ON THE COMBINATION OF INDEPENDENT TESTS

by

TAMÁS LIPTÁK

Introduction

In scientific research there are two types of statistical hypotheses. In the first it is desired to prove a systematic effect in presence of random errors and for this purpose a "null-hypothesis" is being taken in order to disprove it on ground of observations ; having attained it the result is called "significant". In the second type it is desired to prove a statistical law by the "fitting" of observations. According to this there are significance tests and tests of goodness of fit. In the theory of testing statistical hypotheses both kinds of tests are examined the same way. For this purpose the roles of alternative hypotheses are changed. In the following the terminology of significance tests is going to be used.

In the Neyman—Pearson theory the tests serving to decide between accepting or rejecting the null-hypothesis are characterized by their critical set : the null-hypothesis is rejected if the sample observed belongs to this set. The level of a test is the probability of the critical set (if the nullhypothesis is simple) or its upper bound (if it is composite).

The choice of the level of tests is determined by practical considerations : this is one of the reasons why in practice families of tests are used containing a test for each level of significance between 0 and 1 characterized by the following property: a sample being significant on a level α is also significant on any level $\alpha' > \alpha$. The level of significance is often not determined beforehand, instead of this the minimal level is calculated for which the actual result is still significant. This "moving level" is being used for measuring the significancy of results. This practice is supposed to be originating from the early, "preclassical" period of the theory of testing statistical hypotheses, at which period the „deviation" of the results from the null-hypothesis was measured by the actual value of a suitably chosen statistic.

In any case it is obvious that the conscious use of this practice cannot lead to any mistake, it being the same whether the actual value of the statistic is compared to a fixed critical value ("method of fixed level") or whether the moving level is compared to a fixed level („method of moving level"). In the first method this comparison leads only to a statement of significancy or insignificancy, whereas in the second one by noting the actual value of the moving level we obtain some information which can be used in combining more tests. This is just the subject of this paper.

The following may be mentioned as an example. Significance tests were carried out in three experiments for the same problem. On the usual 5% level of significance the results proved insignificant in every case. From this result no further statistical inference can be derived. If it is known, however, that the moving levels in each of the three cases were between 5% and 30%, it may be stated — supposing the independence of the experiments — that the *combined* results of the three experiments are significant on a 5% level.

The paper deals with the suitable combination of moving levels originating from stochastically independent tests (resp. statistics). The corresponding null-hypotheses are in the following connection with each other : either the null-hypothesis is true in each case, or some alternative in all of them.

The connection of null-hypotheses satisfies the above assumptions e. g. in the cases below :

a) the case of observations of different sizes about the same null-hypothesis ;

b) the case of observations made about the same null-hypothesis biased by unknown nuisance parameters (testing of normality etc.) ;

c) the case of moving levels originating from stochastically independent statistics.

This *combination problem* has not been, in such a generality, dealt with in the literature. The problem of combination of independent tests is treated first in R. A. FISHER's *Statistical Methods for Research Workers* (see [3], § 21.1). The problem is dealt with — though not stated — in case of simple hypotheses and of statistics with continous distribution. The theoretical aspects of the problem are not treated here, only a formula for the combination is given. This "omnibus test" may be reduced to the product (or, equivalently, to the geometrical mean) of the levels. The necessary transformation is based on the χ^2-table. K. PEARSON arrived to the same results, independently of FISHER, but in another form [14]. It was E. S. PEARSON who first investigated the efficiency of the above omnibus test, but he did so under the same simple null-hypothesis and sample sizes [12]. The case of tests based on statistics with discrete distribution is treated in many papers (e. g. [8], [13], [17]) I. J. GOOD generalizes the "omnibus test" by introducing different weights of efficiency for the various tests. The transformation necessary to the application of this test is much more complicated than the original transformation and it is not tabulated.

In § 1 of this paper the problem of moving levels is treated quite generally. The moving level is proved to be a random variable having a distribution function which is equal to or less than the uniform distribution function U over the interval [0,1] if the null-hypothesis is true, and, in the unbiased case, it is equal to or greater than U if the alternative hypothesis is true.

In § 2 the class of available combinations is narrowed down by three rational postulates. It is possible to weight the various tests according to the possibly informations relating to their efficiency.

It is proved that the class M of combinations satisfying the mentioned postulates are the combinations generated by the weighted means of the levels. The average function applied here may be any continuous and strictly increasing function defined in the closed interval [0,1]. Here the results of NAGUMO — KOLMOGOROV—DE FINETTI for characterization of mean values are available (see e. g. [7], pp. 158—163). The class M of available combinations is enlarged to the class M^* of all combinations based on weighted means of levels with average function being continuous only in the *open* interval (0,1).

In § 3 two theorems are proved : 1. Every element of M^* is admissible, i. e. a combination problem can be given for every average function and for every system of weights in which the combination generated by the weighted mean of levels based on this function and this system is the optimal solution of the problem. 2. Every element of M^* as a test is unbiased if the levels are originated from unbiased tests.

In § 4 it is proved that the postulate of monotony occurring in § 2 is not too restrictive; namely every Bayes solution of the combination problem for a wide class of "good" tests has this property.

§ 5 is devoted to prove FISHER's "omnibus test" to be the likelihood ratio test of the combination problem for a class of tests being essentially the same as above.

Finally, in § 6 the combination generated by the inverse of the normal distribution function is suggested instead of FISHER's "omnibus test". This combination is optimal for a large class of tests for one-sided hypotheses. In addition, its application needs less numerical work and simpler tables than the "omnibus test" either in original form and even less numerical work and simpler tables in case of its weighted form introduced by GOOD.

§ 1 The Problem of Combination of Moving Levels

In this § the definitions which may be needed later on are given. Then the combination problem will be formulated as a problem of hypothesis testing.

A hypothesis concerning the distribution of a random variable may be tested by observing this variable. The result of observing the variable in any well-defined way is called a *sample*, the set of possible samples is the *sample space*. To the possible distributions of the random variable there corresponds a set of probability measures defined on the sample space, while a *statistical hypothesis* determines a subset of the set of possible probability measures.

Let X be the sample space and $\mathcal{A}$ the σ-algebra of its measurable subsets. The above probability measures defined on the measurable space $(X, \mathcal{A})$, i. e. the possible distributions of the sample $x \in X$, are indexed by an index space Ω. Thus the set of possible distributions of the sample is the system $\mathcal{P}_\Omega = \{P_\theta : \theta \in \Omega\}$. The measures P_θ are assumed to be different, i. e. for every pair $\theta_1 \in \Omega, \theta_2 \in \Omega$, $\theta_1 \neq \theta_2$ a set $A \in \mathcal{A}$ is supposed to exist for which $P_{\theta_1}(A) \neq P_{\theta_2}(A)$.

The statistical hypothesis H_0 (or the "null-hypothesis") determines a subset of $\mathcal{P}_\Omega$ or — being the same — that of Ω, and H_0 is the hypothesis supposing that the ,,true" distribution P_{θ^*} of the sample is such that the index belongs to this subset. Let Ω_0 be the above subset of Ω and, consequently, let $\mathcal{P}_{\Omega_0} = \{P_\theta : \theta \in \Omega_0\}$ be the distributions of the sample consistent with the null-hypothesis. It is assumed that $\Omega_0 \neq \emptyset$ and $\Omega_1 = \Omega - \Omega_0 \neq \emptyset$. The set of distributions consistent with the alternative hypothesis is clearly $\mathcal{P}_{\Omega_1} = \{P_\theta : \theta \in \Omega_1\}$.

H_0 is called a *simple hypothesis* if $\mathcal{P}_{\Omega_0}$ (or Ω_0) consists of a single element. Otherwise, H_0 is a *composite hypothesis*.

A rule by the application of which the acceptance or rejection of H_0 is decided according to the result of sampling is called a *criterion*.[1] Every

[1] Contrary to the usual terminology, the word *test* is reserved for a concept to be introduced later.

criterion may be characterized by its *critical set* i. e. by the set of all samples to which the rejection of H_0 is ordered by the rule. In other words the critical set is the set of *significant* samples. Only rules having measurable critical sets will be called criteria.

In principle, any measurable set may be chosen as the critical set of a criterion. The problem of the theory of statistical hypothesis is to choose a set the use of which relatively seldom leads to false decisions. False decisions can occur with all criteria. An *error of the first kind* is committed if the null-hypothesis gets rejected in spite of it being true ; an *error of the second kind* is arrived at if the null-hypothesis is accepted in spite of it being false. C being the critical set, these two errors might occur in the following way: if θ^* is the "true" index, i. e. the "real" distribution of the sample is $P_{\theta^*} \in \mathcal{P}_\Omega$,

$$\theta^* \in \Omega_0 \quad \text{and} \quad x \in C \tag{1.1}$$

(error of the first kind) ;

$$\theta^* \in \Omega_1 \quad \text{and} \quad x \in X - C \tag{1.2}$$

(error of the second kind).

Thus it follows from the relation $P_\theta(X - C) = 1 - P_\theta(C)$ that a criterion is the better, the smaller is the value of the *power function*

$$p(\theta) = P_\theta(C) \tag{1.3}$$

in the case of $\theta \in \Omega_0$ and the larger it is if $\theta \in \Omega_1$. The probability of committing the error of the first kind is namely $p(\theta^*)$ $(\theta^* \in \Omega_0)$ and that of the error of the second type $1 - p(\theta^*)$ $(\theta^* \in \Omega_1)$.

In case of a simple null-hypothesis the probability of committing an error of the first kind is called the *size* of the criterion. This definition may be immediately transferred to composite null-hypotheses in the case of criteria having critical set C for which

$$P_\theta(C) \equiv \alpha \qquad \theta \in \Omega_0 \tag{1.4}$$

holds. The probability of committing an error of the first kind is determined by the null-hypothesis in this case too. The size of such *similar criteria* is α given in (1.4). In general case the least upper bound of these probabilities, i. e.

$$\alpha = \sup_{\theta \in \Omega_0} p(\theta) = \sup_{\theta \in \Omega_0} P_\theta(C) \tag{1.5}$$

is called the size of the criterion. To a criterion having a size α at most,

$$p(\theta) = P_\theta(C) \leq \alpha \quad \text{for} \quad \theta \in \Omega_0 \tag{1.6}$$

holds.

It is reasonable to require a criterion to reject the null-hypothesis more often when it is false than to do so when it is true. In terms of the power function, it is required for every pair $\theta_0 \in \Omega_0$, $\theta_1 \in \Omega_1$ that

$$p(\theta_0) \leq p(\theta_1) \tag{1.7}$$

should be true. If α denotes the size of the criterion, (1.7) is equivalent to

$$p(\theta) \geqq \alpha \quad \text{for} \quad \theta \in \Omega_0 . \tag{1.8}$$

Those criteria which fullfill the above requirements are called *unbiased criteria*.

A family of criteria is called a *test* if

1° for every *level* α $(0 \leqq \alpha \leqq 1)$ there belongs a criterion having the size α at most ;

2° a sample being significant on a level α_1 is also significant on any level $\alpha_2 > \alpha_1$;

3° the set of samples being significant on an arbitrary level $\alpha' < \alpha$ is identical with the set of samples being significant on the level α.

Denoting the critical set of the criterion, corresponding to the level α, by C_α, the above conditions are equivalent to the following :

1° for every level α $(0 \leqq \alpha \leqq 1)$

$$P_\theta(C_\alpha) \leqq \alpha \quad \text{for} \quad \theta \in \Omega_0 ; \tag{1.9}$$

2° for every pair of levels $0 \leqq \alpha_1 < \alpha_2 \leqq 1$,

$$C_{\alpha_1} \subset C_{\alpha_2} \tag{1.10}$$

holds

3° for every level α $(0 \leqq \alpha \leqq 1)$

$$\bigcup_{\alpha' < \alpha} C_{\alpha'} = C_\alpha \tag{1.11}$$

A test is called *continuous* if for every level α $(0 \leqq \alpha \leqq 1)$ the criterion having C_α as its critical set, is exactly of size α, i. e. instead of (1.9)

$$\sup_{t \in \Omega_0} P_\theta(C_\alpha) = \alpha \tag{1.12}$$

holds for $0 \leqq \alpha \leqq 1$.

The notions defined earlier for criteria may easily be extended to cover tests as well. Thus a family of criteria is called a *similar test* or *unbiased test*, respectively, if it satisfies the conditions 1° — 3° and every one of its criteria is similar or unbiased, respectively.

The real valued measurable functions $T(x)$ defined on the sample space, are called *statistics*. The distribution function of the statistic — in the case of P_θ — is the function

$$F_\theta(t) = P_\theta(\{x : T(x) < t\}) . \tag{1.13}$$

Tests can be derived from any statistic $T(x)$ e. g. in the following manner : the sample x is called significant on the level α if

$$T(x) < t_\alpha . \tag{1.14}$$

The constants t_α $(0 \leqq \alpha \leqq 1)$ must be so selected that the system of the corresponding critical sets

$$C_\alpha^T = \{x : T(x) < t_\alpha\} \qquad 0 \leqq \alpha \leqq 1 \tag{1.15}$$

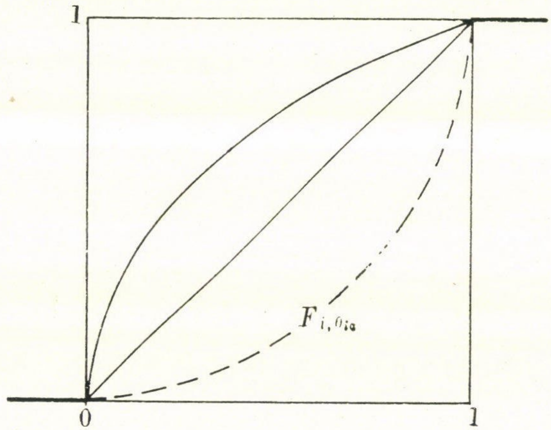
should satisfy the conditions 1°—3°. For this purpose the upper envelope function

$$F(t) = \sup_{\theta \in \Omega_0} F_\theta(t) = \sup_{\theta \in \Omega_0} P_\theta(\{x : T(x) < t\}) \tag{1.16}$$

of the distribution functions of $F_\theta(t)$ according to the null-hypothesis and its generalised inverse,[2] the function $F^{(-1)}(\alpha)$ should be formed.
Then the value of t_α may be defined by the following relation:

$$t_\alpha = F^{(-1)}(\alpha) \qquad 0 \leqq \alpha \leqq 1 . \tag{1.17}$$

It can be easily seen, that the family of criteria of the form (1.15) obtained in such a manner really forms a test.



1. ábra

Supposing that ϱ is a strictly increasing and continuous function defined over the range of T, then the statistic ϱT defined by the relation

$$\varrho T(x) = \varrho(T(x)) \tag{1.18}$$

generates the same test as T. In other words: considering as equivalent the statistics that may be obtained by strictly increasing and continuous transformation from one another, to *every equivalence class of statistics there corresponds uniquely a test which is called the test generated by this equivalence class of statistics.*

The converse of this fact is also true: *to every test there corresponds uniquely an equivalence class of statistics any element of which generates this test in the form of* (1.15).

To demonstrate also the converse of the theorem, let us define the following statistic for the critical sets C_α occurring in the conditions 1°—3°:

$$L(x) = \inf \{\alpha : x \in C_\alpha\} \qquad x \in X . \tag{1.19}$$

---

[2] The generalised inverse function $F^{(-1)}$ of the function F may be determined e. g. in the following way: $F^{(-1)}(\alpha) = 0$ for $\alpha \leqq 0$; $F^{(-1)}(\alpha) = 1$ for $\alpha \geqq 1$ and $F^{(-1)}(\alpha) = \inf \{t : F(t) \geqq \alpha\}$ for $0 < \alpha < 1$.

$L(x)$ is the „smallest" level at which the sample is still significant. This statistic is called the *moving level* of the test having the sets $\{C_\alpha\}$ as critical sets. It is clear that the test generated by this statistic is identical with the original test i. e.

$$C^L_\alpha = C_\alpha \qquad 0 \leqq \alpha \leqq 1 \ , \tag{1.20}$$

(cf. e. g. [11], p. 80). This proves the converse of our assertion.

It should be observed here that in the relation (1.15) the signs $\leqq$, $>$ or $\geqq$ may figure instead of $<$ as well, whereby everything that has been stated above so far would remain valid with suitable modifications. The tests used in practice are derived from statistics by the employment of the sign $\geqq$ in (1.14) (t-test, χ^2-test etc.).

Let us now proceed to the distribution of the moving level L defined by the relation (1.19). Let $F_\theta(t)$ denote the distribution function of $L(x)$ in the case of P_θ. It follows from the definition of $L(x)$, that for every $\theta \in \Omega$ the distribution function $F_\theta(t)$ is *defined in the interval* [0,1], i. e.

$$F_\theta(0) = 0 \quad \text{and} \quad F_\theta(1+0) = 1 \ . \tag{1.21}$$

Furthermore, let U denote the uniform distribution function over the interval [0,1], i. e. put

$$U(t) = \begin{cases} 0 & \text{for} \quad t \leqq 0 \\ t & \text{for} \quad 0 \leqq t \leqq 1 \\ 1 & \text{for} \quad t \geqq 1 \ . \end{cases} \tag{1.22}$$

Then the following theorem is true : *if the null-hypothesis is true the distribution function of the moving level of any test, is everywhere at most as large as the uniform distribution function over the interval* [0,1], *i. e.*

$$F_{\theta_0}(t) \leqq U(t) \qquad \text{for} \quad \theta_0 \in \Omega_0 \ , \quad 0 \leqq t \leqq 1 \ . \tag{1.23}$$

Furthermore: the distribution function of the moving level of any unbiased test, in the case of any possible distribution consistent with the null-hypothesis is everywhere at most as large as the distribution function of it in the case of any possible distribution consistent with the alternative hypothesis, i. e.

$$F_{\theta_0}(t) \leqq F_{\theta_1}(t) \tag{1.24}$$

whenever $\theta_0 \in \Omega_0$ *and* $\theta_1 \in \Omega_1$.

Both statements follow directly from the definition of tests and unbiased tests, respectively.

If the test is continuous and either the null-hypothesis is simple or the test is similar, equality holds identically in the relation (1.23). In other words : *in this case the moving level is uniformly distributed over the interval* [0,1] *assuming the null-hypothesis to be true.*

In the case of unbiased tests it is easy to give an intuitive interpretation of the inequalities (1.23)—(1.24). For this purpose let us define a partial ordering in the set $\mathcal{F}$ of all distribution functions F defined on the interval

[0,1] in the following way: for two elements F_1, F_2 of $\mathcal{F}$ the relation $F_1 \prec F_2$ should mean the following:

$$F_1(t) \geqq F_2(t) \qquad \text{for all } 0 \leqq t \leqq 1 \tag{1.25}$$

and

$$F_1(t_0) > F_2(t_0) \tag{1.26}$$

for some $0 \leqq t_0 \leqq 1$. Without the latter, the fulfillment of (1.25) is indicated as follows: $F_1 \preceq F_2$. The sense of the "direction" of ordering is shown by the following fact: in case of $F_1 \preceq F_2$, two random variables, ξ_1 and ξ_2, can be given in such a way that the distribution function of ξ_1 is F_1, the distribution function of ξ_2 is F_2 with probability 1

$$\xi_1 \leqq \xi_2 \tag{1.27}$$

holds and in case of $F_1 \prec F_2$ the event $\xi_1 < \xi_2$ has a positive probability.

Turning to the problem of the moving level of unbiased tests, the relations (1.23)—(1.24) can be rewritten in the following way: if the null-hypothesis is true i. e. for $\theta_0 \in \Omega_0$, the relation

$$F_{\theta_0} \succeq U \tag{1.28}$$

holds, while for every pair $\theta_0 \in \Omega_0$, $\theta_1 \in \Omega_1$ the relation

$$F_{\theta_0} \succeq F_{\theta_1} \tag{1.29}$$

holds. Intuitively, this means that in case of the null-hypothesis a higher moving level may always be expected than in case of the alternative hypothesis. Thus we are getting back to the intuitively evident fact that the smallness of the moving level indicates the acceptance rather than the alternative hypothesis.

The subject of the present paper is the following: A set of experiments is given in which the statistical hypotheses are in the following connection with each other: either the null-hypothesis is true in each experiment or the alternative one is valid in each case.[3] Each experiment is evaluated by a test i. e. a statistic is selected and the actual value of its moving level is obtained. It is supposed that these statistics — hence their moving levels as well — are independent random variables in each cases.[4] Considering the joint truth of these null-hypothesis as a new null-hypothesis it is desiderable to evaluate this new problem by an "overall" test, in other words: to combine the independent moving levels to one single moving level. This is called the *combination problem.*

---

[3] Examples *a*)—*c*) for such sets of experiments have been given in the Introduction.

[4] This will certainly be the case if the experiments are independent of each other. It may happen, however, that, to evaluate the same experiment, two different tests have been employed and the statistics generating them are stochastically independent of each other in case of each possible distribution. The moving levels of these tests may be treated as different levels for combination.

Let $\Theta_1, \Theta_2, \ldots, \Theta_r$ denote the index sets in the experiments, and let

$$\Theta_i = \Theta_{i0} \cup \Theta_{i1} \qquad (\Theta_{i0} \cap \Theta_{i1} = \varnothing) \tag{1.30}$$

be the decompositions of the index sets into the subset of the indices according to the null-hypothesis and that of indices according to the alternative one. Because of the supposed connection of the null-hypotheses, either one of the indices $(\theta_{10}, \theta_{20}, \ldots, \theta_{r0})$ $(\theta_{i0}, \in \Theta_{i0}, i = 1, 2, \ldots, r)$ or one of the indices $(\theta_{11}, \theta_{21}, \ldots, \theta_{r1})$ $(\theta_{i1} \in \Theta_{i1}, i = 1, 2, \ldots, r)$ is true in the set of experiments. Let the r-dimensional distribution functions

$$F_\theta(t_1, t_2, \ldots, t_r) = \prod_{i=1}^{r} F_{i,\theta_i}(t_i) \tag{1.33}$$

be constructed from the distribution functions F_{i,θ_i} of the moving levels L_i according to the index $\theta_i \in \Theta_i$, but only for the indices $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$ drawn from one of the direct product sets $\Theta_{10} \times \Theta_{20} \times \ldots \times \Theta_{r0}$ and $\Theta_{11} \times \Theta_{21} \times \ldots \times \Theta_{r1}$. The distribution functions which can be obtained in this way will be indexed by a set Ω . Let

$$\Omega = \Omega_0 \cup \Omega_1 \qquad (\Omega_0 \cap \Omega_1 = \varnothing) \tag{1.34}$$

be the decomposition of Ω into the subsets according to the direct sets above. Then the *combination problem is equivalent to the following hypothesis testing problem:* the possible probability distributions are generated by the distribution functions (1.33) defined on the sample space K_r (i. e. on the r-dimensional unit cube) and in case of the null-hypothesis $(\theta_0 = (\theta_{10}, \theta_{20} \ldots, \theta_{r0}) \in \Omega_0)$ the relations

$$F_{i,\theta_{i0}} \geqq U \qquad i = 1, 2, \ldots, r \tag{1.35}$$

hold. Further, in the case of unbiased tests, the relations

$$F_{i,\theta_{i0}} \geqq F_{i,\theta_{i1}} \qquad i = 1, 2, \ldots, r \tag{1.36}$$

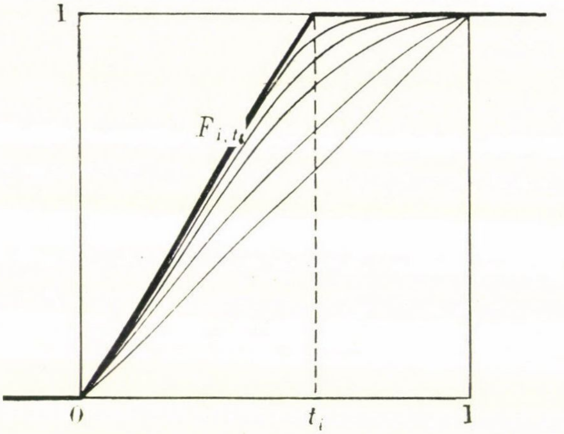
hold for any pair of $\theta_0 = (\theta_{10}, \theta_{20}, \ldots, \theta_{r0})$ and $\theta_1 = (\theta_{11}, \theta_{21}, \ldots, \theta_{r1})$ with $\theta_0 \in \Omega_0$ and $\theta_1 \in \Omega_1$.

§. 2. The Class of Monotone, Compatible and Normed Combinations

The solution of the combination problem is sought in the form of a test, that is to say, in the form of some function of the moving levels. Since the number of experiments evaluated by tests tobe combined may differ from time to time, it is reasonable to seek at once for the solution in the form of a family of combinations, which supplies a combination for every number of experiments. The easiest way obtaining this is to form the empirical distribution of the "sample" consisting of the moving levels permitting in it different weights for the individual levels. For any number of experiments, this function is a discrete distribution function defined on the interval [0,1]. Thus it is reasonable to seek the solution of the combination problem in the form of a functional $\varphi(S)$ defined on the set $\mathcal{S}$ of all finite discrete distribution functions S defined on the interval [0,1]. It may be observed that, owing to the characterisability of tests by statistics as discussed before, it is sufficient to restrict the treatment to the equivalence classes of functionals which are invariant under any strictly increasing and continuous transformation.

Let E_t denote the (degenerated) distribution function of the constant t, i. e. let be

$$E_t(u) = \begin{cases} 0 & \text{for } u \leqq t \\ 1 & \text{for } u > t \, . \end{cases} \tag{2.1}$$



2. ábra

If the weights corresponding to the experiments are $\lambda_1, \lambda_2, \ldots, \lambda_r$ ($\lambda_1 \geqq 0, \lambda_2 \geqq 0, \ldots, \lambda_r \geqq 0$; $\lambda_1 + \lambda_2 + \ldots + \lambda_r = 1$) then the weighted empirical distribution function of the level set $L_1, L_2, \ldots, L_r$ is defined by the relation

$$S = \sum_{i=1}^{r} \lambda_i E_{L_i} \, . \tag{2.2}$$

In an other form this function is the following :

$$S(t) = \sum_{L_i < t} \lambda_i \, . \tag{2.3}$$

By any possible value of the number of levels, the weights and the levels, the functions S defined by (2.2) or (2.3) are finite discrete distribution functions, i.e.

$$S \in \mathcal{S} \, . \tag{2.4}$$

The intuitive interpretation of the "joint" null-hypothesis explained above suggests to require from the functional φ to be monotone according to the partial ordering defined in $\mathcal{S}$. This is contained in the following

Postulate 1. (monotony). For the functional φ

$$\varphi(S_1) < \varphi(S_2) \tag{2.5}$$

should be hold for every pair of distribution functions from $\mathcal{S}$ with

$$S_1 \prec S_2 \, . \tag{2.6}$$

This postulates expresses the evident requirement that a set of levels should be more significant than another if the levels of the former are more significant than the corresponding levels of the other set.

Weighting the elements of the set of levels the following problem arises: if some set of levels with weighted elements are joined to an overall set of levels, how are the levels to be reweighted in this new set? Thus the demand arises for the weighting of the different set of level sets between each other, as well. Let $\lambda_{1j}, \lambda_{2j}, \ldots, \lambda_{r_j,j}$ denote the weights corresponding to the elements of the set of levels $L_{1j}, L_{2j}, \ldots, L_{r_j,j}$ ($\lambda_{1j} \geqq 0, \lambda_{2j} \geqq 0, \ldots, \lambda_{r_j,j} \geqq 0$; $\lambda_{1j} + \lambda_{2j} + \ldots + \lambda_{r_j,j} = 1$). If w_j denotes the weight corresponding to the set $L_{1j}, L_{2j}, \ldots, L_{r_j,j}$ ($w_1 \geqq 0, w_2 \geqq 0, \ldots, w_k \geqq 0$; $w_1 + w_2 + \ldots + w_k = 1$), it is reasonable to choose the weight of the level L_{ij} in the overall set $L_{11}, L_{21}, \ldots, L_{r_1,1}, L_{12}, L_{22}, \ldots, L_{r_2,2}, \ldots, L_{1k}, L_{2k}, \ldots, L_{r_k k}$ to be $\lambda_{ij} w_j$ ($i = 1, 2, \ldots, r_j$; $j = 1, 2, \ldots, k$). According to this, the weighted empirical distribution function S of the overall set of levels is

$$S = \sum_{j=1}^{k} w_j S_j \tag{2.7}$$

where

$$S_j = \sum_{i=1}^{r_j} \lambda_{ij} E_{L_{ij}} \tag{2.8}$$

is the empirical distribution function of the set $L_{1j}, L_{2j}, \ldots, L_{r_j,j}$ ($j = 1, 2, \ldots, k$).

It is reasonable to demand of the combination that in case of joining different sets of levels the value of the combined level should be uniquely determined by the values of the weights corresponding to these sets and the values of the combined levels of the individual sets of levels. This demand is expressed by the following

Postulate 2 (compatibility). The identity

$$\varphi\left(\sum_{j=1}^{k} w_j S'_j\right) = \varphi\left(\sum_{j=1}^{k} w_j S''_j\right) \tag{2.9}$$

should be fulfilled by an arbitrary weight system $w_1, w_2, \ldots, w_k$ if $S'_j \in \mathcal{S}, S''_j \in \mathcal{S}$ and

$$\varphi(S'_j) = \varphi(S''_j) \tag{2.10}$$

($j = 1, 2, \ldots, k$).

The following postulate is devoted to express the evident requirement, that the value of the "combined" level of a "set" consisting of a single level, should agree with the value of this level.

Postulate 3 (normedness). The following identity should be fulfilled:

$$\varphi(E_t) = t \qquad \text{for } 0 \leqq t \leqq 1 \, . \tag{2.11}$$

These three postulates referring to the combination, are equivalent to the conditions occurring in the characterization of the mean values. More precisely the following theorem is true:

Characterization of the mean values (NAGUMO—KOLMOGOROV—DE FINETTI; see e. g. [7] pp. 158—163). *For any functional φ satisfying Postulates 1—3 a strictly increasing and continuous function $\chi(t)$ can be given with the property*

$$\varphi(S) = \chi^{-1}\left(\int_0^1 \chi(t)\, d\, S(t)\right) \qquad \text{for } S \in \mathcal{S} \,. \tag{2.12}$$

Here χ^{-1} is the inverse function of χ.

Conversely: any functional φ of the form (2.12) with average function χ having the above properties satisfies Postulates 1—3.

It should be observed here that from the form

$$S = \sum_{i=1}^{r} \lambda_i E_{L_i} \tag{2.13}$$

of the distribution function S the following formula for φ can be obtained:

$$\varphi(S) = \varphi\left(\sum_{i=1}^{r} \lambda_i E_{L_i}\right) = \chi^{-1}\left(\sum_{i=1}^{r} \lambda_i \chi(L_i)\right) . \tag{2.14}$$

This is equivalent to (2.12). From this form it is clear that these functionals are the weighted means of the levels $L_1, L_2, \ldots, L_r$, weighted by the weights $\lambda_1, \lambda_2, \ldots, \lambda_r$ and averaged by the function $\chi(t)$.

It is worth-while to quote the proof of this theorem, which comes actually from DE FINETTI.

Let us consider first the following family $\mathcal{I} = \{I_t : 0 \leq t \leq 1\}$ of finite discrete distribution functions where

$$I_t = (1-t)\, E_0 + t\, E_1 = E_{1-t}^{(-1)} \in \mathcal{S} \,. \tag{2.15}$$

Evidently

$$I_0 = E_0 \quad \text{and} \quad I_1 = E_1 \,, \tag{2.16}$$

further for every pair of t_1, t_2 with $0 \leq t_1 < t_2 \leq 1$

$$I_{t_1} \prec I_{t_2} \tag{2.17}$$

holds. Moreover, for any weighting system $\lambda_1, \lambda_2, \ldots, \lambda_r$ and for any set $t_1, t_2, \ldots, t_r$ $(0 \leq t_i \leq 1,\ i = 1, 2, \ldots, r)$

$$\sum_{i=1}^{r} \lambda_i I_{t_i} = I_{\sum\limits_{i=1}^{r} \lambda_i t_i} \tag{2.18}$$

holds. Let now be

$$h(t) = \varphi(I_t) \qquad 0 \leq t \leq 1 \,. \tag{2.19}$$

Owing to the relation (2.16) and to the monotony of φ, $h(t)$ is a strictly increasing function of t, further, on account of (2.15),

$$h(0) = 0 \quad \text{and} \quad h(1) = 1 \,. \tag{2.20}$$

It can be proved that φ is continuous as well. Namely, if $h(t)$ would have, for instance, a right discontinuity in a point $t_0 \in [0,1)$, that is to say, $h(t_0) < h(t_0 + 0)$ would hold, there would be such a number h_0 for which

$$h(t_0) < h_0 < h(t_0 + \varepsilon) \tag{2.21}$$

would hold for every arbitrarily small positive value of ε. From this it would follow that

$$E_{h(t_0)} \prec E_{h_0} \prec E_{h(t_0+\varepsilon)} \tag{2.22}$$

and, for an arbitrary $t \in [0,1]$,

$$\tfrac{1}{2} E_{h(t_0)} + \tfrac{1}{2} E_{h(t)} \prec \tfrac{1}{2} E_{h_0} + \tfrac{1}{2} E_{h(t)} \prec \tfrac{1}{2} E_{h(t_0+\varepsilon)} + \tfrac{1}{2} E_{h(t)} \tag{2.23}$$

and in such a way, on account of the monotony of φ,

$$\varphi\left(\tfrac{1}{2} E_{h(t_0)} + \tfrac{1}{2} E_{h(t)}\right) < A_t = \varphi\left(\tfrac{1}{2} E_{h_0} + \tfrac{1}{2} E_{h(t)}\right) < \varphi\left(\tfrac{1}{2} E_{h(t_0+\varepsilon)} + \tfrac{1}{2} E_{h(t)}\right) \tag{2.24}$$

would hold. But

$$\varphi(E_{h(t)}) = h(t) = \varphi(I_t) \ , \tag{2.25}$$

hence, from the compatibility of φ,

$$\varphi\left(\tfrac{1}{2} E_{h(t_0)} + \tfrac{1}{2} E_{h(t)}\right) = \varphi\left(\tfrac{1}{2} I_{t_0} + \tfrac{1}{2} I_t\right) = \varphi\left(I_{\frac{t_0+t}{2}}\right) = h\left(\frac{t_0 + t}{2}\right) \tag{2.26}$$

and, by a similar reasoning,

$$\varphi\left(\tfrac{1}{2} E_{h(t_0+\varepsilon)} + \tfrac{1}{2} E_{h(t)}\right) = \varphi\left(\tfrac{1}{2} I_{t_0+\varepsilon} + \tfrac{1}{2} I_t\right) = \varphi\left(I_{\frac{t_0+t+\varepsilon}{2}}\right) = h\left(\frac{t_0 + t + \varepsilon}{2}\right) \tag{2.27}$$

would follow, consequently, on the basis of (2.24) — (2.27),

$$h\left(\frac{t_0 + t}{2}\right) < A_t < h\left(\frac{t_0 + t + \varepsilon}{2}\right) \tag{2.28}$$

and hence for every possible value of t

$$h\left(\frac{t_0 + t}{2}\right) < h\left(\frac{t_0 + t}{2} + 0\right) \tag{2.29}$$

would hold what is impossible, since a monotone function can have at most a denumerable number of discontinuities.

Thereafter the inverse of the strictly increasing and continuous function $h(t)$ may be formed. Let us denote by it $\chi(t)$. Evidently

$$\varphi(E_t) = t = h(\chi(t)) = \varphi(I_{\chi(t)}) \tag{2.30}$$

consequently on account of the compatibility of φ and because of (2.17)

$$\varphi(S) = \varphi\left(\sum_{i=1}^{r} \lambda_i E_{L_i}\right) = \varphi\left(\sum_{i=1}^{r} \lambda_i I_{\chi(t)}\right) = \varphi\left(I_{\sum_{i=1}^{r} \lambda_i \chi(L_i)}\right) =$$

$$= h\left(\sum_{i=1}^{r} \lambda_i \chi(L_i)\right) = \chi^{-1}\left(\sum_{i=1}^{r} \lambda_i \chi(L_i)\right) = \chi^{-1}\left(\int_0^1 \chi(t)\, dS(t)\right) . \tag{2.31}$$

Since the converse of the theorem is trivial, the proof is ended hereby.

Applying this theorem to the combination problem it may be stated: *the class M of monotone, compatible and normed combinations coincides with the class of those combinations which — as tests — are generated by one of the weighted means of the levels. As an average function any function which is strictly increasing and continuous in the interval* $[0,1]$ *may figure.*

Considering the fact that a statistic is determined in any case by its test only up to a strictly increasing and continuous transformation, the statistic in (2.14) can be replaced by

$$\sum_{i=1}^{r} \lambda_i \chi(L_i) . \tag{2.32}$$

A combination of M may be obtained e. g. if the inverse of the distribution function of a random variable is chosen to be χ the density function of which vanishes outside a finite interval, while it is positive within the same. Example: the variable is uniformly distributed over the interval $[0,1]$, $\chi(t)=t$ and the combination is based on the weighted arithmetic mean of the levels.

The same may not be stated in the relation of an infinite interval: the inverse function of such a distribution function has infinite limiting value either at 0 or at 1 or both, and therefore it cannot be continuous in the *closed* interval $[0,1]$. Accordingly, the functional based on it is monotone only if no level is equal to 0 or 1, respectively. Since in practice these cases are uninteresting, it is reasonable to enlarge the class M to those combinations the generating statistics of which is of the form (2.32) but the function $\chi(t)$ occurring in it is to be supposed continuous only in the *open* interval $(0,1)$. Let M^* denote this class of the compatible, normed and "almost everywhere" monotone combinations.

An element of M^* may be obtained e. g. if the inverse of the distribution function of a random variable is chosen to be χ the density function of which vanishes outside an *arbitrary* (finite or infinite) interval, while within it is positive. Example: the variable is normally distributed with expectation 0 and variance 1, $\chi(t) = \Phi^{-1}(t)$ where

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}}\, du \tag{2.33}$$

is the (standardized) normal distribution function. Also the geometrical, harmonical and arbitrary power means of the levels may be mentioned corresponding to $\chi(t) = \log t$, t and t^a respectively. The (ordinary) geometrical mean is equivalent to FISHER's "omnibus test" [3] while the weighted one to its generalization introduced by GOOD [4].

It is not intended to analyse in this paragraph how the choice of a combination from the class M^* should be made. Here it will be given only how to calculate the combined level in case of a given averaging function and weighting system.

Let us take first the case in which all levels to be combined are uniformly distributed over the interval [0,1] if the null-hypotheses are true (this is the case of continuous tests, everywhere either the null-hypothesis is simple, or the test is similar). The random variables $\chi(L_1)$, $\chi(L_2)$, . . ., $\chi(L_r)$ are independent and, in case of the nullhypotheses, they are identically distributed with the common distribution function $\chi^{-1}(t)$. Consequently, in this case the task — to determine the distribution function of the statistic of the form (2.32) if the null-hypotheses are true — is reduced to determine the distribution function of a linear form of independent and identically distributed random variables. If this distribution function is $G_\chi(t)$, the combined level $\widetilde{L}$ — taking the relations (1.14) — (1.17) into account — will be given by the formula

$$\widetilde{L} = G_\chi\left(\sum_{i=1}^{r} \lambda_i \chi(L_i)\right) . \tag{2.34}$$

In the cases when χ is the inverse of the Cauchy- or the normal distribution function, the determination of G_χ can be very easily achieved. If $\chi(t) = C^{-1}(t)$ where

$$C(t) = \frac{1}{\pi} \int_{-\infty}^{t} \frac{du}{1+u^2} \tag{2.35}$$

is the (standardized) Cauchy distribution function,

$$G_\chi(t) = C(t) \tag{2.36}$$

holds (see e. g. [1], p. 247, equ. (19.2.3)) and therefore we have

$$\widetilde{L} = C\left(\sum_{i=1}^{r} \lambda_i C^{-1}(L_i)\right) . \tag{2.37}$$

If $\chi(t) = \Phi^{-1}(t)$, where $\Phi(t)$ is the normal distribution function given in (2.33),

$$G_\chi(t) = \Phi\left(\frac{t}{\sqrt{\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_r^2}}\right) , \tag{2.38}$$

therefore we have

$$\widetilde{L} = \Phi\left(\frac{\lambda_1 \Phi^{-1}(L_1) + \lambda_2 \Phi^{-1}(L_2) + \ldots + \lambda_r \Phi^{-1}(L_r)}{\sqrt{\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_r^2}}\right) . \tag{2.39}$$

Concerning the general case we observe only that in case of continuous tests the exact level is given by the above equations, otherwise these values may also be used as upper estimates for the exact ones (see concerning this the papers [8], [12] and [17]).

§ 3. Admissibility and Unbiasedness of the Elements of M^*

In this paragraph two good properties of the class of M^* will be shown to support the restrictions introduced in § 2.

First, the following theorem will be proved : *every element of M^* is an admissible solution of the combination problem, i. e. there can be given a hypothesis testing problem for every strictly increasing and continuous function $\chi(t)$ defined on the interval $(0,1)$ and for every weight system $\lambda_1, \lambda_2, \ldots, \lambda_r$ in which the element of M^* corresponding to the function χ and the system $\lambda_1, \lambda_2, \ldots, \lambda_r$ is the optimal solution of the combination problem.*

To prove this theorem a simple alternative testing problem will be given for every element of M^*. A *simple alternative testing problem* is a hypothesis testing having the following structure: there are given only two possible distributions P_0 and P_1 on the sample space X, and the null-hypothesis state P_0 to be the ,,true" distribution of the sample. The moving level of the test generated by the statistic $T(x)$ according to (1.14)—(1.17) is obviously

$$L(x) = G_0(T(x)) , \tag{3.1}$$

where G_0 is the distribution function of T in the case of P_0. $L(x)$ has the distribution function

$$F_0(t) = G_0(G_0^{(-1)}(t)) \quad \text{or} \quad F_1(t) = G_1(G_0^{(-1)}(t)) \tag{3.2}$$

according as to P_0 or P_1 is the ,,true" distribution of the sample X. Here G_1 denotes the distribution function of T according to P_1, and $G_0^{(-1)}$ is the generalized inverse of G_0.

Let r simple alternative testing problems be given where the statistics T_i $(i = 1, 2, \ldots, r)$ generating the tests are mutually independent random variables .The possible distribution functions G_{i0} and G_{i1} of T_i are supposed to be absolutely continuous with density functions g_{i0} and g_{i1} being positive in an interval and vanishing outside of it. Then the moving level L_i is uniformly distributed according to the null-hypothesis, i. e. its distribution function is

$$F_{i0}(t) = U(t) = t \qquad 0 \leqq t \leqq 1\,,\ i = 1, 2, \ldots, r \tag{3.3}$$

in this case and it is

$$F_{i1}(t) = G_{i1}(G_{i0}^{-1}(t)) \qquad 0 \leqq t \leqq 1, i = 1, 2, \ldots, r \tag{3.4}$$

according to the alternative hypothesis. Here G_{i0}^{-1} denotes the ordinary inverse of G_{i0}.

The hypothesis testing problem equivalent to the combination of these independent tests can be solved optimally by the use of the fundamental lemma of NEYMAN and PEARSON (see e. g. [1], pp. 529—531). The optimal test is generated by the statistic

$$T(L_1, \ldots, L_r) = \sum_{i=1}^{r} \log \frac{f_{i0}(L_i)}{f_{i1}(L_i)} . \tag{3.5}$$

Here f_{i0} and f_{i1} denote the density functions of the distribution functions F_{i0} and F_{i1}, respectively. Since

(3.6) $$f_{i0}(t) \equiv 1$$

on account of (3.3) and

(3.7) $$f_{i1}(t) = \frac{d}{dt} G_{i1}(G_{i0}^{-1}(t)) = \frac{g_{i1}(G_{i0}^{-1}(t))}{g_{i0}(G_{i0}^{-1}(t))}$$

on account of (3.4), this optimal statistic is the following:

(3.8) $$T(L_1, L_2, \ldots, L_r) = \sum_{i=1}^{r} \log \frac{g_{i0}(G_{i0}^{-1}(L_i))}{g_{i1}(G_{i0}^{-1}(L_i))}.$$

The combination generated by this statistic is equivalent to the element of M^* characterized by the statistic

(3.9) $$\sum_{i=1}^{r} \lambda_i \chi(L_i)$$

if and only if there exists a strictly increasing and continuous function ϱ for which

(3.10) $$\sum_{i=1}^{r} \log \frac{g_{i0}(G_{i0}^{-1}(L_i))}{g_{i1}(G_{i0}^{-1}(L_i))} = \varrho \left(\sum_{i=1}^{r} \lambda_i \chi(L_i) \right)$$

holds identically. It follows that ϱ may be only a linear function, hence the above condition may be rewritten in the following form: there should exist constants $a > 0, b_1, b_2, \ldots, b_r$ such that

(3.11) $$\log \frac{g_{i0}(G_{i0}^{-1}(L_i))}{g_{i1}(G_{i0}^{-1}(L_i))} = a \lambda_i \chi(L_i) + b_i \qquad i = 1, 2, \ldots, r ,$$

i. e.

(3.12) $$g_{i1}(t) = g_{i0}(t)\, e^{-a \lambda_i \chi(G_{i0}(t)) - b_i} \qquad i = 1, 2, \ldots, r$$

holds identically in t.

Distribution functions G_{i0} resp. G_{i1} may be easily constructed which satisfy the conditions (3.12). To show this let H be any function which is strictly increasing and continuous, further, which maps the interval $(-\infty, \infty)$ into $(0, \infty)$ (e. g. $H(t) = e^t$). Let G_{i0} be defined by

(3,13) $$G_{i0}(t) = G_0(t) = \chi^{-1}(H(t)) \qquad i = 1, 2, \ldots, r$$

where $\chi^{-1}(t) = 0$ for $t \leqq \chi(0)$ and $\chi^{-1}(t) = 1$ for $t \geqq \chi(1)$. Then the functions

(3.14) $$g_{i0}(t) = \begin{cases} g_0(t) = G_0'(t) & \text{for } \chi(0) \leqq t \leqq \chi(1) \\ 0 & \text{for } t < \chi(0) \text{ or } t > \chi(1) \end{cases}$$

are probability density functions and the integrals

(3.15) $$\int_{-\infty}^{t} g_{i0}(u)\, e^{-a \lambda_i H(u) - b_i}\, du \qquad i = 1, 2, \ldots, r$$

define the alternative distribution functions G_{i1}, if the constants $a > 0$ and $b_1, b_2, \ldots, b_r$ are suitable chosen. Namely, let $a = 1$ and

$$b_i = \log \int_{\chi(0)}^{\chi(1)} g_0(u)\, e^{-\lambda_i H(u)}\, du \qquad i = 1, 2, \ldots, r. \tag{3.16}$$

The integrals (3.15) and (3.16) exist according to

$$0 < e^{-\lambda_i H(u)} \leqq 1 \tag{3.17}$$

which follows from $\lambda_i > 0$ and $H(u) \geqq 0$.

To see an example, put $\chi(t) = \Phi^{-1}(t)$. Choosing now

$$G_{i0}(t) = \Phi\left(\frac{t}{\sigma_i/\sqrt{n_i}}\right) \qquad i = 1, 2, \ldots, r \tag{3.18}$$

and

$$a = 1, \qquad b_i = \frac{-\mu_i}{\sigma_i/\sqrt{n_i}} \qquad i = 1, 2, \ldots, r \tag{3.19}$$

with $\sigma_i > 0$, n_i positive integer, $\mu_i < 0$ ($i = 1, 2, \ldots, r$). Then

$$G_{i1}(t) = \Phi\left(\frac{t - \mu_i}{\sigma_i/\sqrt{n_i}}\right) \qquad i = 1, 2, \ldots, r\,. \tag{3.20}$$

The constants n_i, σ_i and μ_i should be chosen so that the relations

$$\lambda_i = \frac{-\mu_i}{\sigma_i/\sqrt{n_i}} \qquad i = 1, 2, \ldots, r \tag{3.21}$$

be valid.

This set of experiments can be interpreted as follows : a sample of size n_i was drawn in the ith experiment for a random variable having known variance σ_i^2 and expectation 0 in the case of null-hypothesis and $\mu_i < 0$ in the alternative case ($i = 1, 2, \ldots, r$). For the combination of these tests the method based on the statistic (3.9) with $\chi = \Phi^{-1}$ is optimal.

In what follows it will be further proved that *combining unbiased tests by an element of M^* this combination — as a test used for the sample consisting of moving levels — is unbiased.* In other words : if the relations (1.34) also are valid for the possible distribution functions F_{i,θ_i} of the levels besides of (1.33), and $\chi(t)$ denotes any strictly increasing and continuous function defined on the interval (0,1) and, further, if H_θ is the distribution function of the combined level $\widetilde{L}$ in the combination (from M^*) generated by the statistic

$$\chi^{-1}\left(\sum_{i=1}^{r} \lambda_i\, \chi(L_i)\right) \tag{3.22}$$

in the case of $\theta \in \Omega$, then besides the relation

$$H_{\theta_0} \geqq U \tag{3.23}$$

being valid for every $\theta_0 \in \Omega_0$, the relations

(3.24) $$H_{\theta_0} \gtrsim H_{\theta_1}$$

i. e.

(3.25) $$H_{\theta_0}(t) \leqq H_{\theta_1}(t) \qquad 0 \leqq t \leqq 1$$

hold for every pair $\theta_1 \in \Omega_1$ and $\theta_0 \in \Omega_0$.

Let G_θ denote the distribution function of the statistic (3.22) in the case $\theta \in \Omega$. The relations (3.25) will follow and so the theorem will be proved if we succeed to show that

(3.26) $$G_{\theta_0}(t) \leqq G_{\theta_1}(t)$$

holds for all $\theta_0 \in \Omega_1$ and $\theta_1 \in \Omega_1$. For this purpose let us consider a random variable η uniformly distributed over [0,1] and, keeping fixed the indices

(3.27) $$\theta_0 = (\theta_{10}, \theta_{20}, \ldots, \theta_{r0}) \in \Omega_0 \text{ and } \theta_1 = (\theta_{11}, \theta_{21}, \ldots, \theta_{r1}) \in \Omega_1$$

occurring in (3.26), let us write :

(3.28) $$\xi_{i0} = F^{(-1)}_{i,\theta_{i0}}(\eta) \qquad i = 1, 2, \ldots, r$$

and

(3.29) $$\xi_{i1} = F^{(-1)}_{i,\theta_{i1}}(\eta) \qquad i = 1, 2, \ldots, r \ .$$

Here $F_{i,\theta_{i0}}$ resp. $F_{i,\theta_{i1}}$ denote the distribution function of the moving level L_i in the case of θ_0 and θ_1, respectively. As it is well-known, ξ_{i0} is then distributed according to $F_{i,\theta_{i0}}$ and ξ_{i1} according to $F_{i,\theta_{i0}}$ (see e. g. [15], p. 183), i. e. the distribution of the moving level L_i coincides with that of ξ_{i0} or ξ_{i1} according to whether θ_0 or θ_1 is the "true" index. Besides the empirical distribution function

(3.30) $$S = \sum_{i=1}^{r} \lambda_i E_{L_i}$$

of the levels $L_1, L_2, \ldots, L_r$, the functions

(3.31) $$R_0 = \sum_{i=1}^{r} \lambda_i E_{\xi_{i0}} \quad \text{and} \quad R_1 = \sum_{i=1}^{r} \lambda_i E_{\xi_{i1}}$$

can be defined, observing that $R_0 \in \mathcal{S}$ and $R_1 \in \mathcal{S}$ as well as $S \in \mathcal{S}$. Denoting by φ_χ the statistic (3.22) in functional form

(3.32) $$\varphi_\chi(S) = \chi^{-1}\left(\int_0^1 \chi(t)\, dS(t)\right) ,$$

it is obvious from the definition of R_0 and R_1 that the distribution of the random variable $\varphi_\chi(S)$ coincides with that of $\varphi_\chi(R_0)$ or $\varphi_\chi(R_1)$ according to whether θ_0 or θ_1 is the "true" index. But the relations (1.33) and (1.34) imply that for every elementary event

(3.33) $$\xi_{i0} \geqq \xi_{i1}$$

and therefore at the same time

$$R_0 \geqq R_1 \tag{3.34}$$

holds, from which

$$\varphi_\chi(R_0) \geqq \varphi_\chi(R_1) \tag{3.35}$$

follows for every elementary event, being the functional φ_χ monotone. The desired relation between G_{θ_0} and G_{θ_1} can be easily derived from (3.35).

§. 4. A Connection Between the Monotone and the Bayes Solutions of the Combination Problem.

It will be proved in this Section that the postulate of monotony of combinations introduced in § 2 is not too restrictive in the following sense: for a rather typical class of tests every Bayes solution of the combination problem is generated by a monotone statistic.

For this purpose let the notion of Bayes solution of a hypothesis testing problem be defined. Let $\mathcal{P}_\Omega = \{P_\theta : \theta \in \Omega\}$ be the system of possible distributions defined on the sample space X (precisely, on the measurable space $(X, \mathcal{A})$), and let the null-hypothesis be characterized by the subset $\Omega_0 \subset \Omega$, where $\Omega_0 \neq \emptyset$ and $\Omega_1 = \Omega - \Omega_0 \neq \emptyset$ is the subset corresponding to the alternative hypothesis. Considering some σ-algebras $\mathcal{B}_0$ and $\mathcal{B}_1$ consisting of subset of the index sets Ω_0 and Ω_1, respectively, *prior distributions* can be introduced, i. e. probability measures Q_0 and Q_1 in the measurable spaces $(\Omega_0, \mathcal{B}_0)$ and $(\Omega_1, \mathcal{B}_1)$, respectively. Supposing the measurability of the functions $P_{\theta_0}(A)$ and $P_{\theta_1}(A)$ for fixed $A \in \mathcal{A}$ with respect to $\mathcal{B}_0$ and $\mathcal{B}_1$, respectively, the *posterior distributions*

$$P_{Q_0}(A) = \int_{\Omega_0} P_{\theta_0}(A)\, d\, Q_0(\theta_0) \tag{4.1}$$

and

$$P_{Q_1}(A) = \int_{\Omega_1} P_{\theta_1}(A)\, d\, Q_1(\theta_1) \tag{4.2}$$

can be formed which are probability measures defined on the original measurable space $(X_1, \mathcal{A})$.

Then a new hypothesis testing problem can be considered the possible distributions of which are P_{Q_0} and P_{Q_1} and P_{Q_0} corresponds to the new null-hypothesis. The optimal solution of this new problem can be determined by the fundamental lemma of NEYMAN and PEARSON; this test is called a *Bayes solution* (corresponding to the prior distributions Q_0, Q_1) of the original hypothesis testing problem.[6]

The theorem to be proved can now be formulated as follows. *Let C be the class of all tests the distribution function of the moving level of which is strictly increasing convex or concave function defined on the interval* [0,1] *according to whether the null-hypothesis or the alternative is true* (Figure 1)[7]. *If a com-*

[6] The definition given here is a simple extension of the usual one (see e. g.[9] p. 5) from criteria to tests (in sense of this paper).

[7] Every uniformly most powerful, uniformly most powerful similar or unbiased test belongs to the class C.

bination problem contains only tests belonging to the class C, then every Bayes solution of this problem is generated by a statistic which is monotone in sense of the definition given in § 2.

The importance of this theorem is clear from the fact that the class of all Bayes solutions in a relatively wide and typical class of hypothesis testing problems are "complete", i. e. for every test there can be given a Bayes solution which is at least as good as this test. (See e. g. [16], pp. 130—138.)

In the proof of the above theorem the fact will be used that the posterior distributions may be derived from the *posterior distribution functions* defined by the relations

$$F_{Q_0}(t_1, t_2, \ldots, t_r) = \int\limits_{\Omega_0} \prod_{i=1}^{r} F_{i,\theta_{i0}}(t_i)\, dQ_0(\theta_{10}, \theta_{20}, \ldots, \theta_{r0}) \tag{4.3}$$

and

$$F_{Q_1}(t_1, t_2, \ldots, t_r) = \int\limits_{\Omega_1} \prod_{i=1}^{r} F_{i,\theta_{i1}}(t_i)\, dQ_1(\theta_{11}, \theta_{21}, \ldots, \theta_{r1}) \tag{4.4}$$

Here Q_0 and Q_1 are the prior distributions introduced on the sets Ω_0 and Ω_1, respectively, in such a way that the functions $F_{i,\theta_{i0}}$ and $F_{i,\theta_{i0}}$ are measurable with respect to the corresponding σ-algebras.

As the functions $F_{i,\theta_{i0}}$ and $F_{i,\theta_{i1}}$ are strictly increasing and continuous, further the functions $F_{i,\theta_{i0}}$ are convex and the functions $F_{i,\theta_{i1}}$ are concave ($i = 1, 2, \ldots, r$), their derivatives

$$f_{i,\theta_{i0}}(t) = F'_{i,\theta_{i0}}(t) \tag{4.5}$$

$$f_{i,\theta_{i1}}(t) = F'_{i,\theta_{i1}}(t) \tag{4.6}$$

$$i = 1, 2, \ldots, r$$

exist and the functions in (4.5) are strictly increasing, while the functions in (4.6) are strictly decreasing. Since the possible density functions of the sample of this posterior hypothesis testing problem are defined by the relations

$$\frac{\partial^r F_{Q_1}(t_1, t_2, \ldots, t_r)}{\partial t_1 \partial t_2 \ldots \partial t_r} = \int\limits_{\Omega_0} \prod_{i=1}^{r} f_{i,\theta_{i0}}(t_i)\, dQ_0(\theta_{10}, \theta_{20}, \ldots, \theta_{r0}) \tag{4.7}$$

and

$$\frac{\partial^r F_{Q_1}(t_1, t_2, \ldots, t_r)}{\partial t_1 \partial t_2 \ldots \partial t_r} = \int\limits_{\Omega_1} \prod_{i=1}^{r} f_{i,\theta_{i1}}(t_i)\, dQ_1(\theta_{11}, \theta_{12}, \ldots, \theta_{r1}) \,, \tag{4.8}$$

the statistic occurring in the fundamental lemma of NEYMAN and PEARSON is given by

$$T(t_1, t_2, \ldots, t_r) = \frac{\int\limits_{\Omega_0} \prod_{i=1}^{r} f_{i,\theta_{i0}}(t_i)\, dQ_0(\theta_{10}, \theta_{20}, \ldots, \theta_{r0})}{\int\limits_{\Omega_1} \prod_{i=1}^{r} f_{i,\theta_{i1}}(t_i)\, dQ_1(\theta_{11}, \theta_{21}, \ldots, \theta_{r1})} \tag{4.9}$$

But this statistic is a strictly increasing function of $t_1, t_2, \ldots, t_r$. This property of T is obviously equivalent to the statement of the theorem.

§. 5. A Property of Fisher's "Omnibus Test"

J. NEYMAN and E. S. PEARSON in their paper [11] developed a principle to determine a "natural" test for composite hypotheses. This "principle of likelihood ratio" is available in case of existence of the "likelihood function" of the sample drawn to test the null-hypothesis. This statistic is the frequency function of the sample, or, in exact terminology, the "generalized (Radon—Nikodym) density" of the sample distribution. Thus the existence of the likelihood function is connected to the assumption that the possible distributions of the sample form a dominated set of measures (see [5]). Under the validity of this assumption let

$$(5.1) \qquad P_\theta(A) = \int_A f_\theta(x)\, d\mu(x) \qquad A \in \mathcal{A}$$

be the possible distrubitions of the sample x. Let as before, Ω_0 resp. Ω_1 denote, the subsets of Ω corresponding to the null-hypothesis resp. to the alternative one. μ is a fixed σ-finite measure defined on $(X, \mathcal{A})$. The measurable function to $f_\theta(x)$ occurring in (5.1) is the *generalized density function* of the sample.

The principle of likelihood ratio means the following method: consider the — so-called *likelihood ratio* — statistic

$$(5.2) \qquad R(x) = \frac{\sup\limits_{\theta_0 \in \Omega_0} f_{\theta_0}(x)}{\sup\limits_{\theta \in \Omega} f_\theta(x)}$$

and use the test generated by R in the sense of (1.14)—(1.17). This is the *likelihood ratio solution* of the problem.

Various authors succeeded in proving several good properties of this principle (see e. g. [9]). Here only the following theorem — related to the combination problem — will be proved: *if the test to be combined is similar and the possible distribution functions of the levels are all belonging to the class C defined in § 4, then the likelihood ratio solution of the combination problem is equivalent to* Fisher'*s "omnibus test*".

To prove this theorem let us remark first that in the combination problem the likelihood function is

$$(5.3) \qquad f_\theta(t_1, t_2, \ldots, t_r) = \prod_{i=1}^{r} f_{i,\theta_i}(t_i)$$

where, as before, f_{i,θ_i} denotes the frequency function of the level L_i in the case of $\theta = (\theta_1, \theta_2, \ldots, \theta_r) \in \Omega$ (see (4.5)—(4.6). On account of the assumption of similarity,

$$(5.4) \qquad f_{i,\theta_{i0}}(t_i) \equiv 1 \qquad 0 \leqq t_i \leqq 1\,, \quad i = 1, 2, \ldots, r$$

if $\theta_0 = (\theta_{10}, \theta_{20}, \ldots, \theta_{r0}) \in \Omega_0$ and thus for the likelihood ratio the relation

$$(5.5) \qquad R(t_1, t_2, \ldots, t_r) = \frac{1}{\sup\limits_{\theta = (\theta_1, \theta_2, \ldots, \theta)\in\Omega} \prod_{i=1}^{r} f_{i,\theta_i}(t_i)}$$

holds. But for any fixed i and t_i $(0 \leqq t_i \leqq 1,\ 1 \leqq i \leqq r)$ it follows from the assumption of concavity of the functions F_{i,θ_i} that

$$\sup f_{i,\theta_i}(t_i) = \frac{1}{t_i} \tag{5.6}$$

where $\frac{1}{t_i}$ is the value, at $t = t_i$, of the left derivative of the distribution function

$$F_{i;t_i}(t) = \begin{cases} \frac{t}{t_i} & \text{for } 0 \leqq t \leqq t_i \\ 1 & \text{for } t_i \leqq t \leqq 1 \end{cases} \tag{5.7}$$

which is the upper envelope of some functions F_{i,θ_i} (see Fig. 2). On account of (5.6), the likelihood ratio is

$$R(t_1, t_2, \ldots, t_r) = t_1 t_2 \ldots t_r \tag{5.8}$$

which is equivalent to FISHER's "omnibus test". this was to be proved.

§ 6. Choice from M^*: Combination by means of Φ^{-1}

The question arises naturally which combination should be used in practice. The answer is not a unique one : the choice depends on the available information concerning the experiments and the tests used in them. As it was shown before, there corresponds a hypothesis testing problem to each combination problem the optimal solution of which — if it exists — gives the optimal solution for the combination problem too.

However, the need for combination of independent tests arises just in those cases in which no or scarce information only is available : the information consists, at most of the number of observations of the individual experiments. Thus the restrictions introduced in § 2 are not too arbitrary : they mean the necessary compromise with the insufficiency of information.

Further, the results of §§ 3 and 4 show that Postulate 1 is a rather natural one, Postulate 2 indicates an intrinsically desirable property of combinations (according to later similar combination) while Postulate 3 means only a norming in the set of combinations. Thus the only thing to be done is to choose an element from M^*.

Any choice has some arbitrariness, but it is suggested by a number of reasons to choose the combination the averaging function of which is Φ^{-1} i. e. the inverse of the normal distribution function

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du \,. \tag{6.1}$$

The formula of the combined level $\tilde{L}$ of the levels $L_1, L_2, \ldots, L_r$ with weights $\lambda_1, \lambda_2, \ldots, \lambda_r$, respectively, is the following :

$$\tilde{L} = \Phi\left(\frac{\lambda_1 \Phi^{-1}(L_1) + \lambda_2 \Phi^{-1}(L_2) + \ldots + \lambda_r \Phi^{-1}(L_r)}{\sqrt{\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_r^2}}\right) \tag{6.1}$$

(cf. (2.39)). This combination needs very easy computational works and tables, and, suitably choosing the weights, is optimal for a rather wide class of one-sided hypothesis testing problems which may be characterized as follows: the possible distributions are generated by densities belonging to the exponential family (see e. g. (8.1) in [10] or (2.3) in [2]) and the null-hypotheses are stating the validity of a one-sided inequality for one of the parameters in the density function. This question and similar ones will be treated in more detail in a forthcoming paper of the author.

The weights occurring in (6.2) should be chosen as to express the efficiencies (or the available information concerning these) of the tests used in the single experiments. Thus λ_i should be chosen proportional to the „expected" difference between the null-hypothesis and the real situation and inversely proportional to the standard deviation of the statistics used for testing in the ith experiment ($i = 1, 2, \ldots, r$). Generally, this standard deviation is inversely proportional to the square root of the number of observations of this experiment. If there is no information available but the number n_i of observations in the experiments ($i = 1, 2, \ldots, r$) the weights may be chosen proportional to the square roots of these numbers, i. e. in this case

$$(6.3) \qquad \tilde{L} = \Phi\left(\frac{\sqrt{n_1}\,\Phi^{-1}(L_1) + \sqrt{n_2}\,\Phi^{-1}(L_2) + \ldots + \sqrt{n_r}\,\Phi^{-1}(L_r)}{\sqrt{n_1 + n_2 + \ldots + n_r}}\right).$$

Formulae (6.2) and (6.3) apply equally to significance tests and to tests for goodness of fit (e. g. for combination testings for normality). In case of significance tests, however, the formulae

$$(6.4) \qquad \tilde{L} = 1 - \Phi\left(\frac{\lambda_1 \Phi^{-1}(1 - L_1) + \lambda_2 \Phi^{-1}(1 - L_2) + \ldots + \lambda_r \Phi^{-1}(L_r)}{\sqrt{\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_r^2}}\right)$$

and

$$(6.5) \qquad \tilde{L} = 1 - \Phi\left(\frac{\sqrt{n_1}\,\Phi^{-1}(1 - L_1) + \sqrt{n_2}\,\Phi^{-1}(1 - L_2) + \ldots + \sqrt{n_r}\,\Phi^{-1}(1 - L_r)}{\sqrt{n_1 + n_2 + \ldots + n_r}}\right)$$

(being equivalent to (6.2) and (6.3), respectively) are more easily applicable as the tables of the function $\Phi^{-1}(t)$ (see e. g. [1] p. 557 or [15], p. 173) are given in detail only for values of t near to 1.

(Received December 20, 1958)

REFERENCES

[1] Cramér, H.: *Mathematical methods of statistics.* Princeton University Press, Princeton, 1946.

[2] Дынкин, Е. В.: "Необходимые и достаточные статистики для семейства распределений вероятностей". *Успехи Математических Наук* **6**:I (1951) 68—90.

[3] Fisher, R. A.: *Statistical methods for research workers.* Oliver and Boyd, London, 1932 (4th edition).

[4] Good, I. J.: "On the weighted combination of significance tests". *Journal of the Royal Statistical Society, Series B* **17** (1955) 264—265.

[5] HALMOS, P. R.—SAVAGE, I. : "Application of the Radon-Nikodym theorem to the theory of sufficient statistics". *The Annals of Mathematical Statistics* **20** (1949) 225—241.
[6] HALMOS, P. R. : *Measure theory*. Van Nostrand, New-York, 1950.
[7] HARDY, G. H.—LITTLEWOOD, J. E.—PÓLYA, G. : *Inequalities*. Cambridge, at the University Press, 1952.
[8] LANCASTER, H. O. : "The combination of probabilities arising from data in discrete distributions". *Biometrika* **36** (1949) 370—382.
[9] LEHMANN, E. L. : "Some principles of the theory of testing hypotheses". *The Annals of Mathematical Statistics* **21** (1950) 1—26.
[10] LEHMANN, E. L.—SCHEFFÉ, H. : "Completeness, similar regions, and unbiased estimation, II.". *Sankhyā* **10** (1950) 305—340.
[11] NEYMAN, J.—PEARSON, E. S. : "On the use and interpretation of certain test criteria for purposes of statistical inference". Biometrika **20** (1928) 175— and 263—
[12] PEARSON, E. S. : "The probability integral transformation for testing goodness of fit and combining independent tests of significance". *Biometrika* **30** (1938) 134—148.
[13] PEARSON, E. S. : "On questions raised by the combination of tests based on discontinuous distributions". *Biometrika* **37** (1950) 383—398.
[14] PEARSON, K. : "On a method of determining whether a sample of size supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random". *Biometrika* **25** (1933) 379—410.
[15] RÉNYI, A. : *Valószínűségszámítás*. Tankönyvkiadó, Budapest, 1954.
[16] WALD, A. : *Statistical decision functions*. Wiley, New York, 1950.
[17] WALLIS, W. A. : "Compounding probabilities from independent significance tests." *Econometrica* **10** (1942) 229—248.

FÜGGETLEN MOZGÓ SZINTES PRÓBÁK ÖSSZEVONT ÉRTÉKELÉSÉRŐL

LIPTÁK T.

Kivonat

E dolgozatban szignifikancia-vizsgálatnál vagy illeszkedés-vizsgálatnál alkalmazható próbák mozgó szintjének nevezzük azt a legkisebb, illetve legnagyobb szintet, melynél az eredményt még szignifikánsnak, illetve illeszkedőnek lehet mondani. A statisztikai gyakorlatban az egyes vizsgálatoknál e mozgó szint megadásával szokták jellemezni a kísérleti eredmények szignifikáns, illetve illeszkedő voltát. Gyakran előfordul, hogy ugyanazon (vagy hasonló) hipotézisre vonatkozólag egymástól függetlenül több kísérletet is végeznek, de a kísérleti feltételek különböző volta és egyéb okok miatt az adatokat egyetlen mintává nem egyesíthetik (normalitásvizsgálat). Célszerű ilyenkor az egész kísérletsorozatot egyetlen mozgó szinttel jellemezni, más szóval az egyes kísérletekben kapott mozgó szinteket összevonni.

A dolgozat 1. §.-ban megállapítjuk, hogy bármely próba mozgó szintje olyan valószínűségi változó, melynek eloszlásfüggvénye a nullhipotézis fennállása esetén mindenütt kisebb vagy legfeljebb akkora , mint a [0,1] intervallumban egyenletes eloszlásfüggvény. Torzítatlan próbák esetén az is igaz, hogy a mozgó szint eloszlásfüggvénye az alternatív hipotézis fennállása esetén mindig nagyobb vagy legfeljebb egyenlő, mint a nullhipotézis alatt. Mozgó szintek összevonásának kérdését így egy hipotézis-vizsgálati problémává lehet átfogalmazni.

A 2. §.-ban kimutatjuk hogy néhány célszerű tulajdonság megkövetelésével az összevonások körét a mozgó szintek valamilyen súlyozott középértéke által

nyerhető összevonásokra lehet leszűkíteni. A szereplő súlyok az egyes mozgó szintek megbízhatóságára, a megfelelő próba efficienciájára vonatkozó esetleges információkból állapíthatók meg. E követelések lényegileg a következők: 1. Az összevonás ,,monoton'' legyen, tehát két szint-sorozat közül okvetlenül az elsőt minősítse szignifikánsabbnak (illetve illeszkedőbbnek), ha annak szintjei rendre szignifikánsabbak (illetve illeszkedőbbek), mint a másik sorozat megfelelő szintjei. 2. Az összevonás ,,kompatilibis'' legyen, tehát tetszőleges szintcsoportok esetén az egyes csoportok összevont szintjei, valamint az egyes csoportoknak tulajdonított súlyok értékei már egyértelműen határozzák meg az egyesített szintcsoport összevont szintjének értékét. 3. Bármely egyelemű szint- ,,sorozat'' összevont szintje egyezzék meg e szint értékével.

Az összevonások fenti M^* osztálya tartalmazza az eddig használt FISHER-féle ,,omnibus test''-et [3], sőt, annak GOOD-féle általánosítását is [4]. A 3. §.-ban megmutatjuk, hogy az M^* osztályba tartozó összevonások valamennyien a szokásos definíció szerint elfogadhatók és torzítatlanok. A 4. §.-ban bebizonyítjuk, hogy amennyiben csak bizonyos tipikusan ,,jó'' próbákra szorítkozunk, az összevonási probléma valamennyi Bayes-megoldása a 2. §.-ban adott definíció értelmében monoton statisztikára alapozott összevonás. A dolgozat 5. §.-át azon tény bizonyításának szenteljük, hogy a FISHER-féle ,,omnibus-test'' lényegében ugyanezen próbák körében az összevonási probléma ,,likelihood-hányados'' megoldása.

Végül a 6. §.-ban az M^* osztály elemei közül a normális eloszlásfüggvény inverze szerint közepelt összevonást ajánljuk, amelynél az összevont szint képlete a következő :

$$\tilde{L} = \Phi\left(\frac{\lambda_1 \Phi^{-1}(L_1) + \lambda_2 \Phi^{-1}(L_2) + \ldots + \lambda_r \Phi^{-1}(L_r)}{\sqrt{\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_r^2}}\right). \tag{1}$$

Itt $L_1, L_2, \ldots, L_r$ az összevonásra kerülő mozgó szintek, $\lambda_1, \lambda_2, \ldots, \lambda_r$ pedig a nekik tulajdonított súlyok értékei. Egyéb információk hiányában λ_i-t az i-edik kisérletben szereplő megfigyelések n_i számának négyzetgyökével arányosnak lehet venni. Ekkor a fenti képletben λ_i helyett egyszerűen $\sqrt{n_i}$-t kell írni. Viszonylag kicsiny szintek esetén célszerűbb a fentiekkel ekvivalens (6.4), illetve (6.5) képletekkel számolni.

О СОВМЕСТНОЙ ОЦЕНКЕ НЕЗАВИСИМЫХ ОПЫТОВ

T. LIPTÁK

Резюме

В настоящей работе движущимся уровнем критерии, применимых при проверке значимости, называется тот наинизший уровень, при котором результат может ещё считаться сигнификантным. В статистической практике при отдельных исследованиях заданием этого движущегося уровня характеризуют сигнификантность результатов опытов. Часто случается, что относительно одной и той же (или похожих) гипотез производится независимо друг от друга несколько опытов, но

из-за различности условий опытов или по другим причинам данные не могут быть объединены в единственной выборке. Целесообразно в этом случае характеризировать серию опытов единственным движущимся уровнем, иначе говоря, совместно оценить движущиеся уровни, полученные в отдельных опытах.

В первом параграфе работы доказывается, что движущийся уровень любой критерии есть такая случайная величина, функция распределения которой, в случае выполнения нуль-гипотезы, не превосходит функции распределения, равномерной на отрезке [0,1]. В случае проб без искажения верно и то, что, в случае выполнения альтернативной гипотезы, имеет место обратное неравенство. С помощью этого вопрос о совместной оценке движущихся уровней может быть переформулирован как проблема об изучении гипотезы.

Во втором параграфе, требуя несколько целесообразных свойств, вопрос сводится к совместных оценкак, получаемых с помощью какого-нибудь средного значения с весом движущихся уровней. Фигурирущие веса могут быть определены из информаций относительно надёжности отдельных движущихся уровней, эффициенции соответствующей критерии. Требования в сущности следующие: 1. совместная оценка должна быть «монотонной», т. е. квалифицирует последовательность уровней более сигнификантной, чем другую, если уровни первой сигнификантней, чем соответствующие уровни последней. 2. совместная оценка должна быть «комбатибильна», т. е. в случае любых групп уровней значения совместного уровней отдельных групп и значения весов, соответствующих отдельным группам, уже единственным образом определяют значение собместного уровня объединенной группы уровней. 3. совместное значение любой одноэлементной «последовательности» уровней должно совпадать со значением уровня.

Указанный выше класс совместных оценок M^* содержит использованное до сих пор «omnibus test» Fisher-а [3] и её обобщение, данное Good-ом [4]. В третьем параграфе доказывается, что все совместные оценки, принадлежащие классу M^*, допустимы и не искажённы. В четвертом параграфе доказывается, что, если ограничиться некоторыми типичными допустимыми критериям, все решения Bayes-а проблемы концентрации монотонны в смысле определения § 2. В параграфе 6. работы доказывается, что «omnibus test» Fisher-а в круге этих же критериев есть решение проблемы «likelihood-частного». Наконец, в параграфе 5 из элементов класса M^* предлагается то в котором усреднение производится с помощью обратной нормальной функции распределения, для которой формула совместного уровня такова:

$$\tilde{L} = \Phi\left(\frac{\lambda_1 \Phi^{-1}(L_1) + \lambda_2 \Phi^{-1}(L_2) + \ldots + \lambda_r \Phi^{-1}(L_r)}{\sqrt{\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_r^2}}\right).$$

Здесь $L_1, L_2, \ldots, L_r$ концентрируемые движущиеся уровни отдельных опытов, а $\lambda_1, \lambda_2, \ldots, \lambda_r$ значения приписанных им весов. Доказывается, что этот способ оптимален в некотором достаточно широком классе изучений односторонных гипотез.