MOST POWERFUL TESTS OF COMPOSITE HYPOTHESES. I. NORMAL DISTRIBUTIONS

By E. L. Lehmann and C. Stein

University of California, Berkeley

Summary. For testing a composite hypothesis, critical regions are determined which are most powerful against a particular alternative at a given level of significance. Here a region is said to have level of significance ϵ if the probability of the region under the hypothesis tested is bounded above by ϵ. These problems have been considered by Neyman, Pearson and others, subject to the condition that the critical region be similar. In testing the hypothesis specifying the value of the variance of a normal distribution with unknown mean against an alternative with larger variance, and in some other problems, the best similar region is also most powerful in the sense of this paper. However, in the analogous problem when the variance under the alternative hypothesis is less than that under the hypothesis tested, in the case of Student's hypothesis when the level of significance is less than $\frac{1}{2}$, and in some other cases, the best similar region is not most powerful in the sense of this paper. There exist most powerful tests which are quite good against certain alternatives in some cases where no proper similar region exists. These results indicate that in some practical cases the standard test is not best if the class of alternatives is sufficiently restricted.

1. Introduction. The problem to be discussed in this paper is that of testing a composite hypothesis against a simple alternative. More specifically let $\mathcal{F} = \{f\}$ be a family of probability density functions defined over a Euclidean space R_n and let g be a probability density function not in $\mathcal{F}$. We wish to test the hypothesis H_0 that the random variable $X = (X_1, \cdots, X_n)$ is distributed according to a density f of $\mathcal{F}$ against the alternative H_1 that X is distributed according to g. By a test we mean a region of rejection, w in R_n.

Neyman and Pearson, in the fundamental paper [1] which laid the groundwork of the theory of optimum tests, restricted their considerations to similar regions. They considered a region (set) w to be optimum for the given level of significance ϵ if it maximizes the power

$$\int_w g(x)\,dx \tag{1}$$

subject to the restriction

$$\int_w f(x)\,dx = \epsilon \quad \text{for all} \quad f \text{ in } \mathcal{F}. \tag{2}$$

As Neyman, Wald and others have pointed out, it is more natural to replace the condition of similarity (2) by the weaker restriction

$$\int_w f(x)\,dx \leq \epsilon \quad \text{for all} \quad f \text{ in } \mathcal{F}. \tag{3}$$

A region w maximizing (1) subject to (3) is called most powerful against the alternative g at the level of significance. Here and throughout the paper, all functions and sets are assumed to be Borel measurable.

In the present paper we shall consider certain composite hypotheses, and derive tests for them which are most powerful against a simple alternative. For the cases in which these tests coincide with the standard similar regions it will thus be established that no further increase in power is possible with tests of fixed sample sizes. In the more usual situation where the most powerful test depends strongly on the specific alternative chosen, no such absolute justification of the standard test is possible. In these cases, any justification must take account of the fact that it is desired to obtain good power against a large class of alternatives. This can be done, for instance, by using Wald's definition of a most stringent test [2] or his concept of minimizing the maximum risk.[1] If, on the other hand, the class of alternatives is sufficiently restricted, the results of the present paper indicate that for small samples there may exist a test which is appreciably better than the standard test.

Frequently the probability of an error of the first kind is an analytic function of a nuisance parameter for every choice of critical region. Hence, if it is known that some nuisance parameter θ lies, say in a certain finite interval I, then any test which is similar for θ in I will be similar for all θ. Consequently, the knowledge concerning θ cannot be used to find a more powerful test. On the other hand, as is indicated at the end of section 5, restrictions of the nuisance parameters may, for small samples, lead to considerably more powerful tests if the condition of similarity is replaced by the weaker condition (3).

There is one class of problems to which it may be desirable to apply the method of the present paper regardless of sample size; namely, if no similar region exists. Suppose, for instance, that $X_1, \cdots, X_n$ are known to be normally and independently distributed, X_i having unknown mean and variance ξ_i and σ_i^2 for $i = 1, \cdots, n$. For testing the hypothesis

$$H_0 : \sigma_i = 1, \qquad (i = 1, \cdots, n)$$

no similar region exists, while it is easy to see that against any simple alternative

$$H_1 : \sigma_i = \sigma_{i1} < 1, \qquad \xi_i = \xi_{i1},$$

there exists a test which satisfies condition (3) and which has good power against H_1 provided the σ_{i1} are sufficiently small.

The present first part of this paper is restricted to hypotheses concerning normal distributions. It is intended to extend the considerations to exponential

[1] In an unpublished paper, it is shown by G. Hunt and C. Stein that the traditional test is most stringent in several cases, including the (univariate) linear hypothesis and the hypothesis specifying the ratio of the variances of two normal distributions. These results can be extended to analogous problems for distributions other than the normal, and similar results can be proved regarding minimization of the maximum risk if the weight function has a certain type of symmetry.

and rectangular distributions, to consider non-parametric problems and possibly also more complicated problems connected with normal distributions, in later parts of the paper.

2. Sufficient conditions for a most powerful test. The method which will be used in this paper to obtain most powerful tests is an adaptation of the fundamental lemma of Neyman and Pearson [1]. At the same time it is essentially a special case of much more general results of Wald [3, 4], although the exact conditions of Wald's investigation are not satisfied in most of our problems.

Let h and g be two functions defined over R_n, let k be a constant and let w be a region in R_n such that

$$g(x) \geq k\, h(x) \text{ in } w; \qquad g(x) \leq k\, h(x) \text{ in } R_n - w. \tag{4}$$

Then if w' is such that

$$\int_{w'} h(x)\, dx \leq \int_w h(x)\, dx, \tag{5}$$

it follows as in the fundamental lemma where in (5) equality is assumed instead of inequality, that

$$\int_{w'} g(x)\, dx \leq \int_w g(x)\, dx. \tag{6}$$

Throughout the present paper we shall be concerned with the special case in which $\mathcal{F}$ is an s-parameter family. We may denote the members of $\mathcal{F}$ by f_θ and we shall obtain all members of $\mathcal{F}$ as θ ranges over a set ω in an s-dimensional Euclidean space. In the theorem which we shall now state, we shall be concerned with point functions λ defined over ω. We shall assume that $\lambda = c\mu$ where c is a positive constant and μ a cumulative distribution function.[2] Also we suppose that $f_\theta(x)$ is a measurable function of x and θ jointly. However, the theorem is also valid if ω is an abstract space and λ a (finite) non-negative additive set function (measure) over ω. Such more general interpretation may be required when applying the theory to non-parametric problems.

THEOREM 1. *Let H_0 be the hypothesis that the random variable X is distributed according to a density function f_θ with θ in ω, and let H_1 denote the alternative that X is distributed according to a density g. Let λ be a function defined over ω and such that*

$$\lambda = c\mu, \tag{7}$$

[2] The introduction of the distribution μ is simply a mathematical device and does not imply that θ is a random variable (see Wald [16] p. 282).

where c is a positive constant and μ a cumulative distribution function. Let k be a constant and let w be a region in R_n such that

$$(8)\qquad \begin{aligned} g(x) &\geq k \int_w f_\theta(x)\, d\lambda(\theta) \quad \text{in} \quad w; \\ g(x) &\leq k \int_w f_\theta(x)\, d\lambda(\theta) \quad \text{in} \quad R_n - w. \end{aligned}$$

Suppose that w is of level of significance ϵ for testing H_0 against H_1, that is that

$$(9)\qquad \int_w f_\theta(x)\, dx \leq \epsilon \quad \text{for all} \quad \theta \text{ in } \omega,$$

and suppose that the subset of ω for which

$$(10)\qquad \int_w f_\theta(x)\, dx < \epsilon$$

has λ-measure zero. Then w is most powerful for testing H_0 against H_1 at level of significance ϵ.

Proof. Without loss of generality we shall assume $c = 1$. Let w' be any test of level of significance ϵ. Then

$$(11)\qquad \int_{w'} f_\theta(x)\, dx \leq \epsilon \quad \text{for all} \quad \theta \text{ in } \omega,$$

and because of (7)

$$(12)\qquad \int_\omega \left\{ \int_{w'} f_\theta(x)\, dx \right\} d\lambda(\theta) \leq \epsilon \int_\omega d\lambda(\theta) = \epsilon.$$

Since λ is of bounded variation we may interchange the order of integration in (12) and obtain

$$(13)\qquad \int_{w'} h(x)\, dx \leq \epsilon,$$

where

$$(14)\qquad h(x) = \int_\omega f_\theta(x)\, d\lambda(\theta).$$

From (9) and the condition surrounding (10) it follows that

$$(15)\qquad \int_\omega \left\{ \int_w f_\theta(x)\, dx \right\} d\lambda(\theta) = \epsilon,$$

and therefore that

$$(16)\qquad \int_w h(x)\, dx = \epsilon.$$

Thus w and w' satisfy conditions (4) and (5), and hence also (6) which completes the proof.

It is useful to notice that, the assumptions of theorem 1 will be satisfied provided

$$\int_w f_\theta(x)\,dx$$

attains its maximum ϵ at all points of increase of λ, and therefore in particular whenever w is a similar region of size ϵ.

We shall in many problems exhibit a function λ which satisfies the conditions of theorem 1 without giving the reasons which led us to this function. However the following comments concerning the tentative process that we used, may be helpful. One may first examine the known most powerful similar region. If there exists a cumulative distribution function λ such that (8) is the most powerful similar region, the problem is solved. If the most powerful similar region cannot even be approximated by (8) with a sequence of λ's, it is reasonable to conclude that the most powerful test is not similar. Because the probability (under the null hypothesis) of any test is in all the problems considered here an analytic function of the parameter, this implies that the probability (under the null hypothesis) of the most powerful test attains its maximum at an at most denumerable (in some cases finite) set of points. In all the cases of this kind which we considered in the present part I, it was then possible to prove the existence of a function λ with a single point of increase, which satisfied the conditions of theorem 1.

A theorem analogous to theorem 1 holds for most powerful similar regions. Let H_0 and H_1 be as before and let λ be a function of bounded variation not necessarily non-decreasing. Let w be a region in R_n such that

$$\begin{aligned} g(x) &\geq k \int_\omega f_\theta(x)\,d\lambda(\theta) \quad \text{in} \quad w; \\ g(x) &\leq k \int_\omega f_\theta(x)\,d\lambda(\theta) \quad \text{in} \quad R_n - w. \end{aligned} \tag{17}$$

Let w be a similar region of level of significance ϵ for testing H_0 against H_1, that is, let

$$\int_w f_\theta(x)\,dx = \epsilon \quad \text{for all} \quad \theta \text{ in } \omega; \tag{18}$$

then w is a most powerful similar region for testing H_0 against H_1.

For all the problems considered in this paper we shall prove the existence of functions λ satisfying the conditions of theorem 1, but we have not investigated the corresponding existence problem in general. On the other hand one verifies easily that for many of the cases treated here in which the most powerful test is not similar, the method for obtaining most powerful similar regions does not apply. However, for all the problems considered in the present paper the most powerful similar tests can be obtained easily by other methods [1, 5, 6, 7, 8]. For most of the problems the corresponding derivations have been carried out in the literature.

Although we restrict ourselves in the present paper to the problem of maximizing the power at a single alternative, theorem 1 clearly also applies to the more general problem of maximizing the average power over surfaces in a space of alternatives. Such problems have been considered from the point of view of similar regions by Wald, Hsu and others [9, 10, 11].

3. Testing the values of one or several variances. Let $X_1, \cdots, X_n$ be a sample from a normal population with mean ξ and variance σ^2, both unknown. We want to test the hypothesis H_0 that $\sigma = \sigma_0$ against the simple alternative that $\sigma = \sigma_1, \xi = \xi_1$. We shall show that the most powerful test for H_0 against H_1 is

$$\Sigma(x_i - \xi_1)^2 \leq k \quad \text{when} \quad \sigma_1 < \sigma_0; \tag{19}$$

$$\Sigma(x_i - \bar{x})^2 \geq c \quad \text{when} \quad \sigma_1 > \sigma_0, \tag{20}$$

where k and c are determined by the level of significance. Thus the best similar region is most powerful if the variance under the alternative is greater than that under the null hypothesis, while the most powerful tests against the other alternatives are not similar. That the region $\Sigma(x_i - \bar{x})^2 \geq c$ $(\leq c')$ is most powerful of all similar regions against $\sigma_1 > \sigma_0$ $(\sigma_1 < \sigma_0)$ was shown by Neyman and Pearson [1].

We consider first the case $\sigma_1 < \sigma_0$, and apply theorem 1 with λ a stepfunction having a single jump at ξ_1, that is,

$$\lambda(\xi) = \begin{matrix} 0 & \text{if} & \xi < \xi_1; \\ 1 & \text{if} & \xi \geq \xi_1. \end{matrix} \tag{21}$$

The region w given by (8) thus becomes

$$\frac{\exp\left[-\frac{1}{2\sigma_1^2}\Sigma(x_i - \xi_1)^2\right]}{\exp\left[-\frac{1}{2\sigma_0^2}\Sigma(x_i - \xi_1)^2\right]} \geq k', \tag{22}$$

which is equivalent to

$$\Sigma(x_i - \xi_1)^2 \leq k, \tag{23}$$

since $\sigma_1 < \sigma_0$. The size of the region (23), that is, its probability under the null hypothesis is a function of ξ and clearly attains its maximum when $\xi = \xi_1$. Thus all conditions of theorem 1 are satisfied provided we choose k so that the maximum size of (23) equals ϵ.

Before considering the case $\sigma_1 > \sigma_0$ we state for later reference the following:

LEMMA 1. *If $\sigma_1 > \sigma_0$ there exists an absolutely continuous non-decreasing function λ of bounded variation such that*

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma_0^2}(t - \xi)^2\right] d\lambda(\xi) = C \exp\left[-\frac{1}{2\sigma_1^2}(t - \xi_1)^2\right]. \tag{24}$$

This follows immediately from the well known representation of $\exp\left(-\frac{1}{a^2}t^2\right)$ as a Laplace transform by applying a translation, and is easily verified directly by substituting

$$\lambda'(\xi) = \exp\left[-\frac{1}{2(\sigma_1^2 - \sigma_0^2)}(\xi - \xi_1)^2\right]. \tag{25}$$

Now let $\sigma_1 > \sigma_0$ and $n > 1$. The region w given by (8) can be expressed in the form

$$\frac{\exp\left[-\frac{1}{2\sigma_1^2}\Sigma(x_i - \bar{x})^2\right]}{\exp\left[-\frac{1}{2\sigma_0^2}\Sigma(x_i - \bar{x})^2\right]} \cdot \frac{\exp\left[-\frac{n}{2\sigma_1^2}(\bar{x} - \xi_1)^2\right]}{\int \exp\left[-\frac{n}{2\sigma_0^2}(\bar{x} - \xi)^2\right] d\lambda(\xi)} \geq k'. \tag{26}$$

By lemma 1 there exists an absolutely continuous function λ for which the second factor is constant. For this λ (26) is equivalent to

$$\Sigma(x_i - \bar{x})^2 \geq c, \tag{27}$$

and since this is a similar region, the conditions of theorem 1 are satisfied provided c is chosen so as to give the correct level of significance.

We next consider the problem in which the random variables X_i $(i = 1, \cdots, n)$ are independently normally distributed with unknown means ξ_i and unknown variances σ_i^2. We wish to test the hypothesis $H_0 : \sigma_i = \sigma_{i0}$ for $i = 1, \cdots, n$ against the alternative $H_1 : \sigma_i = \sigma_{i1}, \xi_i = \xi_{i1}$. Feller [12] showed that there exist no similar regions for this problem. However, as we shall show now, when the critical regions are not required to be similar, non-trivial tests against H_1 do exist provided $\sigma_{i1} < \sigma_{i0}$ for at least one value of i.

Let us assume without loss of generality that $\sigma_{i1} < \sigma_{i0}$ for $i = 1, \cdots, m$; $\sigma_{i1} > \sigma_{i0}$ for $i = m + 1, \cdots, n$ where $n - m$ may be zero but where for the moment we shall assume $m > 0$. With $\lambda(\xi_1, \cdots, \xi_n) = \prod_{i=1}^{n} \lambda_i(\xi_i)$, the region (8) becomes

$$\prod_{i=1}^{m} \frac{\exp\left[-\frac{1}{2\sigma_{i1}^2}(x_i - \xi_{i1})^2\right]}{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma_{i0}^2}(x_i - \xi_i)^2\right] d\lambda_i(\xi_i)} \cdot \prod_{j=m+1}^{n} \frac{\exp\left[-\frac{1}{2\sigma_{j1}^2}(x_j - \xi_{j1})^2\right]}{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma_{j0}^2}(x_j - \xi_j)^2\right] d\lambda_j(\xi_j)} \geq k. \tag{28}$$

For $\lambda_i(i = 1, \cdots, m)$ we take step functions with a single jump at ξ_{i1}, while for the remaining λ's we choose the absolutely continuous functions which make the second factor constant and whose existence is guaranteed by lemma 1. The region (28) thus reduces to

$$\sum_{i=1}^{m}\left(\frac{1}{\sigma_{i1}^{2}} - \frac{1}{\sigma_{i0}^{2}}\right)(x_i - \xi_{i1})^2 \leq c. \tag{29}$$

Since the probability of the region (29) is independent of $\xi_{m+1}, \cdots, \xi_n$ and with varying $\xi_1, \cdots, \xi_m$ takes on its maximum when $\xi_i = \xi_{i1}$ it follows from theorem 1 that this region is most powerful for testing H_0 against H_1.

We still have to consider the case $m = 0$, that is, the case in which $\sigma_{i1} > \sigma_{i0}$ for all i. To treat this problem we adjoin to the variables $X_1, \cdots, X_n$ a random variable Y uniformly distributed between 0 and 1, that is, essentially a table of random numbers. In the space of $n + 1$ random variables we determine a region w according to (8), letting $\lambda(\xi_1, \cdots, \xi_n) = \prod_{i=1}^{n} \lambda_i(\xi_i)$ and choosing the λ's so as to make the left hand side of (8) equal to the right hand side. This is possible by lemma 1 and with this choice of the λ's the inequalities (9) become

$$\begin{aligned} k &\geq k \text{ in } w; \\ k &\leq k \text{ in } R_{n+1} - w, \end{aligned} \tag{30}$$

and hence they impose no restrictions on w. Thus any similar region of the correct size will satisfy the conditions of theorem 1. It follows that the region

$$w: 0 \leq y \leq \epsilon, \tag{31}$$

being a similar region of size ϵ, is most powerful. This result means that we do not use the observations $x_1, \cdots, x_n$ at all but consult a table of random numbers.

The situation just described occurs in other problems to which the same method of proof can be applied. It is therefore convenient for later reference to formulate the following

THEOREM 2. *Let H_0 be the hypothesis that the random variable X is distributed according to a probability density function f_θ with θ in ω, and let H_1 denote the alternative that X is distributed according to the density function g. Let Y be a random variable known to be uniformly distributed over the interval* [0, 1]. *If there exists a real valued function λ satisfying* (7) *for which*

$$g(x) = k \int_{\omega} f_\theta(x)\, d\lambda(\theta), \tag{32}$$

then the critical region $0 \leq y \leq \epsilon$ is most powerful for testing H_0 against H_1 at level of significance ϵ.

4. Testing equality of variances and the value of the circular serial correlation coefficient. For each $i = 1, \cdots, m$ let $X_{ij} (j = 1, \cdots, n_i)$ be a sample from a normal distribution with $E(X_{ij}) = \xi_i$ and $E(X_{ij} - \xi_i)^2 = \sigma_i^2$. We are concerned with the hypothesis H_0 that $\sigma_1 = \sigma_2 = \cdots = \sigma_m$, where first we shall assume the ξ's to be known, so that without loss of generality we may assume them equal to 0. The alternative hypothesis specifies $\sigma_i = \sigma_{i1}$, $i = 1 \cdots m$. Let σ^2 denote the unknown common variance under H_0 and let $\lambda(\sigma)$ be a step function with a single jump at a point σ_0 to be determined later. With $k = \prod_{i=1}^{m} \left(\frac{\sigma_0}{\sigma_{i1}}\right)^{n_i}$, the test (8) takes on the form

$$\frac{\exp\left[-\frac{1}{2}\sum_{i,j}\frac{x_{ij}^2}{\sigma_{i1}^2}\right]}{\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i,j}x_{ij}^2\right]} \geq 1, \tag{33}$$

or equivalently

$$\frac{\sum_{i,j} x_{ij}^2}{\sum_{i,j}\frac{1}{\sigma_{i1}^2}x_{ij}^2} \geq \sigma_0^2. \tag{34}$$

Since the function on the left hand side is homogeneous of degree 0 in the x's, this is a similar region and the conditions of theorem 1 are therefore satisfied provided the region has the correct size. This can be achieved for any level of significance ϵ by proper choice of σ_0^2.

As stated earlier, the conditions of theorem 1 imply that the size of the critical region is equal to ϵ at all points of increase of λ. As a consequence, if the size equals ϵ at only a finite number of points of ω, λ must be a step function. Also if each point of a certain interval is a point of increase of λ, the critical region must be similar over that interval (and, if the functions involved are analytic, the region must be similar over ω). However, the last problem shows that the converse of neither of these two statements is correct. For the region (34) is a similar region although the corresponding λ has only a single point of increase.

Next we consider the hypothesis of equality of variances without assuming the means to be known. For the case $m = 2$ the most powerful similar region was obtained by Neyman and Pearson [1]. We assume first that $n_i > 1$ for all i, and we take $\lambda(\sigma, \xi_1, \cdots, \xi_m) = \lambda_0(\sigma)\prod_{i=1}^{m}\lambda_i(\xi_i)$, with $\lambda_0(\sigma)$ as before a step function with a single jump at a point σ_0 to be determined later. Suppose now that $\sigma_0 > \sigma_{i1}$ for $i = 1, \cdots, s$; $\sigma_0 < \sigma_{i1}$ for $i = s + 1, \cdots, m$, $\sigma_{11} \leq \sigma_{21} \leq \cdots$ where $0 \leq s \leq m$ and s depends on σ_0. Then define

$$\lambda_i(\xi_i) = \begin{cases} 0 & \text{if } \xi_i < \xi_{i1}; \\ 1 & \text{if } \xi_i \geq \xi_{i1} \end{cases} \tag{35}$$

for $i = 1, \cdots, s$ and use lemma 1 for $i = s + 1, \cdots, m$.

For proper choice of k the critical region will then be determined by the inequality

$$(36) \qquad \sum_{i=s+1}^{m} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_{i1}^2}\right) \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 - \sum_{i=1}^{s} \left(\frac{1}{\sigma_{i1}^2} - \frac{1}{\sigma_0^2}\right) \sum_{j=1}^{n_i} (x_{ij} - \xi_{i1})^2 \geq 0.$$

The probability of this region computed under H_0, is independent of $\xi_{s+1}, \cdots, \xi_n$ and for any σ attains its maximum when $\xi_i = \xi_{i1}$ $(i = 1, \cdots, s)$. Since the probability of the region is independent of σ when $\xi_i = \xi_{i1}$ for $i = 1, \cdots, s$, the conditions of theorem 1 are again established. That for $\xi_i = \xi_{i1}$ the size of (36) goes continuously from 0 to 1 with decreasing σ_0 is easily checked since at the only doubtful points $\sigma_0 = \sigma_{i1}$ (where the value of s changes), the corresponding coefficient $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_{i1}^2}$ passes through 0.

We still have to consider the case that some of the n_i are equal to 1. If $n_i = 1$ for some $i \leq s$ there is no change whatever, while if $n_i = 1$ for some $i > s$, the corresponding term in (36) vanishes. It follows easily that if $n_i > 1$ for at least one value of $i > 1$ the solution (36) is valid. On the other hand, if $n_i = 1$ for all $i > 1$, we can apply theorem 2 by taking $\sigma_0 = \sigma_{i1}$, $\lambda_1(\xi_1)$ as a step function with a single jump at ξ_{11} and the remaining $\lambda_i(\xi_i)$ according to lemma 1. It thus follows that for this problem no non-trivial test exists.

The following problem can be reduced to the hypothesis of equality of variances with means assumed known: Under the null hypothesis $X_1, \cdots, X_n$ have a joint multivariate normal distribution with density $C \exp\left[-\frac{1}{2\sigma^2} \Sigma\, a_{ij1}x_ix_j\right]$ where the a's are known and where σ is an unknown scale factor. Under H_1 the X's have a joint multivariate normal distribution with density $C' \exp\left[-\frac{1}{2}\Sigma b_{ij}x_ix_j\right]$. A number of hypotheses specifying the value of one or several correlation coefficients have this form. The most powerful test of H_0 against H_1 is given by

$$(36) \qquad \frac{\Sigma b_{ij}x_ix_j}{\Sigma a_{ij}x_ix_j}$$

as is easily shown by applying a non-singular linear transformation which reduces $\Sigma b_{ij}x_ix_j$ to diagonal form and $\Sigma a_{ij}x_ix_j$ to a sum of squares, or by applying directly the method of proof of the earlier problem.

A corresponding reduction when the X's have a common but unknown mean is usually impossible. One problem of this kind for which the solution is simple is the hypothesis specifying the value of a serial correlation coefficient in a circular population. The most powerful similar region for testing this hypothesis was obtained in [7]. Consider the probability density function

$$(37) \qquad C \exp\left[-\alpha \left\{\sum_{i=1}^{n} (x_i - \xi) - \delta(x_{i+1} - \xi)\right\}^2\right], \qquad (x_{n+1} = x_1), \qquad |\delta| < 1,$$

and let H_0 specify $\delta = \delta_0$ while H_1 assigns to the parameters the values α_1, ξ_1, δ_1. Then the most powerful test of H_0 against H_1 is

$$(38)\qquad \begin{aligned} \frac{\Sigma(x_i - \bar{x})(x_{i+1} - \bar{x})}{\Sigma(x_i - \bar{x})^2} &\geq k \quad \text{if} \quad \delta_1 > \delta_0; \\ \frac{\Sigma(x_i - \xi_1)(x_{i+1} - \xi_1)}{\Sigma(x_i - \xi_1)^2} &\leq k' \quad \text{if} \quad \delta_1 < \delta_0. \end{aligned}$$

We shall omit the proof of this result, since the method is the same as in the other problems considered in this section.

5. Student's hypothesis and some generalizations. As the principal result of the present section we shall prove that for testing Student's hypothesis against a simple alternative the most powerful test is a non-similar region of the form

$$(39)\qquad \Sigma(X_i - \eta)^2 \leq k,$$

if the level of significance ϵ is less than or equal to $\frac{1}{2}$. Here η and k depend on ϵ and on the alternative, and they will not be determined explicitly. It will be shown also that if ϵ is greater than or equal to $\frac{1}{2}$, Student's test is most powerful. These results will be extended rather easily to the general univariate linear hypothesis. The corresponding investigation for similar regions was carried through for Student's hypothesis by Neyman and Pearson [1] while the extension to a general linear hypothesis is contained in a paper by Hsu [13].

The proof of the main result mentioned above is rather lengthy. We shall begin by proving two lemmas.

LEMMA 2. *Let $Y_1, \cdots, Y_n$ be n independent random variables, normally distributed with 0 mean and unit variance, and let*

$$(40)\qquad \begin{aligned} P(a, k) &= P\left\{\sum_{i=1}^{n} (Y_i - a)^2 \leq (n - k)a^2\right\}; \\ \varphi(k) &= \sup_a P(a, k) \quad \textit{for} \quad 0 < k < n, \quad 0 < a. \end{aligned}$$

Then for each k there exists $a(k)$ such that

$$(41)\qquad P(a(k), k) = \varphi(k).$$

PROOF. If $Z_i = Y_i/a$, $(i = 1, \cdots, n)$ the Z's are independently normally distributed with zero mean and variance $1/a^2$ and (40) may be written as

$$(42)\qquad P(a, k) = P\{\Sigma(Z_i - 1)^2 \leq n - k\}.$$

Hence it is seen that for any k, $P(a, k)$ tends to zero as a tends to either zero or infinity. This proves the lemma since for any k, $P(a, k)$ is a continuous function of a.

LEMMA 3. *Given any ϵ, $0 < \epsilon < \frac{1}{2}$ there exists $k(\epsilon)$ between zero and n such that $\varphi(k(\epsilon)) = \epsilon$.*

Proof. The proof will be given in a number of steps.

(i) $\varphi(k) \to \frac{1}{2}$ as $k \to 0$.

Clearly $P(a, k)$ never exceeds $\frac{1}{2}$. The result will therefore follow if we exhibit a sequence a_k such that $P(a_k, k) \to \frac{1}{2}$ as $k \to 0$. Let $a_k = 1/\sqrt{k}$. Then

$$P(a_k, k) = P\{\sqrt{k}\,\Sigma Y_i^2 - 2\Sigma Y_i + \sqrt{k} \le 0\}. \tag{43}$$

The right hand side is a continuous function of k and therefore tends to

$$P\{\Sigma Y_i \ge 0\} = \tfrac{1}{2}, \tag{44}$$

as k tends to zero.

(ii) $\varphi(k) \to 0$ as $k \to n$.

Consider $P(a, k)$ as in (42). Written as an integral of the probability density of the Z's, the region of integration is independent of a and its volume tends to 0 as k tends to n. On the other hand the probability density depends on a but is uniformly bounded over the region of integration if $k > 0$, and hence the result follows.

(iii) If $0 < k_0$, $P(a, k)$ tends to zero uniformly for k in the interval $k_0 \le k \le n$ as a tends to zero or infinity.

This follows from the fact that $0 \le P(a, k) \le P(a, k_0)$ since $P(a, k_0)$ tends to 0 as a tends to zero or infinity.

(iv) Given k_0 and k_1 there exist numbers a_0 and a_1 with $0 < a_0 < a_1 < \infty$ such that $0 < k_0 \le k \le k_1 < n$ implies $a_0 \le a(k) \le a_1$.

If this were not true there would exist a sequence $k^{(i)}$ with $k_0 \le k^{(i)} \le k_1$ and $a(k^{(i)})$ tending to infinity or zero. Then $\varphi(a(k^{(i)}))$ would tend to zero by (iii). On the other hand consider $P(1, k)$ for $k_0 \le k \le k_1$. This is a continuous non-vanishing function of k and hence attains its lower bound m for some k in $k_0 \le k \le k_1$. Therefore m is positive and we have a contradiction.

(v) Given any k_0, k_1 with $0 < k_0 < k_1 < n$, $\varphi(k)$ is continuous on the interval $[k_0, k_1]$.

To see this, select a_0 and a_1 in accordance with (iv). Then $P(a, k)$ is uniformly continuous in the rectangle $a_0 \le a \le a_1$, $k_0 \le k \le k_1$. Given $\eta > 0$ let δ be such that $|k' - k''| < \delta$ implies $|P(a, k') - P(a, k'')| < \eta$. Then $\varphi(k') \ge P(a(k''), k') \ge P(a(k''), k'') - \eta = \varphi(k'') - \eta$, and by symmetry $\varphi(k'') \ge \varphi(k') - \eta$, which establishes the continuity of φ.

The proof of the lemma is now immediate. For let $0 < \epsilon < \frac{1}{2}$. It follows from (i) and (ii) that there exist k_0 and k_1 such that

$$\varphi(k_0) \le \epsilon/2, \qquad \varphi(k_1) \ge \epsilon + \tfrac{1}{2}(\tfrac{1}{2} - \epsilon),$$

and hence by (v) there exists $k(\epsilon)$ for which $\varphi(k(\epsilon)) = \epsilon$.

Let us now consider Student's hypothesis. The random variables $X_1, \cdots, X_n$ are a sample from a normal distribution which under H_0 has mean 0 and unknown variance σ^2, while under H_1 the mean is ξ_1 and the variance σ_1^2. Without loss of generality we shall assume $\xi_1 > 0$. Applying theorem 1 with λ a step-function having a single jump at a point $\sigma_0 > \sigma_1$ to be determined later, we obtain the critical region in the form

(45) $$\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) \Sigma X_i^2 - 2 \frac{\xi_1}{\sigma_1^2} \Sigma X_i \leq c.$$

Let $Y_i = X_i/\sigma$ so that under H_0 the Y's are distributed with zero mean and unit variance. Then (45) becomes

(46) $$\Sigma Y_i^2 - 2 \frac{\xi_1}{\sigma(1 - \sigma_1^2/\sigma_0^2)} \Sigma Y_i \leq \frac{c}{\sigma^2},$$

which may be written as

(47) $$\Sigma(Y_i - a)^2 \leq (n - k)a^2,$$

where

(48) $$a = \frac{\xi_1}{\sigma(1 - \sigma_1^2/\sigma_0^2)}, \qquad k = \frac{-c}{\xi_1^2}\left(1 - \frac{\sigma_1^2}{\sigma_0^2}\right)^2.$$

As σ varies from 0 to ∞, a goes from ∞ to 0. Let $P(a, k)$, $\varphi(k)$ and $a(k)$ be defined as in lemma 2. Given the level of significance ϵ $(0 < \epsilon < \frac{1}{2})$, let k^* and a^* be determined according to lemma 2 and 3 so that

(49) $$\varphi(k^*) = \epsilon \text{ and } P(a^*, k^*) = \varphi(k^*).$$

We now select $\sigma_0 > \sigma_1$ and c so that

(50) $$a^* = \frac{\xi_1}{(1 - \sigma_1^2/\sigma_0^2)\sigma_0} \quad \text{and} \quad k^* = \frac{c}{\xi_1^2}\left(1 - \frac{\sigma_1^2}{\sigma_0^2}\right).$$

We have to show that for this choice of σ_0 and c the size of the critical region attains its maximum when $\sigma = \sigma_0$ and that this maximum size is ϵ. Substituting from (50) we express the region (47) in the form

(51) $$\Sigma\left(Y_i - \frac{\sigma_0}{\sigma} a^*\right)^2 \leq (n - k^*) \frac{\sigma_0^2}{\sigma^2} a^{*2}.$$

Thus the probability of the region is

(52) $$P\left(\frac{\sigma_0}{\sigma} a^*, k^*\right).$$

As σ varies, (52) attains its maximum when $\frac{\sigma_0}{\sigma}a^* = a(k^*) = a^*$, that is, when $\sigma = \sigma_0$ and the maximum value of (52) is $\varphi(k^*) = \epsilon$.

This derivation is valid even when $n = 1$, i.e., when the hypothesis $\xi = 0$ is to be tested by observing only a single random variable X, known to be normally distributed but whose mean ξ and variance are unknown. For this problem no similar region exists. However, critical regions of the form $0 < \xi_1 - a < x < \xi_1 + b$ will give any level of significance $< \frac{1}{2}$ for proper choice of a and b, while the power of such regions will tend to 1 as σ_1 tends to 0. Therefore, the power of the most powerful test will be close to 1 if σ_1 is sufficiently small.

Having completed the discussion of the case $\epsilon < \frac{1}{2}$ let us next suppose that $\epsilon \geq \frac{1}{2}$. We shall need the following

LEMMA 4. *Let c and α_1 be positive constants. Then there exists a function f such that $f(\alpha) = 0$ when $\alpha < \alpha_1$ and such that for all $w > 0$*

$$\int_0^\infty e^{-\alpha w} f(\alpha)\, d\alpha = k e^{-\alpha_1 w - c\sqrt{w}}. \tag{53}$$

This follows from the well known representation of $e^{-c\sqrt{w}}$ as a Laplace transform by applying a translation. (53) can be checked directly by substituting

$$f(\alpha) = \frac{c e^{-(c^2/4(\alpha-\alpha_1))}}{(\alpha - \alpha_1)^{3/2}} \quad \text{for} \quad \alpha \geq \alpha_1. \tag{54}$$

Applying theorem 1 to Student's hypothesis, where again we shall assume ξ_1 to be positive, for proper choice of k we obtain from (9)

$$\frac{\exp\left[-\frac{1}{2\sigma_1^2}\Sigma X_i^2 + \frac{\xi_1}{\sigma_1^2}\Sigma X_i\right]}{\int_0^\infty \exp\left[-\frac{1}{2\sigma^2}\Sigma X_i^2\right]\frac{1}{\sigma^n}\, d\lambda(\sigma)} \geq 1. \tag{55}$$

It follows from lemma 4 that for any positive c there exists a non-decreasing function λ of bounded variation with $\lambda(\sigma)$ constant for $\sigma > \sigma_1$, such that

$$\int_0^\infty \exp\left[-\frac{1}{2\sigma^2}\Sigma X_i^2\right]\frac{1}{\sigma^n}\, d\lambda(\sigma) = \exp\left[-\frac{1}{2\sigma_1^2}\Sigma X_i^2 - c\sqrt{\Sigma x_i^2}\right]. \tag{56}$$

For this choice of λ, (55) reduces to

$$\exp\left[\frac{\xi_1}{\sigma_1^2}\Sigma x_i\right] \geq \exp\left[-c\sqrt{\Sigma x_i^2}\right], \tag{57}$$

and hence to

$$\frac{\Sigma x_i}{\sqrt{\Sigma x_i^2}} \geq c'. \tag{58}$$

This is a similar region and therefore most powerful for testing Student's hypothesis against H_1. By adjusting c, the size of the region can be made equal to any $\epsilon \geq \frac{1}{2}$.

The argument for $\epsilon > \frac{1}{2}$ must be modified slightly in the case $n = 1$, that is, when we want to test Student's hypothesis on the basis of a single observation. Let us adjoin to the variable X a random variable Y known to be uniformly distributed over the interval [0, 1]. Using the same λ and k as before, (58) becomes

(59)

For $c' = -1$ the critical region includes all points (x, y) for which x is positive while (59) places no restriction on which of the remaining points to include in the critical region. The similar region

$$(60) \qquad x \geq 0; \qquad x < 0, \qquad 0 < y < 2(\epsilon - \tfrac{1}{2})$$

therefore satisfies all conditions of theorem 1 and hence is most powerful

In extending these results to the general linear hypothesis, we shall assume the hypothesis reduced to canonical form [14, 15]. We shall therefore assume that $X_1, \cdots, X_n$ are normally distributed with common variance which is unknown under H_0 and has the value σ_1^2 under H_1. Furthermore, under H_0, $E(X_i) = 0$ for $i = 1, \cdots, s, s+1, \cdots, m$; $E(X_i)$ unknown for $i = m+1, \cdots, n$ while under H_1 $E(X_i) = 0$ for $i = s+1, \cdots, m$; $E(X_i) = \xi_{i1}$ for the remaining values of i.

For $\epsilon < \frac{1}{2}$ we shall consider critical regions of the form

$$(61) \qquad \frac{\exp\left\{-\frac{1}{2\sigma_1^2}\left[\sum_{i=1}^{s}(x_i - \xi_{i1})^2 + \sum_{i=s+1}^{m} x_i^2 + \sum_{i=m+1}^{n}(x_i - \xi_{i1})^2\right]\right\}}{\exp\left\{-\frac{1}{2\sigma_1^2}\left[\sum_{i=1}^{s} x_i^2 + \sum_{i=s+1}^{m} x_i^2 + \sum_{i=m+1}^{n}(x_i - \xi_{i1})^2\right]\right\}} \geq k,$$

which are obtained from (8) by substituting for λ a step-function with a single jump at the parameter point $(\sigma_i, \xi_{m+1,1}, \cdots, \xi_{n,1})$. Making an orthonormal transformation from $x_s, \cdots, x_s$ to $y_1, \cdots, y_s$ such that $y_1 = \frac{\sum_{i=1}^{s} \zeta_{i1} x_i}{\sqrt{\Sigma \zeta_{i1}^2}}$ and letting $y_i = x_i$ for $i = s+1, \cdots, m$; $y_i = x_i - \xi_{i1}$ for $i = m+1, \cdots, n$, (61) reduces to

$$(62) \qquad \frac{\exp\left\{-\frac{1}{2\sigma_1^2}\left[\sum_{i=1}^{n} y_i^2 - 2y_1\sqrt{\sum_{i=1}^{s}\xi_{i1}^2}\right]\right\}}{\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n} y_i^2\right\}} \geq c.$$

For $\sigma_0 > \sigma_1$ we can rewrite (62) as

$$(63) \qquad \sum_{i=m+1}^{n} y_i^2 \leq \frac{1}{\sigma_1^2 - \sigma_0^2}\left[c + \frac{2y_1\sqrt{\Sigma\xi_{i1}^2}}{\sigma_1^2}\right] - \sum_{i=1}^{m} y_i^2,$$

and we see that under H_0 for any σ the size of this region considered as a function of the unknown means of $Y_{m+1}, \cdots, Y_n$ takes on its maximum when these means are zero, i.e. when $\xi_i = \xi_{i1}$ for $i = m+1, \cdots, n$. For these maximizing values of the means the existence of a suitable σ_0 and c follows from the corresponding result in connection with Student's hypothesis.

Thus the most powerful test for testing H_0 against H_1 at level of significance $\epsilon = \frac{1}{2}$ has the form

$$(64) \qquad \sum_{i=1}^{s}\left[x_i - \frac{\xi_{i1}}{1 - \sigma_1^2/\sigma_0^2}\right]^2 + \sum_{i=s+1}^{m} x_i^2 + \sum_{i=m+1}^{n}(x_i - \xi_{i1})^2 \leq c.$$

It is interesting that the variables $X_i(i = m + 1, \cdots, n)$ which may be discarded when considerations are restricted to similar regions [18], do contribute to the power when similarity is not required. The same phenomenon also occurs in certain problems considered earlier in this paper.

For the case $\epsilon \geq \frac{1}{2}$, let us take

$$\lambda(\sigma, \xi_{m+1}, \cdots, \xi_n) = \lambda(\sigma) \prod_{i=m+1}^{n} \lambda_i(\xi_i \mid \sigma). \tag{65}$$

We shall select $\lambda(\sigma)$ such that $\lambda(\sigma)$ is constant when $\sigma \geq \sigma_1$. Hence it is enough to define $\lambda_i(\xi_i \mid \sigma)$ for $\sigma < \sigma_1$. For any $\sigma < \sigma_1$ there exists by lemma 1 a function $\lambda_i(\xi_i \mid \sigma)$ such that

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}(x_i - \xi_i)^2\right] d\lambda_i(\xi_i \mid \sigma) = k \exp\left\{-\frac{1}{2\sigma_1^2}(x_i - \xi_{i1})^2\right\}. \tag{66}$$

For this choice of the λ_i, (9) becomes

$$\frac{\exp\left\{-\frac{1}{2\sigma_1^2}\left[\sum_{i=1}^{s}(x_i - \xi_{i1})^2 + \sum_{i=s+1}^{m} x_i^2\right]\right\}}{\int_0^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m} x_i^2\right\} d\lambda\ (\sigma)} \geq k'. \tag{67}$$

Next we chose $\lambda(\sigma)$ according to lemma 4 such that

$$\int_0^{\infty} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{m} x_i^2\right]\frac{1}{\sigma^m}\, d\lambda(\sigma) = \exp\left[-\frac{1}{2\sigma_1^2}\sum_{i=1}^{m} x_i^2 - c\sqrt{\sum_{i=1}^{m} x_i^2}\right], \tag{68}$$

thus, by proper choice of k', reducing (67) to

$$\frac{\sum_{i=1}^{s} \xi_{i1}\, x_i}{\sqrt{\sum_{i=1}^{m} x_1^2}} \geq -c. \tag{69}$$

The probability of this region under H_0 is independent of $\xi_{m+1}, \cdots, \xi_n$ and σ, and hence (69) is most powerful for testing H_0 against H_1.

Let us return once more to the problem of testing Student's hypothesis against a simple alternative $\xi = \xi_1$, $\sigma = 1$ and let us assume as known that $\sigma \leq 1$. No use can be made of this knowledge if consideration is restricted to similar regions. For the probability of first kind error is an analytic function of σ, and consequently, if a test is similar with respect to all values of σ which are ≤ 1, it is similar with respect to all values of σ. Let us now consider this problem without the restriction of similarity. If $\epsilon \geq \frac{1}{2}$, the knowledge concerning σ does not enable us to find a test which is more powerful than that given by (58), since the function $\lambda(\sigma)$ on which (58) was based had all its points of increase for $\sigma \leq 1$.

On the other hand we may expect improvement for $\epsilon < \frac{1}{2}$ since the most powerful test in this case was based on a function λ with a single point of increase

$\sigma_0 > 1$ which is no longer admitted as a possible value of σ. If, instead, we take for λ the step function with a single jump at $\sigma = 1$ we obtain the critical region

$$\frac{\exp\left[-\frac{1}{2}\sum(x_i - \xi_1)^2\right]}{\exp\left[-\frac{1}{2}\sum x_i^2\right]}, \tag{70}$$

which is equivalent to

$$\bar{x} \geq c. \tag{71}$$

Here $c > 0$ since $\epsilon < \frac{1}{2}$, and therefore, when $\xi = 0$ the probability of (71) is an increasing function of σ and hence takes on its maximum at $\sigma = 1$. It follows from theorem 1 that (71) is most powerful under the conditions stated.

In the opposite problem in which it is known that $\sigma \geq 1$, the situation is reversed. For $\epsilon \leq \frac{1}{2}$ no improvement over (45) is possible while for $\epsilon > \frac{1}{2}$ we can use for λ the step function with a single step at $\sigma = 1$ thus obtaining the critical region (70) but this time with $c < 0$. When $\xi = 0$ the probability of this region is a decreasing function of σ and it follows that (70) is most powerful in this case.

Similar remarks apply to other problems. We mention as one further example a modification of the Behrens-Fisher problem. Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ be independently normally distributed, the X's with mean ξ and variance σ^2, the Y's with mean η and variance τ^2, all four parameters being unknown. We wish to test, at level of significance $\epsilon \leq \frac{1}{2}$, the hypothesis $\xi = \eta$ against the simple alternative $\xi = \xi_1$, $\eta = \eta_1$, $\sigma = 1$, $\tau = 1$, where $\xi_1 \neq \eta_1$ and we assume it known that $\sigma \leq 1$, $\tau \leq 1$. Basing the test on a step function λ with a single jump at $\sigma = 1$, $\tau = 1$, $\xi = \dfrac{n\xi_1 + m\eta_1}{m + n}$ we obtain for w the region

$$\frac{\exp\left[-\frac{1}{2}\sum(x_i - \xi_1)^2 - \frac{1}{2}\sum(y_i - \eta_1)^2\right]}{\exp\left[-\frac{1}{2}\sum\left(x_i - \frac{n\xi_1 + m\eta_1}{n + m}\right)^2 - \frac{1}{2}\sum\left(y_i - \frac{n\xi_1 + m\eta_1}{n + m}\right)^2\right]} \geq k, \tag{73}$$

which is equivalent to

$$\bar{y} - \bar{x} \geq c \qquad (c > 0), \tag{74}$$

if we assume, as we may without loss of generality, that $\eta_1 > \xi_1$. When $\eta = \xi_2$, $\bar{Y} - \bar{X}$ is normally distributed with zero mean and variance $\dfrac{\sigma^2}{n} + \dfrac{\tau^2}{m}$. Therefore the probability of (74) is an increasing function of $\dfrac{\sigma^2}{n} + \dfrac{\tau^2}{m}$ and hence attains its maximum when $\sigma = \tau = 1$. It follows from theorem 1 that the region (74) is most powerful for the problem under consideration.

6. Admissibility. The general problem to be considered in this paper has been formulated in section 1: To obtain a region w

(75) $$\text{maximizing} \quad \int_w g(x)\, dx$$

subject to the restriction

(76) $$\int_w f_\theta(x)\, dx \le \epsilon \qquad \text{for all } \theta \in \omega.$$

Since for any particular such problem there may exist several essentially different regions satisfying these conditions, it may happen that there exists a region w' such that

(77) $$\int_{w'} g(x)\, dx = \int_w g(x)\, dx,$$

and

(78) $$\int_{w'} f_\theta(x)\, dx \le \int_w f_\theta(x)\, dx \qquad \text{for all } \theta \in \omega,$$

with inequality holding for some θ. Clearly w' is preferable to w. In this case, following the definition of Wald [4], we say that w is not admissible. We shall rule out this possibility for a large class of problems by proving

THEOREM 3. *If w satisfies the conditions of theorem 1, and if the set of points x for which equality holds in (8) has measure zero, then any region satisfying (75) and (76) differs from w only on a set of measure zero.*

PROOF. Without loss of generality we shall assume λ of theorem 1 to be a distribution function. Then

$$h(x) = \int_\omega f_\theta(x)\, d\lambda(\theta)$$

is a completely specified probability density function, and w is the unique[3]—up to a set of measure zero—most powerful test for testing the simple hypothesis H_0': h against the simple alternative H_1: g. Suppose now that w' satisfies (75) and (76). Then

(79) $$\int_{w'} h(x)\, dx \le \epsilon,$$

and w' is most powerful for testing H_0' against H_1. It follows that w' differs from w at most by a null set.

Earlier we enlarged the problem of testing by adjoining to the original random variable X a random variable with a known distribution. This is equivalent to the following modification of the original problem. Instead of defining a test to be a critical region (of rejection) in the space of x, we define it to be a critical

[3] One sees this easily from Neyman and Pearson's proof of the fundamental lemma [1], by using the assumption that the set of points for which equality holds in (8), has measure zero.

function φ $(0 \leq \varphi(x) \leq 1)$ which with every point x associates a probability of rejection $\varphi(x)$. If x is observed, the hypothesis is rejected with probability $\varphi(x)$ according to a table of random numbers. In the case where random numbers are not employed, φ merely becomes the characteristic function of the set w.

We shall now state a theorem which will prove admissibility for all but one of those problems treated in sections two to five, to which theorem 3 does not apply.

THEOREM 4. *Suppose $\omega = \{\theta\}$ is a subset of an s-dimensional Euclidean space, and that for any measurable function φ and for any set S which has positive measure and is contained in ω*

$$\int \varphi(x) f_\theta(x)\, dx = c \qquad \text{for } \theta \in S \tag{80}$$

implies

$$\int \varphi(x) f_\theta(x)\, dx = c \qquad \text{for } \theta \in \omega. \tag{81}$$

(Here and in all that follows whenever a region of integration is not indicated, the integral extends over the whole x space). Suppose further that φ is a critical function satisfying the conditions of theorem 1 and that the set S_0 of points of increase of λ has positive measure. Then φ is admissible.

PROOF. If φ were not admissible there would exist φ_1 with

$$\int \varphi_1(x)\, g(x)\, dx = \int \varphi(x) g(x)\, dx; \tag{82}$$

$$\int \varphi_1(x) f_\theta(x)\, dx \leq \int \varphi(x) f_\theta(x)\, dx \qquad \text{for all } \theta \in \omega; \tag{83}$$

$$\int \varphi_1(x) f_\theta(x)\, dx < \int \varphi(x) f_\theta(x)\, dx \qquad \text{for some } \theta \in \omega. \tag{84}$$

The set T of points θ for which (84) holds, differs from ω at most by a null set. For

$$\int [\varphi_1(x) - \varphi(x)] f_\theta(x)\, dx = 0 \qquad \text{for } \theta \in \omega - T, \tag{85}$$

and if $\omega - T$ had positive measure, (85) would hold for all $\theta \in \omega$.

Let h and H_0' be defined as in the proof of theorem 3. Since S has positive measure, it follows that

$$\epsilon = \int \varphi(x) h(x)\, dx > \int \varphi_1(x) h(x)\, dx = \eta, \qquad \text{say.} \tag{86}$$

Let $\varphi_2(x) = \min\left[1, \varphi_1(x) + \epsilon - \eta\right]$. Then

$$\int \varphi_2(x) h(x)\, dx \leq \epsilon \tag{87}$$

and

$$\int \varphi_2(x)g(x)\,dx > \int \varphi_1(x)g(x)\,dx. \tag{88}$$

But φ_1 is most powerful for testing H_0' against H_1 and we have a contradiction.

By applying theorems 3 and 4 one can easily show for all but one of the problems treated in sections three to five that the tests obtained there are admissible. The one exception occurs when testing equality of variances. Simplifying the notation, since we are now concerned with a special case, we shall assume that $X_i (i = 1, \cdots, n)$, $Y_1, \cdots, Y_r$ are independently and normally distributed, the X's with mean ξ_0 and variance σ_0^2, Y_i with mean ξ_i and variance σ_i^2, all parameters being unknown. We wish to test the hypothesis of equality of variances against the simple alternative

$$H_1 : \xi_i = \xi_{i1}, \qquad \sigma_i = \sigma_{i1} \qquad (i = 0, \cdots, r),$$

with

$$\sigma_{01} < \sigma_{11} < \cdots < \sigma_{r1}.$$

We shall first consider the case $n = 1$, and prove admissibility of the critical function

$$\varphi(x, y_1, \cdots, y_r) = \epsilon \tag{89}$$

by using a different distribution function for the parameters from the one used earlier. With some specialization of the distribution function, (8) becomes for our problem

$$\frac{\exp\left\{-\frac{1}{2\sigma_{01}}(x-\xi_{01})^2 - \frac{1}{2}\sum_{i=1}^{r}\frac{1}{\sigma_{i1}^2}(y_i-\xi_{i1})^2\right\}}{\int \frac{1}{\sigma^{r+1}}\left\{\int \exp\left[-\frac{1}{2\sigma^2}(x-\xi_0)^2\, d\lambda_\sigma^{(0)}(\xi_0)\right] \cdot \prod_{i=1}^{r}\int \exp\left[-\frac{1}{2\sigma^2}(y_i-\xi_i)^2\right] d\lambda_\sigma^{(i)}(\xi_i)\right\} d\,\mu(\sigma)} \geq k \tag{90}$$

For any $\sigma < \sigma_{01}$ we select the $\lambda_\sigma^{(i)}(\xi_i)$ according to lemma 1. If we then take for μ the uniform distribution over $(\sigma_{01} - 1, \sigma_{01})$ the left hand side of (90) will reduce to k. Admissibility of the critical function (89) then follows from theorem 4.

That a constant critical function is not admissible in the case $n > 1$ is easily seen if one compares it for instance with the critical region

$$\left| \frac{\bar{x} - \xi_{01}}{\sqrt{\Sigma(x_1 - \bar{x})^2}} \right| \leq c. \tag{91}$$

We shall not obtain a complete family of admissible tests (cf. [4]) for the case $n > 1$ but we shall show that this problem is equivalent to the following one: To find a complete class of unbiased admissible tests for the hypothesis specifying the mean and variance of a normal distribution on the basis of a sample from this distribution, the class of alternatives being the totality of univariate normal distributions.

Let $n > 1$ and let φ be any most powerful critical function for testing the hypothesis of equality of variances against H_1. If φ corresponds to the level of significance ϵ and if β_φ denotes the power of φ, we have

$$\beta_\varphi(\sigma, \sigma, \cdots, \sigma, \xi_0, \xi_1, \cdots, \xi_r) \leq \epsilon \tag{92}$$

for all admissible values of the arguments. It also follows from section 4 that

$$\beta_\varphi(\sigma_{01}, \sigma_{11} \cdots, \sigma_{r1}, \xi_{01}, \xi_{11}, \cdots, \xi_{r1}) = \epsilon. \tag{93}$$

Consider for a moment the hypothesis $H_0' : \sigma_i = \sigma_{01} (i = 0, \cdots, r)$, $\xi_0 = \xi_{01}$, ξ_i unspecified for $i = 1, \cdots, r$. It is easily seen that the maximum power for testing H_0' against H_1 is ϵ. Therefore any most powerful test for testing H_0 against H_1 is also most powerful for testing H_0' against H_1, and in particular this holds for φ. Furthermore, it follows easily from theorem 4 that for any most powerful test of H_0' against H_1 the probability of an error of the first kind must be identically equal to ϵ. Therefore

$$\beta_\varphi(\sigma_{01}, \cdots, \sigma_{01}, \xi_{01}, \xi_1, \cdots, \xi_r) = \epsilon \text{ for all } \xi_1, \cdots, \xi_r. \tag{94}$$

But (94) is equivalent to the condition that φ is similar with respect to $\xi_1, \cdots, \xi_r$, and it follows [12] that φ is a function of $x_1, \cdots, x_n$ only. The problem is therefore reduced to that of finding all admissible critical functions $\varphi(x_1, \cdots, x_n)$ satisfying

$$\beta_\varphi(\sigma_{01}, \xi_{01}) = \epsilon; \qquad \beta_\varphi(\sigma_0, \xi_0) \leq \epsilon \text{ for all } \sigma_0, \xi_0. \tag{95}$$

That this problem in turn is equivalent to the one stated above is immediate when one considers the complementary critical functions $1 - \varphi$.

REFERENCES

[1] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. Phil. Trans.*, Ser. A, Vol. 231 (1933), p. 289.

[2] A. Wald, "Test of statistical hypotheses concerning several parameters when the number of observations is large," *Am. Math. Soc. Trans.*, Vol. 54 (1943), p. 426.

[3] A. Wald, "Statistical decision functions which minimize the maximum risk," *Annals of Math.*, Vol. 46 (1945), p. 265.

[4] A. Wald, "An essentially complete class of admissible decision-functions," *Annals of Math. Stat.*, Vol. 18 (1947), p. 549.

[5] J. Neyman, "On a statistical problem arising in routine analysis and in sampling inspection of mass production," *Annals of Math. Stat.*, Vol. 12 (1941), p. 46.

[6] H. Scheffé, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 13 (1942), p. 280.

[7] E. Lehmann, "On optimum tests of composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 18 (1947), p. 473.

[8] E. Lehmann and H. Scheffé, "On the problem of similar regions," *Proc. Nat. Acad. Sci.*, Vol. 33 (1947), p. 382.

[9] A. WALD, "On the power function of the analysis of variance test," *Annals of Math. Stat.*, Vol. 13 (1942), p. 434.
[10] P. L. HSU, "On the power function of the E^2-test and the T^2-test," *Annals of Math. Stat.*, Vol. 16 (1945), p. 278.
[11] H. K. NANDI, "On the average power of test criteria," *Sankhyā*, Vol. 8 (1946), p. 67.
[12] W. FELLER, "Note on regions similar to the sample sapce," *Stat. Res. Memoirs*, Vol. 2 (1938), p. 117.
[13] P. L. HSU, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), p. 62.
[14] S. KOLODZIECZYK, "On an important class of statistical hypotheses," *Biometrika*, Vol. 27 (1935), p. 161.
[15] P. C. TANG, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. 2 (1938), p. 126.
[16] A. WALD, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947) p. 279.