ON TESTING MORE THAN ONE HYPOTHESIS

By J. N. Darroch[1] and S. D. Silvey

University of Manchester

1. Introduction. In two illuminating papers, Lehmann [5], [6] considered some of the problems associated with testing more than one hypothesis within the framework of multiple decision theory and his discussion provides an ideal setting for describing the subject matter of the present paper.

Suppose we wish to test, against a general model G, two hypotheses H_1 and H_2 concerning the distribution underlying certain observed data. This problem induces a multiple decision problem in which the four possible decisions are

$$d^{00}: \quad H_1 \text{ and } H_2 \text{ are both true,}$$
$$d^{01}: \quad H_1 \text{ is true, } H_2 \text{ is false,}$$
$$d^{10}: \quad H_1 \text{ is false, } H_2 \text{ is true,}$$

and

$$d^{11}: \quad H_1 \text{ and } H_2 \text{ are both false.}$$

Now suppose that a test of H_ν is defined by the acceptance region A_ν^0 and the rejection region A_ν^1 $(\nu = 1, 2)$. These separate tests induce a decision procedure for the four-decision problem, this induced procedure being defined by assigning to the decision d^{ij} the region $A_1^i \cap A_2^j$.

This raises the problem: given that the separate tests of H_1 and H_2 are "good", when is the induced procedure also "good"? Lehmann proves that this is so when the loss function for the induced decision problem is the sum of the loss functions for the individual tests and when "good" is interpreted as "having uniformly minimum risk within a wide class of procedures".

While this is a powerful theoretical result there are several reasons why the statistician finds it of little practical value. First he seldom visualises losses explicitly and so finds it difficult to determine whether this additive condition is satisfied for the problem he is facing. Secondly he is often interested not in the complete multiple decision procedure induced by separate tests of the two hypotheses but only in the induced test of the hypothesis $H_1 \wedge H_2$, i.e., in the test with acceptance region $A_1^0 \cap A_2^0$. And thirdly he is not so much concerned that this induced test be optimal as that it should be a reasonably good test.

In the present paper it is the induced test of $H_1 \wedge H_2$ that will be our concern. We shall limit discussion to the case where the separate tests of H_1 and H_2 are likelihood ratio tests and we shall compare the test induced by these with a likelihood ratio test of $H_1 \wedge H_2$ against G. We shall assume that the latter test has the shape of power function that we require of a test of $H_1 \wedge H_2$. If the induced test compares reasonably well with this direct likelihood ratio test, we shall say that it is good: otherwise, that it is poor. We do not however wish to imply it is always desirable that the induced test have this property.

Received July 9, 1962; revised December 27, 1962.

[1] Now at the University of Adelaide.

While many of the points we shall make are sufficiently well exhibited by the case of two hypotheses, there are some which require consideration of three. Again in this case we shall not be concerned with the full induced decision procedure. But now there are four induced tests which may be of interest, those of $H_1 \wedge H_2$, $H_2 \wedge H_3$, $H_3 \wedge H_1$ and $H_1 \wedge H_2 \wedge H_3$: and our problem is to determine when all or some of these are good. Our primary interest, then, is quite different from that of Lehmann [5], [6]. Where he was concerned with theoretical aspects of the induced decision procedure, possibly for an infinite family of hypotheses, we shall be concerned with the essentially practical problem of using data to test a relatively small number of hypotheses which are of interest in combination as well as individually.

2. Size and power of the induced tests. Let α denote the size of the test of $H_1 \wedge H_2$ induced by likelihood ratio tests against G of H_1 and H_2, of sizes α_1 and α_2 respectively. Then it is easily seen that $\alpha \leqq \alpha_1 + \alpha_2$. Moreover, if these separate tests are similar, so also is the induced test of $H_1 \wedge H_2$ and $\alpha \geqq \max\{\alpha_1, \alpha_2\}$. Lastly if the component tests are similar and their test statistics are independent, then $\alpha = 1 - (1 - \alpha_1)(1 - \alpha_2)$. Whether or not this last equality obtains, the preceding inequalities often enable us to pin down α to within a reasonable interval, and the first of them always ensures that the size of the induced test is not too large. Hence we may say that the size of the induced test gives no cause for concern.

It is the power of the induced test that may be questionable. Consider the familiar example of linear regression on two concomitant variables x_1 and x_2 say, H_1 and H_2 being the hypotheses that the two regression coefficients β_1 and β_2 are zero. It is well known that it is "dangerous" to test H_1 and H_2 separately in case x_1 and x_2 are highly correlated. The reason for this is that the induced test of $H_1 \wedge H_2 : \beta_1 = \beta_2 = 0$ has low power. We can illustrate this diagrammatically for two general linear hypotheses, with θ denoting the vector of expected values of the observations, Ω a linear space specified by G and ω_1 and ω_2 subspaces of Ω specified respectively by H_1 and H_2. Suppose these subspaces are such that there exist points of Ω near both ω_1 and ω_2 but distant from $\omega_1 \cap \omega_2$ and suppose that the true θ is such a point, (Figure 1). It is apparent that, with high probability, a likelihood ratio test of $H_1 \wedge H_2$ against G would result in its rejection, whereas the induced test would result in its acceptance. In other words the induced test has low power at this θ, compared with that of a likelihood ratio test of $H_1 \wedge H_2$ against G of comparable size.

We shall now give this notion a more general expression. Let $y = (y_1, y_2, \cdots, y_n)$ denote a set of observations and suppose that their distribution has probability density $f(y, \theta)$, where $\theta = (\theta_1, \theta_2, \cdots, \theta_s)$ is an unknown parameter in R^s. A general model G states that $\theta \varepsilon \Omega$, a subset of R^s, and a hypothesis H that $\theta \varepsilon \omega \subset \Omega$. We shall write

$$L_y(H) = \sup_{\theta\varepsilon\omega} f(y, \theta)/\sup_{\theta\varepsilon\Omega} f(y, \theta),$$

an unconventional notation which, however, proves useful in the ensuing dis-

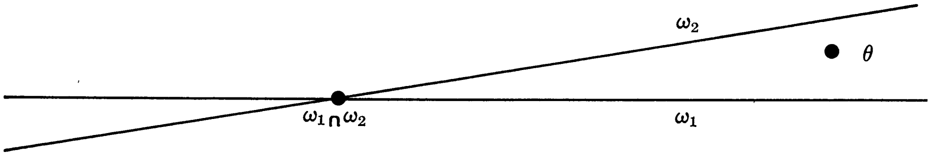

FIG. 1

cussion. Further, if H_1 and H_2 state respectively that $\theta \varepsilon \omega_1$ and ω_2, we shall write

$$L_y(H_2 \mid H_1) = \sup_{\theta \varepsilon \omega_1 \cap \omega_2} f(y, \theta)/\sup_{\theta \varepsilon \omega_1} f(y, \theta) = L_y(H_1 \wedge H_2)/L_y(H_1).$$

The random variable $L(H)$ is the likelihood-ratio statistic for testing H against G, and $L(H_2 \mid H_1)$ that for testing $H_1 \wedge H_2$ against H_1.

Now we can state in general terms what Figure 1 is really illustrating. *Separate tests of H_1 and H_2 may induce a poor test of $H_1 \wedge H_2$ because it is possible that, for some θ, with high probability, $L(H_1)$ and $L(H_2)$ are both "near 1" while $L(H_1 \wedge H_2)$ is small.*

EXAMPLE 1. In a 2×2 contingency table with probabilities $\{\theta_{ij}\}$ $(i, j = 1, 2; \sum_{i,j} \theta_{ij} = 1)$, consider the hypotheses

$$H_1 : \theta_{1\cdot} = \theta_{2\cdot} = \tfrac{1}{2}; \qquad H_2 : \theta_{\cdot 1} = \theta_{\cdot 2} = \tfrac{1}{2},$$

where $\theta_{i\cdot} = \theta_{i1} + \theta_{i2}$ and $\theta_{\cdot j} = \theta_{1j} + \theta_{2j}$. Suppose that the following results are observed:

482	33	515
3	482	485
485	515	1,000

We find that $L_y(H_1) = L_y(H_2) = e^{-0.450}$, while $L_y(H_1 \wedge H_2) = e^{-14.177}$. Now if H_1 is true, $-2 \log L(H_1)$ is distributed approximately as χ_1^2 and the observed value of this random variable is 0.900. Thus H_1 is a "most acceptable" hypothesis when considered by itself. So also is H_2. Hence separate tests of H_1 and H_2 of conventional size induce acceptance of $H_1 \wedge H_2$. However, if $H_1 \wedge H_2$ is true, $-2 \log L(H_1 \wedge H_2)$ is distributed approximately as χ_2^2 and its observed value is 28.354. Thus $H_1 \wedge H_2$ is certainly not acceptable on the basis of a likelihood ratio test, even of very conservative size.

This example illustrates exactly the same possibility as is represented diagrammatically for the case of linear hypotheses in Figure 1. The true θ is, in a likelihood ratio sense, "near" both ω_1 and ω_2 but not "near" $\omega_1 \cap \omega_2$.

We can now see what distinguishes situations in which the induced test has reasonable power from those where it does not. It is possible for $L(H_1 \wedge H_2)$ to be much nearer 0 than min $[L(H_1), L(H_2)]$. If however these random variables are jointly distributed in such a way that "proximity to 1" of both $L(H_1)$ and $L(H_2)$ implies that $L(H_1 \wedge H_2)$ is also "near 1", then the induced test of $H_1 \wedge H_2$

cannot fail in respect of power as it does in the cases just considered; and it is therefore a reasonable test.

These arguments regarding size and power extend to the case of three hypotheses in the following way. So far as the induced tests of $H_1 \wedge H_2$, $H_2 \wedge H_3$ and $H_3 \wedge H_1$ are concerned, there is nothing to add. As regards the size α of the test of $H_1 \wedge H_2 \wedge H_3$ induced by separate tests of sizes α_1, α_2 and α_3 respectively, we may say

(i) $\alpha \leqq \alpha_1 + \alpha_2 + \alpha_3$;

(ii) if the separate tests are similar, $\alpha \geqq \max\{\alpha_1, \alpha_2, \alpha_3\}$;

(iii) if the separate tests are similar and the statistics $L(H_1)$, $L(H_2)$ and $L(H_3)$ are independent, then $\alpha = 1 - (1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_3)$.

The power of the induced test of $H_1 \wedge H_2 \wedge H_3$ will be satisfactory if $L(H_1 \wedge H_2 \wedge H_3)$ cannot be small when $L(H_1)$, $L(H_2)$ and $L(H_3)$ are *all* near 1. This is clearly a requirement which is quite distinct from the requirements concerning the pairs (H_1, H_2), (H_2, H_3) and (H_3, H_1). However, it is basically of the same nature and introduces no essentially new difficulties.

It is apparent that the problem we are considering does not admit an exact solution. On the other hand we may look for situations where exact relationships among the test statistics involved ensure that induced tests are reasonably good. Such exact relationships arise naturally from consideration of nested tests, and we now discuss these in order to motivate the introduction of a relationship of this type.

3. Nested tests. The nested method of constructing tests for two hypotheses H_1 and H_2 is appropriate when we are interested in one of them, H_2 say, only if we have previously accepted the other (Scheffé [8], Lehmann [6]). Using likelihood ratio tests we would proceed as follows. First, test H_1 against G using the statistic $L(H_1)$. Then, if H_1 is accepted, test $H_1 \wedge H_2$ against H_1 using the statistic $L(H_2 \mid H_1)$. This method has recently been considered by Hogg [3] whose primary concern is with the independence of the statistics $L(H_1)$ and $L(H_2 \mid H_1)$; and, for the case of linear hypotheses, by Anderson [2] who describes a set of circumstances in which it is an optimal procedure.

These two tests induce a test of $H_1 \wedge H_2$ against G and it is fairly obvious that this induced test is reasonably good. We can demonstrate this in an economical fashion as follows. First, if the sizes of the component tests are α_1 and α_2 then the remarks concerning the size α of the induced test made at the beginning of Section 2 apply with equal force here and, in addition, $\alpha > \alpha_2$. Further, the power of this induced test is rendered satisfactory by the relationship

$$L(H_1 \wedge H_2) = L(H_1)L(H_2 \mid H_1),$$

which ensures that proximity to 1 of the component test statistics implies $L(H_1 \wedge H_2)$ also near 1. Referring back to Example 1, we note that $-2 \log L(H_2 \mid H_1) = 27.454$, so that the nested method applied to that example would lead to firm rejection of $H_1 \wedge H_2$, a result which verifies the advantage of the test of $H_1 \wedge H_2$ induced by this method over that induced by separate tests of H_1 and H_2. We note, however, that the nested method fails to provide a satisfactory test of H_2 against G, since, as in this case, it may result in firm rejection of H_2, when, as the test of H_2 against G shows, this is a perfectly acceptable hypothesis. Thus nested tests by themselves do not provide a solution to our problem. But they do provide a natural means of identifying certain situations where the test of $H_1 \wedge H_2$ induced by separate tests is satisfactory. For if

$$L(H_2) = L(H_2 \mid H_1) \tag{3.1}$$

this test will share the desirable properties of the test induced by the nested tests. More specifically, (3.1) means that

$$L(H_1 \wedge H_2) = L(H_1)L(H_2) \tag{3.2}$$

and this is just the kind of relationship we are seeking which ensures that $L(H_1)$ and $L(H_2)$ cannot both be near 1 with high probability when $L(H_1 \wedge H_2)$ is small.

When we apply the nested method to three hypotheses there are six possible nesting orders. With the order H_1, H_2, H_3 we test H_1 against G; then if H_1 is accepted, we test $H_1 \wedge H_2$ against H_1; and finally if H_1 and H_2 are accepted we test $H_1 \wedge H_2 \wedge H_3$ against $H_1 \wedge H_2$. As before these tests induce a good test of $H_1 \wedge H_2$ against G; and they also induce a good test of $H_1 \wedge H_2 \wedge H_3$ because of the relation

$$L(H_1 \wedge H_2 \wedge H_3) = L(H_1)L(H_2 \mid H_1)L(H_3 \mid H_1 \wedge H_2),$$

which ensures that $L(H_1 \wedge H_2 \wedge H_3)$ cannot be small when the statistics of the component tests are all near 1. Again the test of $H_1 \wedge H_2 \wedge H_3$ induced by separate tests will share this desirable feature if $L(H_2 \mid H_1) = L(H_2)$ and $L(H_3 \mid H_1 \wedge H_2) = L(H_3)$, so that $L(H_1 \wedge H_2 \wedge H_3) = L(H_1)L(H_2)L(H_3)$.

In the next section we shall investigate the nature of hypotheses which satisfy the kind of relationship we have just introduced.

4. Independent hypotheses. If, for two hypotheses, the Condition (3.1) or the equivalent Condition (3.2) is satisfied we shall say that the hypotheses are *independent* because of the obvious analogy with independent events. The immediate question that arises is: do such hypotheses occur in practice? In fact they do, as we now demonstrate.

Suppose that H_1 and H_2 are hypotheses about essentially different parameters. More precisely, suppose that

(i) $\theta = (\phi_1, \phi_2)$, where $\phi_1 = (\theta_1, \theta_2, \cdots, \theta_{s_1})$ and $\phi_2 = (\theta_{s_1+1}, \theta_{s_1+2}, \cdots, \theta_{s_1+s_2})$ where $s_1 + s_2 = s$;

(ii) ω_i is defined by restrictions on ϕ_i only $(i = 1, 2)$;

(iii) $f(y, \theta)$ can be expressed in the form $f(y, \theta) = g_1(y, \phi_1)g_2(y, \phi_2)$, where

g_1 and g_2 depend only on the variables indicated. Then it is not difficult to verify that H_1 and H_2 are independent.

The most obvious practical situations in which independence of two hypotheses is achieved *directly* by the mechanism just described are those where the observations from two independent experiments depending on totally different parameters are combined to form the set y, and the two hypotheses are about the different parameters. Such situations have two important features: (i) the hypotheses are independent: (ii) the statistics used for testing them, $L(H_1)$ and $L(H_2)$ also are independent. However these are quite distinct concepts and so far as the nature of the induced test of $H_1 \wedge H_2$ is concerned it is the first which is of primary importance.

There are situations where this condition alone is satisfied and, moreover, where the independence of the hypotheses is disguised by the fact that the parameters which are naturally introduced to describe underlying distributions are not those which demonstrate explicitly how this independence is achieved. For this reason we state as a theorem the following result whose proof is straightforward.

THEOREM 1. *If there exists a 1-1 transformation from the parameter space* Ω *to an alternative parameter space* Ω' *such that*

(i) Ω' *is a Cartesian product space* $\Omega_1' \times \Omega_2'$, *where* $\Omega_\nu' = \{\theta_\nu'\}$ $(\nu = 1, 2)$;

(ii) H_ν *imposes restrictions on* θ_ν' *only* $(\nu = 1, 2)$;

(iii) $f(y, \theta)$ *can be expressed in the form*

$$f(y, \theta) = g_1(y, \theta_1')g_2(y, \theta_2'),$$

then H_1 *and* H_2 *are independent hypotheses.*

What this theorem says really amounts to this: if H_1 and H_2 are hypotheses about different parameters and if the observations yield separate discriminatory information about these parameters, then H_1 and H_2 are independent. Put this way the theorem demonstrates the intuitive content of the word independent. In practice the authors have found that when hypotheses are independent (this being discovered by direct calculation of the appropriate statistics), it is possible to find a transformation of the parameter space satisfying the conditions of the theorem: and they are convinced that the theorem does demonstrate the structure of independence of hypotheses in most practical situations, that in fact the stated conditions are almost necessary as well as being sufficient.

EXAMPLE 2. In a 2-factor contingency table with probabilities $\{\theta_{ij}\}$ $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s; \sum_{i,j} \theta_{ij} = 1)$ and marginal probabilities $\theta_{i.} = \sum_j \theta_{ij}$, $\theta_{.j} = \sum_i \theta_{ij}$, consider the three hypotheses:

H_1 : the factors are independent, i.e., $\theta_{ij} = \theta_{i.}\theta_{.j}$,

H_2 : the $\theta_{i.}$'s satisfy certain conditions;

H_3 : the $\theta_{.j}$'s satisfy certain conditions.

It is possible to verify by direct calculation that the following pairs of hypotheses are independent: (i) H_1 and H_2, (ii) H_1 and H_3, (iii) $H_1 \wedge H_2$ and H_3, (iv) $H_1 \wedge H_3$ and H_2. In general it is not true that H_2 and H_3 are independent, a fact which Example 1 has already implicitly demonstrated. However (v) if H_1 is true, then H_2 and H_3 are independent. This is implied, for instance, by (ii) and (iii) since

$$\begin{aligned} L(H_2 \wedge H_3 \mid H_1) &= L(H_1 \wedge H_2 \wedge H_3)/L(H_1) \\ &= L(H_1)\, L(H_2 \mid H_1)\, L(H_3 \mid H_1 \wedge H_2)/L(H_1) \\ &= L(H_2 \mid H_1)\, L(H_3) = L(H_2 \mid H_1)\, L(H_3 \mid H_1). \end{aligned}$$

To demonstrate the applicability of Theorem 1, we consider first (ii) and (iii), and we suppose that no $\theta_{\cdot j}$ is zero. Then we transform from the parameter space $\{\theta\}$ to a new parameter space Ω' by

$$\theta_{ij}/\theta_{\cdot j} = \theta'_{ij}, \qquad \theta_{\cdot j} = \theta'_{\cdot j}.$$

If $\Omega'_1 = \{\theta'_1 = (\theta'_{11}, \theta'_{12}, \cdots, \theta'_{rs}): 0 \leqq \theta'_{ij} \leqq 1, \sum_i \theta'_{ij} = 1, \text{each } j\}$ and $\Omega'_2 = \{\theta'_2 = (\theta'_{\cdot 1}, \theta'_{\cdot 2}, \cdots, \theta'_{\cdot s}): 0 < \theta'_{\cdot j} < 1, \sum_j \theta'_{\cdot j} = 1\}$, then $\Omega' = \Omega'_1 \times \Omega'_2$. The transformation is clearly $1-1$. Further H_1 imposes restrictions only on the θ'_{ij}'s. In fact H_1 says: for fixed i, θ'_{ij} is constant $(= \theta_{i\cdot})$. $H_1 \wedge H_2$ also imposes restrictions only on the θ_{ij}'s. H_3 imposes restrictions only on the $\theta_{\cdot j}$'s. And finally $\theta_{ij} = \theta'_{ij}\, \theta'_{\cdot j}$, i.e., the likelihood function factorises as required in the theorem. Hence by this almost trivial transformation of the parameter space we can demonstrate the truth of (ii) and (iii).

The transformation required to demonstrate the truth of (i) and (iv) is obvious. Finally the transformation required to demonstrate (v) again is not abstruse. For if H_1 is true, Ω has dimension $r + s - 2$. The transformation $\theta_{i\cdot} = \theta'_{i\cdot}$ $(i = 1, 2, \cdots, r)$ and $\theta_{\cdot j} = \theta'_{\cdot j}$ $(j = 1, 2, \cdots, s)$ may easily be seen to satisfy the requirements of Theorem 1, as far as (v) is concerned.

There are three interesting points which emerge from Example 2. First, by definition, independence of two hypotheses means that the likelihood ratio statistic for testing one does not depend on whether or not we assume the other to be true and so, if the likelihood ratio test of given size for that one against G is similar, this *test* is also unaffected by what we assume about the other. In the above example it is not surprising to find that a likelihood ratio test of a hypothesis about one set of marginal probabilities is not changed by knowledge that the factors are independent. At first sight it is more surprising to find that a likelihood ratio test of independence of the factors is unaffected by, for example, knowledge of the true marginal probabilities associated with *one* factor. This of course follows from the pairwise independences just established.

The second point is more general. It follows from (i) and (iii) or from (ii) and (iv), that

$$L(H_1 \wedge H_2 \wedge H_3) = L(H_1)\, L(H_2)\, L(H_3) \tag{4.1}$$

and consequently that separate tests of the three hypotheses induce a good test of $H_1 \wedge H_2 \wedge H_3$. But we cannot deduce this directly from the nesting order H_2, H_3, H_1 because in general we do not have $L(H_3 \mid H_2) = L(H_3)$ nor $L(H_1 \mid H_2$

$\wedge H_3) = L(H_1)$. This illustrates the point that, so far as the induced test of $H_1 \wedge H_2 \wedge H_3$ is concerned it is (4.1) that is the crucial condition and not pairwise independences that lead to it: and it *may* be possible that the induced test of $H_1 \wedge H_2 \wedge H_3$ is good when none of the pairwise induced tests are. However it does seem likely that in nearly all situations where (4.1) is satisfied there will be pairwise independences among hypotheses which enable this result to be derived from *some* nesting order. The authors have not met a counter-example.

Finally Example 2 raises the question of independence of more than two hypotheses. Here the analogy with independence of events continues to be close. For *complete* independence of three hypotheses we require four conditions: $L(H_1 \wedge H_2) = L(H_1)\, L(H_2)$ and the two other pairwise conditions; and also $L(H_1 \wedge H_2 \wedge H_3) = L(H_1)\, L(H_2)\, L(H_3)$. In general no subset of these is sufficient. (In Example 2, three of them are satisfied, but the fourth is not.)

There is an obvious extension of Theorem 1. which establishes sufficient conditions for three hypotheses to be completely independent and we conclude this section with an example illustrating its applicability.

EXAMPLE 3. Let the probabilities in a 3-factor contingency table, with factors α, β and γ be $\{\theta_{ijk}\}$ $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s; k = 1, 2, \cdots, t)$, where all θ's are known to be non-zero. Consider the three hypotheses

H_1 : α is independent of β and γ,

H_2 : β is independent of γ,

H_3 : the γ marginal probabilities satisfy certain conditions. The intuitively appealing result that these hypotheses are independent may be proved by a simple transformation of the parameter space as follows. Let $\theta_{ijk}/\theta_{\cdot jk} = \theta'_{ijk}$, $\theta_{\cdot jk}/\theta_{\cdot\cdot k} = \theta'_{jk}$, and $\theta_{\cdot\cdot k} = \theta'_k$, so the θ's are all positive, $\sum_i \theta'_{ijk} = 1$ for each j and k, $\sum_j \theta'_{jk} = 1$ for each k and $\sum_k \theta'_k = 1$. This establishes a 1-1 transformation from the original parameter space to a new space which is a Cartesian product space $\Omega'_1 \times \Omega'_2 \times \Omega'_3$, Ω'_1 being the space of θ'_{ijk}'s, Ω'_2 that of the θ'_{jk}'s and Ω'_3 that of the θ'_k's.

Now $\theta_{ijk} = \theta'_{ijk}\, \theta'_{jk}\, \theta'_k$, so that the likelihood function is expressible as a product of three functions, one depending only on a variable in Ω'_1, one on a variable in Ω'_2 and one on a variable in Ω'_3. Moreover H_i imposes restrictions on elements of Ω'_i only $(i = 1, 2, 3)$. The extension of Theorem 1 then shows that H_1, H_2 and H_3 are independent.

5. Linear hypotheses with known residual variance. A discussion of the problem with which we are concerned would be incomplete without consideration of linear hypotheses, and it transpires that investigation of this case sheds new light on the role of orthogonality in design.

We suppose, then, that an n-dimensional vector y of observations is generated by the model G: $y = \theta + \epsilon$, where ϵ is an observation on a $N(0, \sigma^2 I_n)$ random vector and θ, a vector of means, lies in a subspace Ω of R^n. A linear hypothesis is one which states that θ belongs to a subspace of Ω. Suppose that H_1 and H_2 are linear hypotheses which specify respectively the subspaces ω_1 and ω_2, and let P_0, P_1, P_2 and P_{12} be the matrix representations of the linear transformations which project R^n orthogonally on Ω, ω_1, ω_2 and $\omega_1 \cap \omega_2$ respectively. Then if all vectors involved are regarded as the column vector representations of elements of R^n with respect to a fixed orthogonal basis, each P is *symmetric* as well as being *idempotent.*

It is convenient for our purposes to consider first the case *when σ^2 is known.* Then it is not difficult to verify that $-2 \log L_y(H_i) = y'(P_0 - P_i)y/\sigma^2$ and $-2 \log L_y(H_1 \wedge H_2) = y'(P_0 - P_{12})y/\sigma^2$. Hence, in this case, H_1 and H_2 are independent hypotheses, i.e., $L(H_1 \wedge H_2) = L(H_1)\, L(H_2)$ if and only if

$$P_0 - P_{12} = P_0 - P_1 + P_0 - P_2 . \tag{5.1}$$

We recall that $P_0 - P_i$ is the matrix representation of the orthogonal projector of R^n on ω_i^p, the orthogonal complement *in* Ω of ω_i. The sum of two orthogonal projectors is an orthogonal projector if and only if their ranges are orthogonal and in this case the range of the sum is the direct sum of the ranges. From these results it follows that (5.1) is true if and only if $\omega_1^p \perp \omega_2^p$ so that we now have established the following result.

THEOREM 2. *With the usual normal model when σ^2 is known, the linear hypotheses H_1 and H_2 are independent if and only if $\omega_1^p \perp \omega_2^p$.*

Now this case of linear hypotheses with σ^2 known has certain special features of considerable importance in large sample theory and these are established in

THEOREM 3. *With the usual normal model when σ^2 is known, pairwise independence of linear hypotheses implies that*

(i) *the hypotheses are completely independent,*

(ii) *their likelihood ratio test statistics are independent.*

PROOF. It is sufficient to consider three hypotheses. Suppose that they specify the subspaces ω_1, ω_2, ω_3; then, by Theorem 2, they are pairwise independent if and only if $\omega_i^p \perp \omega_j^p$, for each i and j. In this case it is not difficult to establish that there exists an orthogonal transformation P and, writing $\phi = P\theta$, a partitioning $[\phi_0', \phi_1', \cdots, \phi_4']$ of ϕ' such that

(i) G says that $\phi_4 = 0$,

(ii) H_i says that $\phi_4 = 0$ and $\phi_i = 0$, $(i = 1, 2, 3)$.

In addition, if $z = Py$ and $[z_0', z_1', \cdots, z_4']$ is the partitioning of z' corresponding to that of ϕ', then for $\theta \varepsilon \Omega$ it can be shown that $f(y, \theta, \sigma^2)$ is expressible in the form

$$f(y, \theta, \sigma^2) = g_4(z_4, \sigma^2) \prod_{i=0}^{3} g_i(z_i, \phi_i, \sigma^2). \tag{5.2}$$

This factorisation in terms of the ϕ's demonstrates, by the extension of Theorem 1 to more than two hypotheses, the complete independence of H_1, H_2 and H_3, when σ^2 is known. The factorisation in terms of the z's, which are independent random variables, demonstrates that, again when σ^2 is known, $L(H_1)$, $L(H_2)$ and $L(H_3)$ are independent. The proof of the theorem is completed by the remark that the above argument very obviously extends to more than these hypotheses

By now it is apparent that there is a connection between "orthogonality of design" and independence of hypotheses. An experimental design is orthogonal relative to a general linear model G and linear hypotheses $H_1, H_2, \cdots, H_k$ if, with this design, the subspaces specified respectively by $G, H_1, H_2, \cdots, H_k$ satisfy the condition $\omega_i^p \perp \omega_j^p$, all $i \neq j$. Thus when σ^2 is known, orthogonality of design implies both independence of hypotheses and independence of their likelihood ratio statistics. We are then dealing effectively with hypotheses relating to separate independent experiments and the situation is ideal.

6. Linear hypotheses with unknown residual variance. We now ask: what does lack of knowledge of σ^2 cost us with reference to the above ideal situation? The presence of an *unknown* σ^2 in each of the g functions in (5.2) prevents our concluding from this identity either that the hypotheses are independent or that their likelihood ratio test statistics are; and it suggests that neither conclusion is valid. Unfortunately this is so, as is readily verified—indeed it is well-known that orthogonality in design certainly does not imply (when σ^2 is unknown) independence of the likelihood ratio statistics for testing the individual hypotheses: they all involve the same "residual sum of squares".

But while we lose these desirable properties by not knowing σ^2, we do *not* lose the characteristic of the orthogonal design which ensures that, for example, good tests of H_1 and H_2 induce a good test of $H_1 \wedge H_2$. We recall that the main function of independence of hypotheses is to prevent the possibility of $L(H_1)$ and $L(H_2)$ taking values near 1 while $L(H_1 \wedge H_2)$ takes very small values, with high probability. Orthogonality of design continues to prevent this even when σ^2 is unknown. Of course this is a generally accepted result, though it has never been stated explicitly in the present form. The mechanism by which this prevention is achieved is demonstrated by the following argument, in which we consider the test of $H_1 \wedge H_2 \wedge H_3$ induced by separate tests of three hypotheses H_1, H_2 and H_3.

We have, in the notation of Section 5,

$$\begin{aligned}[L_y(H_1 \wedge H_2 \wedge H_3)]^{-2/n} - 1 &= y'(I - P_{123})y/y'(I - P_0)y - 1 \\ &= y'(I - P_1 + P_1 - P_{12} + P_{12} - P_{123})y/y' \cdot (I - P_0)y - 1 \\ &= \frac{y'(I - P_1)y}{y'(I - P_0)y} - 1 + \frac{y'(P_1 - P_{12})y}{y'(I - P_0)y} + \frac{y'(P_{12} - P_{123})y}{y'(I - P_0)y}.\end{aligned}$$

Now orthogonality of design implies $P_1 - P_{12} = P_0 - P_2$ and $P_{12} - P_{123} = P_0 - P_3$, because, *in the σ^2 known case*, it implies $L(H_2 \mid H_1) = L(H_2)$ and $L(H_3 \mid H_1 \wedge H_2) = L(H_3)$, from which the matrix identities are easily deduced. Also

$$y'(P_0 - P_i)y/y'(I - P_0)y = y'(I - P_i)y/y'(I - P_0)y - 1 = [L_y(H_i)]^{-2/n} - 1.$$

Hence orthogonality of design implies

$$[L_y(H_1 \wedge H_2 \wedge H_3)]^{-2/n} - 1 = \sum_{i=1}^{3} \{[L_y(H_i)]^{-2/n} - 1\}.$$

Thus if each of $L_y(H_1)$, $L_y(H_2)$ and $L_y(H_3)$ is near 1, so also is $L_y(H_1 \wedge H_2 \wedge H_3)$. And so by likelihood ratio standards separate tests of H_1, H_2 and H_3 induce a test of $H_1 \wedge H_2 \wedge H_3$ of reasonable power.

While the above argument is framed in terms of three hypotheses it is clearly a general one. It would appear to demonstrate in an unambiguous way just why orthogonality of design is generally accepted as desirable. It also lends force to the contention that mild departures from orthogonality are of little consequence, though the question of how far it is possible to depart from orthogonality without seriously affecting the powers of induced tests is one whose answer would involve extensive computation.

7. Large sample theory. The theory of Section 5, as well as having theoretical interest on its own account, is of considerable practical value from the large sample point of view, because it is often possible to interpret large sample problems as essentially linear problems.

Suppose that we have a large number n of independent observations on a random variable (real or vector valued) whose distribution depends on an unknown parameter θ in R^s. Let $\omega_1 = \{\theta: h_{1j}(\theta) = 0, j = 1, 2, \cdots, r_1\}$ and $\omega_2 = \{\theta: h_{2j}(\theta) = 0, j = 1, 2, \cdots, r_2\}$, the h's being well-behaved real-valued functions and let H_i be the hypothesis that the true parameter is in ω_i ($i = 1, 2$). Usually, if the true parameter is not near either of these subsets (if, in fact, its distance from ω_i is more than $O(n^{-\frac{1}{2}})$, this will be so obvious that no test is necessary (Wald [10]). So from the theoretical point of view we may assume that Ω, the set of possible parameters, consists of $\omega_1 \cap \omega_2$ and points of R^s near this set. This, together with consistency of maximum likelihood estimates enables us to treat this large sample problem as essentially a linear problem with *known variance matrix of residuals*, in the kind of way indicated, for example, by Lehmann [7]. In this way the question of independence of H_1 and H_2 is translated into a question of "local orthogonality", which may be treated by a slight modification of the methods of Section 5 in which vectors involved are regarded as representations with respect to a non-orthogonal basis of elements of R^s. We shall not go into details of this argument which uses fairly well-known principles. It transpires that the crucial condition for large sample independence of H_1 and H_2 is that

$$\mathbf{H}_1 \mathbf{B}^{-1} \mathbf{H}_2 = \mathbf{0}, \quad \text{for every } \theta \text{ in } \omega_1 \cap \omega_2. \tag{7.1}$$

Here $\mathbf{H}_i$ is the $s \times r_i$ matrix of partial derivatives at θ of the functions h_{ij}, $j = 1, 2, \cdots, r_i$ ($i = 1, 2$), and $\mathbf{B}$ is the information matrix for θ. Because of the connection with linear theory, σ^2 known, large sample independence of H_1 and H_2 implies large sample independence of their likelihood ratio test statistics also. We note that if H_1 and H_2 are independent for all sample sizes and if, further, the conditions of Theorem 1 are satisfied, then (7.1) is easily seen to be satisfied for every θ in Ω.

In a recent paper Aitchison [1] has given a fuller account of this aspect of large sample theory from a different point of view. In this paper he introduced the notion of the separability of two hypotheses with respect to any method of test construction. Now separability with respect to the likelihood ratio method is what we have called independence. However Aitchison was mainly concerned with large sample likelihood ratio separability and the reader is referred to his paper for further details relating to Condition (7.1).

We now digress briefly from strict likelihood ratio tests to make two remarks about the "partitioning of χ^2" when testing hypotheses about multinomial samples (see Lancaster [4], for example). Because the χ^2 statistic for testing H is asymptotically equivalent to $-2 \log L(H)$, it follows that if H_1 and H_2 are independent in large samples then the χ^2 statistic approximately partitions additively. On the other hand, if H_1 and H_2 are independent in small samples it does not follow that the χ^2 statistic for testing $H_1 \wedge H_2$ against G is expressible as the sum of the χ^2 statistics for testing H_1 and H_2 against G. This is a particular case of a more general result which is readily demonstrated, namely that the Lagrangian multiplier statistic (see Silvey [9]) does not usually partition under the conditions of Theorem 1.

8. Concluding remarks. Up to this point we have concentrated attention on distinguishing the characteristics of situations where good tests of H_1 and H_2 induce a good test of $H_1 \wedge H_2$. We have avoided the question of what to do if we are not in such a desirable situation and we wish to have good tests of all three hypotheses. The trouble is apparent: the induced test must be abandoned and there will be parts of the sample space in which a direct test of $H_1 \wedge H_2$ is not compatible with the tests of H_1 and of H_2. Clearly we must in general modify our demands for, in asking for a four-decision procedure which produces good tests of H_1, H_2 and $H_1 \wedge H_2$ (according to our definition of good), we are demanding more information from the experiment than it is capable of providing. (We exclude the nested situation from these remarks because, if we are interested in H_2, say, only if we have previously accepted H_1, then only three decisions are involved, namely $\tilde{H}_1$, $H_1 \wedge H_2$, $H_1 \wedge \tilde{H}_2$.)

The incompatibility which characterises dependent hypotheses occurs at those points of the sample space which are in the acceptance regions of H_1 and H_2 and in the rejection region of $H_1 \wedge H_2$. An obvious compromise of the four decision procedure is to assign these points to the hypothesis $(H_1 \wedge \tilde{H}_2)v(\tilde{H}_1 \wedge H_2)$. This is the practice in the regression problem of constructing as simple a predictor as possible when it is found that one of two correlated variables x_1, x_2 must be included, but it is not clear which. The choice between $H_1 \wedge \tilde{H}_2$ and $\tilde{H}_1 \wedge H_2$ is then made either arbitrarily or on the basis of some non-statistical consideration.

Acknowledgment. We wish to thank the referee for some helpful comments.

REFERENCES

[1] AITCHISON, JOHN (1962). Large-sample restricted parametric tests. *J. Roy. Statist. Soc. Ser. B* **24** 234–250.

[2] ANDERSON, T. W. (1962). The choice of the degree of a polynomial regression as a multiple decision problem. *Ann. Math. Statist.* **33** 255–265.

[3] HOGG, ROBERT V. (1961). On the resolution of statistical hypotheses. *J. Amer. Statist. Assoc.* **56** 978–989.

[4] LANCASTER, H. O. (1949). The derivation and partition of χ^2 in certain discrete distributions. *Biometrika* **36** 117–129.

[5] LEHMANN, E. L. (1957). A theory of some multiple decision problems, I. *Ann. Math. Statist.* **28** 1–25.

[6] LEHMANN, E. L. (1957). A theory of some multiple decision problems, II. *Ann. Math. Statist.* **28** 547–572.

[7] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[8] SCHEFFÉ, HENRY (1959). *The Analysis of Variance.* Wiley, New York.

[9] SILVEY, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.* **30** 389–407.

[10] WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Ann. Math. Statist.* **14** 462–482.