E 11-10-M

"ON THE TESTING OF COMPOSITE STATISTICAL HYPOTHESES"

by

J. Neyman

(Translation from the Bulletin of the French Mathematical Society, vol. 63, 1935, by W. T. Federer*)

November, 1949

*The translator wishes to acknowledge the helpful comments of J. E. and Gertrude Morton and Dr. Louis DeVries on the translation of this article.

ON THE TESTING OF COMPOSITE STATISTICAL HYPOTHESES

by J. Neyman

1. Historical. It is known that the problems of testing of the hypotheses have been treated since the time of Thomas Bayes [1]. The solutions obtained depended upon the a priori probabilities. These being generally unknown, one was compelled then to make arbitrary hypotheses which gave results not applicable to practical problems.

35 years ago, Karl Pearson [2] published a method for the testing of a particular statistical hypothesis, a method known to us by the name of χ^2. There was no question of the a priori probabilities in this memoir which has played such a remarkable role. It was followed by a series of works by the same author as well as by his successors, of which the principal representatives are W. P. Elderton, R. A. Fisher, E. S. Pearson, "Student", J. Wishart and others. I ought to mention also the works of Lexis and Bortkiewicz concerning the hypothesis on the stability of probabilities.

Although a number of problems of great importance have been resolved by these authors, the origin of the general theory of testing hypotheses is related to extremely interesting remarks of the French geometricians, especially (those) by J. Bertrand [3] and E. Borel [4]. They considered the procedure for the testing of a hypothesis H, which is essentially the same procedure as followed by the English school. We observe a fact E and then we choose as a base for testing the hypothesis H a function f(E) of E. We calculate the probability P, determined by H, for f(E). If the probability P is judged small, we reject H; if, on the other hand, P is large, there isn't sufficient reason for rejecting H.

We know that Bertrand was skeptical about the scientific value of the results of such a procedure. On the other hand, Borel has insisted on the fact that this result can be valuable, provided that the function f(E) of the observed fact, which serves as the basis for testing H, is "in some way remarkable."

This last remark preceded a series of papers [5, 18] beginning the general theory of testing hypotheses in which essentially the effort was made to give a precise definition to the somewhat vague works of Borel concerning the function of the observed fact which is "in some way remarkable."

Before coming to the problem which is the principal subject of this note, it appears desirable to recall briefly the results obtained thus far.

2. Notations and preliminary remarks.--I designate by $P\{E\}$ the probability of any event E, and by $P\{E_1 \mid E_2\}$ of the event E_1 calculated under the assumption that another event E_2 has already taken place. If

$$(1) \qquad X_1, X_2, \dots, X_n$$

designates a system of n random variables and

$$(2) \qquad x_1, x_2, \dots, x_n$$

a system of the particular values of these variables, I am going to designate by E the point in the n-dimensional space W whose coordinates are (2). E will be called the sample point and W the sample space. Let w be any measurable set within W, and $P\{E \varepsilon w\}$ the probability with which the point E falls in w. The probability $P\{E \varepsilon w\}$ considered as a function of the set w will be called the law of the total probability for the variables (1).

Any hypothesis H concerning the nature of the probability distribution $P\{E \varepsilon w\}$ is called a statistical hypothesis. A statistical hypothesis is called simple if it determines the value of $P\{E \varepsilon w\}$ in a unique way for any measurable sub-sets w in W. Any statistical hypothesis which is not simple is called composite; if H is a composite hypothesis, there should be at least one measurable sub-set w' in W such that the value for $P\ E \varepsilon w'$ is not determined uniquely by H. Hence if H is a composite hypothesis, there must exist a set of simple hypotheses, H', which can be obtained from H by adding a few additional assumptions. These hypotheses H' should not contradict H.

In the following I shall consider methods of testing statistical hypotheses. Such a method consists of choosing a region w_0 in W and of following the rule for rejecting the hypothesis H which we are testing when the sample point E which is determined by the observed values [8, 9, 10] falls within w_0. If point E does not fall in w_0, we do not reject H; in this case we will say that one accepts the hypothesis H.

The region which is used in this manner to test the hypothesis H is called a critical region, and its complement $W-w_0$, region of acceptance. If 2 methods of testing a hypothesis H differ, the reason is that the corresponding critical regions are different. To choose a method of testing is to choose a critical region.

The choice of a critical region ought to be based on the consideration of errors which we might commit in testing a hypothesis. There can be one of the two following kinds of errors:

1. we can reject the hypothesis H_0 when, in fact, it is true and 2. we can accept H_0 when it is false.

It should be mentioned that, in testing hypothesis H_0, we admit the the possibility of at least one other hypothesis contradictory to H_0;

this contradictory hypothesis, which may be true, I shall call an alternative hypothesis. I will assume that in any particular case we can define a region Ω of possible simple hypotheses. We can then say that an error of the second kind consists in accepting the tested hypothesis if there is a hypothesis H_1 which is contained in Ω, but contradictory to the true hypothesis, H_0.

In recent literature [8, 9, 10] an attempt was made to find tests (that is to say critical regions) satisfying the two following conditions:

1. that when the hypothesis H_0 which is to be tested is true, the probability that it will be rejected is equal to or smaller than a predetermined α.

2. that when it is an alternative hypothesis H_1 which is true, the probability that we reject H_0 is in general as large as possible.

We see that these two conditions must be quite precisely stated.

3. <u>Uniformly most powerful critical regions.</u>--Let w be some critical region designed to test the statistical hypothesis H_0. Consider the probability $P\{E \varepsilon w \mid H_0\}$ calculated in the assumption that the hypothesis H_0 is true. If this probability is completely determined by H_0 (as is always true when the hypothesis H_0 is simple and in certain cases when it is composite) the value $P\{E \varepsilon w \mid H_0\} = \alpha$ will be called the size of the critical region w. If the probability $P\{E \varepsilon w \mid H_0\}$ is not determinate (which can happen when H_0 is composite) we say that the size of region w is indeterminate.

We see that $P\{E \varepsilon w \mid H_0\}$ is equal to the probability of errors of the first kind calculated on the assumption the hypothesis H_0 which is to be tested is true. Consequently, one tries to use only critical regions which have a completely determined and rather small size, α. I will assume that w is such a region.

Denote by H_1 an admissable simple hypothesis alternative to H_0 and the probability determined by this hypothesis

$$\beta(H_1 \mid w) = P\{E \varepsilon w \mid H_1\} \tag{3}$$

that the sample point E falls in the critical region w. We see that $1 - \beta(H_1 \mid w)$ is the probability of an error of the second kind, calculated in the assumption that the true hypothesis is H_1. We call $\beta(H_1 \mid w)$ the <u>power</u> of the critical region w with respect to the hypothesis H_1.

<u>We say</u> [8, 9] <u>that a region w_0 is a uniformly most powerful region</u>

of size α with respect to a simple alternative hypothesis H_1 if it possesses the following properties:

(a) the size of w_0 is well determined and is equal to α;

(b) the power w_0 with respect to H_1 is at least equal to that of any region w for which the size is well determined and equal to α.

We say that a region w_0 is a critical region of size α, uniformly most powerful with respect to the region Ω of simple alternative hypothesis, if it is a uniformly most powerful critical region of size α with respect to any simple hypothesis alternative to H_0 and belonging to Ω.

It is easy to see that there exists a critical region of size α, uniformly most powerful with respect to Ω, and that in using this region in testing hypothesis H_0 one has the advantage that (1) the probability determined for H_0 for the error of the first kind is equal to α, and (2) for the alternative hypothesis, if it is true, the probability of an error of the second kind is a minimum.

Unfortunately, the uniformly most powerful region exists very seldom. Ordinarily a region of size α which is most powerful with respect to a simple alternative hypothesis H_1 is not for another H_2 and, what is more, the power of this region with respect to H_2 is perhaps smaller than α. Therefore, if we apply such a region for testing H_0 and if it is thought that the true hypothesis is $H_2 \neq H_0$, the hypothesis for testing H_0 will be accepted more often than in the cases in which it is a true hypothesis.

If a uniformly most powerful region does not exist, there is a basis for using a definition of the critical region which does not possess such drawbacks.

4. Unbiased critical regions.--Let H_0, a statistical hypothesis to be tested, and w_0 have a critical region, of a well defined size α. Consider the probability determined by any simple hypothesis H contained in Ω for which the sample point falls in w_0, which we designate as

$$\beta(H|w_0) = P\{E \varepsilon w_0 | H\} . \tag{4}$$

If the hypothesis to be tested, H_0, is simple, then $\beta(H_0|w_0)$ has a meaning and is equal to α. If H_0 is composite and H_0' a simple hypothesis which does not contradict H_0, then $\beta(H_0'|w_0) = \alpha$. Since this equality holds for any simple hypothesis which does not contradict H_0, one may omit the prime designating the simple hypothesis H_0' and consider that

$\beta(H \mid w_0)$ is defined also for the composite hypothesis H_0 and that

(5) $$\beta(H_0 \mid w_0) = \alpha.$$

The function $\beta(H \mid w_0)$ for the hypothesis H thus defined will be called the power function.

The reasoning at the end of the preceding paragraph suggests that it is desirable to find the critical regions w_0 such that the corresponding power function has an absolute minimum at $H = H_0$.

We say that a region w_0 is an unbiased critical region of size α (this term, which corresponds to the English term "unbiased critical set," was suggested by M. Georges Darmois, to whom I am very grateful) when its size is well determined by the hypothesis H_0 to be tested and is equal to α, if the power function $\beta(H \mid w_0)$ has a minimum at $H = H_0$.

Suppose that all the simple admissible hypotheses from the set Ω determiné the probability distribution $P_1\{E\epsilon w\}$ for the variables (1) as a set function with the same analytical form, depending on a certain number l of unknown parameters $\theta_1, \theta_2, \ldots \theta_l$ and which do not differ among themselves but for the values given to parameters.

Then the power function of the critical region w_0 may be considered as a function $\beta(\theta_1, \theta_2, \ldots \theta_l \mid w_0)$ of the parameters $\theta_1, \theta_2, \ldots . \theta_l$.

In a recent paper [10] the case where the number of unknown parameters is equal to one was considered. In this case the hypothesis H_0 to be tested is a simple hypothesis. The critical region was defined there in the following way:

A region w_0 is called an unbiased critical region of size α of type A, if the power function $\beta(\theta \mid w_0)$ has a second derivative and if

(6) $$\frac{d\beta(\theta_0 \mid w_0)}{d\theta} = 0$$

and

(7) $$\frac{d^2\beta(\theta_0 \mid w_0)}{d\theta^2} = \text{maximum}$$

where θ_0 is the value for the parameter θ postulated by the hypothesis H_0 which we are testing.

I will now treat an analogous problem for the case that the hypothesis to be tested is a composite hypothesis.

5. Unbiased critical regions of type B.--Consider the case where there are only two unknown parameters θ_1 and θ_2 and assume that the hypothesis

to be tested, H_0, is composite and postulates the value $\theta_1 = \theta_1^o$ while allowing the value of θ_2 to be undetermined.

Suppose in addition that the values of θ_1 postulated by the hypotheses contained in Ω make up a certain interval containing θ_1^o. [It is interesting that as soon as this condition is fulfilled, there exists no uniformly most powerful region except in exceptional cases [11].]

I say that the region w_0 is an unbiased critical region of size α of type B if:

a. Its size is determined by the hypothesis H_0 and is equal to α, i.e.

$$\beta(\theta_1^o, \theta_2 \mid w_0) = \alpha \tag{8}$$

whatever is assumed for admissible values of θ_2;

b. If the power function $\beta(\theta_1, \theta_2 \mid w_0)$ is defined for the first 2 derivatives with respect to θ_1;

c. If

$$\frac{\delta\beta(\theta_1^o, \theta_2 \mid w_0)}{\delta\theta_1} = 0 \tag{9}$$

whatever is assumed for admissible values of θ_2; and

d. If, whatever is assumed for another region w_1 satisfying a, b, and c, we have

$$\frac{\delta^2\beta(\theta_1^o, \theta_2 \mid w_0)}{\delta\theta_1^2} \geqq \frac{\delta^2\beta(\theta_1^o, \theta_2 \mid w_1)}{\delta\theta_1^2} \tag{10}$$

I have been able to find the solution of the problem of unbiased regions of type B in the particular case which complies with the following conditions:

1. There exists a function $P(E \mid \theta_1 \theta_2)$ which is defined as non-negative and integrable over the entire sample space W, such that, for all measurable regions w in W and for all combinations of values for θ_1 and θ_2 corresponding to an admissible hypothesis H', we have

$$\beta(\theta_1, \theta_2 \mid w) = P\{E \epsilon w \mid H'\} = \int \cdots \int_w p(E \mid \theta_1 \theta_2) dx_1 \cdots dx_n. \tag{11}$$

I will call the function $p(E \mid \theta_1 \theta_2)$ the elementary probability distribution [the terms: distribution of the total and elementary probability has been discussed in an analogous way by P. Levy [20]] of the variables (1).

2. Whatever the measurable region w in W the expression under the integral in (11) possesses a derivative and

$$\frac{\delta^{i+k}\beta(\theta_1, \theta_2 \mid w)}{\delta\theta_1^i \, \delta\theta_2^k} = \int \cdots \int_w \frac{\delta^{i+k} p(E \mid \theta_1 \; \theta_2)}{\delta\theta_1^i \quad \delta\theta_2^k} dx_1 dx_2 \cdots dx_n \tag{12}$$

for $i = 0, 1, 2$ and for all integral non-negative values of k.

3. If we denote by

$$\phi_i = \frac{\delta \log p(E \mid \theta_1 \theta_2)}{\delta\theta_i} \qquad (i = 1,2) \tag{13}$$

$$\phi_{ij} = \frac{\delta\phi_i}{\delta\theta_j} = \phi_{ji} \quad (i, j = 1,2), \tag{14}$$

then we have

$$\phi_{11} = A_0 + A_1\phi_1 + A_2\phi_2, \tag{15}$$

$$\phi_{12} = B_0 + B_1\phi_1 + B_2\phi_2, \tag{16}$$

$$\phi_{22} = C_0 \qquad\qquad + C_2\phi_2, \tag{17}$$

where the coefficients A, B, C are not dependent on $x_1, x_2, \ldots x_n$.

4. The functions ϕ_1 and ϕ_2 are algebraically independent, that is to say that there are two numbers $\underline{l}$ and $\underline{m}$ such that the Jacobian

$$\frac{\delta(\phi_1, \phi_2)}{\delta(x_{\underline{l}}, x_{\underline{m}})} \tag{18}$$

is different from zero in W, except perhaps for a set of points of measure zero.

5. The function ϕ_2 can be considered as a particular value of a random value Φ_2. Let $p(\phi_2)$ be its elementary probability distribution which can be deduced from $p(E \mid \theta_1^o \theta_2)$.

Furthermore let

$$M_k = \int_{-\infty}^{+\infty} \phi_2^k \, p(\phi_2) d\phi_2 \tag{19}$$

as the k^{th} moment of $p(\phi_2)$.

I assume that the moments M_k exist for every integral-non-negative k. It is well known that, from the moments M_k, we can construct a system of orthogonal polynomials π_i of order i in ϕ_2, such that

$$(20)\qquad \int_{-\infty}^{+\infty} \pi_i \, \pi_j p(\phi_2) d\phi_2 = \begin{cases} 0 \text{ if } i \neq j. \\ 1 \text{ if } i = j. \end{cases}$$

I assume that this system is complete, that is, that any integrable function $f(\phi_2)$ is orthogonal to the system of polynomials, that is

$$(21)\qquad \int_{-\infty}^{+\infty} f(\phi_2) \, \pi_i p(\phi_2) d\phi_2 = 0 \qquad (i = 0, 1, 2 \ldots)$$

or, which is the same, such that

$$(22)\qquad \int_{-\infty}^{+\infty} \phi_2^k \, f(\phi_2) p(\phi_2) d\phi_2 = 0 \qquad (k = 0, 1, 2 \ldots)$$

must satisfy

$$(23)\qquad f(\phi_2) = 0,$$

except perhaps for a set of values for ϕ_2 of measure zero.

It is easily seen that, when x satisfies condition 5, then: (a) there exists a single probability distribution $p(\phi_2)$ having the numbers M_k for moments and (b) if y and z are two random variables and $p(y\phi_2)$ and $p(z\phi_2)$ the elementary probability distributions of y and ϕ_2 and for z and ϕ_2, then the equality

$$(24)\qquad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y\phi_2^k \, p(y\phi_2) dy d\phi_2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z\phi_2^k \, p(z\phi_2) dz d\phi_2$$

for $k = 0, 1, 2 \ldots$ leads to

$$(25)\qquad \int_{-\infty}^{+\infty} y p(y\phi_2) \, dy = \int_{-\infty}^{+\infty} z p(z\phi_2) \, dz$$

for almost all values of ϕ_2.

The property (a) results from the fact that if there is another probability distribution $p_1(\phi_2)$ having the same moments M_k, we then write

$$(26)\qquad p_1(\phi_2) = p(\phi_2) \left[1 + \Delta \, (\phi_2) \right],$$

and we have

$$(27)\qquad M_k = \int_{-\infty}^{+\infty} \phi_2^k\, p_1(\phi_2)d\phi_2$$

$$= \int_{-\infty}^{+\infty} \phi_2^k\, p\,(\phi_2)d\phi_2 + \int_{-\infty}^{+\infty} \phi_3^k\, \Delta\,(\phi_2)p(\phi_2)d\phi_2,$$

then

$$(28)\qquad \int_{-\infty}^{+\infty} \phi_2^k\, \Delta\,(\phi_2)p(\phi_2)d\phi_2 = 0$$

for $k = 0, 1, 2 \ldots$ and it follows that $\Delta(\phi_2)$ vanishes almost everywhere.

To demonstrate property (b), we may say that the two members in (25) divided by $p(\phi_2)$ represent the ordinants of the regression lines of y and of z on ϕ_2. If we consider these ordinants as functions of ϕ_2, that is $\bar{y}(\phi_2)$ and $\bar{z}(\phi_2)$, which I will designate as the regression functions, and if we expand these functions in a series of polynomials π_i, we know that the Fourier coefficients for these series are determined by the value of the integral (24). Hence from the inequalities (24) follows the identity of the two developments which proves that $\bar{y}(\phi_2) = \bar{z}(\phi_2)$ almost everywhere where $p(\phi_2) > 0$.

Let $p(\phi_1 \mid \phi_2)$ be the elementary probability distribution of ϕ_1 calculated under the assumption that ϕ_2 is fixed.

6. Solution of the problem of unbiased critical regions of type B.--I say that, if conditions 1, 2, 3, 4 and 5 are fulfilled, the unbiased critical region of size α for type B is defined by the inequality

$$(29)\qquad \phi_1 \leqq k_1(\phi_2) \quad \text{and} \quad k_2(\phi_2) \leqq \phi_1,$$

where $k_1(\phi_2)$ and $k_2(\phi_2)$ satisfy the following equations:

$$(30)\qquad \int_{k_1(\phi_2)}^{k_2(\phi_2)} p(\phi_1|\phi_2)d\phi_1 = 1 - \alpha,$$

and

$$(31)\qquad \frac{1}{1-\alpha}\int_{k_1(\phi_2)}^{k_2(\phi_2)} \phi_1 p(\phi_1 \mid \phi_2)d\phi_2 = \int_{-\infty}^{+\infty} \phi_1 p(\phi_1 \mid \phi_2)d\phi_1.$$

Let us begin by constructing the most general region w, which satisfies both conditions (8) and (9). Assume then that w is some measurable region in W. Therefore equation (8) becomes

(32) $$\beta(\theta_1^\circ \theta_2 \mid w) = \int \cdots \int_w p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = \alpha,$$

and equation (9)

(33) $$\frac{\delta\beta(\theta_1^\circ \theta_2 \mid w)}{\delta\theta_1} = \int \cdots \int_w \phi_1 p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = 0,$$

where

(34) $$\frac{\delta p(E \mid \theta_1^\circ \theta_2)}{\delta\theta_1} = \phi_1 p(E \mid \theta_1^\circ \theta_2).$$

Grant that the integrals in (32) and (33) have values which do not depend on θ_2. Then if we differentiate these expressions with respect to θ_2, we get zero. Upon differentiating (32) once and upon noting that

$$\frac{\delta}{\delta\theta_2} p(E \mid \theta_1^\circ \theta_2) = \phi_2 p(E \mid \theta_1^\circ \theta_2),$$

we have

(35) $$\int \cdots \int_w \phi_2 p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = 0;$$

since this equality exists for all θ_2, we may differentiate again, which gives

(36) $$\int \cdots \int_w (\phi_{22} + \phi_2^2) p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = 0,$$

where, considering (17), (32), and (35)

(37) $$\int \cdots \int_w \phi_2^2 \, p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = \alpha C_0 = \alpha \psi_2,$$

where ψ_2 does not depend on the region w. Since (37) represents an identity, it may be differentiated, yielding after a simple calculation

(38) $$\int \cdots \int_w \phi_2^3 \, p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = \alpha \psi_3,$$

By applying the method of complete induction, it is easy to show that if (32) holds for all values of θ_2, then we have

(39) $$\int \cdots \int_w \phi_2^k \, p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = \alpha \psi_n,$$

where ψ_k is a function of θ_2 which does not depend on the region w. Since the sample space W satisfies (32) for $\alpha = 1$ and (33),

(40) $$M_k = \int \cdots \int_W \phi_2^k \, p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = \psi_k,$$

where M_k is the k^{th} moment of $p(\phi_2)$. We see that if (32) and (33) hold for all values of θ_2, then

(41) $$\frac{1}{\alpha} \int \cdots \int_w \phi_2^k \, p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = M_k \qquad (k = 0, 1, 2 \ldots).$$

The left member of the last equation may also be interpreted as a moment which I have called the k^{th} moment of ϕ_2 relative to w. In fact the probability distribution of ϕ_2 can be calculated by assuming that the sample point E may not fall in the region w. The elementary probability distribution of the variables (1), corresponding to this hypothesis, will be

(42) $$p(E \mid \theta_1^\circ \theta_2 w) = \frac{1}{\alpha} p(E \mid \theta_1^\circ \theta_2)$$

for all $E \varepsilon w$ and

(43) $$p(E \mid \theta_1^\circ \theta_2 w) = 0$$

everywhere else. Now it is easy to see that the left member of (41) represents the moments for the elementary probability distribution of ϕ_2 calculated on the assumption that the sample point E falls in w. If we designate this distribution by $p(\phi_2 \mid w)$ then in consideration of condition (5), we may write

(44) $$p(\phi_2 \mid w) = p(\phi_2)$$

for almost all values of ϕ_2.

As a consequence of the hypothesis all derivatives of (32) with respect to θ_2 are equal to zero. Now, consider (33) which also is an identity. Differentiating (33) with respect to θ_2 and in view of (16), (32), (33) and (35), one obtains easily

(45) $$\int \cdots \int_w (\phi_{12} + \phi_1 \phi_2) p(E \mid \theta_1^\circ \theta_2) dx_1 \cdots dx_n = 0,$$

or

$$(46) \qquad \int\cdots\int_w \phi_1\phi_2 p(E \mid \theta_1^o\ \theta_2)dx_1\cdots dx_n = -\alpha\beta_0 = \alpha\Psi_1$$

where Ψ_1 does not depend on the region w. By differentiating (46) again and by applying the method of mathematical induction, we see easily that, if (32) and (33) hold for all values of θ_2, then

$$(47) \qquad \int\cdots\int_w \phi_1\phi_2^k\, p(E \mid \theta_1^o\ \theta_2)dx_1\cdots dx_n = \alpha\Psi_k,$$

where Ψ_k is a function of θ_2 which does not depend on the region w. In particular

$$(48) \qquad \lambda_k = \int\cdots\int_w \phi_1\phi_2^k\, p(E \mid \theta_1^o\ \theta_2)dx_1\cdots dx_n = \Psi_k \quad (k = 0, 1, \ldots).$$

Hence, whatever the region w such that (32) and (33) hold for all θ_2, it follows that

$$(49) \qquad \frac{1}{\alpha}\int\cdots\int_w \phi_1\phi_2^k\, p(E \mid \theta_1^o\ \theta_2)dx_1\cdots dx_n = \lambda_k \quad (k = 0, 1, \ldots).$$

or, in view of (42) and (43)

$$(50) \qquad \int\cdots\int_w \phi_1\phi_2^k\, p(E \mid \theta_1^o\ \theta_2\ w)dx_1\cdots dx_n = \lambda_k \quad (k = 0, 1, \ldots).$$

Let $p(\phi_1,\phi_2)$ and $p(\phi_1, \phi_2 \mid w)$ be the joint elementary probability distribution of ϕ_1 and ϕ_2, the first calculated from $p(E \mid \theta_1^o\ \theta_2)$ and the second from $p(E \mid \theta_1^o\ \theta_2 w)$. The equality (50) is equivalent to the following

$$(51) \qquad \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \phi_1\phi_2^k\, p(\phi_1\phi_2 \mid w)d\phi_1 d\phi_2$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \phi_1\phi_2^k\, p(\phi_1\phi_2)d\phi_1 d\phi_2 \qquad (k = 0, 1, 2, \ldots).$$

and it follows from condition 5 that, for almost all values of ϕ_2, we have

$$(52) \qquad \int_{-\infty}^{+\infty} \phi_1\, p(\phi_1,\phi_2 \mid w)d\phi_1 = \int_{-\infty}^{+\infty} \phi_1\, p(\phi_1,\phi_2)d\phi_1.$$

Then, in order that the region w satisfy conditions (32) and (33) for all values of θ_2, it is necessary for equations (44) and (52) to hold for almost all values of ϕ_2. Inversely it is easy to see that, when conditions (44) and (52) are fulfilled for almost all values of ϕ_2, all derivatives with respect to θ_2 of the left members of (32) and (33) will vanish, from which it is inferred that these integrals hold for all values of θ_2. Hence, the equalities (44) and (52) form the necessary and sufficient conditions that satisfy equations (32) and (33), whatever the value of θ_2.

This holds when condition 5 is fulfilled. If it does not, equalities (44) and (52) are not sufficient conditions.

In all cases, whether condition 5 is fulfilled or not, the necessary and sufficient conditions that w satisfies (32) and (33), whatever θ_2, are the equalities (41) and (51).

In order to use conditions (44) and (52), transform the sample space W point by point to some other W' by introducing a new system of random variables, $\Phi_1, \Phi_2, Y_3, Y_4, \ldots, Y_n$, whose special values are denoted by $\phi_1, \phi_2, y_3, \ldots, y_n$, where

(53) $$\phi_i = \frac{\delta}{\delta\theta_i} \log p(E \mid \theta_1^o\ \theta_2) \qquad (i = 1,2),$$

and the other variables may be chosen arbitrarily provided that the Jacobean

(54) $$\triangle = \frac{\delta(x_1, x_2, \ldots, x_n)}{\delta(\phi_1, \phi_2, y_3, \ldots, y_n)}$$

esists, is different from zero almost everywhere and does not change its sign in W'. Considering condition 4 this is possible.

Designate E' as a point in the space W' and $p(E'\ \ \theta_1\ \theta_2)$ as the elementary probability distribution of the new variables.

(55) $$p(E \mid \theta_1\ \ \theta_2) = P(E \mid \theta_1\ \ \theta_2) \mid \triangle \mid,$$

where the right member ought to be represented as a function of $\phi_1, \phi_2, y_3, \ldots, y_n$.

If w is any region in W, then w' designates a similar region in W'. A particular w'_0 designates the image of the region w_0 satisfying (32) and (33) whatever the value of θ_2.

Consider within W' the plane ϕ = constant and designate by $W(\phi_2)$

and $w(\phi_2)$ the part of the plane included in W' and w'_0 respectively.

We see that the equality (44) is equivalent to the following:

$$(56) \qquad \int \cdots \int_{w(\phi_2)} p(E' \mid \theta_1^0 \theta_2) d\phi_1 dy_3 \cdots dy_n$$
$$= \alpha \int \cdots \int_{w(\phi_2)} p(E' \mid \theta_1^0 \theta_2) d\phi_1 dy_3 \cdots dy_n.$$

Likewise, formula (51) may be written

$$(57) \qquad \int \cdots \int_{w(\phi_2)} \phi_1 p(E' \mid \theta_1^0 \theta_2) d\phi_1 dy_3 \cdots dy_n$$
$$= \alpha \int \cdots \int_{w(\phi_2)} \phi_1 p(E' \mid \theta_1^0 \theta_2) d\phi_1 dy_3 \cdots dy_n.$$

Let us write finally the integral representing the derivative $\frac{\delta^2}{\delta\theta_1^2} \beta(\theta_1^0 \theta_2 \mid w)$ which ought to be maximized for w_0.

$$(58) \qquad \frac{\delta^2}{\delta\theta_1^2} \beta(\theta_1^0 \theta_2 \mid w_0) = \int \cdots \int_{w_0} (\phi_{11} + \phi_1^2) p(E \mid \theta_1^0 \theta_2) dx_1 \cdots dx_n.$$

Considering the equalities (15), (32), (33) and (35) and introducing the new system of variables, we may write

$$(59) \qquad \frac{\delta^2}{\delta\theta_1^2} \beta(\theta_1^0 \theta_2 \mid w_0)$$
$$= \int_-^+ d\phi_2 \int \cdots \int_{w(\phi_2)} \phi_1^2 \, p(E' \mid \theta_1^0 \theta_2) d\phi_1 dy_3 \cdots dy_n + A_0 \alpha$$

It is evident that the choice of the region w_0 is equivalent to that for w'_0 and in choosing w'_0 it is sufficient to choose $w(\phi_2)$ for all values of ϕ_2. Equations (56) and (57) represent the conditions which must be fulfilled by $w(\phi_2)$ for almost all values of ϕ_2. In satisfying these conditions and by maximizing (59) we see that it is sufficient to find such $w(\phi_2)$ which satisfy (56) and (57) and which maximize

(60) $$\int \cdots \int_{w(\phi_2)} \phi_1^2 \, p(E' \mid \theta_1^o \, \theta_2) d\phi_1 dy_3 \cdots dy_n$$

for all values of ϕ_2. This problem may be solved by applying the general result [10], as follows:

Let

(61) $$F_0, \; F_1, \; \ldots, \; F_m$$

be a series of m + 1 definite and integrable functions in W and let

(62) $$\alpha_1, \; \alpha_2, \; \ldots, \; \alpha_m,$$

be m numbers such that there exists in W at least one region A such that

(63) $$\int \cdots \int_A F_i \, dx_1 \cdots dx_n = \alpha_i \qquad (i = 1, 2, \ldots, m).$$

Then the region A_0 which satisfies condition (63) and which maximizes the integral

(64) $$\int \cdots \int_A F_0 \, dx_1 \cdots dx_n$$

is such that inside A_0

(65) $$F_0 \geqq \sum_{i=1}^{m} a_i F_i,$$

and outside A_0

(66) $$F_0 \leqq \sum_{i=1}^{m} a_i F_i,$$

where the coefficients a_i are constants which we must choose in order that the region A_0 satisfies conditions (63).

Upon applying this general result to the problem of finding the maximum of (60) while fulfilling conditions (56) and (57), we find immediately that for almost all values of ϕ_2 where $p(\phi_2) > 0$, at the interior of $w(\phi_2)$, there should be

(67) $$\phi_1^2 \geqq a_0 + a_1 \phi_1,$$

where a_0 and a_1 are independent of ϕ_2 and must be fixed to satisfy (56) and (57).

Designate by $k_1(\phi_2)$ and $k_2(\phi_2) > k_1(\phi_2)$ the roots of the equation

$$a_0 + a_1\phi_1 - \phi_1^2 = 0. \tag{68}$$

Now the inequality (67) is equivalent to the following two conditions:

$$\phi_1 \leqq k_1(\phi_2) \quad \text{and} \quad k_2(\phi_2) \leqq \phi_1, \tag{69}$$

and instead of finding a_0 and a_1 such that they satisfy (56) and (57), we can determine $k_1(\phi_2)$ and $k_2(\phi_2)$. Since the inequalities (69) which determine $w(\phi_2)$ present no limitation concerning the variables $y_3, y_4, \dots y_n$, we may carry out the integration in (56) and (57) for these variables in the extreme limit and we find

$$\int_{-\infty}^{k_1(\phi_2)} + \int_{k_2(\phi_2)}^{+\infty} p(\phi_1, \phi_2)d\phi_1 = \alpha \int_{-\infty}^{+\infty} p(\phi_1,\phi_2)d\phi_1 = \alpha p(\phi_2) \tag{70}$$

and

$$\int_{-\infty}^{k_2(\phi_2)} + \int_{k_2(\phi_2)}^{+\infty} \phi_1 p(\phi_1,\phi_2)d\phi_1 = \alpha \int_{-\infty}^{+\infty} \phi_1 p(\phi_1,\phi_2)d\phi_1, \tag{71}$$

these two equations may be reduced immediately to (30) and (31) in the statement of the problem, which consequently is solved.

7. Miscellaneous remarks.--Let us note that the hypothesis that ϕ_{11} is of the form (15) and serves only to simplify the form of the solution of the problem. The argument may be applied if this condition is not fulfilled. Let's consider quickly the possibilities which present themselves.

a. ϕ_{11} is a non linear function of ϕ_1 and ϕ_2 which does not depend explicitly on $x_1, x_2, \dots, x_n$.

In this case all the arguments may be repeated and we find that $w(\phi_2)$ is defined for the inequality

$$\phi_{11} + \phi_1^2 \geqq a_0 + a_1\phi_1, \tag{72}$$

where a_0 and a_1 must be determined so as to satisfy (56) and (57), which are reduced to a form analogous to (70) and (71).

b. There exists a system of three indices i, j and k such that

$\dfrac{\delta(\phi_{11}, \phi_1, \phi_2)}{\delta(x_i, x_j, x_k)}$ does not vanish identically. Therefore ϕ_{11} may be introduced as a variable of the system which transforms W into W' for example in place of y_3. By modifying the calculation slightly, we find that $w(\phi_2)$ is determined by the inequality

$$\phi_{11} \geq a_0 + a_1\phi_1 - \phi_1^2 = q(\phi_1, \phi_2), \tag{73}$$

where a_0 and a_1 depend only on ϕ_2 and must be determined so as to satisfy (56) and (57). As to these two equations, they take the form

$$\int_{-\infty}^{+\infty} d\phi_1 \int_{q(\phi_1\phi_2)}^{+\infty} p(\phi_{11},\phi_1,\phi_2)d\phi_{11} = \alpha p(\phi_2) \tag{74}$$

and

$$\int_{-\infty}^{+\infty} \phi_1 d\phi_1 \int_{q(\phi_1,\phi_2)}^{+\infty} p(\phi_{11},\phi_1,\phi_2)d\phi_{11} = \alpha \int_{-\infty}^{+\infty} \phi_1 p(\phi_1,\phi_2)d\phi_1 \tag{75}$$

respectively.

Let's observe finally that the argument does not necessarily require condition 5 for the system of orthogonal polynomials. What is necessary are conclusions (a) and (b) which result from this hypothesis. And even these conclusions need not be true with respect to the probability distribution of ϕ_2 relative to every region w and to the regression function of ϕ_1 on ϕ_2. Important examples are found where hypothesis 5 is not satisfied but where the solution in the form (69) applies because it is possible to show that equations (41) and (51) lead to (44) and (52).

We still must make a remark on the independence of the solution obtained for the value of θ_2.

It is not evident that the inequalities (29) where the limits $k_1(\phi_2)$ and $k_2(\phi_2)$ are determined by (30) and (31) do not depend on the value of θ_2 which is not fixed by the hypothesis H_0 to be tested. If it happened that $k_1(\phi_2)$ and $k_2(\phi_2)$ depend on the value of θ_2, then, strictly speaking, the problem of the unbiased critical region of type B would have no solution.

Now it is easy to show that, thanks to condition (17), the solution given by (30) and (31) does not depend on the value of θ_2. Consider $\theta_1 = \theta_1^0$ as a constant. Condition (17) represents a linear differential equation of first order. Upon integration we obtain

$$\phi_2(\theta_2) = Q(\theta_2) + R(\theta_2)f_1(E), \tag{76}$$

where $Q(\theta_2)$ and $R(\theta_2)$ are not dependent on the sample point E and $f_1(E)$ is independent of θ_2. Hence the distribution $p(E| \theta_1 \theta_2)$ is of the form

$$(77) \qquad p(E \mid \theta_1^o, \theta_2) = C(\theta_2)e^{S(\theta_2)f_1(E)+f_2(E)},$$

where $C(\theta_2)$ and $S(\theta_2)$ do not depend on the sample point E and $f_2(E)$ is independent of θ_2. (We see from (77) that the function $f_1(E)$ is a sufficient statistic [21]).

We see that the locus of the points where $\phi_2(\theta_2)$ = constant = ϕ_2 is the same as where $f_1(E)$ = constant = f_1. The latter is independent of θ_2 in this sense that if we change the value of θ'_2 to θ''_2, we can find a suitable value for ϕ'_2, say ϕ''_2 such that the point $\phi_2(\theta''_2) = \phi''$ be identical with $\phi_2(\theta') = \phi'$. Conditions (70) and (71) which we may interpret in the sample space W apply to surfaces with $f_1(E)$ constant. Consider such a surface. We see easily that if we calculate the integral in (70) and (71) by starting with (77) then we would have on both sides the factor

$$C(\theta_2)e^{S(\theta_2)f_1(E)}$$

which cancels out. Hence if all the possible reductions are made, the inequalities (30) and (31) concerning the locus of the points where $f_1(E)$ has a fixed value, do not depend on the value of θ_2, which proves that the region w_0 determined by (29), (30) and (31) which satisfies conditions (32) and (33), whatever the value of θ_2, gives the maximum value for the second derivative (58) and that whatever the value of θ_2.

One might perhaps think that in general the power function $\beta(\theta_1 \theta_2 | w_0)$ is independent of θ_2. But it follows from reasoning and from numerical results concerning the probability for errors of the second kind published elsewhere [19, in particular pages 127-136] this assumption is not true.

LITERATURE CITED

1. T. Bayes, An essay towards solving a problem in the dectrine of chance (Phil. Trans. Roy. Soc., Vol. 53, 1763).

2. Karl Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling (Phil. Mag., 5th series, Vol. L, 1900).

3. J. Bertrand, Calcul des probabilités, Paris, 1889.

4. E. Borel, Le hasard, Paris, 1932.

5. J. Neyman and E.-S. Pearson, On the use and interpretation.... (Biometrika, Vol. XX-A, 1928).

6. J. Neyman and E.-S. Pearson, On the problem of two samples (Bull. Ac. Polonaise Sci. Lettres, series A, 1930).

7. J. Neyman and E.-S. Pearson, On the problem of K samples (Ibid., 1931).

8. J. Neyman and E.-S. Pearson, On the problem of the most efficient tests of statistical hypotheses (Phil. Trans. Roy. Soc., Vol. 231-A, 1933).

9. J. Neyman and E.-S. Pearson, The testing of statistical hypotheses in relation to probabilities a priori (Cambridge Phil. Soc. Proc., Vol. XXIX, 1933).

10. J. Neyman and E.-S. Pearson, Further contributions to the theory of testing statistical hypotheses (Annals of the Department of Applied Statistics, University College, London, Vol. 1, 1936) (in press).

11. J. Neyman and E.-S. Pearson, Sufficient statistics and uniformly most powerful tests of statistical hypotheses (Ibidem) (in press).

12. J. Neyman, Méthodes nouvelles de vérification des hypothèses statistiques (C. R. Premier Congres Math. Pays Slaves, Varsovie, 1929).

13. E.-S. Pearson, Some notes on sampling tests with two variables (Biometrika, Vol. XXI, 1929).

14. E.-S. Pearson and S.-S. Wilks, Methods of statistical analysis appropriate to K samples of two variables (Biometrika, Vol. XXV, 1933).

15. S. Kolodziejczyk, La vérification de l'hypothèse sur la constance des probabilités (Ann. Soc. Polonaise Math., t. IX, 1930).

16. S. Kolodziejczyk, On an important class of statistical hypotheses (Biometrika, Vol. XXVII, 1935).

17. B.-L. Welch, Some problems in the analysis of regression among K samples of two variables (Biometrika, Vol. XXVII, 1935).

18. S.-S. Wilks, On the independence of K sets of normally distributed statistical variables (Econometrica, Vol. 3, 1935).

19. J. Neyman (with cooperation of K. Iwaszkiewicz and S. Kolodziejczyk), Statistical problems in agricultural experimentation (Supplement to the J. R. S. S., Vol. II, 1935).

20. P. Lévy, Calcul des probabilités, Paris, 1925.

21. G. Darmois, Sur les lois de probabilité à estimation exhaustive C. R., t. 200, 1935).