j) for which $\theta_j \leq \theta_i$ and the second union over all pairs (k, ℓ) for which $\theta_\ell \geq \theta_k$. Let $X_i^* = X_i - \theta_i$, so that the X_i^* are distributed as $N(0, \sigma^2)$. Then $\theta_j \geq \theta_i$ and $\bar{X}_j - \bar{X}_i > C_{ij}S$ imply $X_j^* - X_i^* > C_{ij}S$, and $\theta_\ell \leq \theta_k$ and $\bar{X}_\ell - \bar{X}_k < C_{k\ell}S$ imply $X_\ell^* - X_k^* < C_{k\ell}S$. The probability of the set (10.9) is therefore not decreased if one replaces the X 's by X^* 's, nor if one extends the union over all pairs (i, j) and (k, ℓ) . But this is equivalent to evaluating the probability of (10.9) under the assumption that all of the θ 's are equal, which completes the proof.

11. Decision problems with simple loss functions. As a tool for proving the procedures of Secs. 9 and 10 to be unbiased with uniformly minimum risk, we shall now give an extension of Theorem 2 (Sec. 7), which is valid for a rather general class of decision problems. We shall say that a decision problem \mathcal{P} is *simple* if it satisfies the following two conditions.

(a) Its loss function $W(\theta, d)$, considered for fixed d as a function of θ , has sets of constancy independent of d , that is, there exists a partition Π of the parameter space Ω into sets $\Theta_i, i \in I$, such that $W(\theta, d)$ is independent of θ on each Θ_i . We may then write

$$(11.1) \quad W(\theta, d) = V_i(d) \quad \text{for } \theta \in \Theta_i .$$

(b) With respect to some convergence notion in $\Omega, \theta_n \rightarrow \theta_0$ implies $E_{\theta_n}\psi(X) \rightarrow E_{\theta_0}\psi(X)$ for each integrable ψ or, if all of the functions V_i are bounded, for each bounded ψ .

We shall require the following properties of simple decision problems.

(i) For any procedure δ the risk function $R_\delta(\theta)$ is continuous on each set of the partition Π . The risk function is given by

$$(11.2) \quad R_\delta(\theta) = E_\theta V_i[\delta(X)] \quad \text{for } \theta \in \Theta_i .$$

Hence $\theta_n, \theta_0 \in \Theta_i$ and $\theta_n \rightarrow \theta_0$ imply

$$R_\delta(\theta_n) = E_{\theta_n} V_i[\delta(X)] \rightarrow E_{\theta_0} V_i[\delta(X)] = R_\delta(\theta_0),$$

as was to be proved.

(ii) *Unbiasedness of a procedure δ implies the continuity of its risk function.* By (i) it is enough to prove this for boundary points of Π . Let θ_0 be such a boundary point, and suppose that $\theta_0 \in \Theta_i$ and $\theta_0 = \lim_{n \rightarrow \infty} \theta_n$ with $\theta_n \in \Theta_j$. Unbiasedness implies

$$\begin{aligned} E_{\theta_0} V_i[\delta(X)] &\leq E_{\theta_0} V_j[\delta(X)], \\ E_{\theta_n} V_i[\delta(X)] &\geq E_{\theta_n} V_j[\delta(X)] \quad \text{for } n = 1, 2, \dots \end{aligned}$$

and hence also

$$E_{\theta_0} V_i[\delta(X)] \geq E_{\theta_0} V_j[\delta(X)].$$

It follows that

$$R_\delta(\theta_0) = E_{\theta_0} V_i[\delta(X)] = E_{\theta_0} V_j[\delta(X)] = \lim_{n \rightarrow \infty} R_\delta(\theta_n).$$

(iii) *Any restricted product of simple decision procedures is again simple.*

We can now prove the main result of this section, which provides a sufficient condition for unbiasedness of a product procedure to imply continuity of the risk functions of all of the component procedures.

(iv) Let \mathcal{P} be a restricted product of a finite number of simple decision problems \mathcal{P}_γ , and suppose that the partitions Π_γ of the component problems into sets $\Theta_{i\gamma}$, $i \in I_\gamma$ satisfy the following condition.

(*) Let $\Theta_{i\gamma_0}$, $\Theta_{j\gamma_0}$ be any two sets of Π_{γ_0} with common boundary points, let θ_0 be any such point, and assume without loss of generality that $\theta_0 \in \Theta_{i\gamma_0}$. Then there exists a sequence of points $\theta_n \in \Theta_{j\gamma_0}$ such that $\theta_n \rightarrow \theta_0$ and

$$\theta_n \sim \theta_0(\Pi_\gamma) \quad \text{for all } \gamma \neq \gamma_0,$$

where $\theta \sim \theta'(\Pi)$ indicates that the two points lie in the same set of the partition.

Under these assumptions, if the risk function $R_\delta(\theta)$ of a product procedure δ is continuous, so are the risk functions $R_{\delta_\gamma}(\theta)$ of the components δ_γ of δ .

To see this, let γ_0 be any fixed value of γ , θ_0 any boundary point of Π_{γ_0} , and $\{\theta_n\}$ the sequence guaranteed by (*). Since $R_\delta(\theta) = \sum_\gamma R_{\delta_\gamma}(\theta)$ is continuous, we have

$$\sum_\gamma R_{\delta_\gamma}(\theta_n) \rightarrow \sum_\gamma R_{\delta_\gamma}(\theta_0).$$

Also, for each $\gamma \neq \gamma_0$ all of the points θ_n lie in the same set of the partition Π_γ with θ_0 , so that by (i)

$$R_{\delta_\gamma}(\theta_n) \rightarrow R_{\delta_\gamma}(\theta_0) \quad \text{for all } \gamma \neq \gamma_0.$$

It follows that $R(\delta_{\gamma_0}, \theta_n) \rightarrow R(\delta_{\gamma_0}, \theta_0)$, as was to be proved, where we have written $R(\delta, \theta)$ for $R_\delta(\theta)$.

It is convenient that (*) depends only on the partitions Π_γ , not on the values the loss functions W_γ take on over these partitions. For applications it is further important to note that (*) may be weakened slightly. Let $\Lambda_{i\gamma}$ be the set of common boundary points of $\Theta_{i\gamma}$ and $\Theta_{j\gamma}$ that belongs to $\Theta_{i\gamma}$. Then in order to ensure (iv) it is sufficient if (*) holds on a dense subset of each $\Lambda_{ij\gamma}$. This is an immediate consequence of (i).

Consider now any problem \mathcal{P} , which is a restricted product of a finite number of problems $\mathcal{P}_\pm(\theta_\gamma)$, and which satisfies (*). For the component problems continuity of the risk function is equivalent to similarity on the boundary at level $\alpha_\gamma = b_\gamma/(a_\gamma + b_\gamma)$. Hence a product procedure δ uniformly minimizes the risk among all unbiased procedures of \mathcal{P} provided each component procedure δ_γ uniformly minimizes the risk among all procedures of $\mathcal{P}_\pm(\theta_\gamma)$ that are similar on the boundary at level α_γ . Since $\mathcal{P}_\pm(\theta_\gamma)$ is formally equivalent to the problem of testing $\theta \leq \theta_\gamma$ or $\theta \geq \theta_\gamma$, this is the case in particular if the possible distributions of the observable random variables constitute an exponential family, and the procedures δ_γ are the best unbiased tests of the hypotheses in question. Under

these conditions it is then only necessary to verify (*) in order to establish the desired optimum property for the resulting δ .

As an example consider problem (9.4), which is generated by $\mathcal{P}_{\pm}(\theta_0)$. Here $\mathcal{P}_-(\theta_0)$ induces the partition $\theta \leq \theta_0, \theta > \theta_0$, with θ_0 as its only boundary point. Let θ_n be any sequence of points greater than and tending to θ_0 . Then all the points θ_n and θ_0 lie in the set $\theta \geq \theta_0$, and hence $\theta_n \sim \theta_0$ with respect to the partition induced by $\mathcal{P}_+(\theta_0)$, which completes the verification of (*). The argument is exactly the same for problem (9.6) and (10.1).

In the example of Sec. 10 leading to procedure (10.7), a typical partition is $\theta_2 \leq \theta_1, \theta_2 > \theta_1$. Attention may be restricted to boundary points $\theta^{(0)}$ with coordinates

$$\theta_{i_1}^0 < \dots < \theta_{i_r}^{(0)} < \theta_1^{(0)} = \theta_2^{(0)} < \theta_{j_1}^{(0)} < \dots < \theta_{j_{s-r-2}}^{(0)}.$$

For these, (*) is satisfied by the sequence of points $\theta^{(n)}$ with coordinates $\theta_i^{(n)} = \theta_i^{(0)}$ for $i \neq 2$, and $\theta_2^{(n)}$ between $\theta_1^{(0)}$ and $\theta_{j_1}^{(0)}$ and tending to $\theta_1^{(0)}$.

12. Consecutive decisions in a single experiment. The multiple decision problems treated in the previous sections were generated by the simultaneous consideration of a number of simpler component problems. We shall now suppose that these separate problems arise not in parallel but in sequence. A single sample is available for investigating a number of questions that are potentially of interest and are taken up one by one. Whether a given question is relevant, or which of a number of possible alternative formulations is appropriate at a certain stage, depends on the decisions reached up to that point.

(i) As an example suppose that independent variables X_1, \dots, X_n from a normal distribution $N(\xi, \sigma^2)$ are measurements on an experimental batch of a new product of quality ξ . The product is of no interest unless $\xi > \xi_0$, so that one will wish to test first of all the hypothesis $H_1: \xi \leq \xi_0$. If the quality is found satisfactory, that is, if H_1 is rejected, it becomes necessary to investigate the variability of the product. One will then test $H_2: \sigma \geq \sigma_0$, and in case this hypothesis is accepted one will try to reduce σ , for example by using less variable materials. The problem of testing H_2 arises here only in case H_1 is rejected.

(ii) Suppose that two treatments are being compared on a number of different categories of patients. Let the observed effect of treatment i ($i = 1, 2$) on the k th patient of the j th category be distributed as $N(\xi_{ij}, \sigma^2)$ where

$$\xi_{ij} = \eta + \lambda_i + \mu_j + \nu_{ij} \quad (\sum_i \lambda_i = \sum_j \beta_j = \sum_i \nu_{ij} = \sum_j \nu_{ij} = 0).$$

Here λ_i is the main effect of treatment i and ν_{ij} the interaction between the i th treatment and the j th category. One may believe in the possibility of the interactions being negligible and hence wish first to test the hypothesis $H_1: \nu_{ij} = 0$ for all i, j . In case H_1 is accepted the λ 's are the objects of primary interest, and the problem becomes that of deciding whether $\lambda_2 - \lambda_1$ is $<, =,$ or > 0 , or to estimate this difference either by confidence intervals or by a point estimate. On the other hand, if H_1 is rejected one will be concerned less with the over-all effects of the treatments which is measured by the λ 's than with the treatment differences $\xi_{2j} - \xi_{1j}$ for each category.

More generally, let there be given a first problem \mathcal{P}' , in which the possible decisions are d'_1, \dots, d'_m and the loss function is $W'(\theta, d)$. If decision d_i is taken, a second problem \mathcal{P}''_i arises with possible decisions $d''_{ij}(j = 1, \dots, n_i)$ and loss function $W''_i(\theta, d)$. The combination of these two problems leads to a two-stage problem with decisions $d_{ij} = (d'_i, d''_{ij})$. One may of course continue in this manner and suppose that decision (d'_i, d''_{ij}) gives rise to a further problem \mathcal{P}'''_{ijk} with decisions d'''_{ijk} . However, it is enough to treat the case of two levels since the discussion then extends immediately to the more general situation by induction.

In specifying a loss function we shall assume that even if a wrong decision is taken at the first step, so that the second problem is not the most appropriate one or perhaps need not have been considered at all, it is still desirable to do as well with respect to it as is possible. Thus in example (ii) above, if one has incorrectly decided that the interactions are negligible, one will in the estimation of $\theta_2 - \theta_1$ still wish to obtain as good an estimate as possible, and analogously if H_1 has wrongly been rejected. Whether the assumption holds in Example (i), that is, whether one would wish to control the variability of the new product after having mistakenly judged its quality to be satisfactory, appears to depend on the circumstances of the problem.

With this assumption, a natural loss function for the compound problem is

$$(12.1) \quad W(\theta, d_{ij}) = W'(\theta, d'_i) + W''_i(\theta, d''_{ij}).$$

The possibility of not considering a second problem in case a certain decision d'_i is taken at the first step, is included in this formulation. One need then only take as problem \mathcal{P}''_i the vacuous decision problem, that is, set $n_i = 1$ and $W''_i(\theta, d''_{i1}) = 0$. Suppose in particular, as was the case in Example (i), that a second problem occurs only for one of the decisions of the first stage, say d'_1 . The possible decisions of the compound problem are then

$$d_{ij} = (d'_1, d''_{1j}), \quad j = 1, \dots, n$$

$$\text{and} \quad d_2 = d'_2, \dots, d_m = d'_m,$$

and the loss function is given by

$$(12.2) \quad \begin{aligned} W(\theta, d_i) &= W'(\theta, d'_i), & i &= 2, \dots, m; \\ W(\theta, d_{1j}) &= W'(\theta, d'_1) + W''_1(\theta, d''_{1j}), & j &= 1, \dots, n. \end{aligned}$$

Returning now to the general case, suppose that there exist a satisfactory procedure δ' for \mathcal{P}' , which takes decision d'_i when $X \in D'_i$, and that the problems $\mathcal{P}''_1, \dots, \mathcal{P}''_m$ are all different. It then seems natural to retain δ' as first step of the compound procedure, and to consider the problems at the second level relative to the circumstances in which they occur, namely conditionally given that $X \in D'_1, \dots, X \in D'_m$ respectively. Suppose further that for each $i = 1, \dots, m$ there exists a satisfactory procedure δ''_i for \mathcal{P}''_i when the distribution of X is the conditional distribution given $X \in D'_i$. Such a procedure consists of a partition of the new sample space D'_i into regions D''_{ij} in which the decisions

d''_{ij} are taken. Together, the sets $D''_{ij}(j = 1, \dots, n_i; i = 1, \dots, m)$ form a partition of the original sample space and a solution of the compound problem, with decision $d_{ij} = (d'_i, d''_{ij})$ being taken when $X \in D''_{ij}$. One can of course again include in the formulation the possibility of ruling out some of the decisions d_{ij} , and the resulting compatibility questions can be treated exactly as before. However, this possibility seems to be less important in the present context and in order not to complicate the discussion unnecessarily we shall assume that no restrictions are imposed on the compound decision problem.

As an application consider the case that m and the n_i are equal to two, so that each of the component problems is one of hypothesis testing. Suppose that the hypotheses in question concern the parameters θ_i in an exponential family

$$dP_\theta(x) = C(\theta_1, \dots, \theta_r)e^{\sum \theta_i T_i(x)} d\nu(x).$$

There then exist uniformly most powerful unbiased tests of the hypotheses $\theta_i \leq \theta_i^0$ and more generally of $\sum c_i \theta_i \leq c_0$. (See, for example, [2].) Since after truncation on a fixed set D the family of distributions P_θ retains its property of forming an exponential family, such optimum unbiased tests will in particular also exist at the second stage after a preliminary test of significance has been performed as a first step. This will however not coincide with the standard optimum test for the corresponding problem without truncation.

We have assumed so far that the problems occurring at the second stage are all distinct. Suppose now instead that \mathcal{O}''_1 is appropriate when decisions d'_1, \dots, d'_{r_1} are taken in the first problem, that \mathcal{O}''_2 corresponds to decisions $d'_{r_1+1}, \dots, d'_{r_1+r_2}$, etc. One would then consider the problems $\mathcal{O}''_1, \mathcal{O}''_2, \dots$ conditionally given that $X \in D'_1 + \dots + D'_{r_1}, X \in D'_{r_1+1} + \dots + D'_{r_1+r_2}, \dots$, and otherwise proceed as before. If one is dealing in particular with the product of two decision problems so that all of the problems at the second level are the same, one considers this common problem \mathcal{O}'' given that $X \in D'_1 + \dots + D'_m$, that is, unconditionally. The procedure therefore reduces to the product of the procedures for \mathcal{O}' and \mathcal{O}'' , so that the present theory agrees with that given earlier for products of decision problems.

Unfortunately the properties of the conditional procedures considered in the present section are not as satisfactory as of those discussed in the earlier parts of this paper. To be specific, let \mathcal{O}' and $\mathcal{O}''_1, \dots, \mathcal{O}''_m$ define a two-stage problem, in which the components \mathcal{O}''_i of the second stage are distinct. The risk function of a procedure δ with components $\delta', \delta''_1, \dots, \delta''_m$ is then

$$(12.3) \quad R_\delta(\theta) = R_{\delta'}(\theta) + \sum_{i=1}^m P_\theta(D'_i) R_{\delta''_i|\delta'}(\theta)$$

where the notation $R_{\delta''_i|\delta'}$ is used to indicate that this risk component is computed conditionally given $X \in D'_i$ and hence depends on δ' as well as on δ''_i .

It is clear from (12.3) that unbiasedness of δ' and $\delta''_1, \dots, \delta''_m$ implies that of δ . Also it is again true for most problems of interest that unbiasedness of δ implies either unbiasedness or at least similarity on the boundary for $\delta', \delta''_1, \dots, \delta''_m$. However, the basic comparison of two procedures in terms of their

components is no longer simple. In particular,

$$R_{\delta'}(\theta) \leq R_{\psi'}(\theta), \quad R_{\delta_i''|\delta'}(\theta) \leq R_{\psi_i''|\psi'}(\theta)$$

does not in general imply $R_{\delta}(\theta) \leq R_{\psi}(\theta)$. If one wishes to minimize $R_{\delta}(\theta)$ then, given δ' , one must select δ_i'' to minimize $R_{\delta_i''|\delta'}(\theta)$. But the best choice of δ' is not necessarily that which minimizes $R_{\delta'}(\theta)$ since the choice of δ' influences not only $R_{\delta'}(\theta)$ but also the second components of (12.3) and in particular the conditional risks $R_{\delta_i''|\delta'}$. As a result of these complications it turns out that for the problem under consideration there usually does not exist among the unbiased procedures one that uniformly minimizes the risk. Of the procedures, the components of which have this optimum property we can only say that they are unbiased, and within the class of all unbiased procedures admissible.

13. Some examples of conditional procedures. Although we have found no satisfactory justification for the procedures discussed in the preceding section, they are rather natural from the Neyman-Pearson point of view, and we shall briefly illustrate them here with a few examples leaving a more detailed discussion and comparison with alternative procedures for a later paper.

(i) The problem mentioned at the beginning of Sec. 12 is concerned with testing, on the basis of a normal sample, the two hypotheses $H_1: \xi \leq \xi_0$ and $H_2: \sigma \geq \sigma_0$, where H_2 is assumed to be of interest only in case H_1 is rejected. If without loss of generality we put $\xi_0 = 0$, the best unbiased procedure for testing H_1 is Student's t -test with rejection region

$$(13.1) \quad \bar{X}/S \geq C$$

and size $\alpha_1 = b_1/(a_1 + b_1)$. With this as first step, the condition of unbiasedness implies in the usual way that the rejection region R_2 of H_2 must satisfy

$$(13.2) \quad P_{\sigma_0}(R_2 | S \leq \bar{x}/C | \bar{x}) \stackrel{(\bar{x})}{=} \alpha_2.$$

Applying the fundamental lemma of Neyman and Pearson one sees that the uniformly most powerful unbiased conditional test of H_2 has a rejection region of the form

$$(13.3) \quad S^2 < f(\bar{X}).$$

Here the function f is defined by

$$(13.4) \quad \int_0^{f(u)} p_{\sigma_0}(z) dz = \alpha_2 \int_0^{u^2/C^2} p_{\sigma_0}(z) dz, \quad u > 0,$$

where p_{σ_0} is the probability density of S^2 when $\sigma = \sigma_0$.

The resulting compound procedure then decides between the three possible conclusions

- $d_1: \xi \leq \xi_0$ —the new product is not of satisfactory quality,
- $d_2: \xi > \xi_0, \sigma \geq \sigma_0$ —the quality of the new product is satisfactory but its variability must be reduced,
- $d_3: \xi > \xi_0, \sigma < \sigma_0$ —the new product is satisfactory with regard to both quality and variability,

as the sample falls into the corresponding one of the regions

$$\begin{aligned}
 D_1: \bar{X} \mid S &\leq C, \\
 D_2: \bar{X} \mid S &> C, \quad S^2 \geq f(\bar{X}) \\
 D_3: \bar{X} \mid S &> C, \quad S^2 < f(\bar{X}).
 \end{aligned}$$

These decision regions are illustrated for the case $n = 10, \alpha_1 = \alpha_2 = .05$ in the figure.

(ii) As a somewhat more complex example consider two treatments which are being compared on a number of different categories of patients. Let the observed effect Y_{ijk} of treatment i ($i = 1, 2$) on the k th patient ($k = 1, \dots, n$) in the j th category be distributed as $N(\xi_{ij}, \sigma^2)$, and let

$$\xi_{ij} = \eta + \lambda_i + \mu_j + \nu_{ij} \quad (\sum_i \lambda_i = \sum_j \mu_j = \sum_i \nu_{ij} = \sum_j \nu_{ij} = 0)$$

where λ_i is the main effect of treatment i and ν_{ij} the interaction between the i th treatment and j th category. One may here wish to test first the hypothesis of no interaction

$$H_1: \nu_{ij} = 0 \quad \text{for all } i, j.$$

If H_1 is accepted one will be interested in the difference of the λ 's and wish either to test it or alternatively to estimate it by confidence intervals or point estimate. We shall here suppose that we then want to test

$$H_2: \lambda_2 - \lambda_1 \leq 0.$$

On the other hand, if H_1 is rejected one will usually be concerned less with comparing the over-all effects of the two treatments, which is measured by the λ 's, than with a comparison of the treatment effects ξ_{2j} and ξ_{1j} separately for each category j . In particular one may be interested to test the set of hypotheses

$$H_{3j}: \xi_{2j} - \xi_{1j} = (\lambda_2 + \nu_{2j}) - (\lambda_1 + \nu_{1j}) \leq 0.$$

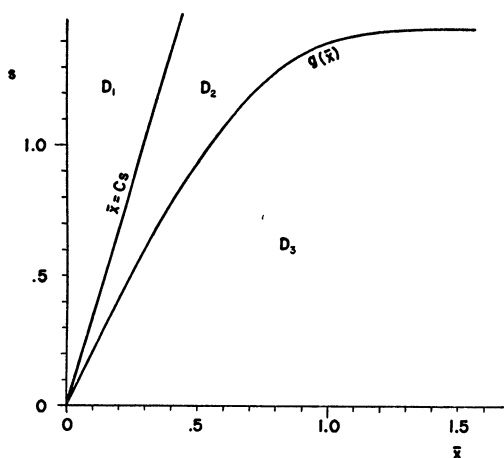


FIG. 1

Although there exists no uniformly most powerful unbiased test of H_1 it seems natural to start out with the standard test of this hypothesis which is uniformly most powerful invariant, and is given by the acceptance region

$$(13.5) \quad S_1^2/S_0^2 \leq C$$

where

$$S_0^2 = \sum \sum \sum (Y_{ijk} - Y_{ij.})^2$$

$$S_1^2 = \sum \sum (Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...})^2.$$

The hypothesis H_2 should then be considered conditionally, given $S_1^2/S_0^2 \leq C$. A routine application of the theory of unbiased tests and similar regions⁴ shows that a necessary condition for the rejection region R_2 to be unbiased is

$$P_{\lambda_2=\lambda_1} \{R_2 \mid s_1^2, s_0^2 + s_2^2, S_1^2/S_0^2 \leq C\} = \alpha_2$$

for all values of s_1^2 and $s_0^2 + s_2^2$ where

$$S_2^2 = \sum_{i=1}^2 (Y_{i..} - Y_{...})^2 = \frac{1}{2}(Y_{2..} - Y_{1..})^2.$$

If one puts

$$U = (Y_{2..} - Y_{1..})/S_0, \quad V = S_0^2 + S_2^2, \quad W = (S_0^2 + S_2^2)/S_1^2,$$

it follows from the fundamental lemma that the uniformly most powerful (conditional) test of H_2 has the rejection region

$$U \geq k(V, W)$$

where k is determined by

$$P_{\lambda_1=\lambda_2} \{U \geq k(V, W) \mid V = v, W = w \text{ and } U^2 \leq Cw - 1\} \stackrel{(v,w)}{=} \alpha_2.$$

Now when $\lambda_1 = \lambda_2$, the variable U is independent of V so that k depends only on w , $k(v, w) = f(w)$ say. The rejection region then becomes

$$(13.6) \quad U \geq f(w),$$

where f is determined by

$$P_{\lambda_1=\lambda_2} \{U < f(w) \mid U^2 \leq Cw - 1\} \stackrel{(w)}{=} (1 - \alpha_2).$$

Since acceptance of H_1 is equivalent to $u^2 \leq Cw - 1$ it implies in particular that $0 \leq Cw - 1$. The defining condition for f may therefore be written as

$$(13.7) \quad \int_{-(Cw-1)^{1/2}}^{f(w)} p_U(u) du = (1 - \alpha_2) \int_{-(Cw-1)^{1/2}}^{(Cw-1)^{1/2}} p_U(u) du$$

where $p_U(u)$ is the probability density of U when $\lambda_1 = \lambda_2$, that is, essentially, the density of a t -distribution with $2m(n - 1)$ degrees of freedom.

⁴ The proof requires the easily shown fact that the family of noncentral χ^2 -distributions with a fixed number of degrees of freedom is boundedly complete.

Let us now consider in a similar manner the hypothesis

$$H_{31}: (\lambda_2 - \lambda_1) + (\nu_{21} - \nu_{11}) \leq 0,$$

conditionally given that $S_1^2/S_0^2 > C$. Let $\hat{\lambda}_i$ and $\hat{\nu}_{ij}$ be the least squares estimates of the corresponding parameters so that

$$S_1^2 = \sum_{i=1}^2 \sum_{j=1}^m \hat{\nu}_{ij}^2$$

and let

$$\begin{aligned} Z_1 &= \frac{1}{2}(\hat{\lambda}_2 - \hat{\lambda}_1 - \hat{\nu}_{21} + \hat{\nu}_{11}) \\ Z_2 &= \frac{1}{2}(\hat{\lambda}_2 - \hat{\lambda}_1 + \hat{\nu}_{21} - \hat{\nu}_{11}) \\ S_1'^2 &= S_1^2 - (\hat{\nu}_{21} - \hat{\nu}_{11})^2 = S_1^2 - (Z_2 - Z_1)^2. \end{aligned}$$

If $\zeta_i = E(Z_i)$, the hypothesis becomes $\zeta_2 \leq 0$, and unbiasedness of the conditional test of H_{31} with rejection region R_{31} implies

$$P_{\zeta_2=0}\{R_{31} \mid z_1, s_0^2 + z_2^2, s_1'^2 \text{ and } S_1^2/S_0^2 > C\} \equiv \alpha_{31}.$$

The condition $S_1^2 > CS_0^2$ is equivalent to $(Z_2 - Z_1)^2 > CS_0^2 - S_1'^2$, which may be rewritten as

$$\left(Z_2 - \frac{Z_1}{1+C} \right)^2 + \frac{CZ_1^2}{(1+C)^2} > \frac{C}{1+C} T^2 - S_1'^2,$$

where $T^2 = S_0^2 + Z_2^2$. This is satisfied for all values of Z_2 if

$$(13.8) \quad \frac{C}{1+C} T^2 - S_1'^2 - \frac{CZ_1^2}{(1+C)^2} \leq 0$$

and otherwise for the values of Z_2 , for which either

$$\frac{Z_2}{T} > \frac{Z_1}{(1+C)T} + \sqrt{\frac{C}{1+C} - \frac{S_1'^2}{T^2} - \frac{C}{(1+C)^2} \frac{Z_1^2}{T^2}} = K_2 \left(\frac{Z_1}{T}, \frac{S_1'}{T} \right)$$

or

$$\frac{Z_2}{T} < \frac{Z_1}{(1+C)T} - \sqrt{\frac{C}{1+C} - \frac{S_1'^2}{T^2} - \frac{C}{(1+C)^2} \frac{Z_1^2}{T^2}} = K_1 \left(\frac{Z_1}{T}, \frac{S_1'}{T} \right).$$

Since Z_2/T is independent of Z_1/T and S_1'/T when $\zeta_2 = 0$, the uniformly most powerful unbiased test of H_{31} given $S_1^2/S_0^2 > C$ is then given by the rejection region

$$\frac{Z_2}{T} \geq K \left(\frac{Z_1}{T}, \frac{S_1'}{T} \right)$$

with the function K defined as follows. When $(Z_1/t, S_1'/t)$ satisfies (13.8), we have

$$K \left(\frac{z_1}{t}, \frac{s_1'}{t} \right) = K$$

where

$$\int_K^1 p_R(r) dr = \alpha_{31},$$

$p_R(r)$ being the probability density of

$$R = \frac{Z_2}{T} = \frac{\frac{1}{2}(\hat{\lambda}_2 - \hat{\lambda}_1 - \hat{\nu}_{21} + \hat{\nu}_{11})}{\sqrt{S_0^2 + \frac{1}{4}(\hat{\lambda}_2 - \hat{\lambda}_1 + \hat{\nu}_{21} - \hat{\nu}_{11})^2}}$$

when $\zeta_2 = 0$. For all other values of $(z_1/t, s'_1/t)$, $K(z_1/t, s'_1/t)$ is given by

$$\int_{K((z_1/t), (s'_1/t))}^1 p_R(r) dr = \alpha_{31} \left[\int_{-1}^{K_1(z_1/t), (s'_1/t)} p_R(r) dr + \int_{K_2((z_2/t), (s'_1/t))}^1 p_R(r) dr \right].$$

(iii) As an example involving more than two stages let us consider the determination of the degree of a regression polynomial. Let Y_1, \dots, Y_n be independently normally distributed with constant variance σ^2 and means

$$\eta_i = E(Y_i) = e_s + e_{s-1}x_i + \dots + e_0x_i^s \quad (i = 1, \dots, n).$$

We shall assume that a polynomial of degree s will in any case be adequate for our purposes, and wish to determine the smallest degree $r \leq s$ that would also be satisfactory. It is convenient for this purpose to express the regression polynomial in terms of the orthogonal polynomials P_i defined by

$$\sum_{k=1}^n P_i(x_k)P_j(x_k) = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases}$$

Writing

$$\eta_i = c_s + c_{s-1}P_1(x_i) + \dots + c_0P_s(x_i),$$

we test successively the hypotheses $H_0:c_0 = 0, H_1:c_1 = 0, \dots$, continuing as long as no rejection occurs.

The problem can be stated in the following canonical form: X_1, \dots, X_s and S_0^2 are independent variables, with X_i being distributed as $N(\xi_i, \sigma^2)$ and with S_0^2/σ^2 having a χ^2 -distribution with $n_0 = n - s$ degrees of freedom. One wishes to test consecutively the hypotheses $H_i:\xi_i = 0, (i = 1, \dots, s)$ continuing as long as no rejection occurs. Slightly more generally one might have variables $X_{ij}:N(\xi_{ij}, \sigma^2), (j = 1, \dots, n_i; i = 1, \dots, s)$ and S_0^2 and consider consecutively the hypotheses $H_i:\xi_{ij} = 0$ for $j = 1, \dots, n_i$. Invariance reduces the problem to the statistics S_0^2 and $S_i^2 = \sum_{j=1}^{n_i} X_{ij}^2$, which are independent and where S_i^2/σ^2 has a χ^2 distribution with n_i degrees of freedom when H_i is true.

Let

$$U_i = \frac{S_i^2}{S_0^2}, \quad V_i = \frac{S_i^2}{S_{i-1}^2}, \quad W_i = V_i \left(1 + \frac{1}{U_i} \right) = \frac{S_i^2 + S_0^2}{S_{i-1}^2}.$$

We shall now show inductively that the best unbiased conditional test of H_i , given the acceptance of H_1, \dots, H_{i-1} , has an acceptance region of the form

$$(13.9) \quad U_i \leq f_i(W_i, V_{i-1}, \dots, V_2).$$

The functions f_i are defined by

$$(13.10) \quad \begin{cases} f_1(w_1) = C, \\ \int_0^{f_i(w_i, v_{i-1}, \dots, v_2)} p_i(r) dr = (1 - \alpha_i) \int_0^{w_i h_{i-1}(v_{i-1}, \dots, v_2) - 1} p_i(r) dr \quad (i \geq 2) \end{cases}$$

where p_i is the probability density of S_i^2/S_0^2 when H_i is true, h_i is defined by

$$(13.11) \quad \begin{cases} h_1 = C, \\ h_i(v_i, \dots, v_2) = \min[v_i h_{i-1}(v_{i-1}, \dots, v_2), g_i(v_i, \dots, v_2)] \quad (i \geq 2) \end{cases}$$

and g_i is a function taking on values in the space of u_i and defined by

$$(13.12) \quad u = g_i(v_i, \dots, v_2) \leftrightarrow v_i = \frac{f_i^{-1}(u, v_{i-1}, \dots, v_2)}{1 + \frac{1}{u}} \quad (i \geq 2)$$

where f_i^{-1} denotes the inverse of f_i considered for fixed values of v_{i-1}, \dots, v_2 as a function of u . It is seen that (13.10), (13.11) and (13.12) define f_i in terms of f_{i-1} .

The best unbiased test of H_1 has the acceptance region $u_1 \leq C$ and hence satisfies (13.9). Suppose now that the acceptance regions A_1, \dots, A_{i-1} have been shown to be given by (13.9). To prove (13.9) for A_i we need the fact that

$$(13.13) \quad A_1 \cap \dots \cap A_{i-1} = \{(u, v) : u_{i-1} \leq h_{i-1}(v_{i-1}, \dots, v_2)\}$$

and since this is true for $i = 2$, we may accept (13.13) as part of our induction hypothesis. The condition of unbiasedness implies that A_i should satisfy

$$P_{H_i}\{A_i \mid s_0^2 + s_1^2, s_1^2, \dots, s_{i-1}^2 \text{ and } U_{i-1} \leq h_{i-1}(V_{i-1}, \dots, V_2)\} = 1 - \alpha_i.$$

Now $U_{i-1} \leq h_{i-1}$ is equivalent to

$$U_i \leq W_i h_{i-1}(V_{i-1}, \dots, V_2) - 1$$

so that the best unbiased acceptance region for H_i is

$$U_i \leq K_i(W_i; S_1^2, \dots, S_{i-1}^2)$$

where

$$P_{H_i}\{U_i \leq K_i \mid s_0^2 + s_1^2, s_1^2, \dots, s_{i-1}^2 \text{ and } U_i \leq W_i h_{i-1}(V_{i-1}, \dots, V_2) - 1\} = 1 - \alpha_i.$$

Since U_i is independent of W_i, V_{i-1}, \dots, V_2 when H_i is true, it is seen that K_i depends only on w_i, v_{i-1}, \dots, v_2 and that A_i is given by (13.9).

To complete the proof, it only remains to verify (13.13) for the set

$A_1 \cap \dots \cap A_i$, which is the intersection of

$$A_1 \cap \dots \cap A_{i-1}: u_{i-1} \leq h_{i-1}(v_{i-1}, \dots, v_2)$$

and

$$A_i: u_i \leq f_i(w_i, v_{i-1}, \dots, v_2).$$

The inequality describing A_i is equivalent to

$$f_i^{-1}(u_i, v_{i-1}, \dots, v_2) \leq w_i = v_i \left(1 + \frac{1}{u_i} \right)$$

and hence to

$$A_i: u_i \leq g_i(v_i, \dots, v_2)$$

by (13.12). Since the inequality describing $A_1 \cap \dots \cap A_{i-1}$ is equivalent to

$$A_1 \cap \dots \cap A_{i-1}: u_i \leq v_i h_{i-1}(v_{i-1}, \dots, v_2),$$

the intersection $A_1 \cap \dots \cap A_i$ is given by (13.13), as was to be proved.

A closely related problem arises in the study of components of variance when one is dealing with a hierarchical classification. Here the problem reduces to independent statistics S_1^2, S_2^2, \dots where S_i^2/σ_i^2 has a χ^2 -distribution with n_i degrees of freedom. It follows from the underlying model that $\sigma_1 \leq \sigma_2 \leq \dots$, and one is interested in testing successively the hypotheses $H_1: \sigma_2 = \sigma_1, H_2: \sigma_3 = \sigma_2, \dots$, continuing only if all the previous hypotheses were accepted. If H_1, \dots, H_i are true, the distribution of S_1^2, \dots, S_i^2 is the same as in the preceding problem, and it is easily seen that exactly the same procedure is applicable to the present situation.

14. Minimizing the maximum risk. For problems of the kind discussed in the preceding section unbiasedness is closely related to the minimax property. We begin by considering the two-decision problem of testing a hypothesis $H: \theta \in \omega$. Let d_0 and d_1 denote the decisions of accepting and rejecting H , and let the losses of false rejection and acceptance be a and b respectively. Then we have

(i) A necessary condition for a procedure δ to be minimax is that it be unbiased.

(ii) This condition is also sufficient if the probability $P_\theta(A)$ of any set A is continuous in θ and if the common boundary of ω and ω^{-1} is nonempty.

To see (i) note that unbiasedness of a procedure δ_0 implies

$$(14.1) \quad P_\theta(d_1) \leq \frac{b}{a+b} \text{ for } \theta \in \omega, \quad P_\theta(d_0) \leq \frac{a}{a+b} \text{ for } \theta \in \omega^{-1}.$$

Since the risk function of any procedure δ is given by

$$(14.2) \quad R_\delta(\theta) = \begin{cases} aP_\theta(d_1) & \text{for } \theta \in \omega \\ bP_\theta(d_0) & \text{for } \theta \in \omega^{-1}, \end{cases}$$

δ_0 satisfies

$$(14.3) \quad \sup R_{\delta_0}(\theta) \leq \frac{ab}{a + b}.$$

Any minimax procedure δ_1 must then also satisfy (14.3) and hence by (14.2) must also be unbiased. Suppose next that the assumptions of (ii) hold, let θ_0 be a common boundary point of ω and ω^{-1} , and let δ_0 be unbiased. Then (14.1) and $P_{\theta}(d_0) = 1 - P_{\theta}(d_1)$ show that

$$P_{\theta_0}(d_1) = \frac{b}{a + b} \text{ and hence } R_{\delta_0}(\theta_0) = ab/(a + b).$$

Thus for any unbiased procedure δ_0 we have

$$(14.4) \quad \sup R_{\delta_0}(\theta) = \frac{ab}{a + b}.$$

By (i), this relation holds in particular for any minimax procedure, so that $ab/(a + b)$ is the minimax risk and (ii) follows from (14.4).

We shall now extend (ii) to the case of a number of hypotheses $H_i: \xi_i(\theta) \in \omega_i$ ($i = 1, \dots, s$) to be tested in sequence, where each hypothesis H_i is tested only if a particular prescribed chain of decisions has been reached on H_1, \dots, H_{i-1} . For these problems we have

THEOREM 3. *Every unbiased procedure minimizes the maximum risk provided (i) $P_{\theta}(A)$ is a continuous function of θ for each A , (ii) the 2^s intersections $\bigcap_{i=1}^s \omega_i^{x_i}$, (each $x_i = \pm 1$), are all nonempty, (iii) there exists at least one parameter point θ_0 which lies on the common boundary of ω_i and ω_i^{-1} for all i , (iv) the losses a_i and b_i for falsely rejecting and accepting H_i satisfy condition (14.10) below.*

PROOF. Let us write a_i^{-1} for b_i , let d_i and \bar{d}_i^{-1} denote the decisions of rejecting and accepting H_i , and let $y_i = +1$ if H_{i+1} is considered in case H_i is rejected, and $y_i = -1$ in the contrary case. If $\theta \in \bigcap_{i=1}^s \omega_i^{x_i}$, the risk of a procedure δ is then given by

$$\begin{aligned} R_{\delta}(\theta) &= a_1^{x_1} P(\bar{d}_1^{x_1}) + P(\bar{d}_1^{y_1}) \{ a_2^{x_2} P(\bar{d}_2^{x_2} | \bar{d}_1^{y_1}) + P(\bar{d}_2^{y_2} | \bar{d}_1^{y_1}) \\ &\quad [a_3^{x_3} P(\bar{d}_3^{x_3} | \bar{d}_1^{y_1} \bar{d}_2^{y_2} + \dots) \} \\ &= a_1^{x_1} P(\bar{d}_1^{x_1}) + a_2^{x_2} P(\bar{d}_1^{y_1}) P(\bar{d}_2^{x_2} | \bar{d}_1^{y_1}) \\ &\quad + a_3^{x_3} P(\bar{d}_1^{y_1}) P(\bar{d}_2^{y_2} | \bar{d}_1^{y_1}) P(\bar{d}_3^{x_3} | \bar{d}_1^{y_1} \bar{d}_2^{y_2}) + \dots \end{aligned}$$

By comparing each of these expressions for which $x_i = 1$ with the corresponding one for which $x_i = -1$ but all the other x 's are the same, it is seen that unbiasedness implies and hence is equivalent to the conditions

$$\begin{aligned} P_{\theta}(d_i | \bar{d}_1^{y_1} \dots \bar{d}_{i-1}^{y_{i-1}}) &\leq \frac{b_i}{a_i + b_i} \text{ for } \theta \in \omega_i \\ P_{\theta}(\bar{d}_i^{-1} | \bar{d}_1^{y_1} \dots \bar{d}_{i-1}^{y_{i-1}}) &\leq \frac{a_i}{a_i + b_i} \text{ for } \theta \in \omega_i^{-1} \end{aligned}$$

for $i = 1, \dots, s$. Putting $r_i = a_i b_i / (a_i + b_i)$ we see that the risk function of any unbiased procedure satisfies

$$\begin{aligned} \sup R_\delta(\theta) \leq r_1 + \frac{a_1^{y_1}}{a_1 + b_1} r_2 + \frac{a_1^{y_1}}{a_1 + b_1} \frac{a_2^{y_2}}{a_2 + b_2} r_3 \\ + \dots + \frac{a_1^{y_1}}{a_1 + b_1} \dots \frac{a_{s-1}^{y_{s-1}}}{a_{s-1} + b_{s-1}} r_s = r^*. \end{aligned}$$

It now remains to show that r^* is the minimax risk.

Consider any procedure δ , and let

$$\alpha_i = P_{\theta_0}(d_i \mid d_1^{y_1} \dots d_{i-1}^{y_{i-1}}).$$

Then it follows from assumptions (i), (ii) and (iii) that

$$(14.6) \quad \sup R_\delta(\theta) \geq \max_{x_1, \dots, x_s} [(a_1 \alpha_1)^{x_1} + \alpha_1^{y_1} (a_2 \alpha_2)^{x_2} + \dots + \alpha_1^{y_1} \dots \alpha_{s-1}^{y_{s-1}} (a_s \alpha_s)^{x_s}]$$

since the 2^s expressions in brackets are the possible values of $\lim R_\delta(\theta)$ as θ tends to θ_0 . Let us now minimize the right-hand side of (14.6). Given any $\alpha_1, \dots, \alpha_{s-1}, x_1, \dots, x_{s-1}$, the maximum of the pair of expressions

$$\begin{aligned} (a_1 \alpha_1)^{x_1} + \dots + \alpha_1^{y_1} \dots \alpha_{s-2}^{y_{s-2}} (a_{s-1} \alpha_{s-1})^{x_{s-1}} + \alpha_1^{y_1} \dots \alpha_{s-1}^{y_{s-1}} (a_s \alpha_s), \\ (a_1 \alpha_1)^{x_1} + \dots + \alpha_1^{y_1} \dots \alpha_{s-2}^{y_{s-2}} (a_{s-1} \alpha_{s-1})^{x_{s-1}} + \alpha_1^{y_1} \dots \alpha_{s-1}^{y_{s-1}} (a_s \alpha_s)^{-1}, \end{aligned}$$

is clearly minimized by minimizing

$$\max[a_s \alpha_s, a_s^{-1} (1 - \alpha_s)].$$

This gives $\alpha_s = \alpha_s^*$ where we put

$$(14.7) \quad \alpha_i^* = \frac{a_i^{-1}}{a_i + a_i^{-1}} = \frac{b_i}{a_i + b_i},$$

so that

$$(14.8) \quad a_i \alpha_i^* = (a_i \alpha_i^*)^{-1}.$$

We can now proceed inductively. Suppose that it has already been shown for any fixed $\alpha_1, \dots, \alpha_i$ that the right-hand side of (14.6) is minimized by $\alpha_j = \alpha_j^*$ for $j = i + 1, \dots, s$. Consider now the minimization of the maximum of the quantities

$$\begin{aligned} (a_1 \alpha_1)^{x_1} + \dots + \alpha_1^{y_1} \dots \alpha_{i-1}^{y_{i-1}} (\alpha_i \alpha_i) + \alpha_1^{y_1} \dots \alpha_i^{y_i} [(a_{i+1} \alpha_{i+1}^*)^{x_{i+1}} \\ + (\alpha_{i+1}^*)^{y_{i+1}} (a_{i+1} \alpha_{i+2}^*)^{x_{i+2}} + \dots], \\ (a_1 \alpha_1)^{x_1} + \dots + \alpha_1^{y_1} \dots \alpha_{i-1}^{y_{i-1}} (a_i \alpha_i)^{-1} + \alpha_1^{y_1} \dots \alpha_i^{y_i} \\ [(a_{i+1} \alpha_{i+1}^*)^{x_{i+1}} + (\alpha_{i+1}^*)^{y_{i+1}} (a_{i+2} \alpha_{i+2}^*)^{y_{i+2}} + \dots], \end{aligned}$$

where the expression in brackets by (14.8) is independent of the x 's and hence equal to

$$(14.9) \quad k_i = a_{i+1}\alpha_{i+1}^* + (\alpha_{i+1}^*)^{y_{i+1}}a_{i+2}\alpha_{i+2}^* + \dots + (\alpha_{i+1}^*)^{y_{i+1}} \dots (\alpha_{s-1}^*)^{y_{s-1}}a_s\alpha_s^*.$$

Eliminating the common terms at the beginning, and the common factor $\alpha_1^{y_1} \dots \alpha_{i-1}^{y_{i-1}}$ of the terms involving α_i , we see that it is enough to minimize

$$\max[a_i\alpha_i + k_i\alpha_i^{y_i}, b_i\alpha_i^{-1} + k_i\alpha_i^{y_i}].$$

This is achieved by equating the two quantities, which gives $\alpha_i = \alpha_i^*$, provided the coefficient of α_i in $b_i + (k_i - b_i)\alpha_i$ is < 0 in case $y_i = 1$, or that the coefficient of α_i in $a_i\alpha_i + k_i(1 - \alpha_i)$ is > 0 in case $y_i = -1$. Thus the right-hand side of (14.6) is minimized by putting $\alpha_i = \alpha_i^*$ for all i , provided

$$(14.10) \quad a_{i+1}\alpha_{i+1}^* + (\alpha_{i+1}^*)^{y_{i+1}}a_{i+2}\alpha_{i+2}^* + \dots + (\alpha_{i+1}^*)^{y_{i+1}} \dots (\alpha_{s-1}^*)^{y_{s-1}}a_s\alpha_s^* < \bar{a}_i^{-y_i}.$$

Putting $\alpha_i = \alpha_i^*$ in the right-hand side of (14.6) one then sees that the 2^s quantities in brackets become equal, and that their common value is r^* . Thus, for any procedure δ , we have

$$\sup R_\delta(\theta) \geq r^*$$

and the desired result now follows from comparison with (14.5).

COROLLARY. *The conclusion of Theorem 3 holds if (14.10) is replaced by either*

$$(14.11) \quad y_i = 1 \text{ for all } i \quad \text{and} \quad b_1 \geq b_2 \geq \dots \geq b_s$$

or

$$(14.12) \quad y_i = -1 \text{ for all } i \quad \text{and} \quad a_1 \geq a_2 \geq \dots \geq a_s.$$

PROOF. It is necessary only to show that each of the conditions (14.11) and (14.12) implies (14.10). If all the y 's are $+1$ the left-hand side of (14.10) may be rewritten as

$$b_{i+1}(\alpha_{i+1}^*)^{-1} + b_{i+2}(\alpha_{i+2}^*)^{-1}\alpha_{i+1}^* + \dots + b_s(\alpha_s^*)^{-1}\alpha_{i+1}^* \dots \alpha_{s-1}^*.$$

If (14.11) holds, this is

$$\begin{aligned} &\leq b_{i+1}[(\alpha_{i+1}^*)^{-1} + \alpha_{i+1}^*(\alpha_{i+2}^*)^{-1} + \dots + \alpha_{i+1}^* \dots \alpha_{s-1}^*(\alpha_s^*)^{-1}] \\ &= b_{i+1}[1 - \alpha_{i+1}^*\alpha_{i+2}^* \dots \alpha_s^*] < b_{i+1} \leq b_i. \end{aligned}$$

Similarly, if (14.12) holds, the left-hand side of (14.10) is

$$\begin{aligned} &\leq a_{i+1}[\alpha_{i+1}^* + (\alpha_{i+1}^*)^{-1}\alpha_{i+2}^* + \dots + (\alpha_{i+1}^* \dots \alpha_{s-1}^*)^{-1}\alpha_s^*] \\ &= a_{i+1}[1 - (\alpha_{i+1}^* \dots \alpha_s^*)^{-1}] < a_i. \end{aligned}$$

If the assumptions of Theorem 3 are not satisfied it is sometimes possible to prove a slightly weaker result. We shall illustrate this with the simplest case of Example (ii) of Sec. 12. Here one is concerned with four means ξ_{ij} ($i, j = 1, 2$) given by

$$\begin{aligned} \xi_{11} &= \lambda + \mu + \nu, & \xi_{12} &= \lambda - \mu - \nu \\ \xi_{21} &= -\lambda + \mu - \nu, & \xi_{22} &= -\lambda - \mu - \nu. \end{aligned}$$

The first hypothesis tested is $H_0: \nu = 0$. In case of acceptance this is followed by $H_1: \lambda \leq 0$, while in case of rejection one is interested in the two hypotheses $H_2: \xi_{11} \leq \xi_{21}$ which is equivalent to $\lambda + \nu \leq 0$ or $H_3: \xi_{12} \leq \xi_{22}$ which is equivalent to $\lambda - \nu \leq 0$. The feature which complicates this problem is that when H_0 is true the three remaining hypotheses must either all be true or all be false. Thus the condition corresponding to (ii) of Theorem 3 is not satisfied.

Let us denote the decision of rejecting or accepting H_i by d_i and d_i^{-1} as before, and consider the class \mathfrak{C}_0 of procedures satisfying

$$(14.13) \left\{ \begin{array}{l} P(d_0) \leq \frac{b_0}{a_0 + b_0} \quad \text{if } \nu = 0, \quad P(d_0^{-1}) \leq \frac{a_0}{a_0 + b_0} \quad \text{if } \nu \neq 0 \\ P(d_1 | d_0^{-1}) \leq \frac{b_1}{a_1 + b_1} \quad \text{if } \lambda + \nu \leq 0, \\ P(d_1^{-1} | d_0^{-1}) \leq \frac{a_1}{a_1 + b_1} \quad \text{if } \lambda + \nu > 0 \\ P(d_i | d_0) \leq \frac{b_i}{a_i + b_i} \quad \text{if } H_i \text{ is true,} \\ P(d_i^{-1} | d_0) \leq \frac{a_i}{a_i + b_i} \quad \text{if } H_i \text{ is false } (i = 2, 3), \end{array} \right.$$

We shall then prove that any element of \mathfrak{C}_0 , which is a subclass of the class of unbiased procedures, is minimax, under the additional assumption that

$$a_i = a \quad \text{and} \quad b_i = b \quad \text{for all } i.$$

For any procedure δ , consider the error probabilities $\alpha_0 = P(d_0), \alpha_1 = P(d_1 | d_0^{-1}), \alpha_i = P(d_i | d_0) (i = 2, 3)$, evaluated for $\lambda = \nu = 0$ and some fixed values μ_0 and σ_0 of μ and σ . Then the possible value of $\lim R_\delta(\theta)$ as $\theta = (\lambda, \mu, \nu, \sigma)$ tends to $\theta_0 = (0, \mu_0, 0, \sigma_0)$ are

$$(14.14) \begin{array}{ll} a_1(1 - \alpha_0)\alpha_1 + \alpha_0[a_0 + a_2\alpha_2 + a_3\alpha_3] & \text{for } \theta \in \omega_0\omega_1\omega_2\omega_3 \\ b_1(1 - \alpha_0)(1 - \alpha_1) + \alpha_0[a_0 + b_2(1 - \alpha_2) + b_3(1 - \alpha_3)] & \text{for } \theta \in \omega_0\omega_1^{-1}\omega_2\omega_3^{-1} \\ (1 - \alpha_0)[b_0 + a_1\alpha_1] + \alpha_0[a_2\alpha_2 + a_3\alpha_3] & \text{for } \theta \in \omega_0^{-1}\omega_1\omega_2\omega_3 \\ (1 - \alpha_0)[b_0 + a_1\alpha_1] + \alpha_0[a_2\alpha_2 + b_3(1 - \alpha_3)] & \text{for } \theta \in \omega_0^{-1}\omega_1\omega_2\omega_3^{-1} \\ (1 - \alpha_0)[b_0 + a_1\alpha_1] + \alpha_0[b_2(1 - \alpha_2) + a_3\alpha_3] & \text{for } \theta \in \omega_0^{-1}\omega_1\omega_2^{-1}\omega_3 \\ (1 - \alpha_0)[b_0 + a_1\alpha_1] + \alpha_0[b_2(1 - \alpha_2) + b_3(1 - \alpha_3)] & \text{for } \theta \in \omega_0^{-1}\omega_1\omega_2^{-1}\omega_3^{-1} \\ (1 - \alpha_0)[b_0 + b_1(1 - \alpha_1)] + \alpha_0[a_2\alpha_2 + a_3\alpha_3] & \text{for } \theta \in \omega_0^{-1}\omega_1^{-1}\omega_2\omega_3 \end{array}$$

$$\begin{aligned}
 &(1 - \alpha_0)[b_0 + b_1(1 - \alpha_1)] + \alpha_0[a_2\alpha_2 + b_3(1 - \alpha_3)] && \text{for } \theta \in \omega_0^{-1}\omega_1^{-1}\omega_2^{-1}\omega_3^{-1} \\
 &(1 - \alpha_0)[b_0 + b_1(1 - \alpha_1)] + \alpha_0[b_2(1 - \alpha_2) + a_3\alpha_3] && \text{for } \theta \in \omega_0^{-1}\omega_1^{-1}\omega_2^{-1}\omega_3 \\
 &(1 - \alpha_0)[b_0 + b_1(1 - \alpha_1)] + \alpha_0[b_2(1 - \alpha_2) + b_3(1 - \alpha_3)] && \text{for } \theta \in \omega_0^{-1}\omega_1^{-1}\omega_2^{-1}\omega_3^{-1}.
 \end{aligned}$$

With $u = \alpha_0, v = (1 - \alpha_0)\alpha_1, w = \alpha_0\alpha_2, z = \alpha_0\alpha_3$ and $a_i = a, b_i = b$ these quantities become

$$\begin{aligned}
 &a[u + v + w + z] \\
 &(a + b)u - bv - bw - bz + b \\
 &-bu + av + aw + az + b \\
 &\quad aw + aw - bz + b \\
 (14.15) \quad &\quad aw - bw + az + b \\
 &\quad bu + av - bw - bz + b \\
 &-2bu - bv + aw + az + 2b \\
 &-bu - bv + aw - bz + 2b \\
 &-bu - bv - bw + az + 2b \\
 &-2bu - bv - bw - bz + 2b.
 \end{aligned}$$

From the first form of these 10 quantities it is seen that they all become equal for $\alpha_i = \alpha_i^* = b_i / (a_i + b_i)$. Let the corresponding values of (u, v, w, z) be (u_0, v_0, w_0, z_0) . We shall now show that if we change from (u_0, v_0, w_0, z_0) to some other point (u_1, v_1, w_1, z_1) at least one of the 10 quantities will be increased so that (u_0, v_0, w_0, z_0) and hence $\alpha_i = \alpha_i^*$ minimizes their maximum. If in (14.15) all possible sign combinations were occurring, this would be obvious. For then in at least one row all of the four increments $\pm(u_1 - u_0), \pm(v_1 - v_0), \pm(w_1 - w_0), \pm(z_1 - z_0)$ would be positive. But of the 16 possible combinations only 12 occur (with rows 4 and 5 each counting for two combinations, the missing combinations being $- + - -, + - + +, + - + -, + - - +$).

As an example let us consider the first possibility. Suppose that $u_1 = u_0 - \xi, v_1 = v_0 + \eta, w_1 = w_0 - \Delta, z_1 = z_0 - \zeta$, with ξ, η, Δ, ζ all nonnegative. The total change in the first and tenth rows of (14.15) will be

$$a(-\xi + \eta - \Delta - \zeta) \quad \text{and} \quad b(2\xi - \eta + \Delta + \zeta).$$

Since both of these are to be ≤ 0 , we have

$$2\xi + \Delta + \zeta \leq \eta \leq \xi + \Delta + \zeta$$

and hence $\xi = 0$. But then the change in the sixth row will be positive unless also $\eta = \Delta = \zeta = 0$. The other three possibilities can be ruled out in a similar manner so that $\alpha_i = \alpha_i^*$ minimizes the maximum of the 10 quantities in question. If r^* denotes the common value $\alpha_0^*[a_0 + a_1\alpha_1^* + a_2\alpha_2^* + a_3\alpha_3^*]$ of these quantities, it only remains to show that (14.13) implies $R_\delta(\theta) \leq r^*$ for all θ . This is clearly the case since for each i the coefficient of α_i in (14.14) is ≥ 0 in the lines corresponding to $\theta \in \omega_i$ and ≤ 0 otherwise.

The result that the natural combination of a number of best unbiased or minimax tests leads to a minimax procedure for the compound problem, does unfortunately not hold even in the simplest cases in which the different hypotheses concern the same parameter. Consider for example problem (i) of section 3, in which the simultaneous consideration of $H_1: \theta \geq \theta_0$ and $H_2: \theta \leq \theta_0$ leads to the choice between the three decisions $d_2: \theta < \theta_0$, $d_0: \theta = \theta_0$ and $d_1: \theta > \theta_0$. If the losses of false rejection and acceptance are a and b for both hypotheses the risk function is given by

$$R_\delta(\theta) = \begin{cases} b P_\theta(d_0) + (a + b)P_\theta(d_1) & \text{for } \theta < \theta_0 \\ a[P_\theta(d_1) + P_\theta(d_2)] & \text{for } \theta = \theta_0 \\ bP_\theta(d_0) + (a + b)P_\theta(d_2) & \text{for } \theta > \theta_0. \end{cases}$$

If $\alpha_i = P_{\theta_0}(d_i)$ we have

$$(14.16) \quad \sup R_\delta(\theta) \geq \max[a\alpha_1 - b\alpha_2 + b, a(\alpha_1 + \alpha_2), a\alpha_2 - b\alpha_1 + b].$$

For a given value of $\alpha_1 + \alpha_2$, the maximum of the first and third terms on the right-hand side is minimized by equating them which gives $\alpha_1 = \alpha_2$ and reduces the right-hand side of (14.16) to

$$\max[(a - b)\alpha + b, 2a\alpha].$$

In the usual case that $a > b$, this is minimized for $\alpha = 0$, that is, for $\alpha_0 = 1$, $\alpha_1 = \alpha_2 = 0$. In the standard case that the family P_θ is homogeneous this implies $P_\theta(d_0) = 1$ for all θ , the risk function of which is

$$R_\delta(\theta) = \begin{cases} b & \text{if } \theta \neq \theta_0 \\ 0 & \text{if } \theta = \theta_0 \end{cases}$$

which is then clearly minimax.

REFERENCES

- [1] E. L. LEHMANN, "A theory of some multiple decision problems. I.," *Am. Math. Stat.*, Vol. 28 (1957), pp. 1-25.
- [2] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation," *Sankhyā*, Vol. 10 (1950), pp. 305-339.
- [3] JOYCE M. MAY, "Extended and corrected tables of the upper percentage points of the Studentized range," *Biometrika*, Vol. 39 (1952), pp. 192-193.
- [4] J. WOLFOWITZ, "Minimax estimates of the mean of a normal distribution with known variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 218-230.