ON OPTIMUM TESTS OF COMPOSITE HYPOTHESES WITH ONE CONSTRAINT[1]

By E. L. Lehmann

University of California, Berkeley

Summary. This paper is concerned with optimum tests of certain composite hypotheses. In section 2 various aspects of a theorem of Scheffé concerning type B_1 tests are discussed. It is pointed out that the theorem can be extended to cover uniformly most powerful tests against a one-sided set of alternatives. It is also shown that the method for determining explicitly the optimum test region may in certain cases be reduced to a simple formal procedure. These results are used in section 3 to obtain optimum tests for the composite hypothesis specifying the value of the circular serial correlation coefficient in a normal distribution. A surprising feature of this example is the fact that for the simple hypothesis obtained by specifying values for the nuisance parameters no test with the corresponding optimum properties exists.

In section 4 the totality of similar regions is obtained for a large class of probability laws which admit a sufficient statistic. Some composite hypotheses concerning exponential and rectangular distributions are treated in section 5. It is proved that the likelihood ratio tests of these hypotheses have various optimum properties.

1. Introduction. In developing tests for a class of hypotheses three phases may be distinguished. First, tests are obtained which are intuitively appealing; next, it is shown that these tests have certain attractive features; finally, it is proved that they are "best possible" tests.

In dealing with parametric hypotheses, the likelihood ratio principle is frequently used to obtain a reasonable test. For many of the tests so derived for normal and exponential distributions, the question of bias has been investigated. In most cases unbiasedness has been established; in the other cases, usually a test based on the same criterion but with the boundaries shifted, can be proved to be unbiased. Other desirable properties which likelihood ratio tests have been shown to possess, relate to the asymptotic behaviour of these tests as the sample sizes tend to infinity. An interesting problem which does not seem to have been treated is the question of admissibility of likelihood ratio tests, a test being admissible if its power can not be improved upon uniformly by any other test of the same level of significance.

Investigations of optimum tests of composite hypotheses have been carried through for many hypotheses concerning normal distributions. When the hypothesis specifies the value of one parameter (hypothesis with one constraint), uniformly most powerful one-sided and type B_1 (uniformly most powerful un-

[1] Presented at a meeting of the Institute of Mathematical Statistics in San Diego, June, 1947.

J. Rojo (ed.), *Selected Works of E. L. Lehmann*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1412-4_12, © Springer Science+Business Media, LLC 2012

biased) tests have been obtained. When the number of constraints is larger than one, not so much can be expected. It has been shown for some of the tests in this class that they have maximum average power uniformly over a family of surfaces in the parameter space, or that they are uniformly most powerful with respect to the subclass of tests whose power depends only on some function of the parameters. (All optimum properties mentioned are relative only to the class of all similar regions. This will be so throughout the paper and will usually not be stated explicitly).

Two methods for finding uniformly most powerful or uniformly most powerful one-sided regions and type B_1 tests, if they exist are known. Neyman and Pearson [1] developed a method for determining all similar regions, and applied it to obtain uniformly most powerful one-sided tests of certain hypotheses. Neyman [2, 3] extended the method to obtain, for certain hypotheses, the class of all bisimilar (unbiased similar) regions, and Scheffé [4], developing the method further, proved the existence of type B_1 tests for an important class of hypotheses.

A different method for obtaining all similar and bisimilar regions was devised by P. L. Hsu and was used by him and other writers to prove various optimum properties of the likelihood ratio tests for the general linear hypothesis, of Hotelling's T^2 and of other tests [5, 6, 7, 8].

In the present paper we are concerned with applications of these two methods to composite hypotheses with one constraint. However, the applicability is not so restricted. In fact, the second method has been used mainly in connection with composite hypotheses with many constraints, and the author believes it to be suitable also for deriving optimum classification procedures. An essential restriction of both methods seems to be that a set of sufficient statistics must exist with respect to the parameters involved: with respect to the nuisance parameters so that all similar regions can be found, with respect to the parameters specified by the hypothesis so that there exists a best of all similar regions.

Extensions of the existing theory based on the first method are obtained in section 2, and the theory is applied in section 3 to a hypothesis concerning a multivariate normal distribution. Sections 4 and 5 are concerned with applications of the second method to problems to most of which the earlier method is not applicable, in particular to hypotheses concerning exponential and rectangular distributions, hitherto only treated from the likelihood ratio point of view.

2. On the theory of optimum tests.

2.1 *One-sided tests.* In an interesting paper [4], Scheffé determined the type B and type B_1 tests of a certain class of composite hypotheses specifying the value θ_0 of a parameter θ in the presence of nuisance parameters.

Scheffé's results can, in an obvious way, be extended to cover one-sided sets of alternatives. To show this, consider the method used in [4]. Under certain assumptions all tests[2] are found which satisfy the two conditions:

[2] The terms "the test w" and "the region [of rejection] w" will be used interchangeably.

(a) The power function β_w at θ_0 has a preassigned value ϵ (the level of significance), independent of the nuisance parameters;

(b) the power function at θ_0 has derivative 0. (Condition of unbiasedness). Then that test w_0 is determined for which, of all those satisfying (a) and (b),

(c) the second derivative at θ_0, $\beta_w''(\theta_0)$, is as large as possible.

By definition w_0 is a type B test. Under a certain additional assumption (this is the convexity assumption $\frac{\partial^2 g}{\partial y_1^2} > 0$ of Scheffé's Theorem 2) it is shown that of all tests satisfying (a) and (b), w_0 has maximum power against all alternatives, i.e. is of type B_1.

If now we want to maximize the power against only the one-sided set of alternatives, $\theta > \theta_0$, we determine that test w_1 of all those satisfying (a), for which

(d) the first derivative at θ_0, $\beta_w'(\theta_0)$, is as large as possible.

Under a certain additional assumption (in Scheffé's notation this would be the monotonicity assumption $\frac{\partial g}{\partial y_1} > 0$) it can then be shown that of all tests satisfying (a), w_1 has maximum power against all alternatives $\theta > \theta_0$, (it also has minimum power against all alternatives $\theta < \theta_0$), i.e. w_1 is uniformly most powerful against alternatives $\theta > \theta_0$. We shall not carry through the discussion in detail since Scheffé's argument applies step by step, with only the obvious changes.

2.2 *Determination of the boundaries.* Let $X_1, \cdots, X_n$ be n random variables with a joint probability density function p, depending on parameters θ_1 and $\theta = (\theta_2, \cdots, \theta_l)$. We shall denote the probability density function of a set of random variables $X_1, \cdots, X_n$ whose distribution depends on a parameter θ by $p(x_1, \cdots, x_n \mid \theta)$ or simply by $p(x_1, \cdots, x_n)$ when the dependence on θ is clear from the context. The set of points $(x_1, \cdots, x_n)$ for which

$$p(x_1, \cdots, x_n \mid \theta)$$

is positive we shall denote by $W_+(\theta)$.

Let

$$\varphi_i(x_1, \cdots, x_n) = \frac{\partial}{\partial \theta_i} \log p(x_1, \cdots, x_n \mid \theta_1, \theta) \big|_{\theta_1 = \theta_1^0}, \qquad (i = 1, \cdots, l), \tag{2.1}$$

and let the random variable Φ_i be defined by

$$\Phi_i = \varphi_i(X_1, \cdots, X_n). \tag{2.2}$$

Then for testing the hypothesis H: $\theta_1 = \theta_1^0$, under the assumptions stated by Scheffé, the type B_1 test w_0 is defined by the inequalities

$$\varphi_1 < k_1, > k_2 \qquad (k_1 < k_2) \tag{2.3}$$

where k_1, k_2 depend on θ_1^0, θ, $\varphi_2, \cdots, \varphi_l$ and are determined by the two equations[3]

$$\int_{k_1}^{k_2} \varphi_1^s p(\varphi_1, \cdots, \varphi_l)\, d\varphi_1 = (1 - \epsilon) \int_{-\infty}^{\infty} \text{same} \qquad (s = 0, 1). \tag{2.4}$$

[3] Although k_1 and k_2 may depend on θ, w_0 is independent of θ, as was shown in [4].

The equations (2.3) and (2.4) are not suitable for the determination of the boundary of w_0. The variables have to be transformed so as to obtain for w_0 an expression from which the calculation of the boundaries becomes feasible, (cf. [9]). This part of the work may be formalized in the following theorem.

THEOREM 1. *Let*

$$U = f(\Phi_1, \Phi_2, \cdots, \Phi_l) \tag{2.5}$$
$$V_i = g_i(\Phi_2, \cdots, \Phi_l), \qquad (i = 2, \cdots, l),$$

be a system of functions, continuously differentiable and with non-vanishing Jacobian almost everywhere, and such that

(i) *U is a linear function of Φ_1*

$$U = a\Phi_1 + b \tag{2.6}$$

with coefficients which may depend on $\Phi_2, \cdots, \Phi_l$ and such that[4] $a(\Phi_2, \cdots, \Phi_l) > 0$,

(ii) *it is possible to solve for $\Phi_2, \cdots, \Phi_l$ in terms of the V's,*

(iii) *under the hypothesis H, U is distributed independently of*

$$V = (V_2, \cdots, V_l).$$

Then the region w_0 is equivalent to the region

$$u < c_1, > c_2 \qquad (c_1 < c_2) \tag{2.7}$$

where c_1, c_2 are determined by

$$\int_{c_1}^{c_2} u^s p(u)\, du = (1 - \epsilon) \int_{-\infty}^{\infty} u^s p(u)\, du \qquad (s = 0, 1). \tag{2.8}$$

PROOF.

$$p(\varphi_1, \varphi_2, \cdots, \varphi_l) = p(u, v_3, \cdots, v_l) \cdot \left| \frac{\partial(u, v_2, \cdots v_l)}{\partial(\varphi_1, \cdots, \varphi_l)} \right| \tag{2.9}$$
$$= p(u) \cdot p(v_2, \cdots, v_l) \frac{\partial u}{\partial \varphi_1} \cdot \left| \frac{\partial(v_2, \cdots, v_l)}{\partial(\varphi_2, \cdots, \varphi_l)} \right|.$$

But

$$u = a(\varphi_2, \cdots, \varphi_l) \cdot \varphi_1 + b(\varphi_2, \cdots, \varphi_l) \tag{2.10}$$
$$= \alpha(v_2, \cdots, v_l) \cdot \varphi_1 + \beta(v_2, \cdots, v_l)$$

so that (2.4) reduces to

$$\int_{c_1(v_2,\cdots,v_l)}^{c_2(v_2,\cdots,v_l)} \left(\frac{u - \beta}{\alpha}\right)^s p(u) p(v_2, \cdots, v_l)\, du \tag{2.11}$$
$$= (1 - \epsilon) \int_{-\infty}^{\infty} \text{same} \qquad (s = 0, 1)$$

---

[4] A similar theorem holds when we assume $a(\Phi_2, \ldots, \Phi_l) < 0$.

and hence to

$$\int_{c_1(v_2,\cdots,v_l)}^{c_2(v_2,\cdots,v_l)} u^s p(u)\,du = (1-\epsilon)\int_{-\infty}^{\infty} \text{same} \qquad (s = 0, 1) \tag{2.12}$$

which shows c_1 and c_2 to be independent of the v's. Also obviously (2.3) transforms into (2.7) which completes the proof.

If U is such that its distribution (when $\theta_1 = \theta_1^0$) is independent of θ, c_1 and c_2 of theorem 1 will depend only on the data of the problem: ϵ, n, θ_1^0. However, the existence of constants c_1 and c_2 satisfying (2.8) still has to be proved. We may show more generally the existence of k_1 and k_2 satisfying (2.4). A proof is immediately supplied by an argument which was used by Neyman [10] and Wald [11] to prove the existence of type A tests, and which may be stated in the following

Lemma. *Let* $0 < \alpha < 1$, *let* $f(x) \geq 0$ *and* $\int_{-\infty}^{\infty} x^s f(x)\,dx < \infty$ *for* $s = 0, 1$. *Then there exist* A, B *such that*

$$\int_A^B x^s f(x)\,dx = \alpha \int_{-\infty}^{\infty} x^s f(x)\,dx \qquad (s = 0, 1). \tag{2.13}$$

3. Testing for circular serial correlation in a normal population. We now apply the results of the previous section to obtain the optimum tests (i.e. uniformly most powerful against the one-sided set of alternatives, type B_1 in the two-sided case) for the hypothesis specifying the value of the circular serial correlation coefficient in the normal population considered by Dixon [12]. (For the literature on testing for non-circular serial correlation in normal populations cf. [12]).

We assume

$$p(x_1, \cdots, x_n) = \frac{1-\delta^n}{(\sqrt{2\pi}\sigma)^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n [(x_i - \xi) - \delta(x_{i+1} - \xi)]^2\right] \tag{3.1}$$

where $x_{n+1} = x_1$ and $|\delta| < 1$, and we test the hypothesis $\delta = \delta_0$. For testing purposes only the value $\delta_0 = 0$ is of interest presumably, however, the family of tests for arbitrary δ_0 is required for estimating δ by means of confidence intervals, and therefore the more general hypothesis is considered.

Making a transformation in one of the parameters we write

$$\begin{aligned} &p(x_1, \cdots, x_n) \\ &= C(\delta, \alpha) \exp\left[\alpha\left[(1+\delta^2)\sum_{i=1}^n (x_i - \xi)^2 - 2\delta \sum_{i=1}^n (x_i - \xi)(x_{i+1} - \xi)\right]\right] \end{aligned} \tag{3.2}$$

where in the notation of the previous section $\theta_1 = \delta$, $\theta_2 = \alpha$, $\theta_3 = \xi$.

Theorem 2. *For testing the hypothesis* $\delta = \delta_0$ *for the distribution* (3.2)

(a) *the type* B_1 *test exists and is given by*

$$r < r_1, > r_2 \tag{3.3}$$

where

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(x_{i+1} - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}; \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3.4}$$

and where r_1 and r_2 are determined by

$$\int_{r_1}^{r_2} \left(\frac{r}{1 + \delta_0^2 - 2\delta_0 r}\right)^s p(r)\, dr = (1 - \epsilon) \int_{-\infty}^{\infty} \text{same} \qquad (s = 0, 1). \tag{3.5}$$

(b) *the uniformly most powerful similar region for testing H against the alternatives $\delta > \delta_0$ exists and is given by*

$$r > r' \tag{3.6}$$

where r' is determined by

$$\int_{-\infty}^{r'} p(r)\, dr = (1 - \epsilon) \int_{-\infty}^{\infty} p(r)\, dr.^5 \tag{3.7}$$

PROOF. We compute

$$\begin{aligned} \varphi_1 &= C_1(\delta_0, \alpha) + 2\alpha[\delta_0 \Sigma(x_i - \xi)^2 - \Sigma(x_i - \xi)(x_{i+1} - \xi)] \\ \varphi_2 &= C_2(\delta_0, \alpha) + (1 + \delta_0^2)\Sigma(x_i - \xi)^2 - 2\delta_0 \Sigma(x_i - \xi)(x_{i+1} - \xi) \\ \varphi_3 &= -2n\alpha(1 - \delta_0^2)(\bar{x} - \xi). \end{aligned} \tag{3.8}$$

There is no difficulty in checking the conditions of Scheffé's theorems [4].

Next we apply Theorem 1 of the previous section, and define

$$\begin{aligned} V_2 &= (1 + \delta_0^2)\Sigma(X_i - \bar{X})^2 - 2\delta_0 \Sigma(X_i - \bar{X})(X_{i+1} - \bar{X}) \\ V_3 &= \bar{X} - \xi \\ U &= \frac{\Sigma(X_i - \bar{X})(X_{i+1} - \bar{X})}{V_2}. \end{aligned} \tag{3.9}$$

Conditions (i) and (ii) of Theorem 1 are easily seen to be satisfied. To show that U is independent of $V = (V_2, V_3)$ we employ arguments which have recently been used by various authors in a number of similar problems (cf. [13, 14, 15]).

It is seen that an orthonormal transformation exists:

$$X_1, \cdots, X_n \to Y_1, \cdots, Y_n$$

such that

$$\begin{aligned} \sqrt{n}\,\bar{X} &= Y_1 \\ \sum_{i=1}^{n} (X_i - \bar{X})(X_{i+1} - \bar{X}) &= \sum_{i=2}^{n} \lambda_i Y_i^2 \\ \sum_{i=1}^{n} (X_i - \bar{X})^2 &= \sum_{i=2}^{n} Y_i^2. \end{aligned} \tag{3.10}$$

[5] A corresponding result holds for the other one-sided case.

Under H the Y's are distributed with probability density

(3.11) $$p(y_1, \cdots, y_n) = C(\delta_0, \alpha) \exp\left[\alpha\left[k(y_1 - \sqrt{n}\xi)^2 + \sum_{i=2}^{n} \mu_i y_i^2\right]\right]$$

where $k, \mu_2, \cdots, \mu_n$ depend on δ_0 and where the μ's are all positive. Introducing new variables

(3.12) $$Z_i = \sqrt{\mu_i}\, Y_i, \qquad (i = 2, \cdots, n),$$

and, then, generalized polar coordinates in the space of the Z's,

(3.13) $$R = \sqrt{\sum_{i=2}^{n} Z_i^2}, \qquad \Psi_1, \cdots, \Psi_{u-2}$$

we see that Y_1, R and $\Psi_1, \cdots, \Psi_{n-2}$ are completely independent. Also

$$V_2 = R^2, \qquad V_3 = \frac{1}{\sqrt{n}}(Y_1 - \xi)$$

while U, being homogeneous of degree 0 in the Z's, is a function of the Ψ's only. This proves that U, V_2 and V_3 are completely independent. The type B_1 test of H is therefore given by

(3.14) $$u = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(x_{i+1} - \bar{x})}{(1 + \delta_0^2)\sum_{i=1}^{n} (x_i - \bar{x})^2 - 2\delta_0 \sum_{i=1}^{n} (x_i - \bar{x})(x_{i+1} - \bar{x})} < c_1, > c_2$$

where c_1 and c_2 are determined by

(3.15) $$\int_{c_1}^{c_2} u^s p(u)\, du = (1 - \epsilon) \int_{-\infty}^{\infty} u^s pu()\, du \qquad (s = 0, 1).$$

We still have to show that this test is equivalent to the one defined by (3.3) and (3.5). For $\delta_0 = 0$ this is trivial. Let us assume $\delta_0 < 0$. (The other case goes through similarly.) The inequality $u < c_1$ is equivalent to

(3.16) $$(1 + 2\delta_0 c_1)\Sigma(x_i - \bar{x})(x_{i+1} - \bar{x}) < (1 + \delta_0^2)\Sigma(x_i - \bar{x})^2$$

and hence to

(3.17) $$\frac{\Sigma(x_i - \bar{x})(x_{i+1} - \bar{x})}{\Sigma(x_i - \bar{x})^2} < w_1$$

provided $1 + 2c_1\delta_0 > 0$. Suppose $1 + 2c_1\delta_0 \leq 0$, i.e. $c_1 \geq -\frac{1}{2\delta_0}$. Then[6]

(3.18) $$P\{U < c_1\} \geq P\left\{U < -\frac{1}{2\delta_0}\right\} = P\{0 < \Sigma(X_i - \bar{X})^2\} = 1$$

[6] We denote the probability of an event A by $P\left\{A\right\}$.

i.e. $P\{U < c_1\} = 1$ which would contradict (3.15). Similarly if $1 + 2c_2\delta_0 \leq 0$ we would have $P\{U > c_2\} = 0$ and hence our test would be one-sided and therefore not unbiased. The inequalities $u < c_1, > c_2$ are thus equivalent to the inequalities $r < r_1, > r_2$ and since

$$u = \frac{r}{1 + \delta_0^2 - 2\delta_0 r},$$

(3.5) also follows.

The existence of type B_1 and uniformly most powerful one-sided tests of the hypothesis H is rather surprising. For when α and ξ are assumed known, neither the type A_1 test nor the uniformly most powerful one-sided test of the simple hypothesis H': $\delta = \delta_0$ exists. This is easily seen by determining the most powerful and the most powerful unbiased test against a specific alternative δ_1 for the hypothesis H' in the population

$$p(x_1, \cdots, x_n) = \frac{1 - \delta^n}{(\sqrt{2\pi})^n} \exp\left[-\tfrac{1}{2}[(1 + \delta^2)\Sigma x_i^2 - 2\delta\Sigma x_i x_{i+1}]\right]. \tag{3.19}$$

The distribution of the criterion R was obtained by R. L. Anderson [16] (see also [17]) for the case $\delta = 0$. Madow [15] using Anderson's result found the distribution for arbitrary δ. (Approximations to the distribution have been studied by various authors; for the literature on this cf. [18]. Recently Hsu [19] obtained an asymptotic expansion.) A direct derivation for arbitrary δ may be based on the following theorem of Cramér, which was communicated to the author by Dr. P. L. Hsu.

THEOREM 3. (*Cramér*)[7]. *If X, Y are two random variables, (not necessarily independent), $Y > 0$, then*

$$P\left\{\frac{X}{Y} \leq x\right\} = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\varphi_x(t) - \psi(t)}{it}\, dt \tag{3.20}$$

where φ_x and ψ are the characteristic functions of $X - xY$ and Y respectively, provided

$$\int_{-\infty}^{\infty} \left|\frac{\varphi_x(t) - \psi(t)}{t}\right| dt < \infty. \tag{3.21}$$

THEOREM 4. *If*

$$\begin{aligned} p(x_1, \cdots, x_n) &= \frac{1 - \delta^n}{(\sqrt{2\pi}\,\sigma)^n} \\ &\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i - \xi) - \delta(x_{i+1} - \xi)]^2\right], \qquad (x_{n+1} = x_1) \end{aligned} \tag{3.22}$$

[7] Differentiated forms of the theorem were given by R. C. Geary [*Jour. Roy. Stat. Soc.* Vol. 107 (1944) p. 56] and H. Cramér [Exercise 6 on p. 317 of *Mathematical Methods of Statistics*. Princeton Univ. Press (1946)].

and if

$$R = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(X_{i+1} - \bar{X})}{\sum_{i=1}^{n} (X_i - X^2)}, \tag{3.23}$$

then

$$P\{R > r\} = \frac{2^{n+1/2}}{n} \frac{1 - \delta^n}{(1 - \delta)(1 + \delta^2 - 2\delta r)} \cdot \sum_j \frac{(-1)^{j+1}(\lambda_j - r)^{n-3/2} \sin \frac{j\pi}{n} \sin \frac{2j\pi}{n}}{1 + \delta^2 - 2\delta\lambda_j} \tag{3.24}$$

where the summation is extended over all integer j, $1 \leq j \leq \frac{n}{2}$, *for which* $\lambda_j > r$, *and where*

$$\lambda_j = 2 \cos \frac{2j\pi}{n}. \tag{3.25}$$

The proof of this theorem from Theorem 3 is straightforward and only will be indicated here. If X and Y denote the numerator and denominator of R respectively, the characteristic functions of Y and $X - rY$ may be obtained by the method of circulants (cf. [12, 17]). The integral on the right hand side of (3.20) is then easily evaluated by the theory of residues when n is odd. In the case that n is even, the integrand has two branchpoints, one in the lower and one in the upper half plane. These may be separated, and then again the method of residues may be applied.

4. Similar regions. The problem of finding all regions similar to the sample space with respect to a parameter θ was solved by Neyman and Pearson [1] for a certain class of probability laws. In a later paper Neyman proved ([20] proposition IX) that if there exists a sufficient statistic T for a parameter θ, then w is similar with respect to θ if it has the following structure: For the intersection $w(t)$ of w with the surface $T = t$, the relative probability of $w(t)$ given $T = t$ has a constant value independent of t. We shall show in this section that for a large class of probability laws which admit a sufficient statistic for θ the regions with the above structure are the only ones that are similar with respect to θ.

We consider samples from a univariate distribution and we distinguish three cases as one, both or neither of the extremes of the range of the distribution depend on the parameter θ. For the first of these cases (cf. Pitman [21]) we consider samples from a distribution with probability density

$$p(x) = \frac{f(x)}{g(\theta)}, \qquad k(\theta) \leq x \leq c, \tag{4.1}$$

where $k(\theta)$ is a strictly monotone continuous function of θ and where c may be infinite. Introducing a new parameter $\delta = k(\theta)$ the distribution of a sample from (4.1) is given by

$$p(x_1, \cdots, x_n) = \frac{f(x_1) \cdots f(x_n)}{b(\delta)}, \qquad \delta \leq x_i \leq c. \tag{4.2}$$

To obtain the totality of regions w similar with respect to δ let us denote by $W_1, \cdots, W_n$ the portions of the sample space where the smallest of the x's is $x_1, \cdots, x_n$ respectively. For any region w denote by w_k' the intersection of w with W_k. Consider a transformation carrying $W_2, \cdots, W_n$ into W_1, letting $y_1 = \min(x_1, \cdots, x_n)$ and letting in W_k:

$$y_2 = x_1, \; y_3 = x_2, \cdots, y_k = x_{k-1}, \; y_{k+1} = x_{k+1}, \cdots, y_n = x_n. \tag{4.3}$$

Denote by w_k the image of w_k' under this transformation. The condition that w be similar with respect to δ,

$$\int_w \frac{f(x_1) \cdots f(x_n)}{b(\delta)} \, dx_1 \cdots dx_n \overset{(\delta)}{\equiv} \epsilon, \tag{4.4}$$

may be written in the form

$$\begin{aligned} &\int_\delta^c \frac{f(y_1)}{b(\delta)} \left\{ \sum_{k=1}^n \int_{w_k(y_1)} f(y_2) \cdots f(y_n) \, dy_2 \cdots dy_n \right\} dy_1 \\ &\qquad \overset{(\delta)}{\equiv} n\epsilon \int_\delta^c \frac{f(y_1)}{b(\delta)} \left\{ \int_{(W(y_1))} f(y_2) \cdots f(y_n) \, dy_2 \cdots dy_n \right\} dy_1 \end{aligned} \tag{4.5}$$

where $W(y_1)$ denotes the region $y_1 \leq y_i \leq c$, $(i = 2, \cdots, n)$, that is, the region of variation of $y_2, \cdots, y_n$ given y_1, and where $w_k(y_1)$ denotes the region of variation of $y_2, \cdots, y_n$ given y_1 and w_k. From (4.5) we obtain

$$\frac{1}{b(\delta)} \int_\delta^c f(y_1) \psi(y_1) \, dy_1 \overset{(\delta)}{\equiv} 0 \tag{4.6}$$

where

$$\begin{aligned} \psi(y_1) = &\sum_{k=1}^n \int_{w_k(y_1)} f(y_2) \cdots f(y_n) \, dy_2 \cdots dy_n \\ &- n\epsilon \int_{y_1}^c \cdots \int_{y_1}^c f(y_2) \cdots f(y_n) \, dy_2 \cdots dy_n. \end{aligned} \tag{4.7}$$

But (4.6) implies

$$\psi(y_1) = 0 \text{ almost everywhere} \tag{4.8}$$

and since we can only determine w up to a set of measure 0, we may omit the qualification in (4.8). Therefore a necessary and sufficient condition for w to be similar is

$$\frac{1}{n\left[\int_{y_1}^{c} f(y)\,dy\right]^{n-1}} \sum_{k=1}^{n} \int_{w_k(y_1)} f(y_2) \cdots f(y_n)\,dy_2 \cdots dy_n = \epsilon \tag{4.9}$$

for all y_1.

To see more clearly the structure of these regions, let us take $n = 2$. Equation (4.9) states that on each of the broken lines of Fig. 1 the relative probability of $w = w_1' + w_2'$ given $Y_1 = y_1$ is ϵ, where the decomposition of this probability into its two components may vary with y_1.

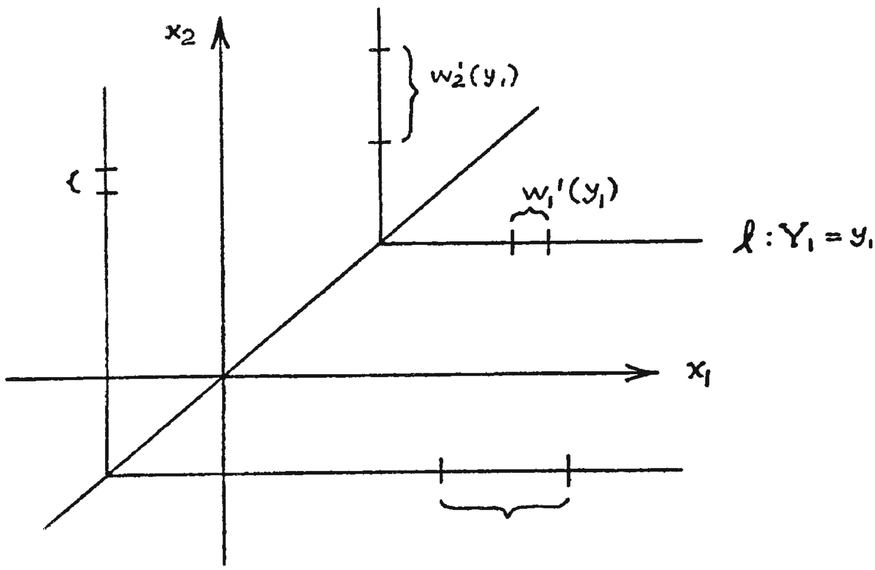

FIG. 1

In general equation (4.9) states that on each hyperplane $Y_1 = y_1$ the relative probability of w is independent of y_1. Since $Y_1 = \min (X_1, \cdots, X_n)$ is a sufficient statistic for θ, Neyman's theorem in this case does give all similar regions.

Next let us consider the case where both extremes of the range of the distribution depend on the parameter. We shall assume (cf. [21]) that $X_1, \cdots, X_n$ are distributed with probability density

$$p(x) = \frac{f(x)}{g(\theta)} \quad \text{in} \quad \theta \leq x \leq b(\theta) \tag{4.10}$$

where b is a strictly decreasing continuous function over an interval $[-\infty, b(-\infty)]$ and where $b[b(-\infty)] = -\infty$. These assumptions insure that there exists a unique number a, $-\infty < a < b(-\infty)$, such that $b(a) = a$.

Denote by W_{ij}, $(i, j = 1, \cdots, n; i \neq j)$, the portion of the sample space where the smallest and the largest of the x's are x_i and x_j respectively. Denote by W_{ij1} and W_{ij2} those portions of W_{ij} where x_i is greater than and less than $b^{-1}(x_j)$ respectively. For any region w denote by w_{ijk}' the intersection of w with

W_{ijk}. Consider a transformation carrying the sample-space into W_{1n}, letting $y_1 = \min(x_1, \cdots, x_n)$, $y_n = \max(x_1, \cdots, x_n)$ and in W_{ij} letting $y_2, \cdots, y_{n-1}$ denote the remaining x's in the order of their subscripts. Next make a transformation carrying W_{1n} into W_{1n1}, letting $z_1 = \max[y_1, b^{-1}(y_n)]$, $z_n = \min[y_1, b^{-1}(y_n)]$ and $z_k = y_k$ for $k = 2, \cdots, n-1$. Denote by w_{ijk} the image of w'_{ijk} in W_{1n1}.

Then Z_n is a sufficient statistic for θ (cf. [21]) and there exist functions f_1, g_1 such that the density of Z_n is given by

$$p(z_n) = \frac{f_1(z_n)}{g_1(\theta)} \quad \text{in} \quad \theta \leq z_n \leq a \tag{4.11}$$

while the distribution of the remaining Z's given Z_n is independent of θ.

The condition that w be a similar region may now be written, analogously to (4.5), in the form

$$\int_a^\theta \frac{f_1(z_n)}{g_1(\theta)} \sum_{i,j,k} \int_{w_{ijk}(z_n)} p(z_1, \cdots, z_{n-1} \mid z_n)\, dz_1 \cdots dz_{n-1}\, dz_n \equiv \epsilon \int_\theta^a \frac{f_1(z_n)}{g_1(\theta)}\, dz_n \tag{4.12}$$

and hence by the argument which led to (4.6), as

$$\sum_{i,j,k} \int_{w_{ijk}(z_n)} p(z_1, \cdots, z_{n-1} \mid z_n)\, dz_1 \cdots dz_{n-1} = \epsilon \qquad \text{for all } z_n. \tag{4.13}$$

Thus in this case also Neyman's theorem gives the most general similar region.

For the case that neither extreme of the range of the distribution depends on the parameter θ, it has been shown by various authors [22, 21, 23] under slightly varying assumptions concerning the regularity of the distribution function, that the existence of a sufficient statistic implies

$$p(x \mid \theta) = \exp[P(\theta) + T(x)Q(\theta) + R(x)]. \tag{4.14}$$

This (cf. [10]) is a special case of that for which Neyman and Pearson determined the totality of similar regions, however under the restriction that the moments of $\Phi = \frac{\partial}{\partial\theta} \sum_{i=1}^n \log p(X_i)$ uniquely determine the distribution of Φ. We shall briefly indicate how this assumption may be avoided.

Let $X_1, \cdots, X_n$ be a sample from (4.14), or, more generally, (this is the case considered by Neyman and Pearson), let $X_1, \cdots, X_n$ be distributed with probability density

$$\begin{aligned} &p(x_1, \cdots, x_n) \\ &\quad = \exp[P(\theta) + u(x_1, \cdots, x_n)Q(\theta) + v(x_1, \cdots, x_n)] \end{aligned} \tag{4.15}$$

in a sample space W_+ which is independent of θ. We shall assume that the set of values which Q takes on contains at least some interval. Introducing $\delta =$

$-Q(\theta)$ as a new parameter, we shall obtain all regions similar to δ (where the set of values of δ contains an interval) for the distribution[8]

$$\begin{aligned} p(x_1, \cdots, x_n) \\ = \exp\,[p_1(\delta) - \delta \cdot u(x_1, \cdots, x_n) + v(x_1, \cdots, x_n)] \end{aligned} \tag{4.16}$$

under the assumption that $\sum_{i=1}^{n} \left(\frac{\partial u}{\partial x_i}\right)^2 \neq 0$ except possibly on a set of measure 0.

Let us for a moment assume that there exist functions $f_i(x_i, \cdots, x_n)$, $(i = 2, \cdots, n)$, with continuous partial derivatives almost everywhere and such that the transformation

$$y_1 = u(x_1, \cdots, x_n); \quad y_i = f_i(x_1, \cdots, x_n), \qquad (i = 2, \cdots, n), \tag{4.17}$$

is one to one on W_+ except possibly on a subset of measure 0. Applying this transformation we may write the condition of similarity in the form

$$\begin{aligned} \int_{-\infty}^{\infty} e^{P_1(\delta)-\delta y_1} \int_{w(y_1)} f(y_1, \cdots, y_n)\, dy_2 \cdots dy_n \cdot dy_1 \\ = \epsilon \int_{-\infty}^{\infty} e^{P_1(\delta)-\delta y_1} \int_{W(y_1)} f(y_1, \cdots, y_n)\, dy_2 \cdots dy_n \cdot dy_1 \end{aligned} \tag{4.18}$$

where $W(y_1)$ denotes the region of variation of $y_2, \cdots, y_n$ given y_1, and where $w(y_1)$ denotes the region of variation of $y_2, \cdots, y_n$ given y_1 and w. Furthermore $f(y_1, \cdots, y_n)$ is independent of δ. From the theory of bilateral Laplace transforms it is known that (4.18) implies that

$$\int_{w(y_1)} f(y_1, \cdots, y_n)\, dy_2 \cdots dy_n = \epsilon \int_{W(y_1)} f(y_1, \cdots, y_n)\, dy_2 \cdots dy_n \tag{4.19}$$

which is the desired result.

More generally it may be shown that our assumption concerning $u(x_1, \cdots, x_n)$ insures the existence of functions f_i, $(i = 2, \cdots, n)$, such that under the transformation (4.17) no point $(y_1, \cdots, y_n)$ has more than a denumerable infinity of counter images in x-space. Our proof can be modified to cover this case. The argument is similar to that used to obtain equations (4.9) and (4.13) which were also arrived at through many to one transformations.

5. Testing exponential and rectangular distributions. In their fundamental 1928 paper [24] on likelihood ratio tests, Neyman and Pearson discussed various hypotheses relating to normal, exponential and rectangular distributions. Later they and other authors developed a theory of similar and bisimilar regions which made it possible to obtain optimum tests of many composite hypotheses with

[8] An assumption that we can solve for θ as a function of δ is not needed since we can determine $P_1(\delta)$ by integrating the density (4.16) over W_+.

one constraint concerning normal populations. This theory however is not applicable to most hypotheses concerning exponential or rectangular distributions. We shall in this section obtain optimum tests of some hypotheses relating to these latter distributions, using the method of the previous section.

Let us first consider a sample $X_1, \cdots, X_n$ from an exponential population, the probability density of the sample being.

$$(5.1) \quad p(x_1, \cdots, x_n) = \frac{1}{a^n} \exp\left[-\frac{1}{a}\sum_{i=1}^{n}(x_i - b)\right] \quad \text{if } x_i > b, \quad (i = 1, \cdots, n)$$

and let us consider the two hypotheses $H_1 : a =_0$, $H_2 : b = b_0$ where, without loss of generality, we shall take $a_0 = 1$, $b_0 = 0$. The likelihood ratio tests of both these hypotheses were shown to be completely unbiased by Paulson [25]. We shall prove

THEOREM 5. *The likelihood ratio tests of H_1 and H_2 are type B_1 and uniformly most powerful, respectively. The one-sided tests based on the likelihood ratio criterion for H_1 are the uniformly most powerful one-sided similar regions for testing this hypothesis.*

PROOF. In order to simplify the argument we shall give a detailed proof only for the restricted class of tests which are symmetric in the variables $X_1, \cdots, X_n$.

For testing H_1 let us make the following transformation introduced by Sukhatme [26]:

$$(5.2) \quad \begin{aligned} Z_1 &= nY_1 \\ Z_i &= (n - i + 1)(Y_i - Y_{i-1}), \qquad (i = 2, \cdots, n), \end{aligned}$$

where Y_i is the ith of the X's in order of magnitude. Then

$$(5.3) \quad p(z_1, \cdots, z_n) = \frac{1}{a^n} \exp\left[-\frac{1}{a}(z_1 - nb) - \frac{1}{a}\sum_{i=2}^{n} z_i\right] \quad \text{if} \quad z_1 \geq nb; z_i \geq 0 \qquad (i = 2, \cdots, n).$$

We want to determine all regions w which under H are similar to the sample space with respect to b, i.e. all regions w satisfying

$$(5.4) \quad \begin{aligned} &\int_w e^{-(z_1 - nb)} \exp\left[-\sum_{i=2}^{n} z_i\right] dz_2 \cdots dz_n \, dz_1 \\ &= \int_{nb}^{\infty} e^{-(z_1 - nb)} \left\{\int_{w(z_1)} \exp\left[-\sum_{i=2}^{n} z_i\right] dz_2 \cdots dz_n\right\} dz_1 \\ &\overset{(b)}{\equiv} \epsilon \overset{(b)}{\equiv} \epsilon \int_{nb}^{\infty} e^{-(z_1 - nb)} \, dz_1 \end{aligned}$$

where $w(z_1)$ denotes the intersection of w with the hyperplane $Z_1 = z_1$. Now (5.4) is equivalent to

(5.5)
$$e^{nb}\int_{nb}^{\infty} e^{-z_1} f(z_1)\,dz_1 \overset{(b)}{\equiv} 0$$

where

(5.6)
$$f(z_1) = \int_{w(z_1)} \exp\left[-\sum_{i=2}^{n} z_i\right] dz_2 \cdots dz_n - \epsilon$$

and this in turn is equivalent to

(5.7)
$$f(z_1) = 0 \text{ for all } z_1 .$$

Of all the regions w satisfying (5.7) we want to determine the one which against a specific alternative, say a_1, has maximum power, i.e. for which

(5.8)
$$\int_{nb}^{\infty} e^{-(1/a_1)(z_1-nb)} \int_{w(z_1)} \exp\left[-\frac{1}{a_1}\sum_{i=2}^{n} z_i\right] dz_2 \cdots dz_n\, dz_1$$

is as large as possible. We thus see that w will have the desired properties if $w(z_1)$ is determined according to the two conditions

(5.9)
$$\int_{w(z_1)} \exp\left[-\sum_{i=2}^{n} z_i\right] dz_2 \cdots dz_n = \epsilon$$

and

(5.10)
$$\int_{w(z_1)} \exp\left[-\frac{1}{a_1}\sum_{i=2}^{n} z_i\right] dz_2 \cdots dz_n = \max.$$

Hence by the Neyman-Pearson fundamental lemma $w(z_1)$ is the set of points satisfying

(5.11)
$$\exp\left[\left(-\frac{1}{a_1}\sum_{i=2}^{n} z_i + \sum_{i=2}^{n} z_i\right)\right] \geq C(a_1, z_1)$$

and therefore according as a_1 is greater or less than 1, $w(z_1)$ is determined by

(5.12)
$$\sum_{i=2}^{n} z_i = \sum_{i=1}^{n} [x_i - \min(x_1, \cdots, x_n)] \geq k(a_1, z_1), \text{ or}$$
$$\sum_{i=2}^{n} z_i = \sum_{i=1}^{n} [x_i - \min(x_1, \cdots, x_n)] \leq k'(a_1, z_1).$$

But $\sum_{i=2}^{n} Z_i$ is independently distributed of Z_1 and under H the distribution of $\sum_{i=2}^{n} Z_i$ does not depend on a_1, in fact it is a chi-square distribution with $2n - 2$ degrees of freedom. Thus k and k', as determined by (5.9) are independent of a_1 and the two tests (5.12) are uniformly most powerful one-sided.

Next we consider the more restricted class of unbiased similar regions. For w to be unbiased we must have

$$\frac{d}{da}\left\{\frac{1}{a^n}\int_w \exp\left[-\frac{z_1-nb}{a}\right]\exp\left[-\frac{1}{a}\sum_{i=2}^n z_i\right]dz_1\cdots dz_n\right\}\Bigg|_{a=1}$$

$$(5.13)\quad = \int_{nb}^{\infty}(z_1-nb-n)\exp\left[-(z_1-nb)\right]\int_{w(z_1)}\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n\,dz_1$$

$$+\int_{nb}^{\infty}\exp\left[-(z_1-nb)\right]\int_{w(z_1)}\left(\sum_{i=2}^n z_i\right)\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n\,dz_1 = 0.$$

The first of the integrals in the middle member equals

$$(5.14)\quad \begin{aligned}&\int_0^{\infty}(z-n)\,e^{-z}\int_{w(z+nb)}\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n\,dz\\ &= \epsilon\int_0^{\infty}(z-n)\,e^{-z}\,dz = -(n-1)\epsilon.\end{aligned}$$

Therefore

$$(5.15)\quad \begin{aligned}&\int_{nb}^{\infty}e^{-(z_1-nb)}\int_{w(z_1)}\left(\sum_{i=2}^n z_i\right)\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n\,dz_1\\ &= (n-1)\epsilon = (n-1)\epsilon\int_{nb}^{\infty}e^{-(z_1-nb)}\,dz_1\end{aligned}$$

or

$$(5.16)\quad \int_{nb}^{\infty}e^{-z_1}\,g(z_1)\,dz_1 \overset{(b)}{\equiv} 0$$

where

$$(5.17)\quad g(z_1) = \int_{w(z_1)}\left(\sum_{i=2}^n z_i\right)\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n - (n-1)\epsilon.$$

Thus finally the condition of unbiasedness reduces to

$$(5.18)\quad \int_{w(z_1)}\left(\sum_{i=2}^n z_i\right)\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n = (n-1)\epsilon$$

and we seek the region $w(z_1)$ which satisfies (5.9), (5.10) and (5.18).

By the fundamental lemma $w(z_1)$ is given by

$$(5.19)\quad \exp\left[-\frac{1}{a_1}\sum_{i=2}^n z_i\right] \geq \left[C_1(a_1, z_1)\sum_{i=2}^n z_i + C_2(a_1, z_1)\right]\cdot\exp\left[-\sum_{i=2}^n z_i\right]$$

which is equivalent to

$$(5.20)\quad \sum_{i=2}^n z_i \leq k_1(a_1, z_1),\ \geq k_2(a_1, z_1)$$

where k_1 and k_2 are determined by (5.9) and (5.18), and are therefore independent of z_1 and a. Thus the region (5.20) which of all unbiased similar regions maximizes the power against the alternative $a = a_1$ is independent of a_1 and hence is a region of type B_1. This completes the proof since it is easily verified that (5.10) is equivalent to the likelihood ratio test.

The proof for regions which are not necessarily symmetric in the variables follows similarly if instead of the transformation (5.2) one uses a transformation $U_i = f_i(X_1, \cdots, X_n)$ which is one to one and such that $U_1 = Z_1$, $U_2 = \sum_{i=2}^{n} Z_i$. The distribution of $U_3, \cdots, U_n$ is then independent of a and b since U_1, U_2 are a pair of sufficient statistics for these parameters, and the proof carries over step by step.

Next we consider the hypothesis $H_2: b = 0$, and again we restrict ourselves to regions which are symmetric in the variables, although as before the proof can be modified to cover also nonsymmetric regions.

We first make the transformation to $Z_1, \cdots, Z_n$ given by (5.2). In the $n - 1$ dimensional space of $Z_2, \cdots, Z_n$, we then transform to new variables $U, \Psi_1, \cdots, \Psi_{n-2}$ where $U = \sum_{i=2}^{n} Z_i$ and where the Ψ's are the generalized polar angles. Obviously the distribution of the Ψ's does not depend on a, since they are homogeneous of degree 0 in the Z's. Furthermore the Ψ's are independently distributed of U since the probability density of the Z's is constant over the hyperplanes $U = u$. Thus

$$p(z_1, u, \psi_1, \cdots, \psi_{n-2}) = \frac{K}{a^n} \exp\left[-\frac{z_1 - nb}{a}\right] u^{n-2} e^{-u/a} p(\psi_1, \cdots, \psi_{n-2}). \tag{5.21}$$

We next introduce new variables

$$V = Z_1 + U \text{ and } T = \frac{Z_1}{Z_1 + U} \tag{5.22}$$

and find

$$p(v, t, \psi_1, \cdots, \psi_{n-2}) = \frac{K}{a^n} \exp\left[-\frac{v - nb}{a}\right] v^{n-1} (1 - t)^{n-2} p(\psi_1, \cdots, \psi_{n-2}) \quad \text{for } v \geq nb, \frac{nb}{v} \leq t \leq 1. \tag{5.23}$$

For w under H_2 to be similar with respect to a, we must have

$$\int_0^\infty \frac{K}{a^n} \exp\left[-\frac{v}{a}\right] v^{n-1} \int_{w_0(v)} (1 - t)^{n-2} p(\psi_1, \cdots, \psi_{n-2})\, dt\, d\psi_1 \cdots d\psi_{n-2} \cdot dv = \epsilon \int_0^\infty \frac{K}{a^n} \exp\left[-\frac{v}{a}\right] v^{n-1}\, dv \tag{5.24}$$

where $w(v)$ designates the intersection of w with the hyperplane $V = v$, and where $w_0(v)$ denotes the part of $w(v)$ lying between the hyperplanes $t = 0$ and $t = 1$.

Hence the condition of similarity may be written as

$$\int_0^\infty \exp\left[-\frac{v}{a}\right] v^{n-1} f(v)\, dv = 0 \qquad \text{for all } a > 0 \tag{5.25}$$

where

$$f(v) = \int_{w_0(v)} (1 - t)^{n-2} p(\psi_1, \cdots \psi_{n-2})\, dt\, d\psi_1 \cdots d\psi_{n-2} - \epsilon. \tag{5.26}$$

By the uniqueness theorem for Laplace transforms, (5.25) implies $f(v) = 0$ for all $v > 0$, so that the condition of similarity finally reduces to

$$\int_{w_0(v)} (1 - t)^{n-2} p(\psi_1, \cdots, \psi_{n-2})\, dt\, d\psi_1 \cdots d\psi_{n-2} = \epsilon. \tag{5.27}$$

Of all similar regions, let us find the one which has maximum power. Obviously we want to include in $w(v)$ all points for which $t < 0$. In addition we want to choose $w_0(v)$ such that

$$\int_{w_b(v)} (1 - t)^{n-2} p(\psi_1, \cdots, \psi_{n-2})\, dt\, d\psi_1 \cdots d\psi_{n-2} = \max \tag{5.28}$$

where $w_b(v)$ is that part of $w(v)$ in which $\max\left(0, \frac{nb}{v}\right) \leq t$.

If, for some alternative b, $w_0(v)$ is contained in $\frac{nb}{v} < t < 1$, then $w_b(v)$ and $w_0(v)$ coincide and hence (5.28) attains its maximum value ϵ whatever the position of $w_0(v)$ in $\frac{nb}{v} \leq t \leq 1$. If on the other hand $\frac{nb}{v}$ is so close to 1 that $\frac{nb}{v} \leq t \leq 1$ is too small to contain $w_0(v)$, then (5.28) attains its maximum for any $w_0(v)$ containing $\frac{nb}{v} \leq t \leq 1$. There exists therefore a unique $w_0(v)$ which maximizes (5.28) for all values of b and v, namely the region defined by

$$C(v) \leq t \leq 1 \tag{5.29}$$

where C is determined by (5.27).

Since under H_2, the statistics V and T are independent, C does not depend on v. The test

$$t \leq 0, \qquad \geq C \tag{5.30}$$

which we have just shown to be uniformly most powerful, is also the likelihood ratio test which completes the proof of the theorem.

We shall finally consider an example of an optimum test in connection with a rectangular distribution. Let $X_1, \cdots, X_n$ be independently and uniformly distributed over $(a, a + \theta)$, where θ is positive. *For testing the hypothesis* $H: a = a_0$, *the test*

$$\frac{Y_1 - a_0}{Y_n - Y_1} \leq 0, \qquad \geq C \tag{5.31}$$

where Y_1 *and* Y_n *are the smallest and the largest of the* X's *respectively, is the uniformly most powerful of all similar regions.*

The proof of this goes through very much like that for H_2 in Theorem 5. Without loss of generality we take $a_0 = 0$. Also again, to simplify the proof, we restrict ourselves to regions which are symmetric in the variables. We need the following lemma.

LEMMA. *Let* $X_1, \cdots, X_n$ *be independently and uniformly distributed over* $(a, a + \theta)$. *Let* Y_i *denote the* ith X *in order of magnitude, and let*

$$T_n = Y_n, \; T_k = \frac{Y_k}{Y_{k+1}}, \qquad (= 1, \cdots, n - 1). \tag{5.32}$$

Then for $a > 0$

$$p(t_1, \cdots, t_n) = \frac{n!}{\theta^n} t_n^{n-1} t_{n-1}^{n-2} \cdots t_2 \tag{5.33}$$

when

$$a \leq t_n \leq a + \theta, \; \frac{a}{t_n \cdot t_{n-1} \cdots t_{k+1}} \leq t_k \leq 1, \qquad (k = 1, \cdots, n - 1).$$

This is easily seen by applying the usual method of Jacobians. The inequalities describing the sample space of the T's are equivalent to the following more convenient ones:

$$a \leq t_n \leq a + \theta, \; \frac{a}{t_n} \leq t_1 t_2 \cdots t_{n-1} \leq 1; \; t_k \leq 1, \qquad (k = 1, \cdots, n - 1). \tag{5.34}$$

Let us denote by $w(t_n)$ the intersection of a region w with the hyperplane $T_n = t_n$, and by $w_0(t_n)$ that part of $w(t_n)$ contained in the cylinder $0 \leq t_k \leq 1$, $(k = 1, \cdots, n - 1)$; then we find as a necessary and sufficient condition for w to be similar with respect to θ (assuming H)

$$(n - 1)! \int_{w_0(t_n)} t_{n-1}^{n-2} t_{n-2}^{n-3} \cdots t_2 \, dt_{n-1} \cdots dt_1 = \epsilon. \tag{5.35}$$

Of all regions satisfying (5.35) we want to find the most powerful one. Let us first consider alternatives $a > 0$. If $w_a(t_n)$ denotes the common part of $w_0(t_n)$ and the region

$$\frac{a}{t_n} \leq t_{n-1} t_{n-2} \cdots t_1 \leq 1, \tag{5.36}$$

we must choose $w_a(t_n)$ such that

$$\int_{w_a(t_n)} t_{n-1}^{n-2} \cdot t_{n-2}^{n-3} \cdots t_2 \, dt_{n-1} \cdots dt_1 = \max. \tag{5.37}$$

From this it follows easily that against alternatives $a > 0$ the uniformly best choice for $w_0(t_n)$ is

$$t_1 t_2 \cdots t_{n-1} = \frac{y_1}{y_n} \geq C'(t_n), \tag{5.38}$$

and since under H, $\frac{Y_1}{Y_n}$ is independently distributed of T_n, $C'(t_n)$ does not depend on t_n.

Consider next alternatives $a < 0$. We include in the region of rejection all points for which $Y_1 \leq 0$. To determine $w_0(t_n)$ we notice that, given $Y_1 > 0$, the X's are uniformly distributed between 0 and $a + \theta$. (Provided $a + \theta > 0$; the case $a + \theta \leq 0$ is trivial). Hence the probability distribution of the T's given $Y_1 > 0$ is

$$p(t_1, \cdots, t_n \mid Y_1 > 0) = \frac{n!}{(a + \theta)^n} t_n^{n-1} \cdots t_2 \tag{5.39}$$

when

$$0 \leq t_n \leq a + \theta, \qquad 0 \leq t_k \leq 1 \qquad \text{for } k = 1, \cdots, n - 1.$$

Thus

$$\frac{p(t_1, \cdots, t_{n-1} \mid t_n, a < 0, Y_1 > 0)}{p(t_1, \cdots, t_{n-1} \mid t_n, a = 0)} \tag{5.40}$$

is independent of $t_1, \cdots, t_{n-1}$ and hence the power of w against alternatives $a < 0$ is independent of the choice of $w_0(t_n)$. Therefore the region

$$y_1 \leq 0, \ \frac{y_1}{y_n} \geq C' \tag{5.41}$$

is uniformly most powerful against all alternatives. But (5.41) is equivalent to

$$\frac{y_1}{y_n - y_1} \leq 0, \geq C. \tag{5.42}$$

It is interesting to compare this result with that for the corresponding simple hypothesis. Let H' be the hypothesis: $a = 0$ when the X's are assumed independently and uniformly distributed over $(a, a + 1)$. There exists no uniformly most powerful test of H'; instead the two uniformly most powerful one-sided tests exist. By analogy with the normal case one might then expect for H′ that of all tests with symmetric power-functions, there be a uniformly most powerful one. This however is not so: there exist infinitely many admissible tests with symmetric powerfunction.

In this and the previous section we restricted ourselves to problems involving only one nuisance parameter. However, the method applies also to problems involving several nuisance parameters.

In the usual way (cf. [20, 9]) the results of this section may be translated to give optimum sets of confidence intervals for estimating the parameters in question. In this connection it is an open question whether the confidence regions based on the type B_1 tests discussed in section 2 will always be intervals; one would expect this to be the case.

The author wishes to acknowledge his indebtedness to Professor P. L. Hsu for many helpful suggestions.

Added in proof: In a joint paper by Professor Henry Scheffé and the present author which has been submitted to the Proceedings of the National Academy of Sciences, a result is given concerning the existence of certain 1:1 transformations. This result bears on Section 4 of the present paper where a question arises concerning the existence of a 1:1 transformation. The existence of such a transformation is now assured and, as a consequence, the last paragraph of Section 4 has become superfluous.

REFERENCES

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

[2] J. Neyman, "Sur la vérification des hypothèses statistiques composées", *Bull. Soc. Math. France*, Vol. 63 (1935), p. 1.

[3] 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.

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

[5] P. L. Hsu, "Analysis of variance from the power function standpoint", *Biometrika*, Vol. 32 (1941), p. 62.

[6] J. B. Simaika, "On an optimum property of two important statistical tests", *Biometrika*, Vol. 32 (1941), p. 70.

[7] A. Wald, "On the power function of the analysis of variance test", *Annals of Math. Stat.*, Vol. 13 (1942), p. 434.

[8] 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.

[9] H. Scheffé, "On the ratio of the variances of two normal populations", *Annals of Math. Stat.*, Vol. 13 (1942), p. 371.

[10] J. Neyman, "L'estimation statistique traitée comme un problème classique de probabilité", *Actualités Sci. et Ind. No. 739*, Hermann et Cie, Paris, 1938.

[11] A. Wald, "Notes on the theory of statistical estimation and testing hypotheses", Columbia University.

[12] W. J. Dixon, "Further contributions to the problem of serial correlation", *Annals of Math. Stat.*, Vol. 15 (1944), p. 119.

[13] J. von Neumann, "Distribution of the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 12 (1941), p. 367.

[14] M. Kac, "A remark on independence of linear and quadratic forms involving independent Gaussian variables", *Annals of Math. Stat.*, Vol. 16 (1945), p. 400.

[15] W. G. Madow, "Note on the distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 16 (1945), p. 308.

[16] R. L. ANDERSON, "Distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 13 (1942), p. 1.
[17] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables", *Annals of Math. Stat.*, Vol. 13 (1942), p. 14.
[18] R. B. LEIPNIK, "Distribution of the serial correlation coefficient in a circularly correlated universe", *Annals of Math. Stat.*, Vol. 18 (1947), p. 80.
[19] P. L. HSU, "On the asymptotic distributions of certain statistics used in testing the independence between successive observations from a normal population", *Annals of Math. Stat.*, Vol. 17 (1946), p. 350.
[20] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability", *Roy. Soc. London Phil. Trans.*, Ser. A, Vol. 236 (1937), p. 333.
[21] E. J. G. PITMAN, "Sufficient statistics and intrinsic accuracy", *Proc. Camb. Phil. Soc.*, Vol. 32 (1936), p. 567.
[22] B. O. KOOPMAN, "On distributions admitting a sufficient statistic", *Trans. Amer. Math. Soc.*, Vol. 39 (1936), p. 399.
[23] D. DUGUÉ, "Application des propriétés de la limite au sens du calcul des probabilités à l'étude des diverses questions d'estimation", *J. École Poly.*, Vol. 3 (1937), p. 305.
[24] J. NEYMAN AND E. S. PEARSON, "On the use and interpretation of certain test criteria for purposes of statistical inference", *Biometrika*, Vol. 20A (1928), p. 175.
[25] E. PAULSON, "On certain likelihood ratio tests associated with the exponential distribution", *Annals of Math Stat.*, Vol. 12 (1941), p. 301.
[26] P. V. SUKHATME, "Tests of significance for samples of the χ^2 population with two degrees of freedom", *Annals of Eugenics*, Vol. 8 (1937), p. 52.