



Soviet-era science, translated into English

MATHEMATICS

A. M. KAGAN, O. V. SHALEVSKII

1964

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196401.36201>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

MATHEMATICS

A. M. KAGAN, O. V. SHALEVSKII

THE BEHRENS-FISHER PROBLEM—EXISTENCE OF SIMILAR REGIONS IN THE ALGEBRA OF SUFFICIENT STATISTICS

(Presented by Academician A. N. Kolmogorov, 3 I 1964)

Consider independent repeated samples x_1, \dots, x_n and y_1, \dots, y_m , drawn from two normal populations whose parameters—the mean a_1 and variance σ_1^2 for the first population and the mean a_2 and variance σ_2^2 for the second—are unknown.

The classical Behrens-Fisher problem deals with tests similar with respect to σ_1, σ_2 and $a = a_1 = a_2$. In the present paper we study to the end one particular case of this problem, namely, the question of the existence of nonrandomized similar tests depending only on the four sufficient statistics

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \bar{y} = \frac{1}{m} \sum_{j=1}^m y_j; \quad s_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2; \quad s_2^2 = \frac{1}{m} \sum_{j=1}^m (y_j - \bar{y})^2.$$

It will be shown that, in fact, similar regions can be found even in a certain proper subalgebra of the algebra generated by $\bar{x}, \bar{y}, s_1^2, s_2^2$. We obtain this result by using the possibility of further generalizing the existing theory of similar regions on the basis of a proposition formulated below in the form of a lemma.

Let $x = x\left(\frac{\bar{x} - \bar{y}}{s_2}, \frac{s_1}{s_2}\right)$ be the smaller and $y = y\left(\frac{\bar{x} - \bar{y}}{s_2}, \frac{s_1}{s_2}\right)$ the larger root of the equation

$$z^2 - z \left[1 + \frac{(\bar{x} - \bar{y})^2}{s_2^2} + \frac{s_1^2}{s_2^2} \right] + \frac{s_1^2}{s_2^2} = 0.$$

Obviously, $0 \leq x \leq 1$ and $1 \leq y < \infty$. Denote this strip by the letter R . It is not difficult to verify that in the strip R the samples induce the family of densities

$$\frac{\Gamma\left(\frac{n+m-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{m-1}{2}\right)} \vartheta^{m/2}(1+\vartheta)^{(n+m-2)/2} \times$$

$$\times \frac{(y-x)(xy)^{(n-3)/2}}{\sqrt{1-x}\sqrt{y-1}[(\vartheta+x)(\vartheta+y)]^{(n+m-1)/2}}, \quad (1)$$

where $\vartheta = m\sigma_2^2/n\sigma_1^2$.

Theorem 1. If $N = (n+m-1)/2$ is an integer, then for the family (1) in the strip R there exists a similar Borel set of any volume $c \in (0, 1)$.

To establish this theorem, we shall need the following

Lemma. Let P_1, \dots, P_s be probability measures on Borel subsets of the strip R , and let all conditional measures $P_i(\cdot|x)$ and $P_i(\cdot|y)$ exist and have no atoms. Then, for any $c \in (0, 1)$, there is a Borel set $A \subset R$ such that

$$P_i(A|x) = P_i(A|y) = c, \quad [P_i], \quad i = 1, \dots, s.$$

Here and below the symbol $[P_i]$ means that the corresponding assertion holds with P_i -probability 1.

The proof of the lemma, in a somewhat different form, was communicated to us by I. V. Romanovskii and V. N. Sudakov. We refer the reader to the publications of these authors (3).

Proof of Theorem 1. Put

$$r(u, v) = \begin{cases} 1, & u < v, \\ 0, & u > v, \end{cases}$$

and introduce on R the probability measures $\{P_u^{k,l}, 0 < u < 1\}$; $k, l = 0, \dots, N-1$, given by the densities

$$G_{k,l}(u) \frac{x^{(n-3)/2+k}}{\sqrt{1-x}} \frac{y^{(n-3)/2+l}}{\sqrt{y-1}} \frac{r(x, u)}{(y-x)^{2N-2}},$$

and the measures $\{Q_u^{k,l}, 1 < u < \infty\}$; $k, l = 0, \dots, N-1$, given by the densities

$$H_{k,l}(u) \frac{x^{(n-3)/2+k}}{\sqrt{1-x}} \frac{y^{(n-3)/2+l}}{\sqrt{y-1}} \frac{r(u, y)}{(y-x)^{2N-2}}.$$

It is easy to see that for the families of measures $P_u^{k,l}$ the statistic $\chi_1(x, y) = x$, and for the families of measures $Q_u^{k,l}$ the statistic $\chi_2(x, y) = y$, is sufficient.

Finally, construct the measures $P^{k,l} = \sum d_t P_{u_t}^{k,l}$ and $Q^{k,l} = \sum d_t P_{u_t}^{k,l}$, where $d_t > 0$, $\sum d_t = 1$, and the summation is taken, for the first measure, over all rational points of the interval $(0, 1)$, and for the second, over all rational points of the interval $(1, \infty)$. It is well known ⁽¹⁾ that

$$\begin{aligned} P_u^{k,l}(\cdot | x) &= P^{k,l}(\cdot | x), \quad [P_u^{k,l}], \quad 0 < u < 1; \\ Q_u^{k,l}(\cdot | y) &= Q^{k,l}(\cdot | y), \quad [Q_u^{k,l}], \quad 1 < u < \infty. \end{aligned} \quad (2)$$

We now indicate, for a given $c \in (0, 1)$, a Borel set $A \subset R$ such that

$$\begin{aligned} P^{k,l}(A | x) &= c, \quad [P^{k,l}], \quad k, l = 0, \dots, N-1; \\ Q^{k,l}(A | y) &= c, \quad [Q^{k,l}], \quad k, l = 0, \dots, N-1. \end{aligned}$$

Such a set exists by virtue of our lemma, since all its conditions are satisfied here. From (2) it follows that

$$\begin{aligned} P_u^{k,l}(A | x) &= c, \quad [P_u^{k,l}], \quad k, l = 0, \dots, N-1; \\ Q_u^{k,l}(A | y) &= c, \quad [Q_u^{k,l}], \quad k, l = 0, \dots, N-1. \end{aligned} \quad (3)$$

If $\chi_A(x, y)$ is the characteristic function of the set A , then for $0 < u$, $u \neq 1$ equations (3) lead to the relation

$$\begin{aligned} &\iint_{x < u < y} \chi_A(x, y) \frac{x^{(n-3)/2+k}}{\sqrt{1-x}} \frac{y^{(n-3)/2+l}}{\sqrt{y-1}} \frac{dx dy}{(y-x)^{2N-2}} = \\ &= c \iint_{x < u < y} \frac{x^{(n-3)/2+k}}{\sqrt{1-x}} \frac{y^{(n-3)/2+l}}{\sqrt{y-1}} \frac{dx dy}{(y-x)^{2N-2}}, \quad k, l = 0, \dots, N-1. \end{aligned}$$

Multiplying both sides of this relation by $(-1)^{k+l} C_{N-1}^k C_{N-1}^l u^{2N-2-k-l}$ and summing over k, l from 0 to $N-1$, we obtain

$$\iint_{x < u < y} \chi_A(x, y) \frac{(xy)^{(n-3)/2}}{\sqrt{1-x}\sqrt{y-1}} \left(\frac{u-x}{y-x}\right)^{N-1} \left(1 - \frac{u-x}{y-x}\right)^{N-1} dx dy$$

$$= C \iint_{x < u < y} \frac{(xy)^{(n-3)/2}}{\sqrt{1-x}\sqrt{y-1}} \left(\frac{u-x}{y-x}\right)^{N-1} \left(1 - \frac{u-x}{y-x}\right)^{N-1} dx dy. \quad (4)$$

Subject both sides of (4) to the Laplace transform with respect to u , multiply the result by the transform parameter raised to the power $2N-1$, and again subject it to the Laplace transform with respect to this parameter. Straightforward calculations, whose legitimacy is easily justified, give us the result

$$\begin{aligned} & \iint_R \chi_A(x, y) \frac{(y-x)(xy)^{(n-3)/2} dx dy}{\sqrt{1-x}\sqrt{y-1} [(\vartheta+x)(\vartheta+y)]^N} = \\ & = C \iint_R \frac{(y-x)(xy)^{(n-3)/2} dx dy}{\sqrt{1-x}\sqrt{y-1} [(\vartheta+x)(\vartheta+y)]^N}, \end{aligned}$$

where ϑ is the parameter of the last Laplace transform. Comparing this with (1), we see that A is indeed a similar region of volume C .

Examining the proof of Theorem 1 carefully, one can be convinced that a stronger assertion holds.

Theorem 2. *If the pairs (n, m) , taken in any finite number, are such that all the corresponding numbers N are integers, then for any level $c \in (0, 1)$ there exists in the strip R a Borel set that is similar for all the families (1) under consideration simultaneously.*

Analogous theorems could also be proved for the multidimensional Behrens-Fisher problem, as well as for some of its generalizations.

In conclusion, let us note that in 1940 in [2] there appeared a communication asserting the nonexistence of functions $\Phi(\bar{x}, \bar{y}, s_1^2, s_2^2, a_1 - a_2)$, measurable in the Borel sense, whose distribution does not depend on $a_1, a_2, \sigma_1, \sigma_2$. However, neither in this communication nor anywhere else was this fact proved. We see that it is altogether false.

Received
31 I 1964

REFERENCES

- ¹ P. Halmos, L. Savage, *Ann. Math. Statist.*, **20**, No. 2 (1949). ² S. S. Wilks, *Ann. Math. Statist.*, **11**, No. 4, 475 (1940). ³ I. V. Romanovskii, V. N. Sudakov, *Tr. Matem. inst. im. V. A. Steklova AN SSSR* (1964).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.