MATHEMATICS

Corresponding Member of the Academy of Sciences of the USSR
Yu. V. LINNIK

1962

SovietRxiv

Abstract

Full Text

MATHEMATICS

Corresponding Member of the Academy of Sciences of the USSR Yu. V. LINNIK

ON STATISTICALLY SIMILAR ZONES OF LINEAR TYPE

We shall consider similar zones for a family of measures $\{P_\theta\}$ with one and the same σ-algebra on the Euclidean space E_n. A set A from this σ-algebra will be a similar zone for the family of measures $\{P_\theta\}$ if $P_\theta(A)$ does not depend on θ. If $P_\theta(A) \neq 0$ or 1, the similar zone A will be nontrivial and may serve for testing the corresponding statistical hypothesis. The theory of similar zones was founded by J. Neyman and E. Pearson in 1933; a considerable literature is devoted to it, as indicated in Neyman' s survey report ([1]). The further development and problems of the theory of similar zones were presented by J. Neyman in his survey report on analytical problems of mathematical statistics at the Fourth All-Union Mathematical Congress in Leningrad.

If the space E_n is partitioned into disjoint similar zones, then from them one can construct a statistic t, the distribution of which does not depend on θ (assuming measurability with respect to P_θ). We shall call such a statistic zonal. Conversely, if such a statistic t is given, then the zones $C_1 < t < C_2$, for arbitrary C_1, C_2, will be similar. If θ is a scalar parameter and there exists a probability density $L(x, \theta) = L$ with respect to some common dominating measure not depending on θ, then, provided certain simple analytical requirements are observed, for the statistic t to be zonal it is necessary and sufficient that $E(I/t) = 0$, where

$$I = \frac{1}{L}\frac{\partial L}{\partial \theta},$$

i.e. the regression of I on t must be zero* (note that EI^2 is the Fisher information measure for our distributions). In the present note we study the case in which P_θ determines a random vector with independent components $x_1, \dots, x_n$ (the case of repeated sampling), and t is a linear zonal statistic

$$t = a_1 x_1 + \cdots + a_{nx} n.$$

The zonality of t is expressed by the fact that $P_\theta(t < \xi)$ does not depend on θ for all ξ. The theory of linear zonal statistics turns out to be very closely connected with the theory of identically distributed linear statistics considered in ([2,3]). Using the methods developed in the cited papers, one can derive certain theorems on linear zonal statistics.

We shall call a linear zonal statistic t the simplest if $t = x_i - x_j$, where $i \neq j$. If θ is a scalar parameter and the distribution of x_i depends only on $x_i - \theta$ (θ is a shift parameter), then, obviously, $t = x_i - x_j$ will be a zonal statistic. However, $t = x_i - x_j$ may be a zonal statistic in other cases as well. For example, if the x_i are infinitely divisible for all θ, and the spectral function of the corresponding distributions can be decomposed into an "even" and an "odd" part, where its "even" part does not depend on the parameter θ, while the "odd" part depends on θ, then it is easy to see that $x_i - x_j$ will be a zonal statistic.

Let

$$t = a_1 x_1 + \dots + a_{nx} n$$

be a linear zonal statistic, and suppose not all a_i vanish. If all nonzero a_i have the same $|a_i|$, then it is easy to establish that $x_i - x_j$ will also be a zonal statistic, by considering the characteristic function of t. In view of this, we shall assume that not all numbers $b_j = |a_j|$ are equal to zero or to one another. Following the method of ([2,3]),

* With probability 1.

we form the "determining function"

$$\sigma(z) = |a_1|^z + \dots + |a_n|^z = b_1^z + \dots + b_r^z. \tag{1}$$

By virtue of the conditions imposed on the a_i, the real and complex zeros of $\sigma(z)$ will lie in a vertical strip of finite width (see ([1])). We denote the upper bound of the abscissas of the zeros of $\sigma(z)$ by γ.

Theorem. *Let the linear zonal statistic $t = a_1 x_1 + \dots + a_{nx} n$ have not all $|a_i|$ equal to 0 or to one another, so that $\gamma \neq \infty$.*

Suppose that for all θ there exists the $2m$-th moment of x_i, where $m = [\gamma/2 + 1]$. If the characteristic function of x_i, $f(u, \theta) \neq 0$ on the entire real axis of values of u, then the simplest linear statistic $x_i - x_j$ will also be zonal.

We note that the condition $f(u, \theta) \neq 0$ may be replaced by the condition of quasi-analyticity of $f(u, \theta)$ in a neighborhood of zero (see ([4])). If no conditions are imposed on the existence of moments of x_j, then the assertion just stated will turn out to be false, as will be seen from the example set forth below.

Theorem 1 can be derived by following the arguments of paper ([2]) (see the proof of Theorem II′ on pp. 208–234). Let θ_0 be some value of the parameter θ, and let (y_1, y_n) be a random vector with independent components distributed as x_j; $\varphi(u, \theta) = f(u, \theta) f(-u, \theta)$ is the characteristic function of $x_j - y_j$; by assumption, $\varphi(u, \theta) \neq 0$ on the entire real axis, and therefore $\varphi(u, \theta) > 0$. Put $\psi(u, \theta) = \ln \varphi(u, \theta) - \ln \varphi(u, \theta_0)$. We have:

$$\sum_{j=0}^{n} \psi(b_j u, \theta) = 0. \tag{2}$$

Here $\psi(u,\theta)$ is continuous on the entire axis. The solution of equation (2) by means of the Laplace transform was carried out and investigated in paper ([2]) (pp. 208–234). The difference from the case considered here consists in the fact that $\psi(u,\theta)$ is not the logarithm of a characteristic function, but the difference of such logarithms. Following the indicated arguments, we arrive at the conclusion that from the existence of the $2m$-th moment it follows that

$$\psi(u,\theta)=\sum_{k=1}^{K}A_k(\theta)u^{2k}, \tag{3}$$

where $A_k(\theta)$ are certain functions of θ; K is a constant. In view of the fact that, for $k=1,2,\dots,K$,

$$\sum_{j=1}^{n}b_j^{2k}>0,$$

it follows immediately from (2) that $A_k(\theta)\equiv 0$ $(k=1,2,\dots,K)$, so that $\psi(u,\theta)\equiv 0$, whence it follows that x_i-x_j will be a zonal statistic, as was required to prove.

We now consider families of distributions explaining the existence of cases in which linear zonal statistics exist, but x_i-x_j is not such a statistic. For this it is sufficient to apply a certain modification of Theorem IV of paper ([3]), proved in the same way as that theorem itself. Let positive numbers $\gamma_1<\gamma_2<\dots<\gamma_l$, $0<\gamma_1<2$, real numbers $\tau_2,\dots,\tau_{l-1}$, a positive number A_l, and a set of real numbers $E_2(\theta),E_3(\theta),\dots,E_{l-1}(\theta)$ be given, with $|E_j(\theta)|\le E_j$ for all θ, where E_j are prescribed positive numbers. Then the function

$$f(u,\theta)=\exp\left(-A|u|^{\gamma_1}+\sum_{j=2}^{l-1}E_j(\theta)\big(|u|^{\gamma_j+it_j}+|u|^{\gamma_j-it_j}\big)-A_l|u|^{\gamma_l}\right)$$

for sufficiently large $A>0$ will be a characteristic function.

If the numbers $\rho_j=\gamma_j+it_j$ are chosen so that they are roots of the function

$$\sum_{j=1}^{n}|a_j|^z,$$

then $t=\sum_{j=1}^{n}a_{jx}j$ will be a zonal statistic for the dis–

with distribution function $f(u,\theta)$. In this case, of course, $x_i^*-x_j$ will not always be a zonal statistic.

In the present example there is no variance of x_j, but it is easy to give examples of the same kind in which the x_j will have any prescribed number of moments. For this one may use certain modifications of Lemma XV of paper ([3]) (p. 274).

Thus we see that the families of distributions corresponding to linear similar zones have a rather complicated structure.

Received
12 III 1962

References

[1] J. Neyman, Current Problems of Mathematical Statistics. Survey reports at the International Mathematical Congress in Amsterdam, 1954.
[2] Yu. V. Linnik, Ukr. Math. J., 5, No. 2, 207 (1953).
[3] Yu. V. Linnik, Ukr. Math. J., 5, No. 3, 247 (1953).
[4] I. L. Romanovskaya, Vestn. LGU, No. 13 (1962).

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