ON THE ELIMINATION OF NUISANCE PARAMETERS IN STATISTICAL PROBLEMS

YU. V. LINNIK
MATHEMATICAL INSTITUTE OF THE ACADEMY OF SCIENCES, LENINGRAD

1. Introduction

Consider a family of the distributions $\mathcal{P}_\theta$ characterized by the probability densities $\ell(x, \theta)$ with respect to a dominating measure $\mu(x)$ on the σ_n-algebra of a measurable space $(\mathfrak{X}, \mathfrak{A})$ with a parameter $\theta \in \Omega$. Later on $\mathfrak{X} \subset E_n$ will be a parallelepiped of the n-dimensional Euclidean space, $\mathfrak{A}$ the Borel σ-algebra; $\Omega \subset E_s$, in general, a compact in the s-dimensional Euclidean space, $\ell(x, \theta)$ being a function continuous with respect to θ for a fixed x, and for a given θ—almost everywhere continuous with respect to $\mu(x)$, the Lebesgue measure.

We shall consider the problems of hypothesis testing and unbiased estimation. The first class of problems will be formulated as follows.

Let $\Pi_1(\theta), \cdots, \Pi_r(\theta)$ be continuous functions of $\theta \in \Omega \subset E_s$; $r < s$. The hypothesis H_0 to be tested is composite and consists of the equations

$$\Pi_1(\theta) = 0, \cdots, \Pi_r(\theta) = 0 \tag{1.1}$$

which determine the set Ω_0 in the set Ω.

The alternative H_1 to H_0 consists of the inclusion $\theta \in \Omega \backslash \Omega_0$. Sometimes a Bayes distribution $B(\theta)$ on $\mathrm{y} \backslash \mathrm{y}_0$ is given; this converts H_1 into a simple hypothesis. The last set-up is perhaps not quite natural, but it is convenient for the primary investigation of the composite hypothesis H_0.

We study the tests of H_0 against the alternative H_1.

The problem of unbiased estimation will consist in the investigation of the behavior of the statistics $\xi(x)$ possesing the mathematical expectation $E(\xi|\theta) = F(\theta)$ unbiased with respect to this function in the presence of the relations (1.1).

The way the question is presented above does not, of course, cover all the important problems of hypotheses testing and unbiased estimation. For instance, the problems of sequential analysis are not covered in this way. But in the set-up described above we can find a series of problems which are very interesting and deep from the analytical point of view; some of these will be considered below. The proofs of the theorems formulated below are rather long and complicated; therefore it is not possible to exhibit them in this article. (See [12] for the simplest cases.)

2. Verifiable and nonverifiable functions

In general, let $\phi(x)$ be any randomized test of the hypothesis H_0. Form the power function $\varphi(\theta) = E(\phi(x)|\theta)$. The hypothesis H_0 relates only to the values of the functions $\Pi_1(\theta), \cdots, \Pi_r(\theta)$. If the function $\varphi(\theta)$ is not trivial (that is not constant), it is characterized only by the deviations of $\Pi_1, \cdots, \Pi_r$ from the values prescribed by H_0 for all $\theta \in \Omega_0$. In this case we shall say that the hypothesis H_0 is verifiable in an invariant way by means of the test ϕ. If such a test $\phi(x)$ with a nontrivial power function $\varphi(\theta)$ exists for the vector function $(\Pi_1(\theta), \cdots, \Pi_r(\theta))$, we shall say that the function $(\Pi_1(\theta), \cdots, \Pi_r(\theta))$ is verifiable in an invariant way (more concisely, verifiable). The question which now arises, given a family $\mathcal{P}_0 = \{\ell(x, \theta)\}$, is how to describe the verifiable functions $(\Pi_1(\theta), \cdots, \Pi_r(\theta))$.

This set-up for the first time appeared in 1940 in the work of G. Dantzig [1]. He studied Student's problem (for the repeated normal sample $x_1, \cdots, x_n \in N(a, \sigma^2)$ to test the hypothesis H_0: $a = a_0$), and proved that, in the terminology introduced above, $\Pi_1 = a$ is nonverifiable. In 1945 a well-known work of Charles Stein appeared; this showed that the function $\Pi_1 = a$ becomes verifiable in the setup of sequential analysis.

At present we have little information on verifiable and nonverifiable functions. For a repeated normal sample $x_1, \cdots, x_n \in N(a, \sigma^2)$, with the constant sample size n we can prove the following theorem.

THEOREM 2.1. *The functions* $\Pi_1 = (a/\sigma^\rho)$ *are nonverifiable for* $\rho < 1$.

The result of G. Dantzig follows from this theorem for $\rho = 0$. On the other hand, it is easy to prove that for $\rho = 1$ this property fails to hold; the function $\Pi_1 = a/\sigma$ is verifiable.

To test the hypothesis H_0: $(a/\sigma) = \gamma_0$ in an invariant way, it is sufficient to apply the nonrandomized test ϕ with the critical zone depending only upon the ratio

$$\frac{\bar{x}}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^{1/2}}. \tag{2.1}$$

Further, for the Behrens-Fisher problem, given two independent repeated normal samples H_0: $a_1 - a_2 = 0$, the function $a_1 - a_2$ proves to be nonverifiable. However, besides these separate results nothing more is known about the verifiable functions. For the case where $\mathcal{P}_\theta$ is an exponential family (cf. [3], pp. 50–59) the problem on the verifiable functions is reducible to certain rather peculiar questions of the theory of the multiple Laplace transforms.

3. Similar tests

The above results, however fragmentary, lead to the conjecture that the verifiable functions are seldom encountered in the usual problems, and if encountered, the corresponding tests $\phi(x)$ form a narrow class among all the tests and do not have desirable properties with respect to their power function. In view of this we can try to find tests for H_0, eliminating the nuisance parameters from the power function $E(\phi(x)|\theta) = \varphi(\theta)$ only for the null hypothesis H_0, so that $\varphi(\theta)$ becomes constant due to the relations (0, 1) implied by the null hypothesis H_0. We can pose the problem of describing a sufficiently complete class of these tests and of singling out the optimal ones in some sense.

This problem is linked in a direct way to the theory of similar regions introduced into statistics in a well-known work of J. Neyman and E. Pearson [4], and the Neyman structures [5], [3]. The last ones can be constructed if, after introducing the relations (1.1), the family $\mathcal{P}_\theta$ is converted into the family $\mathcal{P}_\theta^0$ depending upon the $(s - r)$ dimensional nuisance parameter in the Euclidean space E_{s-r} and admitting for this parameter suitable sufficient statistics. In particular cases there are no relations $(r = 0)$, and all the parameters are nuisance parameters. This problem appears in a natural way if it is required to test whether a family of distributions $\{\ell(x_1, \theta)\}$ has a density with the function ℓ of a given type on the evidence of N independent observations $x_1, \cdots, x_N$, where the parameter values θ are different for each observation (the problem of N small samples). For instance, suppose that we are given N samples from an exponential family, characterized by the density

$$p_\theta(x) = \exp - (\theta_1 T_1(x) + \cdots + \theta_s T_s(x))h(x) \tag{3.1}$$

where $x \in E_n$, and $T_1(x), \cdots, T_s(x)$ are sufficient statistics, and $s < n$. If there are no relations between the parameters $\theta_1, \cdots, \theta_s$, we have the complete exponential family, and all the similar regions have Neyman structure (see E. Lehmann, H. Scheffé [7], E. Lehmann [3]). In the problem of N small samples we can take as an alternative the hypothesis H_1 that the family is of the type

$$p_\theta(x) = \exp - [(\theta_1 T_1(x) + \cdots + \theta_s T_s(x)) + \epsilon V(x)]h(x) \tag{3.2}$$

where ϵ is a small number and $V(x)$ a suitable statistic. Here we shall have the same sufficient statistics $T_1(x), \cdots, T_s(x)$.

The case where the parameters are those of the affine transformations of the repeated sample and similar statistics—the affine invariants—was considered in detail by A. A. Petrov [13].

Note the possibility of construction of similar domains for certain mixtures of distributions when we have only trivial sufficient statistics (see [8], [10]). These mixtures are of the type

$$p_\theta(x) = R_1(T_1(x), \theta)r_1(x) + \cdots + R_q(T(x), \theta)r_q(x) \tag{3.3}$$

where $x \in E_m$, $\mathcal{P}(x) \in E_n$, $(n < m)$, and the densities $p_\theta(x)$ are taken with respect to a dominating measure $\mu(x)$; $q \geq 1$ is an integer, and R_j, r are measurable for $j = 1, 2, \cdots, q$, and

$$\int_{E_m} |R_j(T(x), \theta)(r_j(x)| d\mu(x) < \infty, \qquad (j = 1, 2, \cdots, n). \tag{3.4}$$

For $q = 1$, the $p_\theta(x)$ become the well-known families possessing nontrivial sufficient statistics.

We return now to the family $\mathcal{P}_\theta$ subject to the relations (1.1). If we can construct under such requirements nontrivial similar regions, they will lead us to the nonrandomized similar tests $\phi_1(x)$ of a given level $\alpha \in (0, 1)$ for which we have

$$E(\phi_1(x)|H_0) = \alpha. \tag{3.5}$$

For a randomized similar test $\phi(x)$ only this condition is obligatory; in general, its distribution may depend on nuisance parameters. As is well known (see [3]), the randomized similar test can be converted into a nonrandomized one of the same level α, if a new random variable $\mathfrak{U}$ is introduced which is independent of the observation x and uniformly distributed on the segment $[0, 1]$. Then the test ϕ^* with the critical region $\mathfrak{U} - \phi(x) < 0$ will be nonrandomized in the space $\{\mathfrak{U}\} \times \{x\}$ and will be similar with the same level α as the initial test.

Let our family $\mathcal{P}_\theta = \{p_\theta(x)\}$ admit nontrivial sufficient statistics

$$T = (T_1, \cdots, T_k) \tag{3.6}$$

where $k < n$ (we suppose that no pathological situations as described by D. Basu in [9] arise).

Then the transition from the randomized tests $\phi(x)$ to the nonrandomized ones does not require the supplementary random and uniformly distributed variable $\mathfrak{U}$; it can be constructed by means of the observations relevant to the problem.

Suppose we are given a randomized test $\phi(x)$ of level α. Form the expression

$$\phi_1(T) = E(\phi(x)|T). \tag{3.7}$$

It will also be a similar level α test measurable with respect to the σ-algebra of sufficient statistics. Moreover, take a measurable scalar function $V(x)$ and form the conditional distribution

$$F(y|T) = P(V < y|T). \tag{3.8}$$

This distribution does not depend on θ. Suppose that for almost all the values of T, the function $F(y|T)$ is strictly monotone with respect to y. Then, as is well known, the transformation $\mathfrak{U} = F(V|T)$ gives a random variable $\mathfrak{U}$ which for almost all given T, is uniformly distributed on $[0, 1]$.

If we define now the nonrandomized test $\phi^*(x)$ with the critical region $\mathfrak{U} - \phi_1(T) < 0$, then it will depend only on the observations $x \in \mathfrak{X}$ and will be a similar level α test. In general, its power function will be the same as that of the initial randomized test $\phi(x)$.

Hence, in the construction of the tests which do not depend only on sufficient statistics, there are no essential differences between the randomized and nonrandomized tests; one can easily pass (in principle, at least) from the former to the latter without changing the power.

However, if we restrict ourselves to the tests depending only on sufficient statistics, this difference becomes essential, and the construction of nonrandomized similar tests becomes difficult, as we shall see further on in the example of the Behrens-Fisher problem.

If there are nontrivial sufficient statistics $T = (T_1, \cdots, T_n)$, then instead of arbitrary tests $\phi(x)$ we can consider the tests $\phi_1(T)$ equivalent to them with respect to power. These are determined by the formula (3.7) and depend only on sufficient statistics. In what follows, we shall consider only such tests, and this basically for incomplete exponential families.

4. Similar and unbiased tests for incomplete exponential families

For incomplete exponential families, the principal analytical tool for describing the similar and unbiased tests and unbiased estimates can be obtained from the theory of the ideals of holomorphic functions and the theory of analytical sheaves connected with them. Similar tests generate an analytical sheaf of ideals, and their description can be effectuated by means of "theorem B" of H. Cartan on the behavior of the first cohomology group [11].

To single out the cases where the similar tests are described rather simply, we first impose certain requirements on the structure of exponential families and the relations formed by the null hypothesis, and then we shall weaken these requirements.

Conditions upon the exponential family. The exponential family is given by the density, with respect to the Lebesgue measure, of its sufficient statistics $T_1, \cdots, T_s$,

$$P_\theta(T_1, \cdots, T_s) = C(\theta) \exp \{\theta_1 T_1 + \cdots + \theta_s T_s\} h(T_1, \cdots, T_s). \tag{4.1}$$

Denote by $\mathfrak{I} \subset E_s$ the range of values of the sufficient statistics $(T_1, \cdots, T_s)$. The conditions required are the following.

(I) There is a number $s_1 \leq s$ (s_1 might be equal to 0) such that $h(T_1, \cdots, T_s) = 0$ if at least one of the variables $T_j < 0$, $(j = 1, 2, \cdots, s_1)$. This defines the carrier $\mathfrak{I}$ of the function $h(T_1, \cdots, T_s)$.

(II) In all interior points of $\mathfrak{I}$, the function $h(T_1, \cdots, T_s)$ does not vanish and has there continuous partial derivatives. Moreover, in the domain $\mathfrak{I}_\epsilon \subset \mathfrak{I}$ defined by the inequalities $T_1 \leq \epsilon$ and $T_s \geq \epsilon$ for any $\epsilon > 0$, we have the estimate

$$\left|\frac{\partial \ln h}{\partial T_1}\right| + \cdots + \left|\frac{\partial \ln h}{\partial T_s}\right| = 0\left(\frac{1}{\epsilon^a} + 1\right) \tag{4.2}$$

where $a \leq 1$ is a constant.

(III) The integral $\int \underset{\mathfrak{I}}{\cdots} \int P_\theta(T_1, \cdots, T_s)\, dT_1 \cdots dT_s$ is absolutely convergent for $\theta = (\theta_1, \cdots, \theta_s) \in \mathcal{P}$ where $\mathcal{P}$ is the product $\mathcal{P} = R_1 \times \cdots \times R_{S_1} \times S_{S_1+1} \times \cdots \times S_s$ of s_1 right half-planes $\operatorname{Re} \theta_j > 0$ and $(s - s_1)$ strips $0 < \operatorname{Re} \theta_j < A_j$. (Of course, any open vertical strip can be reduced to the type of the strips S_j by shifting the parameter values.)

Conditions upon the null hypothesis H_0. For real points $\theta = (\theta_1, \cdots, \theta_s) \in \mathcal{P}$, the null hypothesis H_0 is determined by $r < s$ relations

$$\Pi_1(\theta_1, \cdots, \theta_s) = 0; \qquad \Pi_r = (\theta_1, \cdots, \theta_s) = 0. \tag{4.3}$$

The functions $\Pi_1, \cdots, \Pi_r$ must be real for real $(\theta_1, \cdots, \theta_s)$. After multiplying by $[(\theta_1 + 1) \cdots (\theta_s + 1)]^{-N}$ where N is a suitable number, the functions $\Pi_1, \cdots, \Pi_r$ must become functions of $(1/\theta_1 + 1), \cdots, (1/\theta_s + 1)$ holomorphic on the closure of $\mathcal{P}$, that is on $\bar{\mathcal{P}}$ (including the points with $\theta_j = \infty$).

(Such conditions are always fulfilled, for instance, for the case of polynomial relations, which often appear in statistics.)

The null hypothesis H_0 consists in the fulfillment of the relations (4.3) on the compact Ω of real numbers, defined by the inequalities $\epsilon_j \leq \theta_j \leq E_j$; $E_j < A_j$, $(j = 1, 2, \cdots, s)$, for ϵ_j sufficiently small. The corresponding set of points will be denoted by Ω_0. The alternatives to H_0 consist in the inclusion $(\theta_1, \cdots, \theta_s) \in \Omega\backslash\Omega_0$. The alternatives are provided with a Bayes probability measure $B(\theta)$ defined on $\Omega\backslash\Omega_0$ which converts them into a simple hypothesis. For the test ϕ similar with respect to H_0 we introduce the Bayes gain

$$W(\phi|B) = \int_{\Omega\diagdown\Omega_0} E(\phi|\theta)\, dB(\theta). \tag{4.4}$$

Further conditions on the relations. (I) The equations (4.3) considered in the complex domain must generate there an analytical set of points $V_{\pi_1\cdots\pi_r}$ which can be decomposed into a finite number of disjoint components $V^q_{\pi_1\cdots\pi_r}$, each of complex dimension $(s - r)$ and each containing inside a connected set $\mathcal{R}^q_{\pi_1\cdots\pi_r}$ of real points entering into Ω_0 and having a real dimension $(s - r)$.

(II) rank $\left\| \dfrac{\partial \Pi_i}{\partial \theta_j} \right\| = r$ inside $\mathcal{P}$ $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s)$.

We can consider the similar tests $\phi(T_1, \cdots, T_s)$ of the null hypothesis H_0 depending upon the sufficient statistics only. As was explained above, any similar test is equivalent to such a test. We can now give a description of an "everywhere dense" family of similar tests.

A simple case of an analogous set-up was considered by Robert A. Wijsman [27].

THEOREM 4.1. *For a given* $\epsilon > 0$ *as small as we please and for a given level* α *similar test* $\phi = \phi(T_1, \cdots, T_s)$, *we can indicate the similar test* $\phi_\epsilon = \phi_\epsilon(T_1, \cdots, T_s)$ *such that*

$$|W(\phi_\epsilon|B) - W(\phi|B)| \leq \epsilon \tag{4.5}$$

for which we have the representation

$$\phi_\epsilon(T_1, \cdots, T_s) = \alpha + \frac{1}{h}(A_1^*H_1 + \cdots + A_r^*H_r). \tag{4.6}$$

Here $A_j = A_j(T)$, $(j = 1, 2, \cdots, r)$ *are pre-images of the functions*

$$\Pi_j(\theta_1, \cdots, \theta_s)^{-(N+1)}$$

*for a one-sided Laplace transform and the asterisk * is the convolution sign. The functions* $H_1, \cdots, H_r$ *have a prescribed number of partial derivatives and have the estimate*

$$H_j(T_1, \cdots, T_s) = 0(\exp \zeta(|T_1| + \cdots + |T_s|); \qquad j = 1, 2, \cdots, r \tag{4.7}$$

where $\zeta > 0$ *is as small as we please. If* i *is one of the numbers* $1, 2, \cdots, s_1$ *and* $T_i < 0$, *then* $H_j(T_1, \cdots, T_s)$ *vanishes.*

This theorem enables us to solve, at least in principle, the problem of the choice of the "ϵ-optimal" similar test for a given Bayes distribution B. To find an ϵ_n-optimal "cotest" $\psi_\epsilon = \phi_\epsilon - \alpha$, we look for $H_1, \cdots, H_r$ such that

$$\psi_\epsilon = \frac{1}{h}(A_1^* H_1 + \cdots + A_r^* H_r) \tag{4.8}$$

under the restrictions

$$-\alpha \le \psi_\epsilon \le 1 - \alpha \tag{4.9}$$

gives the largest possible value to

$$W(\psi_\epsilon | B) = \int_{\Omega \setminus \Omega_0} E(\psi_\epsilon | B)\, dB(\theta). \tag{4.10}$$

We thus obtain a variational problem with restrictions. For its solution one can apply the methods of linear programming (see [12]).

We return now to the numerous requirements imposed upon $h(T_1, \cdots, T_s)$ and the relations (4.3), implied by the null hypothesis H_0. Among these conditions the rather restrictive one is the requirement that $h(T_1, \cdots, T_s) \neq 0$ inside $\mathfrak{J}$.

If we reject this condition, the difference will be only that the functions $H_1, \cdots, H_r$ in the formula should be chosen so that the expression

$$A_1^* H_1 + \cdots + A_r^* H_r \tag{4.11}$$

vanishes at the points where $h(T_1, \cdots, T_s) = 0$. In this domain we put $\phi = \alpha$.

We pass now to the structure of unbiased tests of H_0 against the alternative H_1. By virtue of the known theorems of test theory (see [3]) in our set-up, the unbiased tests will form a part of the set of similar tests. Under the conditions of theorem 4.1 we can describe an everywhere dense set among all the unbiased tests—the set of all sufficiently smooth unbiased tests $\mathcal{G}$. Namely, for any unbiased test ϕ there exists a $\phi_\epsilon \in \mathcal{G}$ such that the condition (4.5) will be fulfilled. The general form of the tests of the set $\mathcal{G}$ is given by the following theorem.

THEOREM 4.2. *Under the conditions of theorem* 4.1, *any sufficiently smooth unbiased test can be represented in the form*

$$\phi = \alpha + \frac{1}{h} \sum_{i,j=1}^{r} A_i^* A_j^* H_{ij}, \tag{4.12}$$

A_i *being the functions defined above,* H_{ij}, $(i, j = 1, 2, \cdots, r)$ *functions of the same type as the* H_j *introduced above.*

If the condition about the nonvanishing of h in the domain $\mathfrak{Z}$ is violated, we must require that $\sum_{i,j=1}^{r} A_i^* A_j^* H_{ij}$ *vanish at the points where* $h = 0$, *and put* $\phi = \alpha$ *at these points.*

5. Unbiased estimates

As is well known from the theorem of C. R. Rao [15] and D. Blackwell, in a sense the unbiased estimates cannot deteriorate if "projected" into the space of sufficient statistics. We shall consider the exponential families (4.1) and sufficiently smooth statistics depending only upon the sufficient statistics $T_1, \cdots, T_s$ and fulfilling the condition

$$\xi(T_1, \cdots, T_s) = O(\exp \zeta(|T_1| + \cdots + |T_s|) \tag{5.1}$$

for any $\zeta > 0$. Each such statistic will be an unbiased estimate of $E(\xi|\theta) = \ell(\theta)$. If there are no relations, so that the family is complete, the unbiased estimate of $\ell(\theta)$ is unique with probability 1. If there are relations (4.3), then all the unbiased estimates of $\ell(\theta)$ differ by unbiased estimates of zero χ, that is, the statistics satisfying the conditions $E(\chi|\theta) = 0$ and called U.E.Z. for short. The set of smooth U.E.Z. with the growth condition (5.1) are described by the following theorem.

THEOREM 5.1. *Under the conditions of theorem* 4.1, *U.E.Z., which are sufficiently smooth and fulfill the growth condition* (5.1) *are described by the formula*

$$\chi = \frac{1}{h}(A_1^* H_1 + \cdots + A_r^* H_r) \tag{5.2}$$

in the notation of section 4.

If h vanishes inside $\mathfrak{Z}$, we must, as indicated earlier, choose H_j so that the numerator of the fraction (5.2) vanishes at the corresponding points. Taking into account the description of all sufficiently smooth unbiased estimates of zero (5.2) for the case of incomplete exponential families considered by us, one can establish certain cases of inadmissibility of unbiased estimates. For instance, A. M. Kagan established that for a repeated sample $x_1, \cdots, x_n$ of a one-parameter family

$$P_\theta(x) = C_0 \exp - (x - \theta)^{2k}, \qquad k \geq 2 \tag{5.3}$$

the sample moments $\bar{x}$, $a_m = (1/n) \sum_{i=1}^{n} x_i^m$ for $2 \leq m \leq 2k - 2$ are unbiased estimates of the corresponding moments of the distribution which are inadmissible on any compact set of values of the parameter θ. (For the estimate of $\bar{x}$ this follows from the well-known theorem of C. R. Rao [15].)

6. Investigations on the Behrens-Fisher problem

In this section are expounded the investigations of Leningrad statisticians in the period of 1963–1965 on the Behrens-Fisher problem—a classical problem on the elimination of nuisance parameters.

First, we shall present the results obtained by applying to the Behrens-Fisher problem the theory of similar tests for incomplete exponential families expounded above. The corresponding normal samples will be denoted $(x_1, \cdots, x_{n_1}) \in N(a_1, \sigma_1^2)$, $(y_1, \cdots, y_{n_2}) \in N(a_2, \sigma_2^2)$, and the sufficient statistics $\bar{x}$, $\bar{y}$, s_1^2, s_2^2. Consider the similar tests ϕ depending only on $|\bar{x} - \bar{y}|$, s_1^2, s_2^2.

Introducing instead of $\bar{x} - \bar{y}$, s_1^2, s_2^2 the proportional variables

$$X = (\bar{x} - \bar{y})\sqrt{n_1 n_2}; \qquad u = n_1 s_1^2; \qquad v = n_2 s_2^2 \tag{6.1}$$

and putting

$$m_1 = \frac{n_1 - 3}{2}, \qquad m_2 = \frac{n_2 - 3}{2}, \qquad n_1, n_2 \geq 3 \tag{6.2}$$

$$F_0 = n_2 u + n_1 v - x,$$

we can represent all sufficiently smooth cotests $\phi - \alpha$ in the form

$$\begin{aligned} \psi(x, u, v) &= x^{1/2} u^{-m_1} v^{-m_2} F_0^* H \\ &= x^{1/2} u^{-m_1} v^{-m_2} \int_0^x \int_0^u \int_0^v F_0(x - \xi, u - \eta, v - \zeta) \cdot H(\xi, \eta, \zeta)\, d\xi\, d\eta\, d\zeta \end{aligned} \tag{6.3}$$

where $H = H(\xi, \eta, \zeta)$ is a sufficiently smooth function of the three variables. The cotest $\psi(x, u, v)$ is to fulfill the restrictions

$$-\alpha \leq \psi(x, u, v) \leq 1 - \alpha. \tag{6.4}$$

For a given Bayes probability measure $B(\theta)$ on the alternatives, the formula (6.3) gives an ϵ-complete family of tests.

We now pass to the properties of similar tests $\phi = \phi(\bar{x}, \bar{y}, s_1^2, s_2^2)$ depending on the sufficient statistics only.

In his well-known article [16], A. Wald considers nonrandomized tests for the Behrens-Fisher problem, subject to four axioms. The first one requires that the tests depend on the sufficient statistics only. The second one requires the invariance of the critical zone with respect to one and the same shift of all the sample elements. The third axiom requires the invariance of the critical zone with respect to the contraction or expansion of all the sample elements by the same scale factor. The fourth axiom of Wald will be formulated later; we consider now the first three. It is easy to deduce from them that the nonrandomized test ϕ must be of the form:

$$\phi = \phi\left(\frac{\bar{x} - \bar{y}}{s_2}, \frac{s_1}{s_2}\right). \tag{6.5}$$

We shall call the general (randomized) test of this form homogeneous; the description of all homogeneous randomized or nonrandomized tests (tests of the form (6.5)) we shall call the homogeneous Behrens-Fisher problem.

The fourth axiom of Wald leads to the conclusion that the critical zone of the nonrandomized test is of the form

(6.6) $$\frac{|\bar{x} - \bar{y}|}{s_2} > \psi\left(\frac{s_1}{s_2}\right)$$

where ψ is a Lebesgue measurable function. The tests of this type were studied by R. A. Fisher [17] and B. Welch [18]. Therefore, we shall call them nonrandomized Fisher-Welch-Wald tests. The problem of the existence of nontrivial similar tests of this type is not yet solved.

In the work [16] cited above, A. Wald considers the tests of the type (6.6) and constructs approximately similar tests of this type. Raising the question of the existence of exact similar tests, he makes an attempt to construct tests with analytical boundary for the critical zone; his calculations are made for samples of the same size.

The investigations expounded in [19] prove that there are no such tests. Denote $\xi = (\bar{x} - \bar{y}/s_2)$, $\eta = (s_1/s_2)$. Then the boundary of the critical zone is of the type

(6.7) $$|\xi| = \psi(\eta).$$

THEOREM 6.1. *For the case of two samples of equal sample size $n \geq 4$ there exists no nonrandomized Fisher-Welch-Wald test with boundary for the critical zone* (6.7) *possessing a finite first derivative in the open interval* (0, 1) *and fulfilling the Lipschitz condition in a sufficiently large segment of the type* $[0, \eta_0]$.

Here one can define the number η_0 in the following way: $\eta_0 > 1$ so that the function $\psi(\eta)$ is continuous in $[0, 1]$. Denote $\sup_{0 \leq \eta \leq 1} \psi(\eta) = M$; then one can take $\eta_0 = 2M + 1$.

In article [19] the condition $n \geq 4$ was omitted by an oversight.

I. L. Romanovskaia [20] transferred this result to the case of the samples of unequal sizes. She proved the nonexistence of the nonrandomized Fisher-Welch-Wald test with critical zone of the type

(6.8) $$\frac{|\bar{x} - \bar{y}|\sqrt{n_1 n_2}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \geq \psi(\eta)$$

where $\eta = (n_2 s_1^2/n_1 s_2^2)$ and $n_2 \geq 4$. The test is supposed to be similar with respect to a bounded countable set of values of (σ_1/σ_2). The function $\psi(\eta)$ must be continuous and must satisfy the Lipschitz condition on the segment $[0, \eta_0]$ and have a finite first derivative in the open segment (0, 1). The number $\eta_0 > 0$ is defined as in theorem 6.1. It is not known whether these conditions upon $\psi(\eta)$ can be replaced by continuity or measurability only; the results expounded below cause one to have doubts about it. The method applied in [19] to prove theorem 6.1 to all appearances can also be applied to study this question, but the question still remains unanswered.

At any rate, if the nonrandomized test of the Fisher-Behrens-Welch type exists, it must evidently have a "pathological structure" and bad statistical properties.

However, for the equal sample sizes one can construct a randomized homogeneous test of the type which is similar to the one mentioned above and has good statistical characteristics. This test can be obtained by projecting the well-known Bartlett test (paired sample test) on the space of sufficient statistics and therefore has the properties of the Bartlett test. There exists a whole family of such tests. Denote $\xi = (\bar{x} - \bar{y}/s_2)$, $\eta = (s_1/s_2)$ as was done earlier; let $c \geq 0$ be any constant; form the expression

$$z = \tfrac{1}{2}\left(\eta + \frac{1}{\eta} - \frac{1}{c^2}\frac{\xi^2}{\eta}\right). \tag{6.9}$$

The test $\phi = \phi(\xi, \eta)$ is constructed in the following way: if $|\xi| \leq c|\eta - 1|$, let $\phi = 0$ (the null hypothesis is accepted with probability 1). If $|\xi| > c|\eta - 1|$, then $z < 1$. In this case we introduce the function

$$\phi(\xi, \eta) = \int_{\max(z,-1)}^{1} \ell_n(r)\, dr \tag{6.10}$$

where

$$\ell_n(r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)\sqrt{\pi}}(1 - r^2)^{\frac{n-4}{2}}. \tag{6.11}$$

For the test obtained in this way, by changing the constant c, we can get any level $\alpha = E(\phi|H_0)$. Note that the zone where $\phi = 0$ is bounded by segments of straight lines passing through the point (0, 1).

We can consider the families of nonrandomized homogeneous tests with the critical regions of the type

$$t\left(\frac{\bar{x} - \bar{y}}{s_2}, \frac{s_1}{s_2}\right) \geq c \tag{6.12}$$

where the constant c is arbitrary. For the case where t is a continuous function of both variables, the best results were obtained by O. V. Shalaevsky [21].

THEOREM 6.2. *There exist no continuous functions t of two variables which generate for each c a similar test of the type* (6.12), *except for the trivial case $t =$* const.

It would seem that the continuity condition of the function t could be replaced by a measurability condition. But in 1964, several Leningrad statisticians proved simultaneously [22], [23] that this is not so and that for any level $\alpha \in (0, 1)$ there exist measurable nonrandomized homogeneous similar tests for the Behrens-Fisher problem. More exactly, the following theorem was proved.

THEOREM 6.3. *For any level $\alpha \in$* (0, 1) *and pairs of samples of sizes n_1 and n_2, one even and another uneven, there exists a nonrandomized similar homogeneous test for the Behrens-Fisher problem with the critical zone depending only on* $(\bar{x} - \bar{y}/s_2)$, (s_1/s_2).

Theorem 6.3 can be improved a little.

THEOREM 6.4. *Suppose that we are given a finite number K of pairs of samples of sizes n_{1i}, n_{2i}, $(i = 1, 2, \cdots, K)$, one even and another uneven. Then there exists a measurable nonrandomized homogeneous test*

$$\phi = \phi\left(\frac{|\bar{x} - \bar{y}|}{s_2}, \frac{s_1}{s_2}\right) \tag{6.13}$$

which is similar for all these sample pairs simultaneously and has a prescribed level $\alpha \in (0, 1)$.

Note that the sufficient statistics $\bar{x}$, $\bar{y}$, s_1, s_2 are chosen for the given sample pair so that the test of theorem 6.4 will depend on the number i of the sample pair through this.

The articles [22] and [23] constructing different variants of tests of theorems 6.3 and 6.4 are based on a lemma proved by I. V. Romanovsky and V. N. Sudakov.

LEMMA OF I. V. ROMANOVSKY AND V. N. SUDAKOV. *Suppose that on a rectangle* θ: $a \leq \alpha \leq b$; $c \leq y \leq d$ *is given a finite number of measurable probability densities* $p_m(x, y)$, $(m = 1, 2, \cdots, M)$. *Then for any given* $\alpha \in (0, 1)$ *there exists a measurable function* $I(x, y)$ *taking only the values* 0 *and* 1 *such that for almost all values of* x *(correspondingly, almost all values of* y*)*

$$E^{(m)}(I(x, y)|x) = \alpha; \quad E^{(m)}(I(x, y)|y) = \alpha \qquad \text{for} \quad m = 1, 2, \cdots, M. \tag{6.14}$$

Here $E^{(m)}(\cdot|\cdot)$ *is the symbol for the conditional expectation for the densities* $p_m(x, y)$.

The homogeneous similar tests constructed have apparently bad statistical properties.

For the sample sizes of equal parity the question on the existence of the tests of the described type remains unanswered.

We remark that the requirement of similarity of the test (and this is all the more true for the stronger requirement of unbiasedness) cannot always be conciliated with some other conditions of statistical expediency. In particular, it is natural to require that the homogeneous test ϕ accepts the null hypothesis H_0 with probability 1 if $|\bar{x} - \bar{y}|$ is small in comparison with $\sqrt{s_1^2 + s_2^2}$. The following theorem (see [24]) shows that such a condition cannot be conciliated with the similarity property of the test.

THEOREM 6.5. *There exists no randomized homogeneous nontrivial similar test* ϕ, *accepting the null hypothesis* H_0 *with probability* 1 *if*

$$\frac{|\bar{x} - \bar{y}|}{\sqrt{s_1^2 + s_2^2}} < \epsilon_0 \tag{6.15}$$

where $\epsilon > 0$ *is a given arbitrarily small constant.*

In [24] this theorem was proved in a somewhat stronger form. We require from the test ϕ to accept H_0 if in addition to (6.15) we also have $(|\bar{x} - \bar{y}|/s_2) \leq 1 + \eta_0$ where $\eta_0 > 0$ is a given arbitrarily small constant.

7. Characterization of the tests of Bartlett-Scheffé type

The Bartlett-Scheffé tests for the Behrens-Fisher problem (see [25]) are based on the introduction of a new random object which for the null hypothesis H_0: $a_1 = a_2$, admits the description of all similar tests as Neyman's structures.

Namely, linear forms depending on the observations $x_1, \cdots, x_{n_1}; y_1, \cdots, y_n$ are introduced which are independent and whose variances differ only by constant factors for all values of the variances σ_1^2 and σ_2^2. For the simplest case of the samples of equal size ($n_1 = n_2$), we can take, for instance,

$$\chi = \bar{x} - \bar{y}; \qquad \ell_i = (x_i - \bar{x}) - (y_i - \bar{y}), \qquad (i = 1, 2, \cdots, n) \tag{7.1}$$

and form a test with the critical zone

$$|\chi| \left(\sum_{i=1}^{n} \ell_i^2\right)^{-1/2} \geq C_0 \tag{7.2}$$

where C_0 is chosen corresponding to a prescribed level α (the Bartlett test). We see that $\mathfrak{D}(\bar{x} - \bar{y}) = (1/n)(\sigma_1^2 + \sigma_2^2)$; $\mathfrak{D}(\ell_i) = (1 - (1/n))(\sigma_1^2 + \sigma_2^2)$. Hence, the random vector $(X, \ell_1, \cdots, \ell_n)$ generates an exponential family of distributions with one parameter $(\sigma_1^2 + \sigma_2^2)$ and one sufficient statistic which is a quadratic form and the test (6.2) is a Neyman structure test for this family. The more general Scheffé tests possess the same properties. This proves that such tests can be characterized by the requirements which are very simple and natural from the statistical point of view (see [26]).

Let a randomized test $\phi = \phi(x_1, \cdots, x_{n_1}; y_1, \cdots, y_{n_2})$ be defined in the space of the linear forms $\chi, \ell_1, \ell_2, \cdots, \ell_\mu$ where $\mu \leq n_1 + n_2 - 1$; moreover, these forms are statistically independent and

$$E(\chi | H_0) = E(\ell_i | H_0) = 0, \qquad (i = 1, 2, \cdots, \mu). \tag{7.3}$$

THEOREM 7.1. *Let there exist a small $\epsilon_0 > 0$ such that the test ϕ accepts the null hypothesis H_0 with probability 1 if for at least one of the numbers $i = 1, 2, \cdots, \mu$, we have*

$$\frac{|\chi|}{|\ell_i|} \leq \epsilon_0 \tag{7.4}$$

(but it can also accept H_0 in other cases).

Then the fractions $(\mathfrak{D}(\chi)/\mathfrak{D}(\ell_i)) = a_i$, $(i = 1, 2, \cdots, \mu)$ do not depend on σ_1 and σ_2; $\theta = \chi^2 + \sum_{i=1}^{\mu} (\ell_i^2/a_i)$ is the sufficient statistic in the space $(\chi, \ell_1, \cdots, \ell_\mu)$, and the test ϕ is a Neyman structure for this space.

We see that the conditions of theorem 7.1 are fulfilled for the Bartlett-Scheffé tests. Hence, their characterization as Neyman structures follows from a simple property connected with (7.3).

REFERENCES

[1] G. DANTZIG, "On the non-existence of tests of 'Student's' hypothesis having power function independent of σ," *Ann. Math. Statist.*, Vol. 11 (1940), pp. 186–191.

[2] C. STEIN, "A two-sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Statist.*, Vol. 16 (1945), pp. 243–258.

[3] E. LEHMANN, *Testing Statistical Hypotheses*, New York, Wiley, 1959.

[4] J. NEYMAN and E. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London Ser. A*, Vol. 231 (1933), pp. 289–337.

[5] J. NEYMAN, "Sur la vérification des hypotheses statistiques composées," *Bull. Soc. Math. France*, Vol. 63 (1935), pp. 346–366.

[6] ———, "Current problems of mathematical statistics," *Proc. Internat. Congress of Mathematicians*, Amsterdam, 1954, Vol. 1 (1957), pp. 1–22.

[7] E. LEHMANN and H. SCHEFFÉ, "Completeness, similar regions and unbiased estimates," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[8] A. M. KAGAN and YU. V. LINNIK, "A class of families admitting similar zones," *Vestnik Leningrad. Univ.*, Vol. 13 (1914), pp. 25–27. (In Russian.)

[9] D. BASU, "On statistics independent of sufficient statistics," *Sankhyā*, Vol. 20 (1958), pp. 3–4.

[10] YU. V. LINNIK, "On the theory of statistically similar zones," *Dokl. Akad. Nauk SSSR*, Vol. 146 (1962), pp. 300–302. (In Russian.)

[11] H. CARTAN, "Variétés analytiques complexes et cohomologie," Colloque de Liege, 1959.

[12] YU. V. LINNIK, "On the construction of optimal similar solutions for the Behrens-Fisher problem," *Trudy Mat. Inst. Steklov*, Vol. 79 (1965), pp. 26–32.

[13] A. A. PETROV, "Testing statistical hypotheses on the distribution type by use of small samples," *Teor. Verojatnost. i Primenen.*, Vol. I (1956), pp. 248–270.

[14] C. R. RAO, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81–91.

[15] ———, "Some theorems on the minimum variance estimates," *Sankhyā*, Vol. 12 (1952), pp. 1–2, 27–42.

[16] A. WALD, "Testing the difference between the means of two normal populations with unknown standard deviations," *Selected Papers in Probability and Statistics*, New York, McGraw-Hill (1955), pp. 669–695.

[17] R. A. FISHER, "The comparison of samples with possibly unequal variances," *Ann. of Eugenics*, Vol. 9 (1939), pp. 174–180.

[18] B. L. WELCH, "The generalization of 'Student's' problem when several different populations are involved," *Biometrika*, Vol. 34 (1947), pp. 28–35.

[19] YU. V. LINNIK, "On A. Wald's test for comparison of two normal samples," *Teor. Verojatnost. i Primenen.*, Vol. 9 (1964), pp. 16–30. (In Russian.)

[20] J. L. ROMANOVSKAIA, "On the Fisher-Welch-Wald test," *Sibirsk. Mat. Ž.*, Vol. 5 (1964), pp. 1344–1350. (In Russian.)

[21] O. V. SHALAEVSKY, "On the non-existence of the regularly changing tests for the Behrens-Fisher problem," *Dokl. Akad. Nauk SSSR*, Vol. 151 (1950), pp. 509–510. (In Russian.)

[22] A. M. KAGAN and O. V. SHALAEVSKY, "The Behrens-Fisher problem: similar zones in the algebra of sufficient statistics," *Dokl. Akad. Nauk SSSR*, Vol. 155 (1964), pp. 1250–1252. (In Russian.)

[23] YU. V. LINNIK, I. V. ROMANOVSKY, and V. N. SUDAKOV, "Non-randomized homogeneous test in the Behrens-Fisher problem," *Dokl. Akad. Nauk SSSR*, Vol. 155 (1964), pp. 1262–1264. (In Russian.)

[24] ———, "On randomized homogeneous tests for the Behrens-Fisher problem," *Izv. Akad. Nauk SSSR*, Vol. 28 (1964), pp. 1–12. (In Russian.)

[25] H. SCHEFFÉ, "On solutions of the Behrens-Fisher problem based on the t-distribution," *Ann. Math. Statist.*, Vol. 14 (1943), pp. 35–44.

[26] YU. V. LINNIK, "On the characterization of the tests of Bartlett-Scheffé type," *Trudy Mat. Inst. Steklov*, Vol. 79 (1965), pp. 32–39. (In Russian.)

[27] R. A. WIJSMAN, "In complete statistics and similar tests," *Ann. Math. Statist.*, Vol. 29 (1958), pp. 1028–1075.