ON THE RATIO OF THE VARIANCES OF TWO NORMAL POPULATIONS

By Henry Scheffé

Princeton University

CONTENTS

1. Introduction and summary. Suppose that we have two samples E_1 and E_2 from normal populations π_1 and π_2 with unknown means and variances. Let us designate by θ the ratio of the variance of π_1 to that of π_2. *The two problems discussed in this paper are to formulate in terms of E_1 and E_2, and to compare,*

(*i*) *significance tests for the hypothesis that the unknown ratio θ is equal to a given positive number θ_0*, and

(*ii*) *confidence intervals for θ.*

Since, on the one hand, these problems are of considerable importance to the practical statistician and the teacher of statistics, and on the other, they cry for the application of recently developed theory which is unfortunately not yet familiar to many practical workers and teachers, the development has been divided into two parts: Part I, it is hoped, will be intelligible to the above class of readers; part II, slanted toward a smaller circle, is more esoteric, general, and condensed.

More specifically, in part I it is pointed out that any choice of limits on the F-distribution satisfying the condition that the sum of the areas in the tails be equal to a prescribed number, leads to solutions of problems (*i*) and (*ii*). After considering and then ruling out the "one-sided" situations in which it is appropriate to use only one tail, two conditions are proposed (*ad hoc* and on an intuitive basis) for the "two-sided" case,—a symmetry condition, and a condition for logarithmically shortest confidence intervals. The second condition leads to a choice of limits on the F-distribution. From other considerations,—

reciprocal limits, likelihood ratio, and equal tails,—other choices are advanced. It is found that all four of these choices satisfy the first condition, and that furthermore if $N_1 = N_2$, where N_i is the number of variates in E_i, then the four choices become identical. If $N_1 \neq N_2$ which of the four tests is "best"? which of the four sets of confidence intervals? For defining and answering the first question in a logically satisfactory way just a little of the Neyman-Pearson theory of testing hypotheses suffices. For the second, Neyman's theory of confidence intervals is called for, and because of its greater difficulty, this has been relegated to part II. However, the limits determined by the criterion that the test be unbiased turn out to be the same as those which yield optimum confidence intervals from the elementary viewpoint of §5. Their numerical values are unfortunately laborious to calculate accurately if $N_1 \neq N_2$, and part I concludes with some numerical evidence indicating the loss of efficiency in using instead the easily found "equal tails" limits. For N_1 and $N_2 \geqq 10$ this loss is seen to be quite small. It will perhaps bear repeating that if $N_1 = N_2$, the "equal tails" limits on the F-distribution are the same as those associated with the unbiased test and that hence in this case all the advantages uncovered in parts I and II for the unbiased test and the related confidence intervals are obtained by using the easily available "equal tails" limits.

In part II we drop the restriction that the tests be based on a one or two-tailed use of the F-distribution. By a slight extension of results of Neyman and Pearson, common best critical regions for testing the hypothesis $\theta = \theta_0$ against alternatives $\theta < \theta_0$, or $\theta > \theta_0$, are found. Since the regions are always distinct for these two "one-sided" cases, there is no uniformly most powerful test. In order to find the most efficient unbiased test some recently published theorems of the writer are applied to prove that the critical region of the unbiased test proposed in part I is of type B_1.

The fact that the results summarized in the above paragraph are obtained for arbitrary positive θ_0 will immediately suggest to the reader familiar with Neyman's theory of confidence intervals that it may be easy on the basis of those results to draw conclusions about the existence of Neyman's various categories of confidence intervals. It is. In particular we find that the set of confidence intervals arrived at in §5 constitutes Neyman's short unbiased set.

The writer is aware that not all the results of this paper are new, and hopes he has given credit where it is due, but believes it desirable to bring together all the results, old and new, in this attempt to clean up the problems (*i*) and (*ii*). He is pleased to acknowledge his debt to Mr. David Votaw for aiding in the calculations for fig. 1 and for finding the formulas (6).

Part I. Significance Tests and Confidence Intervals Based on the F-Distribution

2. The F-distribution. The sample $E_i : (x_{i1}, x_{i2}, \cdots, x_{iN_i})$, $i = 1, 2$, is assumed to be from a normal population π_i with mean a_i and variance σ_i^2. We write $\theta = \sigma_1^2/\sigma_2^2$, and might regard the statistic T as an estimate[1] of θ, where $T = s_1^2/s_2^2$ and

$$s_i^2 = \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)^2/n_i, \qquad \bar{x}_i = \sum_{j=1}^{N_i} x_{ij}/N_i, \qquad n_i = N_i - 1.$$

It will be convenient to consider θ, σ_2^2, a_1, a_2 as the population parameters, σ_1^2 being eliminated from the joint p.d.f. (probability density function) of E_1 and E_2 by the substitution $\sigma_1^2 = \theta\sigma_2^2$. For any given positive number θ_0 we define the composite hypothesis

$$H_0: \quad \theta = \theta_0, \quad 0 < \sigma_2^2 < +\infty, \quad -\infty < a_1 < +\infty, \quad -\infty < a_2 < +\infty.$$

In Hotelling's apt terminology the last three parameters are nuisance parameters.

It is well known that U_1 and U_2, where $U_i = n_i s_i^2/\sigma_i^2$, are independently distributed according to χ^2-laws with n_1 and n_2 degrees of freedom respectively, and that hence the quotient $F = (U_1/n_1) \div (U_2/n_2) = T/\theta$ has the F-distribution $h_{n_1 n_2}(F)\, dF$ with n_1 and n_2 degrees of freedom, where

$$h_{n_1 n_2}(u) = \frac{(n_1/n_2)^{\frac{1}{2}n_1}}{\mathrm{B}(\frac{1}{2}n_1, \frac{1}{2}n_2)} u^{\frac{1}{2}n_1 - 1}\left(1 + \frac{n_1}{n_2} u\right)^{-\frac{1}{2}(n_1+n_2)}, \qquad 0 \leqq u \leqq \infty.$$

For later reference we note that if we define the variable x from

$$F = \frac{n_2}{n_1} \frac{x}{1 - x}, \tag{1}$$

then the cumulative distribution function of x is the incomplete Beta function[2] $I_x(\frac{1}{2}n_1, \frac{1}{2}n_2)$.

Let α be any number such that $0 < \alpha < 1$ (α will be the significance level for (i); $1 - \alpha$, the confidence coefficient for (ii)). The symbols $A_{n_1 n_2}$, $B_{n_1 n_2}$ will always denote a pair of numbers for which[3]

$$\int_{A_{n_1 n_2}}^{B_{n_1 n_2}} h_{n_1 n_2}(u)\, du = 1 - \alpha. \tag{2}$$

Every choice of the pair A, B leads to a solution of problems (i) and (ii):

(i). *A test of H_0 at significance level α consists of rejecting H_0 if $T < A_{n_1 n_2}\theta_0$ or $T > B_{n_1 n_2}\theta_0$.*

The probability of rejecting H_0 if it is true is

$$1 - Pr(A\theta_0 \leqq T \leqq B\theta_0 \mid \theta_0) = 1 - Pr(A < T/\theta_0 < B \mid \theta_0) = \alpha,$$

independently of the true values of the nuisance parameters.

---

[1] Biased.

[2] All the results of this paper pertaining to the F-distribution could of course be stated in terms of Fisher's z-distribution [2] or the incomplete Beta distribution; the first is used here because of its popularity in applied statistics, and because it permits the simplest statements for solutions of problems (i) and (ii).

[3] Superscripts on A, B will signify that a further condition has been laid on the pair A, B. The subscripts will be dropped when there is no danger of confusion. We permit $B = \infty$ as a possible choice.

(ii). A set of confidence intervals for θ with confidence coefficient $1 - \alpha$ *is*[4]

$$T/B_{n_1n_2} \leqq \theta \leqq T/A_{n_1n_2} .$$

The probability that the true value of θ will be covered by the above random interval is

$$Pr(T/B \leqq \theta \leqq T/A \mid \theta) = Pr(A \leqq T/\theta \leqq B \mid \theta) = 1 - \alpha,$$

whatever be the true values of θ and the nuisance parameters.

It will be convenient to adopt a brief notation for the tests and confidence intervals determined by certain choices of the limits A, B. In the sequel we shall denote these choices by $A^i_{n_1n_2}$, $B^i_{n_1n_2}$, where $i = \text{I, II}, \cdots, \text{VI}$. We shall call the significance test based on the pair A^i, B^i the *test i*, and the set of confidence intervals based on this pair, the *set i of confidence intervals*, or sometimes more briefly, the *confidence intervals i*.

3. Use of one tail. Suppose a situation in which we do not mind accepting H_0 if the true value of θ exceeds θ_0, but we desire a test which is as sensitive as possible in rejecting H_0 when $\theta < \theta_0$. It can be shown (for $n_2 > 2$) that the expected value of T is $\mathcal{E}(T) = n_2\theta/(n_2 - 2)$, and hence when the true value of θ is small compared with θ_0, so is $\mathcal{E}(T)$. By the usual intuitive considerations we are led to rejecting H_0 if $F = T/\theta_0$ falls in the left tail of the F-distribution. To make the significance level equal to α we take the limits A, B so that

$$\int_0^{A^{\text{I}}_{n_1n_2}} h_{n_1n_2}(u)\, du = \alpha, \qquad B^{\text{I}}_{n_1n_2} = \infty.$$

Similarly, to test H_0 against alternatives $\theta > \theta_0$ we define test II by

$$A^{\text{II}}_{n_1n_2} = 0, \qquad \int_{B^{\text{II}}_{n_1n_2}}^{\infty} h_{n_1n_2}(u)\, du = \alpha.$$

Why test I is best for testing H_0 against alternatives $\theta < \theta_0$, and test II for $\theta > \theta_0$, will be explained more convincingly in §9.

The confidence intervals I and II are then semi-infinite. It is apparent that if we are not loath to accept large values of θ but wish to exclude the largest possible interval of small values $(0, T/B)$, we should use the set II. Indeed, the set II is optimum in the case where we are willing to accept values of θ larger than the true value but desire the highest possible probability of excluding any values less than the true value; however, the precise formulation and proof of this statement must be postponed to part II. Analogous remarks apply to the set I and a willingness to accept values of θ less than the true value.

For $\alpha = .05$ or $.01$ the values of $B^{\text{II}}_{n_1n_2}$ are given in Snedecor's F-tables [12;

[4] If $B = \infty$ we omit the equality sign to the left of θ, if $A = 0$, the equality sign to the right of θ.

same n_1, n_2 as ours], and the values of $A^{\mathrm{I}}_{n_1n_2}$ may be calculated from the same tables by using the relation

$$A^{\mathrm{I}}_{n_1n_2} = 1/B^{\mathrm{II}}_{n_2n_1}. \tag{3}$$

$A^{\mathrm{I}}_{n_1n_2}$ for $\alpha = .50, .25, .10, .025, .005$ may be obtained by use of the transformation (1) and Thompson's new tables [13] of percentage points for the incomplete Beta distribution. $B^{\mathrm{II}}_{n_1n_2}$ for these values of α can then be found from (3).

4. Symmetry condition. We now restrict our attention (until §9) to the "two-sided" situation in which we are interested in all alternatives to $\theta = \theta_0$ on the range $0 < \theta < \infty$. Let us contemplate the following *symmetry condition*:

$$A_{n_1n_2} = 1/B_{n_2n_1} \tag{4}$$

for all positive integers n_1, n_2. The desirability of this condition and that of §5 follows not from mathematical principles but from practical considerations which might be relevant whenever significance tests or confidence intervals are considered for a parameter θ which is the quotient of two other positive parameters θ_1 and θ_2, and the estimate of θ is the quotient of the estimates of θ_1 and θ_2.

Suppose that given the samples E_1 and E_2, computer $\mathcal{C}_1$ labels them 1, 2, the same way we have, and using our test of §2, rejects the hypothesis that $\sigma_1^2/\sigma_2^2 = k$ unless

$$A_{n_1n_2}k \leqq s_1^2/s_2^2 \leqq B_{n_1n_2}k;$$

while computer $\mathcal{C}_2$ labels them 2, 1, and following a similar rule rejects $\sigma_2^2/\sigma_1^2 = 1/k$ (in our notation) unless

$$A_{n_2n_1}/k \leqq s_2^2/s_1^2 \leqq B_{n_2n_1}/k.$$

It will be seen that (4) is merely the condition that they reach the same conclusion. This makes life simpler, at least for computers and consulting statisticians. Likewise, if $\mathcal{C}_1$ and $\mathcal{C}_2$ use the confidence intervals of §2, then they will make numerically equivalent statements about σ_1^2/σ_2^2 and σ_2^2/σ_1^2 if (4) is satisfied.

5. Logarithmically shortest confidence intervals. The length of the confidence intervals of §2 is $L = T(A^{-1} - B^{-1})$. We might consider choosing A, B in such a way that $\mathcal{E}(L)$ is minimum. This leads to the problem of minimizing $A^{-1} - B^{-1}$ subject to (2). It might seem just as desirable, however, to minimize the expected length of the confidence interval for $\theta^{\frac{1}{2}}$,

$$(T/B)^{\frac{1}{2}} \leqq \sigma_1/\sigma_2 \leqq (T/A)^{\frac{1}{2}}.$$

This leads to a different problem with a different solution.

The condition on confidence intervals for θ which appears intuitively desirable to the writer, is that the limits $\underline{\theta}$, $\bar{\theta}$ of the confidence interval $\underline{\theta}(E_1, E_2) \leqq \theta \leqq \bar{\theta}(E_1, E_2)$ be such that $\mathcal{E}(\log \bar{\theta} - \log \underline{\theta})$ is minimum. For the confidence intervals of §2 this is equivalent to minimizing B/A, and by using the method of Lagrange's multipliers we easily find that

$$[uh_{n_1n_2}(u)]_{u=A}^{B} = 0 \tag{5}$$

and (2) must be satisfied. Denote the solution[5] by $A_{n_1n_2}^{\mathrm{III}}$, $B_{n_1n_2}^{\mathrm{III}}$. It is evident that the same condition (5) is obtained if we ask for logarithmically shortest confidence intervals (based on the F-distribution) for θ^k where $k > 0$.

The numerical values of the limits A^{III}, B^{III} are difficult to calculate if $n_1 \neq n_2$. The best procedure seems to be to transform to the incomplete Beta distribution by means of (1) and to calculate the corresponding points $a_{n_1n_2}^{\mathrm{III}}$, $b_{n_1n_2}^{\mathrm{III}}$ from the equations

$$[I_x(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2)]_{x=a}^{b} = [I_x(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2)]_a^b = 1 - \alpha. \tag{6}$$

The points a, b can be found to two decimals by inspection of Pearson's tables [9]. Unfortunately, in the many cases where a is close to 0, or b to 1, A^{III}, B^{III} are then subject to enormous error when calculated from (1).

6. Reciprocal limits. While the problems (*i*) and (*ii*) are closely related, the last choice of limits was suggested solely by our consideration of (*ii*). Later we will reconsider this choice from the standpoint of (*i*),—the reader may anticipate that it will again be found advantageous in some respect. For the present, we proceed to three further choices, these arising from various approaches to (*i*).

The procedure recommended in several statistics manuals (see §8) for testing the hypothesis $\theta = 1$ is to refer the quotient of the larger of s_1^2, s_2^2 by the smaller to tables. This suggests the introduction of a statistic M defined as the maximum of T, T^{-1}. Its distribution[6] under the hypothesis $\theta = 1$ is easily found: Let $g_{n_1n_2}(M)$ be its p.d.f. Then for $1 \leqq u \leqq \infty$,

$$\begin{aligned} g_{n_1n_2}(u)\,du &= Pr(u < M < u + du \mid \theta = 1) \\ &= Pr(u < T < u + du \text{ or } u < T^{-1} < u + du) \\ &= Pr(u < T < u + du) + Pr(u < T^{-1} < u + du), \end{aligned}$$

since the last two terms are the probabilities of mutually exclusive events. Furthermore, the first term is $h_{n_1n_2}(u)\,du$, and because of the symmetry induced by $\theta_0 = 1$ we can evaluate the second term by merely interchanging subscripts. Hence the desired distribution is

$$g_{n_1n_2}(u) = h_{n_1n_2}(u) + h_{n_2n_1}(u),$$

regardless of the true values of the nuisance parameters.

[5] It can be shown by elementary methods that the solution of these equations exists and is unique; likewise for the solutions later denoted by superscripts IV and V.

[6] Considered by K. Pearson [8].

If we reject the hypothesis $\theta = 1$ if $M > M_{n_1 n_2}$, where

$$\int_{M_{n_1 n_2}}^{\infty} g_{n_1 n_2}(u)\, du = \alpha,$$

then this significance test is easily shown to be the same as that of §2 with $\theta_0 = 1$ and

$$A_{n_1 n_2} = B_{n_1 n_2}^{-1}.$$

We remark that again these limits are not easy to compute if $n_1 \neq n_2$. While this choice of A, B, which we shall call $A_{n_1 n_2}^{\mathrm{IV}}$, $B_{n_1 n_2}^{\mathrm{IV}}$, has been motivated only for the case $\theta_0 = 1$, it leads of course to a test IV for any θ_0 and a set IV of confidence intervals.

7. The likelihood ratio. Since the properties of λ-criteria in general have received much attention in the literature, and since in particular the λ-test for H_0 is equivalent to a certain choice of A, B, we shall mention it here, and see whether it has any advantages in §9. λ for H_0 in the case $\theta_0 = 1$ was given by Pearson and Neyman [7; their H_1, n_i, s_i^2, θ, λ_{H_1} are our H_0, N_i, $s_i^2(N_i - 1)/N_i$, $N_1(N_2 - 1)/\{N_2(N_1 - 1)T\}$, λ]; for any θ_0 it may be shown to be

$$\lambda = C_{n_1 n_2} F^{3/2}\left(1 + \frac{n_1}{n_2} F\right)^{-1} h_{n_1 n_2}(F).$$

On considering the (bell-shaped) graph of λ against F we see that $\lambda < \lambda_0$ corresponds to two intervals, say $0 \leqq F < F'$ and $F'' < F \leqq \infty$. The λ-test, which consists of rejecting H_0 when $\lambda < \lambda_0$, where λ_0 is determined so that the significance level is α, is thus equivalent to test V with $A_{n_1 n_2}^{\mathrm{V}}$, $B_{n_1 n_2}^{\mathrm{V}}$ satisfying (2) and

$$\left[u^{3/2}\left(1 + \frac{n_1}{n_2} u\right)^{-1} h_{n_1 n_2}(u)\right]_{u=A}^{B} = 0.$$

8. Equal tails. Perhaps the most venerable procedure for determining limits on a distribution for a significance test in a "two-sided" case is to choose them so that the tails of the distribution have equal areas. Define $A_{n_1 n_2}^{\mathrm{VI}}$, $B_{n_1 n_2}^{\mathrm{VI}}$ from

$$\int_0^{A_{n_1 n_2}^{\mathrm{VI}}} h_{n_1 n_2}(u)\, du = \int_{B_{n_1 n_2}^{\mathrm{VI}}}^{\infty} h_{n_1 n_2}(u)\, du = \tfrac{1}{2}\alpha.$$

The values of $B_{n_1 n_2}^{\mathrm{VI}}$ for $\alpha = .10$ and $.02$ are given in the F-tables [12; same n_1, n_2 as ours] as 5% and 1% points. The relation

$$A_{n_1 n_2}^{\mathrm{VI}} B_{n_2 n_1}^{\mathrm{VI}} = 1 \tag{7}$$

is easy to get, and hence $A_{n_1 n_2}^{\mathrm{VI}}$ for these values of α may also be calculated from the F-tables. The limits for $\frac{1}{2}\alpha = .25$, $.10$, $.025$, $.005$ can be calculated from (1), (7), and Thompson's tables [13].

Since test VI will later be seen to have some merit we will discuss it somewhat further at this point: In several statistics texts [e.g., 3, 14] the student is told to take the quotient of the larger by the smaller of s_1^2, s_2^2, refer it to the F-table, taking the n_1 of the table to be the n_i of the numerator, and to reject the null hypothesis $\theta = 1$ if the sample value is larger than the tabulated. It is then further stated without proof that in using the 5% or 1% points of the F-table, the significance level is actually 10% or 2%. Since the quotient thus referred to the table is precisely the statistic M of §6, it would seem logical to refer it to an M-table rather than the F-table! However, the above procedure can be justified[7] as follows: The equation (7) tells us that test VI fulfills the symmetry condition (4). It makes no difference then in his conclusions whether the computer uses the statistic s_1^2/s_2^2 and the distribution $h_{n_1n_2}(F)$ or s_2^2/s_1^2 and $h_{n_2n_1}(F)$. In particular he may always use the larger ratio and $h_{mn}(F)$, where m and n are the "degrees of freedom" of numerator and denominator, respectively. Since this statistic cannot fall in the lower tail, he need consider only whether the calculated value exceeds the tabulated. *But in using the value tabulated as the upper p% point of the F-distribution, he makes his test at the 2p% significance level.*

9. Comparison of the tests and confidence intervals. We now have at hand two one-tailed and four two-tailed tests, and corresponding sets of confidence intervals, all based on the F-distribution. We note at this point that all four of the two-tailed tests satisfy the symmetry condition (4), and that in the special case $n_1 = n_2$, these four tests become identical. In comparing any two tests, an instrument which makes their relative advantages completely *anschaulich* is the power curve (surface in a more complicated case). The definition and interpretation of the power curve of a test are based on the insight of Neyman and Pearson [5] that two types of error are possible in applying a test: We may (I) reject the hypothesis when it is true, or (II) accept it when it is false.

We see immediately that for any test of the class considered in §2, the probability of a type I error is the same, namely α. To find the probability of a type II error, let us introduce a little more terminology: We denote by E the sample point (E_1, E_2) and by w the region of sample space defined by

$$T < A\theta_0 \quad \text{and} \quad T > B\theta_0 . \tag{8}$$

w is called the *critical region* of the test: the test rejects H_0 if and only if E falls in w. The probability of this, which is called the *power* of the test, is

$$1 - Pr(A\theta_0/\theta \leqq T/\theta \leqq B\theta_0/\theta \mid \theta, \sigma_2^2, a_1, a_2).$$

Since in the present case this happens to be completely independent of the true values of the nuisance parameters, even for $\theta \neq \theta_0$, let us write it as $P(w \mid \theta)$. Then

[7] The writer is indebted to Mr. T. W. Anderson, Jr. for pointing out to him that it is not necessary to use the M-distribution.

$$P(w \mid \theta) = 1 - \int_{A\theta_0/\theta}^{B\theta_0/\theta} h_{n_1 n_2}(u)\, du. \tag{9}$$

Finally, by the *power curve* of the test we mean simply the graph of the power $P(w \mid \theta)$ as a function of θ.

We may now state the probability of a type II error: it is $1 - P(w \mid \theta)$, where necessarily $\theta \neq \theta_0$. Hence the ordinate on the power curve for $\theta \neq \theta_0$ is the probability of avoiding a type II error, while for $\theta = \theta_0$ it is the probability of making a type I error. By inspection of equation (9) we find that, barring the cases $B = \infty$ or $A = 0$ (tests I and II), $P(w \mid \theta) \to 1$ as $\theta \to 0$ or ∞. We calculate the derivative to be

$$P'(w \mid \theta) = [uh_{n_1 n_2}(u)/\theta]_{u=A\theta_0/\theta}^{B\theta_0/\theta}, \tag{10}$$

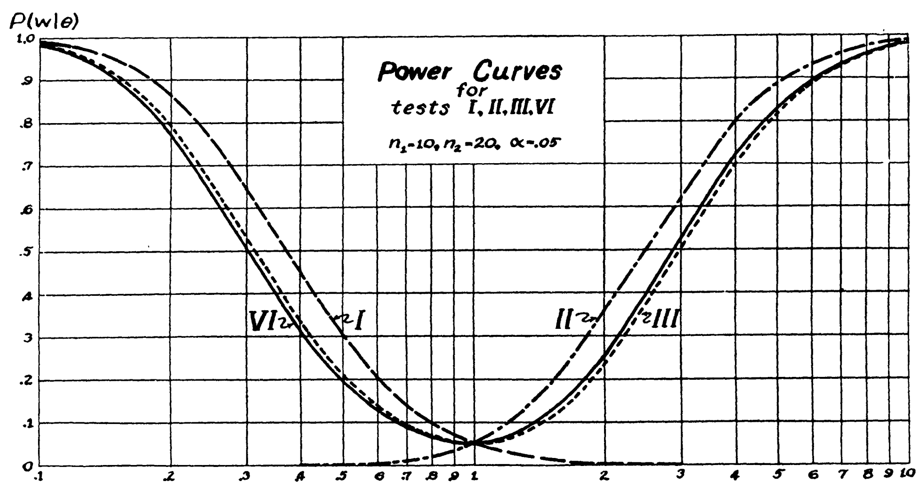

FIG. 1

which is obviously continuous for $0 < \theta < \infty$. If we equate this to zero we find a unique solution for θ, and hence the power curve has a single minimum point. In the exceptional case $B = \infty$ we see from (9) that $P(w \mid \theta)$ decreases monotonically from 1 to 0 as θ increases from 0 to ∞; in the case $A = 0$, $P(w \mid \theta)$ increases monotonically from 0 to 1. Some power curves[8] are plotted in fig. 1.

Always understanding by w a region of the set defined by (8), and recalling the above interpretation of the ordinate on the power curve, we are led to ask whether there is not a w, say w_0, whose power curve nowhere drops below any other curve $P = P(w \mid \theta)$. (They all pass through (θ_0, α).) The test based on such a region w_0 would be called *uniformly most powerful* (UMP) of the class considered, and obviously would be preferred under any circumstances. Alas,

[8] Power curves for test V may be found in a paper by Brown [1]. It did not seem worthwhile to construct curves for test IV, since the limits are hard to compute, the test is biased, and has little historical interest.

it does not exist. Perhaps some insight into the fact of the general non-existence of UMP tests can be gained by returning to fig. 1. While fig. 1 is for the case $n_1 = 10$, $n_2 = 20$, and $\alpha = .05$, the following remarks are valid for any n_1, n_2, α: We note that for testing H_0 against alternatives $\theta < \theta_0$ test I is far superior to the other three, indeed it is superior to any of the tests of the class defined by (8) in the sense that its power curve lies above that of any of the other tests.[9] But for alternatives $\theta > \theta_0$, test I is seen to be very poor (the worst possible, it can be shown). Similar remarks apply to test II and the complementary alternatives. This constitutes the more convincing explanation promised in §2 of the superiority of tests I and II in the "one-sided" cases. Since the power curve of test I lies above all other power curves for $\theta < \theta_0$, and that of test II above all for $\theta > \theta_0$, it is now clear that there is no UMP test of the class considered.

To cope with the commonly occurring situation where there is no UMP test, Neyman and Pearson [5] defined an *unbiased* test,—one whose power curve has an absolute minimum at θ_0. The desirability of an unbiased test in the "two-sided" case is evident when we note that if a test is biased, the probability that we accept the hypothesis $\theta = \theta_0$ is greater if θ has certain values $\theta \neq \theta_0$ than if $\theta = \theta_0$. To find which, if any, of our tests is unbiased, we equate expression (10) to zero for $\theta = \theta_0$. As a result we find[10] the condition (5) which determines test III.

We see now that the limits A^{III}, B^{III} yield the preferred test in the "two-sided" case, as well as the logarithmically shortest confidence intervals. However, as pointed out in §5, the numerical values of these limits are difficult to calculate, and the question then arises, do we lose much by using instead the easily obtained "equal tails" limits A^{VI}, B^{VI}? In the case $n_1 = 10$, $n_2 = 20$, $\alpha = .05$, fig. 1 shows that the power curves of tests III and VI differ very little. The extent of the bias of test VI for other values of n_1, n_2, and $\alpha = .05$, .01 is indicated in table I. (The missing diagonal entries are all 1,5 or 1,1). Let us call the entries β, 100 $\bar{\alpha}$, where $\beta = \theta_{\min}/\theta_0$, $\bar{\alpha} = P(w^{\mathrm{VI}} \mid \theta_{\min})$. From (10) and (1) we get the following formula for computing β:

$$\beta = (\mathcal{B} - \mathcal{A}Q^{n_1/(n_1+n_2)})/(Q - 1),$$

where

$$Q = \mathcal{B}/\mathcal{A}, \qquad \mathcal{A} = a/(1 - a), \qquad \mathcal{B} = b/(1 - b),$$

and a and $1 - b$ are the $100(\frac{1}{2}\alpha)\%$ points on the incomplete Beta distribution for $\nu_2 = n_1$, $\nu_1 = n_2$, and $\nu_1 = n_1$, $\nu_2 = n_2$, respectively, in the notation of Thompson's tables [13]. $\bar{\alpha}$ may then be computed by transforming (9),

$$\bar{\alpha} = 1 - \Big[I_x(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2) \Big]_{x = (1 + \beta/\mathcal{A})^{-1}}^{(1 + \beta/\mathcal{B})^{-1}},$$

[9] The reader may prove this from (9) or note that it is a special case of the results of §10.

[10] The equivalent condition on the incomplete Beta distribution was given by Pitman [10] for the case $\theta_0 = 1$.

TABLE I

Minimum points of power curves of test VI

The entries are $\theta_{\min}/\theta_0$, $100\,P(w^{VI} \mid \theta_{\min})$,

Roman type for $\alpha = .05$, **bold face for $\alpha = .01$**

n_1 \ n_2	1	2	3	5	10	20	40	∞
1		.634, 4.75 **.631, .946**	.576, 4.47 **.577, .883**	.559, 4.17 **.571, .808**	.565, 3.89 **.595, .740**	.574, 3.75 **.617, .705**	.581, 3.68 **.630, .687**	.588, 3.61 **.645, .670**
2	1.578, 4.75 **1.585, .946**		.861, 4.93 **.855, .982**	.779, 4.69 **.776, .928**	.745, 4.44 **.749, .853**	.737, 4.26 **.749, .804**	.735, 4.15 **.753, .778**	.735, 4.05 **.760, .751**
3	1.735, 4.47 **1.734, .883**	1.161, 4.93 **1.170, .982**		.895, 4.92 **.889, .978**	.838, 4.70 **.835, .917**	.819, 4.51 **.821, .867**	.812, 4.41 **.819, .837**	.808, 4.29 **.820, .804**
5	1.789, 4.17 **1.752, .808**	1.284, 4.69 **1.289, .928**	1.117, 4.92 **1.124, .978**		.927, 4.92 **.924, .975**	.898, 4.78 **.896, .934**	.886, 4.67 **.887, .903**	.877, 4.54 **.882, .864**
10	1.771, 3.89 **1.682, .740**	1.342, 4.44 **1.335, .853**	1.194, 4.70 **1.198, .917**	1.079, 4.92 **1.083, .975**		.965, 4.96 **.964, .987**	.949, 4.89 **.949, .964**	.941, 4.76 **.937, .925**
20	1.742, 3.75 **1.622, .705**	1.357, 4.26 **1.335, .804**	1.221, 4.51 **1.217, .867**	1.114, 4.78 **1.116, .934**	1.036, 4.96 **1.038, .987**		.983, 4.98 **.983, .993**	.967, 4.88 **.968, .960**
40	1.722, 3.68 **1.587, .687**	1.360, 4.15 **1.327, .778**	1.231, 4.41 **1.221, .837**	1.129, 4.67 **1.127, .903**	1.053, 4.89 **1.054, .964**	1.017, 4.98 **1.018, .993**		.984, 4.94 **.984, .980**
∞	1.700, 3.61 **1.549, .670**	1.360, 4.05 **1.315, .751**	1.238, 4.29 **1.219, .804**	1.140, 4.54 **1.134, .864**	1.063, 4.76 **1.067, .925**	1.034, 4.88 **1.034, .960**	1.017, 4.94 **1.017, .980**	

and using Pearson's tables [9], or, when x is very close to 0 or 1, using a few terms of the series

$$I_\delta(\tfrac{1}{2}m, \tfrac{1}{2}n) = 1 - I_{1-\delta}(\tfrac{1}{2}n, \tfrac{1}{2}m) = \frac{\delta^{\frac{1}{2}m}}{B(\tfrac{1}{2}m, \tfrac{1}{2}n)}\left[\frac{2}{m} - \frac{n-2}{2^0(m+2)}\frac{\delta}{1!} + \frac{(n-2)(n-4)}{2^1(m+4)}\frac{\delta^2}{2!} - \frac{(n-2)(n-4)(n-6)}{2^2(m+6)}\frac{\delta^3}{3!} + \cdots\right].$$

In computing β, $\bar{\alpha}$ it is perhaps simplest to take $n_1 > n_2$ and use the relationships

$$\beta_{n_1 n_2} = 1/\beta_{n_2 n_1}, \qquad \bar{\alpha}_{n_1 n_2} = \bar{\alpha}_{n_2 n_1}.$$

When sample sizes $n_1 + 1$, $n_2 + 1$ are such that table I indicates a large bias' it might be worthwhile to get limits for an unbiased test from the "equal tails', limits as follows: The limits $\bar{A}^{\text{III}}$, $\bar{B}^{\text{III}}$ for an unbiased test III may be obtained by taking

$$\bar{A}^{\text{III}} = A^{\text{VI}}/\beta, \qquad \bar{B}^{\text{III}} = B^{\text{VI}}/\beta,$$

but *the test will then be at significance level* $\bar{\alpha}$. The gain in using $\bar{A}^{\text{III}}$, $\bar{B}^{\text{III}}$ instead of A^{VI}, B^{VI} is more apparent when we consider confidence intervals: The sets associated with $\bar{A}^{\text{III}}$, $\bar{B}^{\text{III}}$, and A^{VI}, B^{VI} have the same logarithmic lengths, but the confidence coefficients are $1 - \bar{\alpha}$ and $1 - \alpha$, respectively.

This seems to be about as far as it is worthwhile to carry the developments at the elementary level of part I. Some inadequacies may already have disturbed the reader: Why not consider in place of the interval (A, B) on the range of F any measurable region[11] R such that the integral of $h_{n_1 n_2}(F)$ over R is $1 - \alpha$? Under the transformation $T = \theta_0 F$ the complement of R, just as the complement of (A, B), would lead to critical regions w for which $P(w \mid \theta_0) = \alpha$ for all values of the nuisance parameters. Critical regions satisfying the last condition are said to be *similar* to the sample space with regard to the nuisance parameters. More generally, how would our preferred test I, II, III stand up if we admit for comparison, tests based on any similar regions whatever? Finally, how can one formulate in a general way conditions for optimum confidence intervals, and would a more general formulation still lead to the preference of the sets I, II, III? Answers to these questions will be found in part II.

Part II. Significance Tests and Confidence Intervals Based on any Similar Regions

10. Common best critical regions. For the case $\theta_0 = 1$, Neyman and Pearson [6] have shown that the critical region of test I is the common best critical (CBC) region for testing H_0 against alternatives $\theta < \theta_0$. This result is easily extended to any θ_0 by a simple device. We consider the following 1:1 transformations of variables and parameters:

[11] Our intuitions may balk at the notion of using sets R more general than intervals, but it would nevertheless be reassuring to find that our tests can meet this competition.

$$(11) \qquad x_{1j} = \theta_0^{\frac{1}{2}} x'_{1j}, \qquad x_{2k} = x'_{2k}, \qquad j = 1, 2, \cdots, N_1; k = 1, 2, \cdots, N_2,$$

$$(12) \qquad \theta = \theta_0 \theta', \qquad \sigma_2^2 = (\sigma_2')^2, \qquad a_1 = \theta_0^{\frac{1}{2}} a_1', \qquad a_2 = a_2'.$$

Denote by E_1', E_2', E' the points corresponding to E_1, E_2, E, respectively, under the transformation (11), by ϑ any point in the space of the three nuisance parameters, and by ϑ' its correspondent under the transformation (12), by H_0' the transformed hypothesis, H_0': $\theta' = 1$; ϑ', unspecified. If w is any Borel-measurable region of the space of E, and w' the map of w under (11), then $Pr(E \epsilon w \mid \theta, \vartheta) = Pr(E' \epsilon w' \mid \theta', \vartheta')$, which we shall write as

$$(13) \qquad P(w \mid \theta, \vartheta) = P(w' \mid \theta', \vartheta').$$

We note that the coordinates of E_i' are normally distributed with mean a_i' and variance $(\sigma_i')^2$ where $(\sigma_1')^2 = \theta'(\sigma_2')^2$, all $N_1 + N_2$ coordinates being statistically independent. Designating the critical region of test I by w_0, and its map under (11) by w_0', the result of Neyman and Pearson may then be stated as follows: w_0' is a CBC region for H_0' and alternatives $\theta' < 1$. Now suppose w_0 were not a CBC region for H_0 and alternatives $\theta < \theta_0$. Then there would exist a region w_1, a value $\theta_1 < \theta_0$, and a point ϑ_1 such that $P(w_1 \mid \theta_1, \vartheta_1) > P(w_0 \mid \theta_1, \vartheta_1)$, while $P(w_1 \mid \theta_0, \vartheta) = \alpha$ for all ϑ. Let w_1', θ_1', ϑ_1' correspond to w_1, θ_1, ϑ_1 under (11) and (12). Then from (13) we would have that $P(w_1' \mid \theta_1', \vartheta_1') > P(w_0' \mid \theta_1', \vartheta_1')$, where $\theta_1' < 1$, while $P(w_1' \mid 1, \vartheta') = \alpha$ for all ϑ'. But this would contradict the fact that w_0' is a CBC region for H_0' and alternatives $\theta' < 1$.

The proof that the critical region of test II is a CBC region for testing H_0 against alternatives $\theta > \theta_0$ is of course completely analogous. This establishes the non-existence of a UMP test for H_0, and so we consider next the existence of a "best" unbiased test.

11. Type B_1 region. This section is a direct application of a recent paper "On the theory of testing composite hypotheses with one constraint" to which we shall refer as [11]. Since it is not feasible to restate here the definitions, assumptions, and theorems of [11], we shall refer to them by their numbers there. It is convenient to transform the parameters of the p.d.f. of E by putting

$$(14) \qquad \theta = 1/\psi, \qquad \theta_0 = 1/\psi_0, \qquad \sigma_2^2 = 1/h.$$

Then

$$(15) \quad p(E \mid \psi, h, a_1, a_2) = (2\pi)^{-\frac{1}{2}N} \psi^{\frac{1}{2}N_1} h^{\frac{1}{2}N} \cdot \exp\{-\tfrac{1}{2}\psi h[N_1(\bar{x}_1 - a_1)^2 + S_1] + h[N_2(\bar{x}_2 - a_2)^2 + S_2]\},$$

where

$$N = N_1 + N_2, \qquad S_i = n_i s_i^2.$$

We note that type B and type B_1 regions (definitions 1, 2 in [11]) are invariant under certain transformations of parameters: Suppose new parameters θ', ϑ'

are introduced by 1:1 transformations $\theta = \theta(\theta')$, $\vartheta = \vartheta(\vartheta')$. Let θ_0' correspond to θ_0, and consider the transformed hypothesis $H_0' : \theta' = \theta_0' ; \vartheta'$, unspecified. Sufficient conditions that a region be of type B for testing H_0 if it is of type B for testing H_0' are that the function $\theta(\theta')$ have first and second derivatives and that the first not vanish at θ_0'. The last statement remains true if B is replaced by B_1. Since the transformations (14) satisfy these sufficient conditions, we define

$$H_0' : \quad \psi = \psi_0 ; \quad \vartheta' = (h, a_1, a_2), \text{ unspecified,}$$

and propose to show that there exists a type B_1 region for testing H_0', and that it is the critical region of test III.

For later reference we now note that the four functions of variables and parameters defined in Table II are mutually independently distributed as indicated there.

TABLE II

Function	Distribution
$U_1 = \psi h S_1 = S_1/\sigma_1^2$	χ^2, with n_1 degrees of freedom
$U_2 = hS_2 = S_2/\sigma_2^2$	" " n_2 " " "
$u_3 = (\psi h N_1)^{\frac{1}{2}}(\bar{x}_1 - a_1) = N_1^{\frac{1}{2}}(\bar{x}_1 - a_1)/\sigma_1$	normal, with zero mean and unit variance
$u_4 = (hN_2)^{\frac{1}{2}}(\bar{x}_2 - a_2) = N_2^{\frac{1}{2}}(\bar{x}_2 - a_2)/\sigma_2$	" " " " " " "

Let us first verify the critical assumption 3^0 of [11]: Identifying our ψ, h, a_1, a_2 with $\theta_1, \theta_2, \theta_3, \theta_4$ of [11], we find from (15) that

$$\begin{aligned}
\phi_1 &= \tfrac{1}{2}\{N_1/\psi - h[N_1(\bar{x}_1 - a_1)^2 + S_1]\},\\
\phi_2 &= \tfrac{1}{2}\{N/h - \psi[N_1(\bar{x}_1 - a_1)^2 + S_1] - [N_2(\bar{x}_2 - a_2)^2 + S_2]\},\\
\phi_3 &= \psi h N_1(\bar{x}_1 - a_1),\\
\phi_4 &= hN_2(\bar{x}_2 - a_2),
\end{aligned} \tag{16}$$

and then check 3^0 by differentiating equations (16).

To verify assumption 4^0, let x_1, x_2, x_3, x_4 of [11] be our $x_{11}, x_{12}, x_{21}, x_{22}$, respectively. We calculate

$$\frac{\partial(\phi_1, \phi_2, \phi_3, \phi_4)}{\partial(x_1, x_2, x_3, x_4)} = \psi h^3(x_1 - x_2)(x_4 - x_3),$$

which vanishes only on the same set of probability zero for all admissible values of the parameters. The validity of assumption 5^0 follows from §5 of [11], and there is no difficulty in verifying 1^0 and 2^0.

To apply theorem 1 of [11] we must find functions $k_i(\phi_2, \phi_3, \phi_4 ; \psi_0, \vartheta')$, $i = 1, 2$, such that

$$\int_{k_1}^{k_2} \phi_1^t p(\phi_1, \phi_2, \phi_3, \phi_4 \mid \psi_0, \vartheta')\, d\phi_1 = (1 - \alpha) \int_{-\infty}^{+\infty} \text{same,} \tag{17}$$

for $t = 0, 1$, where the *symbols* ϕ_i *henceforth are understood to stand for the functions* (16) *with* ψ *replaced by* ψ_0. If the functions k_i exist, then the region in sample space defined by

$$\phi_1 < k_1 \quad \text{and} \quad \phi_1 > k_2 \tag{18}$$

is independent of ϑ' and of type B.

From equations (16) and Table II we see that

$$\begin{aligned} \phi_1 &= \tfrac{1}{2}(N_1 - u_1)/\psi_0, & \phi_2 &= \tfrac{1}{2}(N - u_2)/h, \\ \phi_3 &= (\psi_0 h N_1)^{\frac{1}{2}} u_3, & \phi_4 &= (hN_2)^{\frac{1}{2}} u_4, \end{aligned} \tag{19}$$

where

$$u_1 = U_1 + u_3^2, \qquad u_2 = U_1 + U_2 + u_3^2 + u_4^2,$$

and ψ is put equal to ψ_0 in U_1, u_3. Furthermore, for fixed u_2, u_3, u_4, the range of u_1 is

$$u_3^2 \leqq u_1 \leqq u_2 - u_4^2.$$

Transforming the integrals in (17) by substituting (19) and

$$p(\phi_1, \phi_2, \phi_3, \phi_4 \mid \psi_0, \vartheta') = \frac{p(U_1, U_2, u_3, u_4 \mid \psi_0, \vartheta')}{\dfrac{\partial(\phi_1, \phi_2, \phi_3, \phi_4)}{\partial(u_1, u_2, u_3, u_4)} \cdot \dfrac{\partial(u_1, u_2, u_3, u_4)}{\partial(U_1, U_2, u_3, u_4)}},$$

where the p.d.f. in the numerator is, from Table II,

$$CU_1^{\frac{1}{2}n_1 - 1} U_2^{\frac{1}{2}n_2 - 1} \exp\left(-\tfrac{1}{2}u_2\right),$$

we get as the equivalent of (17)

$$\int_{K_1}^{K_2} (N_1 - u_1)^t (u_1 - u_3^2)^{\frac{1}{2}n_1 - 1} (u_2 - u_4^2 - u_1)^{\frac{1}{2}n_2 - 1}\, du_1 = (1 - \alpha) \int_0^1 \text{same}$$

with

$$K_i(u_2, u_3, u_4; \psi_0, \vartheta') = k_i(\phi_2, \phi_3, \phi_4; \psi_0, \vartheta').$$

Finally, we let

$$x = (u_1 - u_3^2)/(u_2 - u_3^2 - u_4^2), \tag{20}$$

and get

$$\int_{\kappa_1}^{\kappa_2} [N_1 - u_3^2 - (u_2 - u_3^2 - u_4^2)x]^t x^{\frac{1}{2}n_1 - 1} (1 - x)^{\frac{1}{2}n_2 - 1}\, dx = (1 - \alpha) \int_0^1 \text{same},$$

where $\kappa_i(u_2, u_3, u_4; \psi_0, \vartheta')$ are the values of x obtained by setting u_1 equal to the function K_i in (20). The last condition is equivalent to

$$\int_{\kappa_1}^{\kappa_2} x^{\frac{1}{2}n_1 - 1 + t} (1 - x)^{\frac{1}{2}n_2 - 1}\, dx = (1 - \alpha) \int_0^1 \text{same}, \qquad t = 0, 1. \tag{21}$$

Since x is a continuous monotonic function of ϕ_1, (18) becomes

$$x < \kappa_1 \quad \text{and} \quad x > \kappa_2 . \tag{22}$$

Solutions for the functions κ_1, κ_2 satisfying (21) exist in the form κ_i = constant. Indeed, if we now note that the x defined by (20) is the same as that defined in (1), and let $\kappa_1 = a$, $\kappa_2 = b$, we see that the conditions (21) are identical with (6), and that our method of finding type B regions has led us to the critical region of test III.

To show that the type B region obtained from Theorem 1 of [11] is also of type B_1, we appeal to Theorem 2: From (15) we have

$$p(E \mid \psi, \vartheta')/p(E \mid \psi_0, \vartheta') = (\psi/\psi_0)^{\frac{1}{2}N_1} \exp\{(\psi - \psi_0)(\phi_1 - \tfrac{1}{2}N_1/\psi_0)\}.$$

Since for $\psi \neq \psi_0$ this function is convex in ϕ_1, Theorem 2 is applicable. The result of this section is the conclusion that the critical region of test III is of type B_1 for testing H_0.

12. Neyman's categories of confidence intervals. The concepts and terminology of this section are those formulated in a basic paper [4] by Neyman. Suppose a distribution depends on a parameter θ, and on further parameters $\theta_2, \theta_3, \cdots, \theta_l$ which we shall symbolize by ϑ. The hypothesis

$$H(\theta_0): \qquad \theta = \theta_0 ; \qquad \vartheta, \text{ unspecified},$$

may be called a composite hypothesis with one constraint [11]. Let E be the sample point, W be the sample space, and w be any Borel-measurable region in W. Write $Pr\{E \in w \mid \theta, \vartheta\} = P\{w \mid \theta, \vartheta\}$. The condition that a critical region $w(\theta_0)$ for testing $H(\theta_0)$ be similar to W with respect to ϑ is

$$P\{w(\theta_0) \mid \theta_0, \vartheta\} = \alpha \text{ for all } \vartheta, \tag{23}$$

where α is fixed throughout our discussion. Suppose for every admissible θ_0 there exists a similar region $w(\theta_0)$. The complementary region $A(\theta_0) = W - w(\theta_0)$ we may call a region of acceptance. For any E we next define the linear set $\delta(E)$ of points on the θ-axis as the totality of points θ such that $E \in A(\theta)$. The probability [4] that the random set $\delta(E)$ cover a value θ'' if the true value of θ is θ' is

$$Pr\{\theta'' \in \delta(E) \mid \theta', \vartheta\} = 1 - P\{w(\theta'') \mid \theta', \vartheta\}, \tag{24}$$

and hence from (23),

$$Pr\{\theta' \in \delta(E) \mid \theta', \vartheta\} = 1 - \alpha \tag{25}$$

for all θ', ϑ, and we might call the aggregate $\{\delta(E)\}$ a set of confidence regions with confidence coefficient $1 - \alpha$. Now if all $\delta(E)$ are intervals, then they form a set of confidence intervals.

We have now shown that if $H(\theta_0)$ is a composite hypothesis with one constraint, if for every admissible θ_0 there exists a similar region $w(\theta_0)$ for testing

$H(\theta_0)$, and if the aggregate $\{\delta(E)\}$ determined by the family $\{w(\theta_0)\}$ consists of *intervals* $\delta(E)$, then $\{\delta(E)\}$ is a set of confidence intervals. By similar use of (24) the reader may prove that if furthermore each $w(\theta_0)$ of the family has the property P of the table below, then the corresponding set $\{\delta(E)\}$ of confidence intervals is of Neyman's category C:

P: property of $w(\theta_0)$	C: category of $\{\delta(E)\}$
gives UMP test	shortest
CBC for $\theta > \theta_0$ (or $\theta < \theta_0$)	best one-sided
gives unbiased test	unbiased
of type B	short unbiased
of type B_1	shortest unbiased

We have taken the liberty of calling a set of one-sided confidence intervals

$$\delta(E): \qquad \underline{\underline{\theta}}(E) \leqq \theta \text{ (or } \theta \leqq \bar{\bar{\theta}}(E)),$$

where $\underline{\underline{\theta}}(E)$ and $\bar{\bar{\theta}}(E)$ are Neyman's *unique lower and upper estimates*, respectively, *best one-sided*, and of calling a set $\{\delta_0(E)\}$ *shortest unbiased* if for all θ', ϑ it satisfies (25) and

$$(26) \qquad [\partial Pr\{\theta' \in \delta_0(E) \mid \theta, \vartheta\}/\partial\theta]_{\theta=\theta'} = 0,$$

while for any other set $\{\delta_1(E)\}$ satisfying (25) and (26), and all θ'', θ', ϑ,

$$Pr\{\theta'' \in \delta_0(E) \mid \theta', \vartheta\} \leqq Pr\{\theta'' \in \delta_1(E) \mid \theta', \vartheta\}.$$

It follows immediately from this discussion that our sets II and I of confidence intervals are the best one-sided, and that the set III is not only a short, but the shortest, unbiased set.

In conclusion, we remark that Neyman's concept of the "shortness" of a set of confidence intervals strikes one at first as indirect,—to fully appreciate its elegance it is perhaps necessary to attempt the formulation of a general theory from a more naive approach,—and that it is then of interest to discover that in the present case his short unbiased set coincides with that reached by the direct intuitive (but obviously extremely limited) method of §5.

REFERENCES

[1] G. W. Brown, "On the power of the L_1 test for equality of several variances," *Annals of Math. Stat.*, Vol. 10 (1939), p. 127.

[2] R. A. Fisher, "On a distribution yielding the error function of several well known statistics," *Proc. Int. Math. Congress, Toronto, 1924*, Vol. 2, p. 808.

[3] J. F. Kenney, *Mathematics of Statistics*, part 2, N. Y., 1939, p. 144.

[4] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc. London*, ser. A, Vol. 236 (1937), pp. 333–380.

[5] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses, part I," *Stat. Res. Mem.*, Vol. 1 (1936), pp. 1–37.

[6] J. NEYMAN and E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc. London*, ser. A, Vol. 231 (1933), pp. 289–337.
[7] E. S. PEARSON and J. NEYMAN, "On the problem of two samples," *Bull. Int. Acad. Polon. Sc. Let.*, ser. A, 1930, p. 82.
[8] K. PEARSON, S. A. STOUFFER, and F. N. DAVID, "Further applications in statistics of the $T_m(x)$ Bessel function," *Biometrika*, Vol. 24 (1932), pp. 306, 339, 340.
[9] K. PEARSON (Editor), *Tables of the Incomplete Beta Function*, Cambridge, 1934.
[10] E. J. G. PITMAN, "Tests of hypotheses concerning location and scale parameters," *Biometrika*, Vol. 31 (1939), p. 207.
[11] H. SCHEFFÉ, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 280–293.
[12] G. W. SNEDECOR, *Statistical Methods*, Ames, 1940, pp. 184–187.
[13] C. M. THOMPSON, "Tables of percentage points of the incomplete Beta function," *Biometrika*, Vol. 32 (1941), pp. 168–181.
[14] L. H. C. TIPPETT, *The Methods of Statistics*, London, 1937, p. 118.