IX. *On the Problem of the most Efficient Tests of Statistical Hypotheses.*

By J. NEYMAN, *Nencki Institute, Soc. Sci. Lit. Varsoviensis, and Lecturer at the Central College of Agriculture, Warsaw, and* E. S. PEARSON, *Department of Applied Statistics, University College, London.*

(*Communicated by* K. PEARSON, *F.R.S.*)

(Received August 31, 1932.—Read November 10, 1932.)

CONTENTS.

INTRODUCTORY.

The problem of testing statistical hypotheses is an old one. Its origin is usually connected with the name of THOMAS BAYES, who gave the well-known theorem on the probabilities *a posteriori* of the possible " causes " of a given event.* Since then it has been discussed by many writers of whom we shall here mention two only, BERTRAND† and BOREL,‡ whose differing views serve well to illustrate the point from which we shall approach the subject.

BERTRAND put into statistical form a variety of hypotheses, as for example the hypothesis that a given group of stars with relatively small angular distances between them as seen from the earth, form a " system " or group in space. His method of attack, which is that in common use, consisted essentially in calculating the probability, P, that a certain character, x, of the observed facts would arise if the hypothesis tested were true. If P were very small, this would generally be considered as an indication that the hypothesis, H, was probably false, and *vice versa*. BERTRAND expressed the pessimistic view that no test of this kind could give reliable results.

BOREL, however, in a later discussion, considered that the method described could be applied with success provided that the character, x, of the observed facts were properly chosen—were, in fact, a character which he terms " en quelque sorte remarquable."

We appear to find disagreement here, but are inclined to think that, as is so often the case, the difference arises because the two writers are not really considering precisely the same problem. In general terms the problem is this : Is it possible that there are any efficient tests of hypotheses based upon the theory of probability, and if so, what is their nature ? Before trying to answer this question, we must attempt to get closer to its exact meaning. In the first place, it is evident that the hypotheses to be tested by means of the theory of probability must concern in some way the probabilities of the different kinds of results of certain trials. That is to say, they must be of a statistical nature, or as we shall say later on, they must be statistical hypotheses.

Now what is the precise meaning of the words " an efficient test of a hypothesis ? " There may be several meanings. For example, we may consider some specified hypothesis, as that concerning the group of stars, and look for a method which we should hope to tell us, with regard to a particular group of stars, whether they form a system, or are grouped " by chance," their mutual distances apart being enormous and their relative movements unrelated.

If this were what is required of " an efficient test," we should agree with BERTRAND in his pessimistic view. For however small be the probability that a particular grouping of a number of stars is due to " chance," does this in itself provide any evidence of another " cause " for this grouping but " chance ? " " Comment définir, d'ailleurs, la singularité dont on juge le hasard incapable ? "§ Indeed, if x is a continuous variable—as for example is the angular distance between two stars—then any value of x is a singularity of relative probability equal to zero. We are inclined to think that

* 'Phil. Trans.,' vol. 53, p. 370 (1763).

† " Calcul des Probabilités," Paris (1907).

‡ 'Le Hasard,' Paris (1920).

§ BERTRAND, *loc. cit.*, p. 165.

as far as a particular hypothesis is concerned, no test based upon the theory of probability* can by itself provide any valuable evidence of the truth or falsehood of that hypothesis.

But we may look at the purpose of tests from another view-point. Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong. Here, for example, would be such a "rule of behaviour": to decide whether a hypothesis, H, of a given type be rejected or not, calculate a specified character, x, of the observed facts; if $x > x_0$ reject H, if $x \leq x_0$ accept H. Such a rule tells us nothing as to whether in a particular case H is true when $x \leq x_0$ or false when $x > x_0$. But it may often be proved that if we behave according to such a rule, then in the long run we shall reject H when it is true not more, say, than once in a hundred times, and in addition we may have evidence that we shall reject H sufficiently often when it is false.

If we accept the words "an efficient test of the hypothesis H" to mean simply such a rule of behaviour as just described, then we agree with BOREL that efficient tests are possible. We agree also that not any character, x, whatever is equally suitable to be a basis for an efficient test,† and the main purpose of the present paper is to find a general method of determining tests, which, from the above point of view would be the most efficient.

In common statistical practice, when the observed facts are described as "samples," and the hypotheses concern the "populations" from which the samples have been drawn, the characters of the samples, or as we shall term them criteria, which have been used for testing hypotheses, appear often to have been fixed by a happy intuition. They are generally functions of the moment coefficients of the sample, and as long as the variation among the observations is approximately represented by the normal frequency law, moments appear to be the most appropriate sample measures that we can use. But as FISHER‡ has pointed out in the closely allied problem of Estimation, the moments cease to be efficient measures when the variation departs widely from normality. Further, even though the moments are efficient, there is considerable choice in the particular function of these moments that is most appropriate to test a given hypothesis, and statistical literature is full of examples of confusion of thought on this choice.

A blind adoption of the rule,

* Cases will, of course, arise where the verdict of a test is based on certainty. The question "Is there a black ball in this bag?" may be answered with certainty if we find one in a sample from the bag.

† This point has been discussed in earlier papers. *See* (*a*) NEYMAN and PEARSON. 'Biometrika,' vol. 20A, pp. 175 and 263 (1928); (*b*) NEYMAN. 'C. R. Premier Congrès Math., Pays Slaves,' Warsaw, p. 355 (1929); (*c*) PEARSON and NEYMAN. 'Bull. Acad. Polonaise Sci. Lettres,' Série A, p. 73 (1930).

‡ 'Phil. Trans.,' vol. 222, A, p. 326 (1921).

Standard Error of

$$(x - y) = \sqrt{(\text{Standard Error of } x)^2 + (\text{Standard Error of } y)^2}, \quad . \quad . \quad . \quad (1)$$

has lead to frequent inconsistencies. Consider, for example, the problem of testing the significance of a difference between two percentages or proportions; a sample of n_1 contains t_1 individuals with a given character and an independent sample of n_2 contains t_2. Following the rule (1), the standard error of the difference $d = t_1/n_1 - t_2/n_2$ is often given as

$$\sigma_d = \sqrt{\frac{t_1}{n_1^2}\left(1 - \frac{t_1}{n_1}\right) + \frac{t_2}{n_2^2}\left(1 - \frac{t_2}{n_2}\right)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

But in using $\frac{t_1}{n_1^2}\left(1 - \frac{t_1}{n_1}\right)$ and $\frac{t_2}{n_2^2}\left(1 - \frac{t_2}{n_2}\right)$ as the squares of the estimates of the two standard errors, we are proceeding on the supposition that sample estimates must be made of two different population proportions p_1 and p_2. Actually, it is desired to test the hypothesis that $p_1 = p_2 = p$, and it follows that the best estimate of p is obtained by combining together the two samples, whence we obtain

Estimate of

$$\sigma_d = \sqrt{\frac{t_1 + t_2}{n_1 + n_2}\left(1 - \frac{t_1 + t_2}{n_1 + n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

A rather similar situation arises in the case of the standard error of the difference between two means. In the case of large samples there are two forms of estimate—*

$$\sigma_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$\sigma'_d = \sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The use of the first form is justified if we believe that in the populations sampled there are different standard deviations, σ_1 and σ_2; while if $\sigma_1 = \sigma_2$ the second form should be taken. The hypothesis concerning the two means has not, in fact, always to be tested under the same conditions, and which form of the criterion is most appropriate is a matter for judgment based upon the evidence available regarding those conditions.

The role of sound judgment in statistical analysis is of great importance and in a large number of problems common sense may alone suffice to determine the appropriate method of attack. But there is ample evidence to show that this has not and cannot always be enough, and that it is therefore essential that the ideas involved in the process of testing hypotheses should be more clearly understood.

* s_1 and s_2 are the observed standard deviations in independent samples of n_1 and n_2. The expression (5) is the limiting form of $\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ when n_1 and n_2 are large, based upon the common estimate of the unknown population variance σ^2.

In earlier papers we have suggested that the criterion appropriate for testing a given hypothesis could be obtained by applying the principle of likelihood.* This principle was put forward on intuitive grounds after the consideration of a variety of simple cases. It was subsequently found to link together a great number of statistical tests already in use, besides suggesting certain new methods of attack. It was clear, however, in using it that we were still handling a tool not fully understood, and it is the purpose of the present investigation to widen, and we believe simplify, certain of the conceptions previously introduced.

We shall proceed to give a review of some of the more important aspects of the subject as a preliminary to a more formal treatment.

II.—Outline of General Theory.

Suppose that the nature of an event, E, is exactly described by the values of n variates,

$$x_1, x_2, \dots x_n. \qquad (6)$$

For example, the series (6) may represent the value of a certain character observed in a sample of n individuals drawn at random from a given population. Or again, the x's may be the proportions or frequencies of individuals in a random sample falling into n out of the $n+1$ categories into which the sampled population is divided. In any case, the event, E, may be represented by a point in a space of n dimensions having (6) as its co-ordinates; such a point might be termed an Event Point, but we shall here speak of it in statistical terms as a Sample Point, and the space in which it lies as the Sample Space. Suppose now that there exists a certain hypothesis, H_0, concerning the origin of the event which is such as to determine the probability of occurrence of every possible event E. Let

$$p_0 = p_0 (x_1, x_2, \dots x_n) \qquad (7)$$

be this probability—or if the sample point can vary continuously, the elementary probability of such a point. To obtain the probability that the event will give a sample point falling into a particular region, say w, we shall have either to take the sum of (7) over all sample points included in w, or to calculate the integral

$$P_0 (w) = \int \dots \int_w p_0 (x_1, x_2, \dots x_n)\, dx_1\, dx_2 \dots dx_n. \qquad (8)$$

The two cases are quite analogous as far as the general argument is concerned, and we shall consider only the latter. That is to say, we shall assume that the sample points may fall anywhere within a continuous sample space (which may be limited or not), which we shall denote by W. It will follow that

$$P_0 (W) = 1. \qquad (9)$$

* 'Biometrika,' vol. 20A.

We shall be concerned with two types of hypotheses, (*a*) simple and (*b*) composite. The hypothesis that an event E has occurred subject to a completely specified probability law $p_0(x_1, x_2, \ldots x_n)$ is a simple one; while if the functional form of p_0 is given, though it depends upon the value of c unspecified parameters, H_0 will be called a composite hypothesis with c degrees of freedom.* The distinction may be illustrated in the case where H_0 concerns the population Π from which a sample, Σ, has been drawn at random. For example, the normal frequency law,

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad \ldots \ldots \ldots \ldots \quad (10)$$

represents an infinite set of population distributions. A simple hypothesis is that Σ has been sampled from a definite member of this set for which $a = a_0$, $\sigma = \sigma_0$. A composite hypothesis with one degree of freedom is that the sampled population, Π, is one of the sub-set for which $a = a_0$ but for which σ may have any value whatever. "STUDENT'S" original problem consisted in testing this composite hypothesis.†

The practice of using observational data to test a composite hypothesis is a familiar one. We ask whether the variation in a certain character may be considered as following the normal law; whether two samples are likely to have come from a common population; whether regression is linear; whether the variance in a number of samples differs significantly. In these cases we are not concerned with the exact value of particular parameters, but seek for information regarding the conditions and factors controlling the events.

It is clear that besides H_0 in which we are particularly interested, there will exist certain admissible alternative hypotheses. Denote by Ω the set of all simple hypotheses, which in a particular problem we consider as admissible. If H_0 is a simple hypothesis, it will clearly belong to Ω. If H_0 is a composite hypothesis, then it will be possible to specify a part of the set Ω, say ω, such that every simple hypothesis belonging to the sub-set ω will be a particular case of the composite hypothesis H_0. We could say also that the simple hypotheses belonging to the sub-set ω, may be obtained from H_0 by means of some additional conditions specifying the parameters of the function (7) which are not specified by the hypothesis H_0.

In many statistical problems the hypotheses concern different populations from which the sample, Σ, may have been drawn. Therefore, instead of speaking of the sets Ω or ω of simple hypotheses, it will be sometimes convenient to speak of the sets Ω or ω of populations. A composite hypothesis, H_0, will then refer to populations belonging to the sub-set ω, of the set Ω. Every test of a statistical hypothesis in the sense described above, consists in a rule of rejecting the hypothesis when a specified

* The idea of degrees of freedom as defined above, though clearly analogous, is not to be confused with that introduced by FISHER.

† 'Biometrika,' vol. 6, p. 1 (1908).

character, x, of the sample lies within certain critical limits, and accepting it or remaining in doubt in all other cases. In the n-dimensional sample space, W, the critical limits for x will correspond to a certain critical region w, and when the sample point falls within this region we reject the hypothesis. If there are two alternative tests for the same hypothesis, the difference between them consists in the difference in critical regions.

We can now state briefly how the criterion of likelihood is obtained. Take any sample point, Σ, with co-ordinates $(x_1, x_2, \ldots x_n)$ and consider the set A_Σ of probabilities $p_H(x_1, x_2, \ldots x_n)$ corresponding to this sample point and determined by different simple hypotheses belonging to Ω. We shall suppose that whatever be the sample point the set A_Σ is bounded. Denote by $p_\Omega(x_1, x_2, \ldots x_n)$ the upper bound of the set A_Σ, then if H_0 is a simple hypothesis, determining the elementary probability $p_0(x_1, x_2, \ldots x_n)$, we have defined its likelihood to be

$$\lambda = \frac{p_0(x_1, x_2, \ldots x_n)}{p_\Omega(x_1, x_2, \ldots x_n)}. \qquad (11)$$

If H_0 is a composite hypothesis, denote by $A_\Sigma(\omega)$ the sub-set of A_Σ corresponding to the set ω of simple hypotheses belonging to H_0 and by $p_\omega(x_1, x_2, \ldots x_n)$ the upper bound of $A_\Sigma(\omega)$. The likelihood of the composite hypothesis is then

$$\lambda = \frac{p_\omega(x_1, x_2, \ldots x_n)}{p_\Omega(x_1, x_2, \ldots x_n)}. \qquad (12)$$

In most cases met with in practice, the elementary probabilities, corresponding to different simple hypotheses of the set Ω are continuous and differentiable functions

$$p(\alpha_1, \alpha_2, \ldots \alpha_c, \alpha_{c+1}, \ldots \alpha_k;\ x_1, x_2, \ldots x_n), \qquad (13)$$

of the certain number, k, of parameters $\alpha_1, \alpha_2, \ldots \alpha_c, \alpha_{c+1}, \ldots \alpha_k$; and each simple hypothesis specifies the values of these parameters. Under these conditions the upper bound, $p_\Omega(x_1, x_2, \ldots x_n)$, is often a maximum of (13) (for fixed values of the x's), with regard to all possible systems of the α's. If H_0 is a composite hypothesis with c degrees of freedom, it specifies the values of $k - c$ parameters, say $\alpha_{c+1}, \alpha_{c+2}, \ldots \alpha_k$ and leaves the others unspecified. Then $p_\omega(x_1, x_2, \ldots x_n)$ is often a maximum of (13) (for fixed values of the x's and of $\alpha_{c+1}, \alpha_{c+2}, \ldots \alpha_k$) with regard to all possible values of $\alpha_1, \alpha_2, \ldots \alpha_c$.

The use of the principle of likelihood in testing hypotheses, consists in accepting for critical regions those determined by the inequality

$$\lambda \leq C = \text{const.} \qquad (14)$$

Let us now for a moment consider the form in which judgments are made in practical experience. We may accept or we may reject a hypothesis with varying degrees of confidence ; or we may decide to remain in doubt. But whatever conclusion is reached the following position must be recognised. If we reject H_0, we may reject it when it is true ; if we accept H_0, we may be accepting it when it is false, that is to say, when really some alternative H_t is true. These two sources of error can rarely be eliminated completely ; in some cases it will be more important to avoid the first, in others the second. We are reminded of the old problem considered by LAPLACE of the number of votes in a court of judges that should be needed to convict a prisoner. Is it more serious to convict an innocent man or to acquit a guilty ? That will depend upon the consequences of the error ; is the punishment death or fine ; what is the danger to the community of released criminals ; what are the current ethical views on punishment ? From the point of view of mathematical theory all that we can do is to show how the risk of the errors may be controlled and minimised. The use of these statistical tools in any given case, in determining just how the balance should be struck, must be left to the investigator.

The principle upon which the choice of the critical region is determined so that the two sources of errors may be controlled is of first importance. Suppose for simplicity that the sample space is of two dimensions, so that the sample points lie on a plane. Suppose further that besides the hypothesis H_0 to be tested, there are only two alternatives H_1 and H_2. The situation is illustrated in fig. 1, where the cluster of spots round the point O, of circles round A_1, and of crosses round A_2 may be taken to

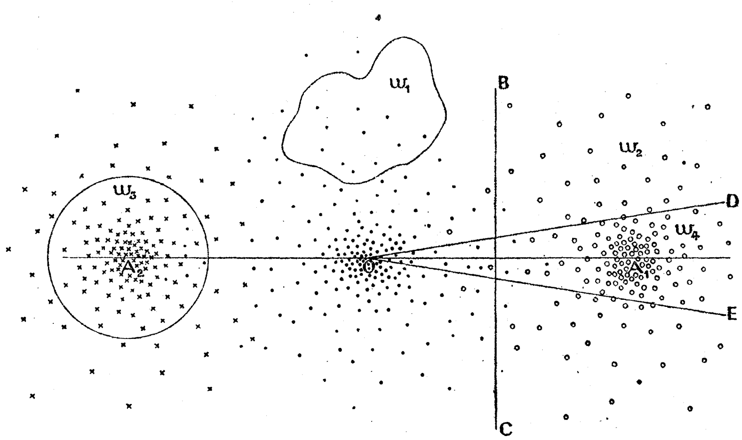

FIG. 1.

represent the probability or density field appropriate to the three hypotheses. The spots, circles and crosses in the figure suggest diagrammatically the behaviour of the functions $p_i(x_1, x_2)$, $i = 0, 1, 2$, in the sense that the number of spots included in any region w is proportional to the integral of $p_0(x_1, x_2)$ taken over this region, etc. Looking at the diagram we see that, if the process of sampling was repeated many times, then, were the hypothesis H_0 true, most sample points would lie somewhere near the point O. On the contrary, if H_1 or H_2 were true, the sample points would be close to O in comparatively rare cases only.

In trying to choose a proper critical region, we notice at once that it is very easy to control errors of the first kind referred to above. In fact, the chance of rejecting the hypothesis H_0 when it is true may be reduced to as low a level as we please. For if w is any region in the sample space, which we intend to use as a critical region, the chance, $P_0(w)$, of rejecting the hypothesis H_0 when it is true, is merely the chance determined by H_0 of having a sample point inside of w, and is equal either to the sum (when the sample points do not form a continuous region) or to the integral (when they do) of $p_0(x_1, x_2)$ taken over the region w. It may be easily made $\leq \varepsilon$, by chosing w sufficiently small.

Four possible regions are suggested on the figure; (1) w_1; (2) w_2, *i.e.*, the region to the right of the line BC; (3) w_3, the region within the circle centred at A_2; (4) w_4, the region between the straight lines OD, OE.

If the integrals of $p_0(x_1, x_2)$ over these regions, or the numbers of spots included in them, are equal, we know that they are all of equal value in regard to the first source of error; for as far as our judgment on the truth or falsehood of H_0 is concerned, if an error cannot be avoided it does not matter on which sample we make it.* It is the frequency of these errors that matters, and this—for errors of the first kind—is equal in all four cases.

It is when we turn to consider the second source of error—that of accepting H_0 when it is false—that we see the importance of distinguishing between different critical regions. If H_1 were the only admissible alternative to H_0, it is evident that we should choose from w_1, w_2, w_3 and w_4 that region in which the chance of a sample point falling, if H_1 were true, is greatest; that is to say the region in the diagram containing the greatest number of the small circles forming the cluster round A_1. This would be the region w_2, because for example,

$$P_1(w_2) > P_1(w_1) \text{ or } P_1(W - w_2) < P_1(W - w_1).$$

This we do since in accepting H_0 when the sample point lies in $(W - w_2)$, we shall be accepting it when H_1 is, in fact, true, less often than if we used w_1. *We need indeed to pick out from all possible regions for which* $P_0(w) = \varepsilon$, *that region*, w_0, *for which* $P_1(w_0)$ *is a maximum and* $P_1(W - w_0)$ *consequently a minimum*; this region (or regions if more than one satisfy the condition) we shall term the Best Critical Region for H_0 with regard to H_1. There will be a family of such regions, each member corresponding to a different value of ε. The conception is simple but fundamental.

It is clear that in the situation presented in the diagram the best critical region with regard to H_1 will not be the best critical region with regard to H_2. While the first may be w_2, the second may be w_3. But it will be shown below that in certain problems there is a common family of best critical regions for H_0 with regard to the whole class

* If the samples for which H_0 is accepted are to be used for some purpose and those for which it is rejected to be discarded, it is possible that other conceptions of relative value may be introduced. But the problem is then no longer the simple one of discriminating between hypotheses.

of admissible alternative hypotheses Ω.* In these problems we have found that the regions are also those given by using the principle of likelihood, although a general proof of this result has not so far been obtained, when H_0 is composite.

In the problems where there is a *different* best critical region for H_0 with regard to each of the alternatives constituting Ω, some further principle must be introduced in fixing what may be termed a Good Critical Region with regard to the set Ω. We have found here that the region picked out by the likelihood method is the envelope of the best critical regions with regard to the individual hypotheses of the set. This region appears to satisfy our intuitive requirements for a good critical region, but we are not clear that it has the unique status of the common best critical region of the former case.

We have referred in an earlier section to the distinction between simple and composite hypotheses, and it will be shown that the best critical regions may be found in both cases, although in the latter case they must satisfy certain additional conditions. If, for example in fig. 1, H_0 were a composite hypothesis with one degree of freedom such that while the centre of the cluster of spots were fixed at O, the scale or measure of radial expansion were unspecified, it is clear that w_4 could be used as a critical region, since $P_0(w_4) = \varepsilon$ would remain constant for any radial expansion or contraction of the field. Neither w_1, w_2 nor w_3 satisfy this condition. "STUDENT'S" test is a case in which a hyperconical region of this type is used.

III.—SIMPLE HYPOTHESES.

(*a*) *General Theory.*

We shall now consider how to find the best critical region for H_0 with regard to a single alternative H_1; this will be the region w_0 for which $P_1(w_0)$ is a maximum subject to the condition that

$$P_0(w_0) = \varepsilon. \qquad (15)$$

We shall suppose that the probability laws for H_0 and H_1, namely, $p_0(x_1, x_2, \ldots x_n)$ and $p_1(x_1, x_2, \ldots x_n)$, exist, are continuous and not negative throughout the whole sample space W; further that

$$P_0(W) = P_1(W) = 1. \qquad (16)$$

Following the ordinary method of the Calculus of Variations, the problem will consist in finding an unconditioned minimum of the expression

$$P_0(w_0) - kP_1(w_0) = \int \ldots \int_{w_0} \{p_0(x_1, x_2, \ldots x_n) - kp_1(x_1, x_2, \ldots x_n)\}\, dx_1 \ldots dx_n, \qquad (17)$$

k being a constant afterwards to be determined by the condition (15). Suppose that the region w_0 has been determined and that S is the hypersurface limiting it. Let s_1

* Again as above, each member of the family is determined by a different value of ε.

be any portion of S such that every straight line parallel to the axis Ox_n cuts s_1 not more than once; denote by σ_1 the orthogonal projection of s_1 on the prime or hyperplane $x_n = 0$. Clearly σ_1 will be a region in the $(n - 1)$ dimensioned space of

$$x_1, x_2, \dots x_{n-1}, \quad \dots\dots \quad (18)$$

for which there will exist a uniform function

$$_{s_1}x_n = \psi(x_1, \dots x_{n-1}), \quad \dots\dots \quad (19)$$

whose value together with the corresponding values of (18) will determine the points on the hypersurface s_1.

Consider now a region $w_0(\alpha)$ bounded by

(1) The hypersurface s_2 with equation

$$_{s_2}x_n = \psi(x_1, x_2, \dots x_{n-1}) + \alpha\theta(x_1, x_2, \dots x_{n-1}), \quad \dots\dots \quad (20)$$

where α is a parameter independent of the x's, and θ any uniform and continuous function.

(2) By the hypercylinder projecting s_1 on to σ_1.

(3) By any hypersurface s_3 with equation,

$$_{s_3}x_n = \phi(x_1, \dots x_{n-1}), \quad \dots\dots \quad (21)$$

having the following properties :—

(*a*) It is not cut more than once by any straight line parallel to the axis Ox_n.

(*b*) It lies entirely inside the region w_0.

(*c*) Every point on a straight line parallel to Ox_n lying between the points of its intersection with s_1 and s_3 belongs to w_0.

For the case of $n = 2$, the situation can be pictured in two dimensions as shown in fig. 2. Here $w_0(\alpha)$ is the shaded region bounded by the curves s_2 and s_3, and the two vertical lines through A and B. σ_1 is the portion AB of the axis Ox_1. In general

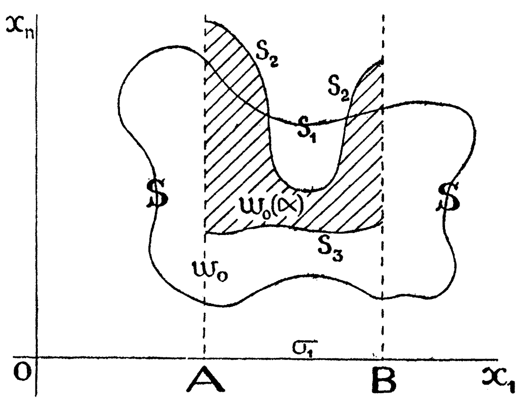

FIG. 2.

if the region w_0 is such as to give a minimum value to the expression (17), it follows that

$$\mathrm{I}(\alpha) = \mathrm{P}_0(w_0(\alpha)) - k\mathrm{P}_1(w_0(\alpha)) = \int \cdots \int_{\sigma_1} dx_1 \ldots dx_{n-1} \int_{s_1 x_n}^{s_2 x_n} (p_0 - kp_1)\, dx_n, \tag{22}$$

considered as a function of the varying parameter α must be a minimum for $\alpha = 0$. Hence differentiating

$$\frac{d\mathrm{I}}{d\alpha} = \pm \int \cdots \int_{\sigma_1} \theta \,\{p_0(x_1, \ldots x_{n-1}, {}_{s_2}x_n) - kp_1(x_1, x_2, \ldots x_{n-1}, {}_{s_2}x_n)\}\, dx_1 \ldots dx_{n-1} = 0, \tag{23}$$

whatever be the form of the function θ. This is known to be possible only if the expression within curled brackets on the right hand side of (23) is identically zero. It follows that if w_0 is the best critical region for H_0 with regard to H_1, then at every point on the hypersurface s_1 and consequently at every point on the complete boundary S, we must have

$$p_0(x_1, x_2, \ldots x_n) = kp_1(x_1, x_2, \ldots x_n), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{24}$$

k being a constant. This result gives the necessary boundary condition. We shall now show that the necessary and sufficient condition for a region w_0, being the best critical region for H_0 with regard to the alternative hypothesis, H_1, consists in the fulfilment of the inequality $p_0(x_1, x_2, \ldots x_n) > kp_1(x_1, x_2, \ldots x_n)$, k being a constant, at any point outside w_0; that is to say that w_0 is defined by the inequality

$$p_0(x_1, x_2, \ldots x_n) \leq kp_1(x_1, x_2, \ldots x_n). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{25}$$

Denote by w_0 the region defined by (25) and let w_1 be any other region satisfying the condition $\mathrm{P}_0(w_1) = \mathrm{P}_0(w_0) = \varepsilon$ (say). These regions may have a common part, w_{01}. The situation is represented diagrammatically in fig. 3.

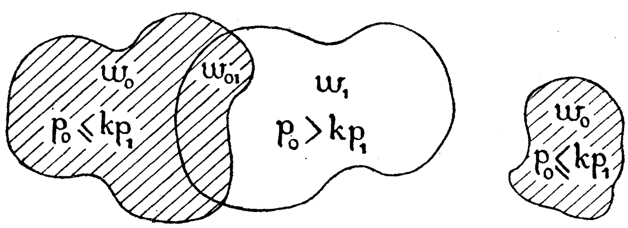

FIG. 3.

It will follow that,

$$\mathrm{P}_0(w_0 - w_{01}) = \varepsilon - \mathrm{P}_0(w_{01}) = \mathrm{P}_0(w_1 - w_{01}), \quad . \quad . \quad . \quad . \quad . \quad . \tag{26}$$

and consequently

$$k\mathrm{P}_1(w_0 - w_{01}) \geq \mathrm{P}_0(w_0 - w_{01}) = \mathrm{P}_0(w_1 - w_{01}) \geq k\mathrm{P}_1(w_1 - w_{01}). \quad . \tag{27}$$

If we add $k\mathrm{P}_1(w_{01})$ to both sides of the inequality, we obtain

$$k\mathrm{P}_1(w_0) \geq k\mathrm{P}_1(w_1)$$

or

$$\mathrm{P}_1(w_0) \geq \mathrm{P}_1(w_1). \quad \ldots \ldots \ldots \ldots \quad (28)$$

From the considerations advanced above it follows that w_1 is less satisfactory as a critical region than w_0. That is to say, of the regions w, for which $\mathrm{P}_0(w) = \varepsilon$, satisfying the boundary condition (24), the region w_0 defined by the inequality (25) is the best critical region with regard to the alternative H_1. There will be a family of such best critical regions, each member of which corresponds to a different value of ε.

As will appear below when discussing illustrative examples, in certain cases the family of best critical regions is not the same for each of the admissible alternatives $\mathrm{H}_1, \mathrm{H}_2, \ldots$; while in other cases a single common family exists for the whole set of alternatives. In the latter event the basis of the test is remarkably simple. If we reject H_0 when the sample point, Σ, falls into w_0, the chance of rejecting it when it is true is ε, and the risk involved can be controlled by choosing from the family of best critical regions to which w_0 belongs, a region for which ε is as small as we please. On the other hand, if we accept H_0 when Σ falls outside w_0, we shall sometimes be doing this when some H_t of the set of alternatives is really true. But we know that whatever be H_t, the region w_0 has been so chosen as to reduce this risk to a minimum. In this case even if we had precise information as to the *a priori* probabilities of the alternatives $\mathrm{H}_1, \mathrm{H}_2, \ldots$ we could not obtain any improved test.*

It is now possible to see the relation between best critical regions and the region defined by the principle of likelihood described above. Suppose that for a hypothesis H_t belonging to the set of alternatives Ω, the probability law for a given sample is defined by

(1) an expression of given functional type $p(x_1, x_2, \ldots x_n)$

(2) the values of c parameters contained in this expression, say

$$\alpha_t^{(1)}, \alpha_t^{(2)}, \ldots \alpha_t^{(c)}. \quad \ldots \ldots \ldots \ldots \quad (29)$$

This law for H_t may be written as $p_t = p_t(x_1, x_2, \ldots x_n)$. The hypothesis of maximum likelihood, $\mathrm{H}(\Omega \text{ max.})$, is obtained by maximising p_t with regard to these c parameters, or in fact from a solution of the equations,

$$\frac{\partial p}{\partial \alpha^{(i)}} = 0, \qquad i = 1, 2, \ldots c. \quad \ldots \ldots \ldots \ldots \quad (30)$$

The values of the α's so obtained are then substituted into p to give $p(\Omega \text{ max.})$. Then

* For properties of critical regions given by the principle of likelihood from the point of view of probabilities *a posteriori*, see NEYMAN, " Contribution to the Theory of Certain Test Criteria," 'Bull. Inst. int. Statist.,' vol. 24, pp. 44 (1928).

the family of surfaces of constant likelihood, λ, appropriate for testing a simple hypothesis H_0 is defined by

$$p_0 = \lambda p\,(\Omega \text{ max.}). \qquad (31)$$

It will be seen that the members of this family are identical with the envelopes of the family

$$p_0 = kp_t, \qquad (32)$$

which bound the best critical regions. From this it follows that,

(*a*) If for a given ε a common best critical region exists with regard to the whole set of alternatives, it will correspond to its envelope with regard to these alternatives, and it will therefore be identical with a region bounded by a surface (31). Further, in this case, the region in which $\lambda \leq \lambda_0 =$ const. will correspond to the region in which $p_0 \leq \lambda_0 p_t$. The test based upon the principle of likelihood leads, in fact, to the use of best critical regions.

(*b*) If there is not a common best critical region, the likelihood of H_0 with regard to a particular alternative H_t will equal the constant, k, of equation (32). It follows that the surface (31) upon which the likelihood of H_0 with regard to the whole set of alternatives is constant, will be the envelope of (32) for which $\lambda = k$. The interpretation of this result will be seen more clearly in some of the examples which follow.

(*b*) *Illustrative Examples.*

(1) *Sample Space Unlimited; Case of the Normal Population.—Example* (1). Suppose that it is known that a sample of n individuals, $x_1, x_2, \ldots x_n$ has been drawn randomly from *some* normally distributed population with standard deviation $\sigma = \sigma_0$, but it is desired to test the hypothesis H_0 that the mean in the sampled population is $a = a_0$. Then the admissible hypotheses concern the set of populations for which

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad (33)$$

the mean, a, being unspecified, but σ always equal to σ_0. Let H_1 relate to the member of this set for which $a = a_1$. Let $\bar{x}$ and s be the mean and standard deviation of the sample. The probabilities of its occurrence determined by H_0 and by H_1 will then be

$$p_0(x_1, \ldots x_n) = \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n e^{-n\frac{(\bar{x}-a_0)^2+s^2}{2\sigma_0^2}} \qquad (34)$$

$$p_1(x_1, \ldots x_n) = \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n e^{-n\frac{(\bar{x}-a_1)^2+s^2}{2\sigma_0^2}}, \qquad (35)$$

and the equation (24) becomes

$$\frac{p_0}{p_1} = e^{-n\frac{(\bar{x}-a_0)^2-(\bar{x}-a_1)^2}{2\sigma_0^2}} = k. \qquad (36)$$

From this it follows that the best critical region for H_0 with regard to H_1, defined by the inequality (25), becomes

$$(a_0 - a_1)\,\bar{x} \leq \tfrac{1}{2}(a_0^2 - a_1^2) + \frac{\sigma_0^2}{n}\log k = (a_0 - a_1)\,\bar{x}_0 \quad \text{(say)}. \quad . \quad . \quad (37)$$

Two cases will now arise,

(*a*) $a_1 < a_0$, then the region is defined by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i) \leq \bar{x}_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

(*b*) $a_1 > a_0$, then the region is defined by

$$\bar{x} \geq \bar{x}_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

We see that whatever be H_1 and a_1, the family of hypersurfaces corresponding to different values of k, bounding the best critical region, will be the same, namely,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i) = \bar{x}_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

These are primes in the n-dimensioned space lying at right angles to the line,

$$x_1 = x_2 = \ldots = x_n. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

If, however, the class of admissible alternatives includes both those for which $a < a_0$ and $a > a_0$, there will not be a single best critical region; for the first it will be defined by $\bar{x} \leq \bar{x}_0$ and for the second by $\bar{x} \geq \bar{x}_0$, where $\bar{x}_0$ is to be chosen so that $p_0(\bar{x} \lesssim \bar{x}_0) = \varepsilon$.* This situation will not present any difficulty in practice. Suppose $\bar{x} > a_0$ as in fig. 4. We deal first with the class of alternatives for which $a > a_0$. If $\varepsilon = 0{\cdot}05$; $\bar{x}_0 = a_0 + 1{\cdot}6449\,\sigma_0/\sqrt{n}$, and if $\bar{x} < \bar{x}_0$, we shall probably decide to accept the hypothesis H_0 as far as this class of alternatives is concerned. That being so, we shall certainly not reject H_0 in favour of the class for which $a < a_0$, for the risk of rejection when H_0 were true would be too great.

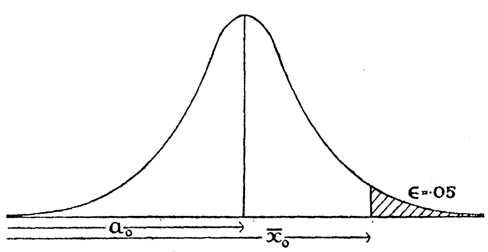

FIG. 4.

* In this example $P_0(\bar{x} \geq \bar{x}_0) = \frac{1}{\sigma_0}\sqrt{\frac{n}{2\pi}}\int_{-\infty}^{x_0} e^{-\frac{n(\bar{x}-a_0)^2}{2\sigma_0^2}}\,d\bar{x}$, $P_0(\bar{x} \geq \bar{x}_0) = \frac{1}{\sigma_0}\sqrt{\frac{n}{2\pi}}\int_{x_0}^{+\infty} e^{-\frac{n(\bar{x}-a_0)^2}{2\sigma_0^2}}\,d\bar{x}$.

The test obtained by finding the best critical region is in fact the ordinary test for the significance of a variation in the mean of a sample; but the method of approach helps to bring out clearly the relation of the two critical regions $\bar{x} \leq \bar{x}_0$ and $\bar{x} \geq \bar{x}_0$. Further, it has been established that starting from the same information, the test of this hypothesis could not be improved by using any other form of criterion or critical region.

Example (2).—The admissible hypotheses are as before given by (33), but in this case the means are known to have a given common value a_0, while σ is unspecified. We may suppose the origin to be taken at the common mean, so that $a = a_0 = 0$. H_0 is the hypothesis that $\sigma = \sigma_0$, and an alternative H_1 is that $\sigma = \sigma_1$. In this case it is easy to show that the best critical region with regard to H_1 is defined by the inequality,

$$\frac{1}{n} \sum_{i=1}^{n} (x_i^2) (\sigma_0^2 - \sigma_1^2) = (\bar{x}^2 + s^2) (\sigma_0^2 - \sigma_1^2) \leq v^2 (\sigma_0^2 - \sigma_1^2), \quad . \quad . \quad . \quad (42)$$

where v is a constant depending only on ε, σ_0, σ_1. Again two cases will arise,

(*a*) $\sigma_1 < \sigma_0$; then the region is defined by

$$\bar{x}^2 + s^2 \leq v^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

(*b*) $\sigma_1 > \sigma_0$ when it is defined by

$$\bar{x}^2 + s^2 \geq v^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

The best critical regions in the n-dimensioned space are therefore the regions (*a*) inside and (*b*) outside hyperspheres of radius $v\sqrt{n}$ whose centres are at the origin of co-ordinates. This family of hyperspheres will be the same whatever be the alternative value σ_1; there will be a common family of best critical regions for the class of alternatives $\sigma_1 < \sigma_0$, and another common family for the class $\sigma_1 > \sigma_0$.

It will be seen that the criterion is the second moment coefficient of the sample about the known population mean,

$$m'_2 = \bar{x}^2 + s^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

and not the sample variance s^2. Although a little reflection might have suggested this result as intuitively sound, it is probable that s^2 has often been used as the criterion in cases where the mean is known. The probability integral of the sampling distributions of m'_2 and s^2 may be obtained from the distribution of $\psi = \chi^2$, namely,

$$p(\psi) = c\psi^{\frac{1}{2}f-1} e^{-\frac{1}{2}\psi}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

by writing

$$m'_2 = \sigma_0^2 \psi/n, \qquad f = n, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

and

$$s^2 = \sigma_0^2 \psi/n, \qquad f = n - 1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

It is of interest to compare the relative efficiency of the criteria m'_2 and s^2 in avoiding errors of the second type, that is of accepting H_0 when it is false. If it is false, suppose the true hypothesis to be H_1 relating to a population in which

$$\sigma_1 = h\sigma_0 > \sigma_0. \quad \ldots \ldots \ldots \quad (49)$$

In testing H_0 with regard to the class of alternatives for which $\sigma > \sigma_0$, we should determine the critical value of ψ_0 so that

$$P_0(\psi \geq \psi_0) = \int_{\psi_0}^{+\infty} p(\psi)\, d\psi = \varepsilon, \quad \ldots \ldots \ldots \quad (50)$$

and would accept H_0 if $\psi < \psi_0$. But if H_1 is true, the chance of finding $\psi < \psi_0$, ψ_0 being determined from (50), that is of accepting H_0 (though it is false), will be

$$P_1(\psi \leq \psi_0) = \int_0^{\psi_0 h^{-2}} p(\psi)\, d\psi. \quad \ldots \ldots \ldots \quad (51)$$

The position is shown in fig. 5. Suppose that for the purpose of illustration we take $\varepsilon = 0{\cdot}01$ and $n = 5$.

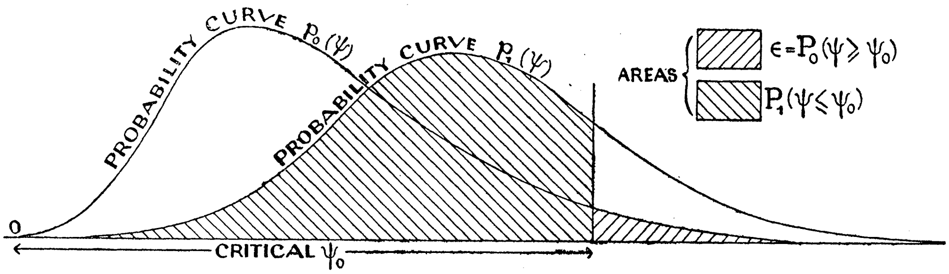

FIG. 5.

(*a*) Using m'_2 and thus the best critical region, we shall put $f = 5$ in (46), and from (50) entering the tables of the χ^2 integral with 5 degrees of freedom, find that $\psi_0 = 15{\cdot}086$. Hence from (51),

$$\begin{cases} \text{if } h = 2, \quad (\sigma_1 = 2\sigma_0), & P_1(\psi \leq \psi_0) = 0{\cdot}42 \\ \text{if } h = 3, \quad (\sigma_1 = 3\sigma_0), & P_1(\psi \leq \psi_0) = 0{\cdot}11. \end{cases}$$

(*b*) On the other hand, if the variance, s^2, is used as criterion, we must put $f = 4$ in (46) and find that $\psi_0 = 13{\cdot}277$. Hence

$$\begin{cases} \text{if } h = 2, \quad (\sigma_1 = 2\sigma_0), & P_1(\psi \leq \psi_0) = 0{\cdot}49 \\ \text{if } h = 3, \quad (\sigma_1 = 3\sigma_0), & P_1(\psi \leq \psi_0) = 0{\cdot}17. \end{cases}$$

In fact for $h = 2$, 3 or any other value, it is found that the second test has less power of discrimination between the false and the true than the test associated with the best critical region.

Example (3).—The admissible hypotheses are given by (33), both a and σ being in this case unspecified. We have to test the simple hypothesis H_0, that $a = a_0$, $\sigma = \sigma_0$. The best critical region with regard to a single alternative H_1, with $a = a_1$, $\sigma = \sigma_1$, will be defined by

$$\frac{p_0}{p_1} = \left(\frac{\sigma_1}{\sigma_0}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left\{\left(\frac{x_i - a_0}{\sigma_0}\right)^2 - \left(\frac{x_i - a_1}{\sigma_1}\right)^2\right\}} \leq k. \quad \quad (52)$$

This inequality may be shown to result in the following

(*a*) If $\sigma_1 < \sigma_0$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \alpha)^2 = (\bar{x} - \alpha)^2 + s^2 \leq v^2. \quad \quad (53)$$

(*b*) If $\sigma_1 > \sigma_0$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \alpha)^2 = (\bar{x} - \alpha)^2 + s^2 \geq v^2, \quad \quad (54)$$

where

$$\alpha = \frac{a_0\sigma_1^2 - a_1\sigma_0^2}{\sigma_1^2 - \sigma_0^2}, \quad \quad (55)$$

and v is a constant, whose value will depend upon a_0, a_1, σ_0, σ_1 and ε. It will be seen that a best critical region in the n-dimensioned space is bounded by a hypersphere of radius $v\sqrt{n}$ with centre at the point $(x_1 = x_2 = \ldots = x_n = \alpha)$. The region will be the space inside or outside the hypersphere according as $\sigma_1 < \sigma_0$ or $\sigma_1 > \sigma_0$. If $a_1 = a_0 = 0$ the case becomes that of example (2).

Unless the set of admissible hypotheses can be limited to those for which α = constant, there will not be a common family of best critical regions. The position can be seen most clearly by taking $\bar{x}$ and s as variables; the best critical regions are then seen to be bounded by the circles

$$(\bar{x} - \alpha)^2 + s^2 = v^2. \quad \quad (56)$$

If $p_0(v)$ be the probability law for v, then the relation between ε and v_0, the radius of the limiting circles is given by

$$\int_0^{v_0} p_0(v)\,dv = \varepsilon \quad \text{if} \quad \sigma_1 < \sigma_0, \quad \quad (57)$$

and

$$\int_{v_0}^{+\infty} p_0(v)\,dv = \varepsilon \quad \text{if} \quad \sigma_1 > \sigma_0. \quad \quad (58)$$

By applying the transformation

$$\bar{x} = \alpha + v\cos\phi, \quad s = v\sin\phi, \quad \quad (59)$$

to

$$p_0(\bar{x}, s) = c\, s^{n-2}\, e^{-\frac{1}{2}n(\bar{x}^2 + s^2)}, \quad \quad (60)$$

it will be found that

$$p_0(v) = c\, e^{-\frac{1}{2}n\alpha^2}\, v^{n-1}\, e^{-\frac{1}{2}nv^2}\int_0^{\pi} e^{-n\alpha v\cos\phi}\sin^{n-2}\phi\, d\phi. \quad \quad (61)$$

This integral may be expressed as a series in ascending powers of v,* but no simple method of finding v_0 for a given value of ε has been evolved.

* It is a series containing a finite number of terms if n be odd, and an infinite series if n be even.

The relation between certain population points (a, σ), and the associated best critical regions is shown in fig. 6. A single curve of the family bounding the best critical regions is shown in each case.

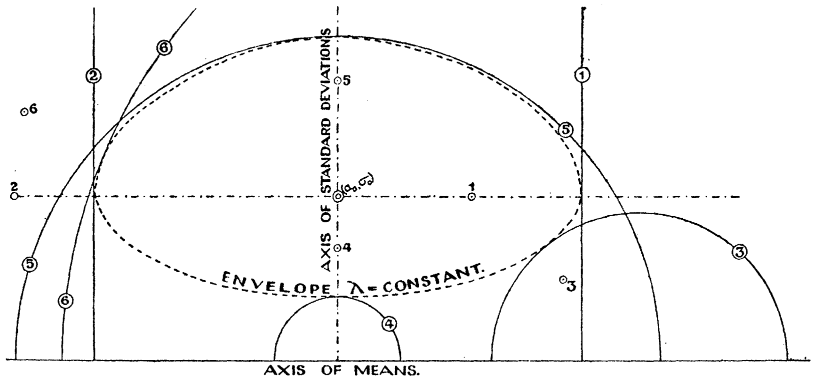

FIG. 6.

N.B.—The same number is assigned to a point (a, σ) and to the boundary of the corresponding best critical region.

Cases (1) and (2). $\sigma_1 = \sigma_0$, then $\alpha = \pm \infty$. The B.C.R. (best critical region) will be to the right of straight line (1), $(a_1 > a_0)$, or to the left of straight line (2), $(a_1 < a_0)$.

Case (3). $\sigma_1 < \sigma_0$. Suppose $\sigma_1 = \frac{1}{2}\sigma_0$, then $\alpha = a_0 + \frac{4}{3}(a_1 - a_0)$ and the B.C.R. lies inside the semi-circle (3).

Case (4). $\sigma_1 < \sigma_0$ and $a_1 = a_0$. $\alpha = a_0$. The B.C.R. lies inside the semi-circle (4).

Case (5). $\sigma_1 > \sigma_0$ and $a_1 = a_0$. $\alpha = a_0$. The B.C.R. lies outside the semi-circle (5).

Case (6). $\sigma_1 > \sigma_0$. Suppose $\sigma_1 = \frac{3}{2}\sigma_0$, then $\alpha = a_0 - \frac{4}{5}(a_1 - a_0)$, and for $a_1 < a_0$, $\alpha > a_0$. In the diagram the B.C.R. lies outside the large semi-circle, part of which is shown as curve (6).

It is evident that there is no approach to a common best critical region with regard to all the alternatives H_t, of the set Ω represented by equation (33). If $w_0(t)$ is the best critical region for H_t, then $W - w_0(t)$ may be termed the region of acceptance of H_0 with regard to H_t. The diagram shows how these regions of acceptance will have a large common part, namely, the central space around the point $a = a_0$, $\sigma = \sigma_0$. This is the region of acceptance picked out by the criterion of likelihood. It has been pointed out above that if λ be the likelihood of H_0 with regard to the set Ω, then the hypersurfaces $\lambda = k$ are the envelopes of the hypersurfaces $p_0/p_t = k = \lambda$ considered as varying with regard to a_t and σ_t. The equation of these envelopes we have shown elsewhere to be,*

$$\left(\frac{\bar{x} - a_0}{\sigma_0}\right)^2 - \log\left(\frac{s}{\sigma_0}\right)^2 = 1 - \frac{2}{n}\log\lambda. \quad \ldots \ldots \ldots \quad (62)$$

* 'Biometrika,' vol. 20A, p. 188 (1928). The ratio p_0/p_t is given by equation (52) if we write a_t and σ_t for a_1 and σ_1. It should be noted that the envelope is obtained by keeping $\lambda = k$ = constant, and since k is a function of a_t and σ_t, this will not mean that ε = constant for the members of the system giving the envelope.

The dotted curve shown in fig. 6 represents one such envelope. The region in the $(\bar{x}, s)$ plane outside this curve and the corresponding region in the n-dimensioned space may be termed good critical regions, but have not the unique status of the best critical region common for all H_t. Such a region is essentially one of compromise, since it includes a part of the best critical regions with regard to each of the admissible alternatives.

It is also clear that considerations of *a priori* probability may now need to be taken into account in testing H_0. If a certain group of alternatives were more probable than others *a priori*, we might be inclined to choose a critical region more in accordance with the best critical regions associated with the hypotheses of that group than the λ region. Occasionally it happens that *a priori* probabilities can be expressed in exact numerical form,* and if this is so, it would at any rate be possible theoretically to pick out the region w_0 for which $P_0(w_0) = \varepsilon$, such that the chance of accepting H_0 when one of the weighted alternatives H_t is true is a minimum. But in general, we are doubtful of the value of attempts to combine measures of the probability of an event if a hypothesis be true, with measures of the *a priori* probability of that hypothesis. The difficulty seems to vanish in this as in the other cases, if we regard the λ surfaces as providing (1) a control by the choice of ε of the first source of error (the rejection of H_0 when true); and (2) a good compromise in the control of the second source of error (the acceptance of H_0 when some H_t is true). The vague *a priori* grounds on which we are intuitively more confident in some alternatives than in others must be taken into account in the final judgment, but cannot be introduced into the test to give a single probability measure.†

(2) *The Sample Space Limited ; Case of the Rectangular Population.*—Hitherto we have supposed that there is a common sample space, W, for all admissible hypotheses, and in the previous examples this has been the unlimited n-dimensioned space. We must, however, consider the case in which the space W_0, in which $p_0 > 0$, associated with H_0, does not correspond exactly with the space W_1, associated with an alternative H_1 where $p_1 > 0$. Should W_0 and W_1 have no common part, then we are able to discriminate absolutely between H_0 and H_1. Such would be the case for example if $p_t(x) = 0$ when $x < a_t$ or $x > b_t$, and it happened that $a_1 > b_0$. But more often W_0 and W_1 will have a common part, say W_{01}. Then it is clear that $W_1 - W_{01}$ should be *included* in the best critical region for H_0 with regard to H_1. If this were the whole critical region, w_0, we should never reject H_0 when it is true, for $P_0(w_0) = 0$, but it is possible that we should accept H_0 too often when H_1 is true. Consequently we may wish to make up w_0 by adding to $W_1 - W_{01}$ a region w_{00} which is a part of W_{01} for which $P_0(w_{00}) = P_0(w_0) = \varepsilon$. The method of choosing the appropriate w_{00} with regard to H_1 will be as before, except that the sample space for which it may be chosen

* As for example in certain Mendelian problems.

† Tables and diagrams to assist in using this λ-test have been given in 'Biometrika,' vol. 20A, p. 233 (1928), and are reproduced in "Tables for Statisticians and Biometricians," Part II.

is now limited to W_{01}. If, however, a class of alternatives exists for which the space W_{0t} varies with t, there will probably be no common best critical region. The position may be illustrated in the case of the so-called rectangular distribution, for which the probability law can be written,

$$\left.\begin{array}{lll} p(x) = 1/b & \text{for} & a - \tfrac{1}{2}b \leq x \leq a + \tfrac{1}{2}b \\ p(x) = 0 & \text{for} & x < a - \tfrac{1}{2}b \text{ and } x > a + \tfrac{1}{2}b \end{array}\right\}, \quad (63)$$

a will be termed the mid-point and b the range of the distribution.

Example (4).—Suppose that a sample of n individuals $x_1, x_2, \ldots x_n$ is known to have been drawn at random from *some* population with distribution following (63), in which $b = b_0$, and it is wished to test the simple hypothesis H_0 that in the sampled population, $a = a_0$. For the admissible set of alternatives, $b = b_0$, but a is unspecified. For H_0 the sample space W_0 is the region within the hypercube, defined by

$$a_0 - \tfrac{1}{2}\, b_0 \leq x_i \leq a_0 + \tfrac{1}{2}\, b_0. \quad (64)$$

If H_1 be a member of the set of alternatives for which $a = a_1$, then

$$p_0(x_1, x_2, \ldots x_n) = p_1(x_1, x_2, \ldots x_n) = \frac{1}{b^n}, \quad (65)$$

provided the sample point lies within W_{01}. It follows that at every point in W_{01} $p_0/p_1 = k = 1$, and that $P_0(w_0) = \varepsilon$ for any region whatsoever within W_{01}, the content of which equals ε times the content, $b_0{}^n$, of the hypercube W_0.

There is in fact no single best critical region with regard to H_1. Fig. 7 illustrates the

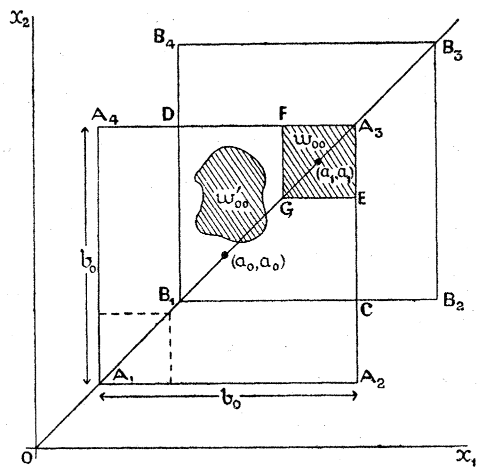

Fig. 7.

position for the case of samples of 2. The sample spaces W_0 and W_1 are the squares $A_1 A_2 A_3 A_4$ and $B_1 B_2 B_3 B_4$ respectively. A critical region for H_0 with regard to H_1 will consist of—

(1) The space $W_1 - W_{01} = A_3CB_2B_3B_4D$;

(2) Any region such as w'_{00} lying wholly inside the common square $W_{01} = B_1CA_3D$, containing an area $\varepsilon\, b_0^2$.

The value of ε is at our choice and may range from O to $(a_0 - a_1 + b_0)^2$, according to the balance it is wished to strike between the two kinds of error. We shall not allow any part of w_0 to lie outside B_1CA_3D in the space $W_0 - W_{01}$, for this would lead to the rejection of H_0 in cases where the alternative H_1 could not be true.

For different alternatives, H_t, of the set, the mid-point of the square $B_1B_2B_3B_4$ will shift along the diagonal OA_1A_3. For a fixed ε we cannot find a region that will be included in W_{0t} for every H_t, but we shall achieve this result as nearly as possible if we can divide the alternatives into two classes—

(*a*) $a_1 > a_0$. Take w_{00} as the square GEA_3F with length of side $= b_0\sqrt{\varepsilon}$ lying in the upper left hand corner of W_0.

(*b*) $a_1 < a_0$. Take a similar square with corner at A_1.

In both cases the whole space outside W_0 must be added to make up the critical region w_0. In the general case of samples of n, the region w_{00} will be a hypercube with length of side $b_0\sqrt[n]{\varepsilon}$ fitting into one or other of the two corners of the hypercube of W_0 which lie on the axis $x_1 = x_2 = \ldots = x_n$. The whole of the space outside W_0 within which sample points can fall will also be added to w_{00} to make up w_0.*

Example (5). Suppose that the set of alternatives consists of distributions of form (63), for all of which $a = a_0$, but b may vary. H_0 is the hypothesis that $b = b_0$. The sample spaces, W_t, are now hypercubes of varying size all centred at the point $(x_1 = x_2 = \ldots = x_n = a_0)$. A little consideration suggests that we should make the critical region w_0 consist of—

(1) The whole space outside the hypercube W_0.

(2) The region w_{00} inside a hypercube with centre at $(x_1 = x_2 = \ldots = x_n = a_0)$, sides parallel to the co-ordinate axes and of volume εb_0^n. This region w_{00} is chosen because it will lie completely within the sample space W_{0t} common to H_0 and H_t for a larger number of the set of alternatives than any other region of equal content.

Example (6).—H_0 is the hypothesis that $a = a_0$, $b = b_0$, and the set of admissible alternatives is given by (63) in which both a and b are now unspecified. Both the mid-point $(x_1 = x_2 = \ldots = x_n = a_t)$ and the length of side, b_t, of the alternative sample spaces W_t can therefore vary. Clearly we shall again include in w_0 the whole space outside W_0, but there can be no common region w_{00} within W_0.

Fig. 8A represents the position for $n = 2$. Four squares W_1, W_2, W_3, and W_4 correspond to the sample spaces of possible alternatives H_1, H_2, H_3, and H_4, and the smaller shaded squares w_1, w_2, w_3, and w_4 represent possible critical regions for H_0 with regard to these. What compromise shall we make in choosing a critical region with

* If the set is limited to distributions for which $b = b_0$, no sample point can lie outside the envelope of hypercubes whose centres lie on the axis $x_1 = x_2 = \ldots = x_n$.

regard to the whole set Ω? As we have shown elsewhere* the method of likelihood fixes for the critical region that part of the space that represents samples for which the range (the difference between extreme variates) is less than a given value, say $l \leq l_0$. For samples of 2, $l = x_1 - x_2$ if $x_1 > x_2$, and $x_2 - x_1$ if $x_1 < x_2$, and the critical region w_{00} will therefore lie between two straight lines parallel to and equidistant from the axis $x_1 = x_2$. A pair of such lines will be the envelope of the small squares w_1, w_2, etc., of fig. 8A. In fact, the complete critical region will be as shown in fig. 8B, the belt w_{00} being chosen so that its area is ϵb_0^2.

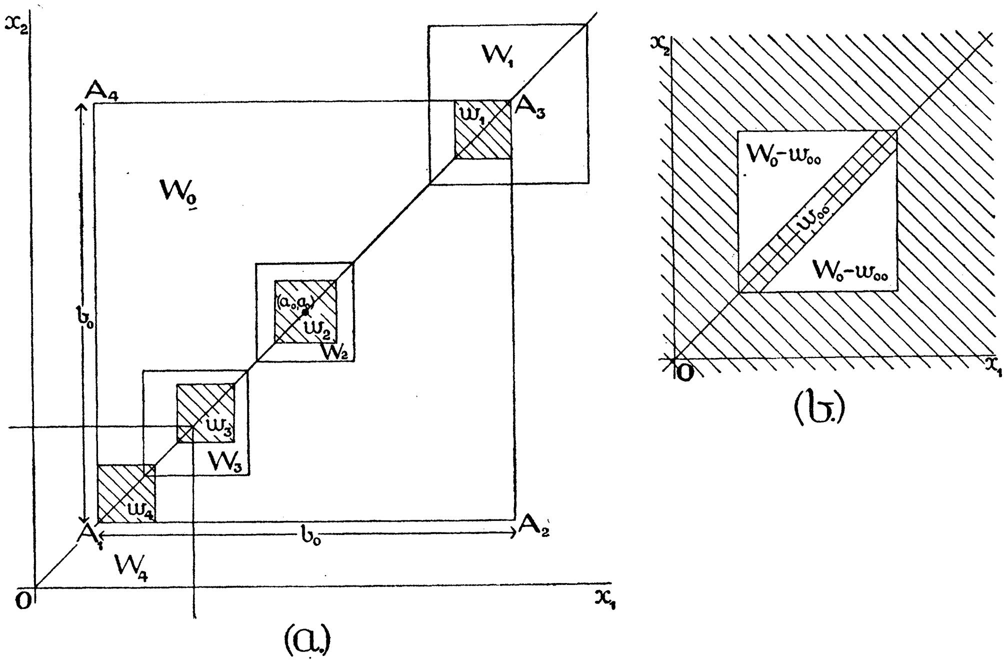

FIG. 8.

For $n = 3$ the surface $l = l_0$ is a prism of hexagonal cross-section, whose generating lines are parallel to the axis $x_1 = x_2 = x_3$. The space, w_{00}, within this and the whole space outside the cube W_0 will form the critical region w_0. In general for samples of n the critical region of the likelihood method will consist of the space outside the hypercube W_0, and the space of content ϵb_0^n within the envelope of hypercubes having centres on the axis $x_1 = x_2 = \ldots = x_n$, and edges parallel to the axes of co-ordinates.

It will have been noted that a correspondence exists between the hypotheses tested in examples (1) and (4), (2) and (5), (3) and (6), and between the resulting critical regions. Consider for instance the position for $n = 3$ in example (3); the boundary of the critical region may be obtained by rotating fig. 6 in 3-dimensioned space about the axis of means. The region of acceptance of H_0 is then bounded by a surface analogous to an anchor ring surrounding the axis $x_1 = x_2 = x_3$, traced out by the rotation of the dotted curve λ = constant. Its counterpart in example (6) is the region inside a cube from which the hexagonal sectioned prism w_{00} surrounding the diagonal

* 'Biometrika,' vol. 20A, p. 208 (1928). Section on Samples from a Rectangular Population.

$x_1 = x_2 = x_3$ has been removed. A similar correspondence may be traced in the case of sampling from a distribution following the exponential law. It continues to hold in the higher dimensioned spaces with $n > 3$.

The difference between the normal, rectangular and exponential laws is of course, very great, but the question of what may be termed the stability in form of best critical regions for smaller changes in the frequency law, $p(x_1, x_2, \dots x_n)$, is of considerable practical importance.

IV.—Composite Hypotheses.

(a) *Introductory.*

In the present investigation we shall suppose that the set Ω of admissible hypotheses defines the functional form of the probability law for a given sample, namely—

$$p(x_1, x_2, \dots x_n), \qquad (66)$$

but that this law is dependent upon the values of $c + d$ parameters

$$\alpha^{(1)}, \alpha^{(2)}, \dots \alpha^{(c)};\ \alpha^{(c+1)}, \dots \alpha^{(c+d)}. \qquad (67)$$

A composite hypothesis, H'_0, of c degrees of freedom is one for which the values of d of these parameters are specified and c unspecified. We shall denote these parameters by

$$\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(c)};\ \alpha_0^{(c+1)}, \dots \alpha_0^{(c+d)}. \qquad (68)$$

This composite hypothesis consists of a sub-set ω (of the set Ω) of simple hypotheses. We shall denote the probability law for H'_0 by

$$p_0 = p_0(x_1, x_2, \dots x_n), \qquad (69)$$

associating with (69) in any given case the series (68). An alternative simple hypothesis which is definitely specified will be written as H_t, and with this will be associated

(1) a probability law

$$p_t = p_t(x_1, x_2, \dots x_n). \qquad (70)$$

(2) a series of parameters

$$\alpha_t^{(1)}, \alpha_t^{(2)}, \dots \alpha_t^{(c+d)}. \qquad (71)$$

We shall suppose that there is a common sample space W for any admissible hypothesis H_t, although its probability law p_t may be zero in some parts of W.

As when dealing with simple hypotheses we must now determine a family of critical regions in the sample space, W, having regard to the two sources of error in judgment. In the first place it is evident that a necessary condition for a critical region, w, suitable for testing H'_0 is that

$$P_0(w) = \iint \dots \int_w p_0(x_1, x_2, \dots x_n)\, dx_1\, dx_2 \dots dx_n = \text{constant} = \varepsilon \qquad (72)$$

for every simple hypothesis of the sub-set ω. That is to say, it is necessary for $P_0(w)$ to be independent of the values of $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$. If this condition is satisfied we shall speak of w as a region of "size" ε, similar to W with regard to the c parameters $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$.

Our first problem is to express the condition for similarity in analytical form. Afterwards it will be necessary to pick out from the regions satisfying this condition that one which reduces to a minimum the chance of accepting H'_0 when a simple alternative hypothesis H_t is true. If this region is the same for all the alternatives H_t of the set Ω, then we shall have a common best critical region for H'_0 with regard to the whole set of alternatives. The fundamental position from which we start should be noted at this point. It is assumed that the only possible critical regions that can be used are similar regions; that is to say regions such that $P(w) = \varepsilon$ for every simple hypothesis of the sub-set ω. It is clear that were it possible to assign differing measures of *a priori* probability to these simple hypotheses, a principle might be laid down for determining critical regions, w, for which $P(w)$ would vary from one simple hypothesis to another. But it would seem hardly possible to put such a test into working form.

We have, in fact, no hesitation in preferring to retain the simple conception of control of the first source of error (rejection of H'_0 when it is true) by the choice of ε, which follows from the use of similar regions. This course seems necessary as a matter of practical policy, apart from any theoretical objections to the introduction of measures of *a priori* probability.

(b) *Similar Regions for Case in which* H'_0 *has One Degree of Freedom.*

We shall commence with this simple case for which the series (68) becomes

$$\alpha^{(1)};\ \alpha_0^{(2)}, \alpha_0^{(3)}, \ldots \alpha_0^{(1+d)}. \qquad (73)$$

We have been able to solve the problem of similar regions only under very limiting conditions concerning p_0. These are as follows :—

(*a*) p_0 is indefinitely differentiable with regard to $\alpha^{(1)}$ for all values of $\alpha^{(1)}$ and in every point of W, except perhaps in points forming a set of measure zero. That is to say, we suppose that $\dfrac{\partial^k p_0}{\partial (\alpha^{(1)})^k}$ exists for any $k = 1, 2, \ldots$ and is integrable over the region W.

Denote by

$$\phi = \frac{\partial \log p_0}{\partial \alpha^{(1)}} = \frac{1}{p_0}\frac{\partial p_0}{\partial \alpha^{(1)}};\ \phi' = \frac{\partial \phi}{\partial \alpha^{(1)}}. \qquad (74)$$

(*b*) The function p_0 satisfies the equation

$$\phi' = A + B\phi, \qquad (75)$$

where the coefficients A and B are functions of $\alpha^{(1)}$ but are independent of $x_1, x_2, \ldots x_n$.

This last condition could be somewhat generalised by adding the term $C\phi^2$ to the right-hand side of (75), but this would introduce some complication and we have not found any practical case in which p_0 does satisfy (75) in this more general form and does not in the simple form. We have, however, met instances in which neither of the two forms of the condition (*b*) is satisfied by p_0.

If the probability law p_0 satisfies the two conditions (*a*) and (*b*), then it follows that a necessary and sufficient condition for w to be similar to W with regard to $\alpha^{(1)}$ is that

$$\frac{\partial^k P_0(w)}{\partial (\alpha^{(1)})^k} = \iint \ldots \int_w \frac{\partial^k p_0}{\partial (\alpha^{(1)})^k}\, dx_1\, dx_2 \ldots dx_n = 0, \qquad k = 1, 2, \ldots \quad (76)$$

Taking in (76) $k = 1$ and 2 and writing

$$\frac{\partial p_0}{\partial \alpha^{(1)}} = p_0\, \phi \quad (77)$$

$$\frac{\partial^2 p_0}{\partial (\alpha^{(1)})^2} = \frac{\partial}{\partial \alpha^{(1)}} (p_0 \phi) = p_0 (\phi^2 + \phi'), \quad (78)$$

it will be found that

$$\frac{\partial P_0(w)}{\partial \alpha^{(1)}} = \iint \ldots \int_w p_0 \phi\, dx_1\, dx_2 \ldots dx_n = 0 \quad (79)$$

$$\frac{\partial^2 P_0(w)}{\partial (\alpha^{(1)})^2} = \iint \ldots \int_w p_0 (\phi^2 + \phi')\, dx_1\, dx_2 \ldots dx_n = 0. \quad (80)$$

Using (75) we may transform this last equation into the following

$$\frac{\partial^2 P_0(w)}{\partial (\alpha^{(1)})^2} = \iint \ldots \int_w p_0 (\phi^2 + A + B\phi)\, dx_1\, dx_2 \ldots dx_n = 0. \quad (81)$$

Having regard to (72) and (79) it follows from (81) that

$$\iint \ldots \int_w p_0\, \phi^2\, dx_1 \ldots dx_n = -A\varepsilon = \varepsilon \psi_2 (\alpha^{(1)}) \text{ (say)}. \quad (82)$$

The condition (76) for $k = 3$ may now be obtained by differentiating (81). We shall have

$$\frac{\partial^3 P_0(w)}{\partial (\alpha^{(1)})^3} = \iint \ldots \int_w p_0 (\phi^3 + 3B\phi^2 + (3A + B^2 + B')\, \phi + A' + AB)\, dx_1 \ldots dx_n = 0, \quad (83)$$

which, owing to (72), (79) and (82) is equivalent to the condition

$$\iint \ldots \int_w p_0\, \phi^3\, dx_1 \ldots dx_n = (3AB - A' - AB)\, \varepsilon = \varepsilon \psi_3 (\alpha^{(1)}) \quad \text{(say)}. \quad (84)$$

As it is easy to show, using the method of induction, this process may be continued indefinitely and we shall find

$$\iint \ldots \int_w p_0 \phi^k \, dx_1 \, dx_2 \ldots dx_n = \varepsilon \psi_k (\alpha^{(1)}) \qquad k = 1, 2, \ldots \quad . \quad . \quad . \quad . \quad . \quad . \quad (85)$$

where $\psi_k(\alpha^{(1)})$ is a function of $\alpha^{(1)}$ but independent of the sample x's, since the quantities A, B and their derivatives with regard to $\alpha^{(1)}$ are independent of the x's. $\psi_k(\alpha^{(1)})$ is also independent of the region w, and it follows that whatever be w, and its size ε, if it be similar to W with regard to $\alpha^{(1)}$, the equation (85) must hold true for every value of k, *i.e.*, 1, 2, … . Since the complete sample space W is clearly similar to W and of size unity, it follows that

$$\frac{1}{\varepsilon} \iint \ldots \int_w p_0 \, \phi^k \, dx_1 \, dx_2 \ldots dx_n = \iint \ldots \int_W p_0 \, \phi^k \, dx_1 \, dx_2 \ldots dx_n, \qquad k = 1, 2, 3, \ldots \quad . \quad . \quad (86)$$

Now $p_0 \ (x_1, x_2, \ldots x_n)$ is a probability law of n variates $x_1, x_2, \ldots x_n$, defined in the region W; similarly $\frac{1}{\varepsilon} p \ (x_1, x_2, \ldots x_n)$ may be considered as a probability law for the same variates under the condition that their variation is limited to the region w. We may regard ϕ as a dependent variate which is a known function of the n independent variates x_i. The integral on the right-hand side of (86) is the k-th moment coefficient of this variate ϕ obtained on the assumption that the variation in the sample point $x_1, x_2, \ldots x_n$ is limited to the region W, while the integral on the left-hand side is the same moment coefficient obtained for variation of the sample point, limited to the region w. Denoting these moment coefficients by μ_k (W) and $\mu_k(w)$, we may rewrite (86) in the form—

$$\mu_k (w) = \mu_k (\mathrm{W}), \qquad k = 1, 2, 3, \ldots \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (87)$$

It is known that if the set of moment coefficients satisfy certain conditions, the corresponding frequency distribution is completely defined.* Such, for instance, is the case when the series $\Sigma\{\mu_k (it)^k/k!\}$ is convergent, and it then represents the characteristic function of the distribution.

We do not, however, propose to go more closely into this question, and shall consider only the cases in which the moments coefficients of ϕ satisfy the conditions of H. HAMBURGER†. In these cases, to which the theory developed below only applies‡, it follows from (87) that when ϕ, which is related to $p_0(x_1, x_2, \ldots x_n)$ by

* HAMBURGER; 'Math. Ann.,' vol. 81, p. 4 (1920).

† We are indebted to Dr. R. A. FISHER for kindly calling our attention to the fact that we had originally omitted to refer to this restriction.

‡ It may easily be proved that these conditions are satisfied in the case of Examples (7), (8), (9), (10) and (11) discussed below.

(74), is such as to satisfy the equation (75), the identity of the two distributions of ϕ is the necessary (and clearly also sufficient) condition for w being similar to W with regard to the parameter $\alpha^{(1)}$.

The significance of this result may be grasped more clearly from the following consideration. Every point of the sample space W will fall on to one or other of the family of hypersurfaces

$$\phi = \text{constant} = \phi_1. \qquad (88)$$

Then if

$$P_0(w(\phi)) = \iint \dots \int_{w(\phi_1)} p_0 \, dw(\phi_1) \qquad (89)$$

$$P_0(W(\phi)) = \iint \dots \int_{W(\phi_1)} p_0 \, dW(\phi_1) \qquad (90)$$

represent the integral of p_0 taken over the common parts, $w(\phi)$ and $W(\phi)$, of $\phi = \phi_1$ and w and W respectively, it follows that if w be similar to W and of size ε,

$$P_0(w(\phi)) = \varepsilon\, P_0(W(\phi)), \qquad (91)$$

whatever be ϕ_1.

Whatever be ε, a similar region is, in fact, built up of pieces of the hypersurfaces (88) for which (91) is true.

We shall give at this stage only a single example of this result, which will be illustrated more fully when dealing with the best critical regions.

Example (7).—*A single sample of n from a normal population* ; σ *unspecified.*

$$p(x_1, x_2, \dots x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-n\frac{(\bar{x}-a)^2+s^2}{2\sigma^2}}. \qquad (92)$$

For H'_0,

$$\alpha^{(1)} = \sigma, \qquad \alpha_0^{(2)} = a_0 \qquad (93)$$

$$\phi = \frac{\partial \log p_0}{\partial \sigma} = -\frac{n}{\sigma} + n\frac{(\bar{x}-a_0)^2+s^2}{4\sigma^3} \qquad (94)$$

$$\phi' = \frac{n}{\sigma^2} - 3n\frac{(\bar{x}-a_0)^2+s^2}{\sigma^4} = -\frac{2n}{\sigma^2} - \frac{3}{\sigma}\phi. \qquad (95)$$

Equation (95) shows that the condition (75) is satisfied. Further, ϕ is constant on any one of the family of hypersurfaces

$$n\{(\bar{x}-a_0)^2 + s^2\} = \sum_{i=1}^{n}(x_i - a_0)^2 = \text{const.} \qquad (96)$$

Consequently the most general region w similar to W (which in this case is the whole n-dimensioned space of the x's) is built up of pieces of the hyperspheres (96) which satisfy the relation (91). Since $p_0(x_1, x_2, \dots x_n)$ is constant upon each hypersphere, the content of the "piece" $w(\phi)$ must be in a constant proportion, $\varepsilon : 1$ to the content of the complete hyperspherical shell $W(\phi)$. The possible similar regions may be of infinite variety in form. They need not be hypercones, but may be of irregular shape as suggested in fig. 9 for the case $n = 3$. It is out of these possible forms that the best critical region has to be chosen.

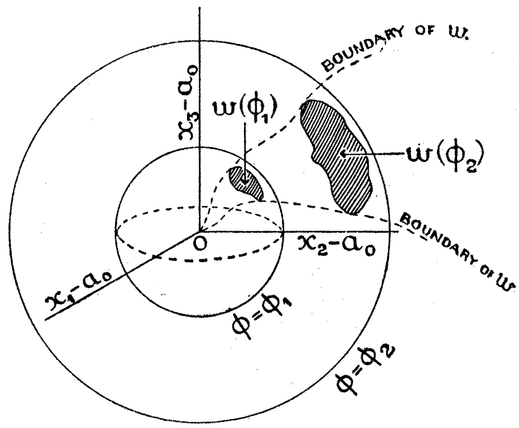

FIG. 9.

(c) Choice of the Best Critical Region.

Let H_t be an alternative simple hypothesis defined by the relations (70) and (71). We shall assume that regions similar to W with regard to $\alpha^{(1)}$ do exist. Then w_0, the best critical region for H_0 with regard to H_t, must be determined to maximise

$$P_t(w_0) = \iint \ldots \int_w p_t(x_1, x_2, \ldots x_n)\, dx_1 \ldots dx_n, \quad \ldots \ldots \quad (97)$$

subject to the condition (91) holding for all values of ϕ, which implies the condition (72). We shall now prove that if w_0 is chosen to maximise $P_t(w_0)$ under the condition (72), then except perhaps for a set of values of ϕ of measure zero, the region $w_0(\phi)$ will maximise

$$P_t(w(\phi)) = \int \ldots \int_{w(\phi)} p_t(x_1, x_2, \ldots x_n)\, dw(\phi), \quad \ldots \ldots \quad (98)$$

under the condition (91). That is to say, we shall prove that whatever be the $(n-1)$-dimensioned region, say $v(\phi)$, being a part of the hypersurface $\phi =$ const. and satisfying the condition

$$P_0(v(\phi)) = \varepsilon P_0(W(\phi)), \quad \ldots \ldots \quad (99)$$

we should have

$$P_t(v(\phi)) \leq P_t(w_0(\phi)), \quad \ldots \ldots \quad (100)$$

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

Suppose in fact that the proposition is not true and that there exists a set E of values of ϕ of positive measure for which it is possible to define the regions $v(\phi)$ satisfying (99), and such that

$$P_t(v(\phi)) > P_t(w_0(\phi)). \quad \ldots \ldots \quad (101)$$

Denote by CE the set of values of ϕ complementary to E. We shall now define a region, say v, which will be similar to W with regard to $\alpha^{(1)}$ and such that

$$P_t(v) > P_t(w_0), \quad \ldots \ldots \ldots \ldots \quad (102)$$

which will contradict the assumption that w_0 is the best critical region with regard to H_t.

The region v will consist of parts of hypersurfaces $\phi = \text{const.}$ For ϕ's included in CE, these parts, $v(\phi)$, will be identical with $w_0(\phi)$ and for ϕ's belonging to E, they will be $v(\phi)$ satisfying (101). Now,

$$P_t(v) = \int_{E+CE} P_t(v(\phi))\, d\phi,$$

$$P_t(w_0) = \int_{E+CE} P_t(w_0(\phi))\, d\phi,$$

and, owing to the properties of v,

$$P_t(v) - P_t(w_0) = \int_E (P_t(v(\phi)) - P_t(w_0(\phi))\, d\phi > 0. \quad \ldots \quad (103)$$

It follows that if w_0 is the best critical region, then (101) may be true at most for a set of ϕ's of measure zero. It follows also that if (100) be true for every ϕ and every $v(\phi)$ satisfying (99), then the region w_0, built up of parts of hypersurfaces $\phi = \text{const.}$ satisfying (91), is the best critical region required.

Having established this result the problem of finding the best critical region, w_0, is reduced to that of finding parts, $w_0(\phi)$, of $W(\phi)$, which will maximise $P(w(\phi))$ subject to the condition

$$P_0(w_0(\phi)) = \varepsilon\, P_0(W(\phi)) \quad \ldots \ldots \ldots \quad (104)$$

where ϕ is fixed. This is the same problem that we have treated already when dealing with the case of a simple hypothesis (*see* pp. 298–301), except that instead of the regions w_0 and W, we have the regions $w_0(\phi)$ and $W(\phi)$, and a space of one dimension less. The inequality

$$p_t \geq k(\phi)\, p_0 \quad \ldots \ldots \ldots \ldots \quad (105)$$

will therefore determine the region $w_0(\phi)$, where $k(\phi)$ is a constant (whose value may depend upon ϕ) chosen subject to the condition (104).

The examples which follow illustrate the way in which the relations (104) and (105) combine to give the best critical region. It will be noted that if the family of surfaces bounding the pieces $w_0(\phi)$ conditioned by (105) is independent of the parameters $\alpha_t^{(1)}, \alpha_t^{(2)} \ldots \alpha_t^{(1+d)}$, then a common best critical region will exist for H_0 with regard to all hypotheses H_t of the set Ω.

(*d*) *Illustrative Examples.*

(1) *Example* (8).—*The Hypothesis concerning the Population Mean* ("STUDENT'S" *Problem*).—A sample of n has been drawn at random from some normal population, and H'_0 is the composite hypothesis that the mean in this population is $a = a_0$, σ being unspecified. We have already discussed the problem of determining similar regions for H'_0 in example (7). H_t is an alternative for which

$$\alpha_t^{(1)} = \sigma_t, \alpha_t^{(2)} = a_t. \quad \ldots \quad (106)$$

The family of hypersurfaces, $\phi =$ constant, in the n-dimensioned space are hyperspheres (96) centred at $(x_1 = x_2 = \ldots = x_n = a_0)$; we must determine the nature of the pieces defined by condition (105), to be taken from these to build up the best critical region for H'_0 with regard to H_t.

Using (92), it is seen that the condition $p_t \geq kp_0$ becomes

$$\frac{1}{\sigma_t^n} e^{-\frac{n[(\bar{x}-a_t)^2+s^2]}{2\sigma_t^2}} \geq k \frac{1}{\sigma^n} e^{-\frac{n[(\bar{x}-a_0)^2+s^2]}{2\sigma^2}}. \quad \ldots \quad (107)$$

As we are dealing with regions similar with regard to $\alpha^{(1)}$, that is, essentially independent of the value of the parameter $\alpha^{(1)} = \sigma$, we may put $\sigma = \sigma_t$ and find that (107) reduces to,

$$\bar{x}(a_t - a_0) \geq \frac{1}{n}\sigma_t^2 \log k + \tfrac{1}{2}(a_t^2 - a_0^2) = (a_t - a_0)\, k_1(\phi) \quad \text{(say)}. \quad \ldots \quad (108)$$

Two cases must be distinguished in determining $w_0(\phi)$—

$$(a)\ a_t > a_0, \quad \text{then} \quad \bar{x} \geq k_1(\phi) \quad \ldots \quad (109)$$

$$(b)\ a_t < a_0, \quad \text{then} \quad \bar{x} \leq k_1(\phi), \quad \ldots \quad (110)$$

where $k_1(\phi)$ has to be chosen so that (91) is satisfied. Conditions (109) and (110) will determine the pieces of the hyperspheres to be used. In the case $n = 3$, $\bar{x} = \frac{1}{3}(x_1 + x_2 + x_3)$ is a plane perpendicular to the axis $x_1 = x_2 = x_3$, and it follows that $w_0(\phi)$ will be a "polar cap" on the surface of the sphere surrounding this axis. The pole is determined by the condition $a_t > a_0$ or $a_t < a_0$. The condition (91) implies that the area of this cap must be ε times the surface area of the whole sphere. The position is indicated in fig. 10. For all values of ϕ, that is to say, for all the concentric spherical shells making up the complete space, these caps must subtend a constant angle at the centre. Hence the pieces, $w_0(\phi)$, will build up into a cone of circular cross-section, with vertex at (a_0, a_0, a_0) and axis $x_1 = x_2 = x_3$. For each value of ε there will be a cone of different vertical angle. There will be two families of these cones containing the best critical regions—

(*a*) For the class of hypotheses $a_t > a_0$; the cones will be in the quadrant of positive x's.

(*b*) For the class of hypotheses $a_t < a_0$; the cones will lie in the quadrant of negative x's.

It is of interest to compare the general type of similar region suggested in fig. 9 with the special best critical region of fig. 10.

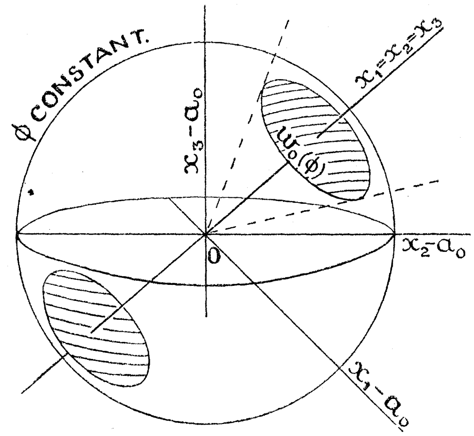

Fig. 10.

For the cases $n > 3$ we may either appeal to the geometry of multiple space, or proceed analytically as follows.

If $m'_2 = (\bar{x}_0 - a_0)^2 + s^2$, then it can be deduced from the probability law (92) that

$$p_0(\bar{x}, m'_2) = c_1 \sigma^{-n} \{m'_2 - (\bar{x} - a_0)^2\}^{\frac{n-3}{2}} e^{-\frac{nm'_2}{2\sigma^2}} \quad \ldots \ldots \quad (111)$$

$$p_0(m'_2) = c_2 \sigma^{-n} (m'_2)^{\frac{n-2}{2}} e^{-\frac{nm'_2}{2\sigma^2}}, \quad \ldots \ldots \quad (112)$$

where c_1 and c_2 are constants depending on n only. Taking the class of alternatives $a_t > a_0$, $w_0(\phi)$ is that portion of the hypersphere on which $m'_2 =$ constant, for which $\bar{x} \geq k_1(\phi)$. Consequently the expression (91) becomes

$$\int_{k_1}^{a_0+\sqrt{m'_2}} p_0(\bar{x}, m'_2)\, d\bar{x} = \varepsilon p_0(m'_2), \quad \ldots \ldots \quad (113)$$

or

$$\int_{k_1}^{a_0+\sqrt{m'_2}} \{m'_2 - (\bar{x} - a_0)^2\}^{\frac{n-3}{2}} d\bar{x} = \varepsilon \frac{c_2}{c_1} (m'_2)^{\frac{n-2}{2}}. \quad \ldots \ldots \quad (114)$$

Make now the transformation

$$\bar{x} - a_0 = \frac{z\sqrt{m'_2}}{\sqrt{1+z^2}}, \quad \ldots \ldots \quad (115)$$

from which it follows that

$$d\bar{x} = \frac{\sqrt{m'_2}}{\sqrt{(1+z^2)^3}}\, dz, \quad \ldots \ldots \quad (116)$$

and the relation (114) becomes

$$\int_{z_0}^{+\infty} (1+z^2)^{-\frac{n}{2}} dz = \varepsilon \sqrt{\pi} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \quad \ldots \ldots \quad (117)$$

the constant multiplying ε necessarily assuming this value so that $\varepsilon = 1$ when $z_0 = -\infty$. But it is seen from (115) that $z = \frac{\bar{x} - a_0}{s}$; consequently the boundary of the partial region $w_0(\phi)$ lies on the intersection of the hypersphere, $m'_2 =$ constant, and the hypercone, $(\bar{x} - a_0)/s = z_0$. This is independent of ϕ; its axis is the line $x_1 = x_2 = \ldots = x_n$ and its vertical angle is $2\theta = 2 \cot^{-1} z_0$.

If the admissible alternatives are divided into two classes, there will therefore be for each a common best critical region of size ε.

$$(a) \text{ Class } a_t > a_0;\ \text{region } w_0 \text{ defined by } z = \frac{\bar{x} - a_0}{s} \geq z_0 \quad (118)$$

$$(b) \text{ Class } a_t < a_0;\ \text{region } w_0 \text{ defined by } z = \frac{\bar{x} - a_0}{s} \leq z'_0 = -z_0, \quad (119)$$

where z_0 is related to ε by (117), and z'_0 by a similar expression in which the limits of the integral are $-\infty$ and $z'_0 = -z_0$.

This is "STUDENT'S" test.* It is also the test reached by using the principle of likelihood. Further, it has now been shown that starting with information in the form supposed, there can be no better test for the hypothesis under consideration.

(2) *Example* (9).—*The Hypothesis concerning the Variance in the Sampled Population.*—The sample has been drawn from *some* normal population and H'_0 is the hypothesis that $\sigma = \sigma_0$, the mean a being unspecified. We shall have for H'_0

$$\alpha^{(1)} = a;\ \alpha_0^{(2)} = \sigma_0, \quad (120)$$

while for an alternative H_t the parameters are as in (106).
Further

$$\phi = \frac{\partial \log p_0}{\partial a} = \frac{n(\bar{x} - a)}{\sigma_0^2} \quad (121)$$

$$\phi' = -n/\sigma_0^2 \quad (122)$$

satisfying the condition (75) with $B = 0$. We must therefore determine on each of the family of hypersurfaces $\phi = \phi_1$ (that is, from (121), $\bar{x} =$ constant) regions $w_0(\phi)$ within which $p_t \geq k(\phi_1) p_0$, where $k(\phi_1)$ is chosen so that

$$P_0(w_0(\phi_1)) = \varepsilon P_0(W(\phi_1)). \quad (123)$$

Since we are dealing with regions similar with regard to the mean a, we may put $a = a_t$, and consequently find that

$$s^2(\sigma_0^2 - \sigma_t^2) \leq -(\bar{x} - a_t)^2(\sigma_0^2 - \sigma_t^2) + 2\sigma_0^2\sigma_t^2\{\log\frac{\sigma_0}{\sigma_t} - \frac{1}{n}\log k\}$$
$$= (\sigma_0^2 - \sigma_t^2)\, k'(\phi_1) \quad \text{(say)}. \quad (124)$$

* 'Biometrika,' vol. 6, p. 1 (1908).

The admissible alternatives must again be broken into two classes according as $\sigma_t > \sigma_0$ or $< \sigma_0$, and since $\bar{x}$ is constant on $\phi = \phi_1$, the regions $w_0(\phi)$ will be given by the following inequalities :—

$$(a) \text{ Case } \sigma_t > \sigma_0, \quad s^2 \geq k'(\phi) \quad \ldots \ldots \ldots \quad (125)$$

$$(b) \text{ Case } \sigma_t < \sigma_0, \quad s^2 \leq k'(\phi). \quad \ldots \ldots \ldots \quad (126)$$

But since for samples from a normal distribution $\bar{x}$ and s^2 are completely independent the values of $k'(\phi)$ that determine the regions $w_0(\phi)$ so as to satisfy (123), will be functions of ε and n only. It follows that the best critical regions, w_0, for H'_0 will be—

(*a*) for the class of alternatives $\sigma_t > \sigma_0$, defined by $s^2 \geq s_0{}^2$. . . . . . (127)

(*b*) for the class of alternatives $\sigma_t < \sigma_0$, defined by $s^2 \leq s'_0{}^2$. . . . (128)

These regions lie respectively outside and inside hypercylinders in the n-dimensioned space. The relation between ε and the critical values $s_0{}^2$ and $s'_0{}^2$ may be found from equations (46), (48) and (50) of example 2.*

V.—Composite Hypotheses with C Degrees of Freedom.

(*a*) *Similar Regions.*

We shall now consider a probability function depending upon c parameters α

$$\left.\begin{array}{c} p(x_1, x_2, \ldots x_n) \\ \{\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}\} \end{array}\right\} \quad \ldots \ldots \ldots \quad (129)$$

This will correspond to a composite hypothesis H'_0 with c degrees of freedom.

Every function (129) having specified values of the α's, will correspond to some simple hypothesis belonging to H'_0. Let w be any region in the sample space W, and denote by $P(\{\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}\} w)$ the integral of (129) over the region w. Generally it will depend upon the values of the α's.

Fix any system of values of the α's

$$\alpha_A{}^{(1)}, \alpha_A{}^{(2)}, \ldots \alpha_A{}^{(c)}. \qquad (A) \quad \ldots \ldots \ldots \quad (130)$$

If the region w has the property, that

$$P(\{\alpha_A{}^{(1)}, \ldots \alpha_A{}^{(c)}\} w) = \varepsilon = \text{constant}, \quad \ldots \ldots \ldots \quad (131)$$

whatever be the system A, we shall say that it is similar to the sample space W with regard to the set of parameters $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$ and of size ε.

* The difference between the two cases should be noted: In example (2) the population mean is specified, H_0 is a simple hypothesis and m'_2 is the criterion. In example (9) the mean is not specified, H'_0 is composite and the criterion is s^2.

This is the natural generalisation of the notion of similarity with regard to one single parameter previously introduced.

Let us first consider regions w, which are similar to W with regard to some single parameter, $\alpha^{(i)}$, for some fixed values of other parameters $\alpha^{(j)}$ ($j = 1, 2, \ldots c$, but $j \neq i$). Clearly there may be regions which are similar to W with regard to $\alpha^{(i)}$ when the other α's have some definite values, but which cease to be similar when these values are changed.

We shall now prove the following proposition.

The necessary and sufficient condition for w being similar to W with regard to the set of parameters $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$, is that it should be similar with regard to each one of them separately for every possible system of values of the other parameters.

The necessity of this condition is evident. We shall have to prove that it is also sufficient. This we shall do assuming $c = 2$, since the generalisation follows at once from this.

The conditions of the theorem mean that, (a), whatever be the values of $\alpha_A^{(1)}$, $\alpha_B^{(1)}$ and $\alpha_C^{(2)}$ we shall have

$$P(\{\alpha_A^{(1)}, \alpha_C^{(2)}\} w) = P(\{\alpha_B^{(1)}, \alpha_C^{(2)}\} w) = \varepsilon, \quad \ldots \ldots \quad (132)$$

and (b), whatever be $\alpha_B^{(1)}, \alpha_C^{(2)}, \alpha_D^{(2)}$, then

$$P(\{\alpha_B^{(1)}, \alpha_C^{(2)}\} w) = P(\{\alpha_B^{(1)}, \alpha_D^{(2)}\} w) = \varepsilon. \quad \ldots \ldots \quad (133)$$

It follows that whatever be $\alpha_A^{(1)}, \alpha_B^{(1)}$; $\alpha_C^{(2)}, \alpha_D^{(2)}$ we shall have

$$P(\{\alpha_A^{(1)}, \alpha_C^{(2)}\} w) = P(\{\alpha_B^{(1)}, \alpha_D^{(2)}\} w) = \varepsilon, \quad \ldots \ldots \quad (134)$$

and thus that the region w is similar to W with regard to the set $\alpha^{(1)}, \alpha^{(2)}$.

We shall now introduce a conception which may be termed that of the independence of a family of hypersurfaces from a parameter.

Let

$$f_i(\alpha, x_1, x_2, \ldots x_n) = C_i \quad (i = 1, 2, \ldots k < n), \quad \ldots \ldots \quad (135)$$

be the equations of certain hypersurfaces in the n-dimensioned space, α and C_i being parameters. Denote by $S(\alpha, C_1, C_2, \ldots C_k)$ the intersection of these hypersurfaces, or if $k = 1$, the hypersurface corresponding to the equation (135). Consider the family of hypersurfaces $S(\alpha, C_1, C_2, \ldots C_k)$ corresponding to a fixed value of α and to all possible values of $C_1, C_2, \ldots C_k$. This will be denoted by $F(\alpha)$. Take any hypersurface $S(\alpha_1, C'_1, \ldots C'_k)$ from any family $F(\alpha_1)$. If whatever be α_2 it is possible to find suitable values of the C's, for example $C''_1, C''_2, \ldots C''_k$, such that the hypersurface $S(\alpha_2, C''_1, C''_2, \ldots C''_k)$ is identical with $S(\alpha_1, C'_1, \ldots C'_k)$, then we shall say that the family $F(\alpha)$ is independent of α. A simple illustration in 3-dimensioned space may be helpful. Let

$$f_1(\alpha, x_1, x_2, x_3) = (x_1 - \alpha)^2 + (x_2 - \alpha)^2 + (x_3 - \alpha)^2 = C_1 \quad \ldots \quad (136)$$

$$f_2(\alpha, x_1, x_2, x_3) = x_1 + x_2 + x_3 = C_2. \quad \ldots \ldots \quad (137)$$

These equations represent families of spheres and of planes. For given values of α, C_1 and C_2, $S(\alpha, C_1, C_2)$ will be a circle lying at right angles to the line $x_1 = x_2 = x_3$ and having its centre on that line. The family $F(\alpha)$ obtained by varying C_1 and C_2 consists of the set of all possible circles satisfying these conditions. This family is clearly independent of α—that is, of the position of the centre of the sphere on the line $x_1 = x_2 = x_3$—so that $F(\alpha)$ may be described as independent of α.

It is now possible to solve the problem of finding regions similar to W with regard to a set of parameters $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$, but we are at present only able to do so under rather limiting conditions. In the first place we shall have conditions analogous to those assumed when dealing with the case $c = 1$ (*see* p. 313).

(A) We shall assume the existence of $\dfrac{\partial^k p_0}{\partial (\alpha^{(i)})^k}$ in every point of the sample space except perhaps in a set of measure zero, and for all values of the α's.

(B) Writing

$$\phi_i = \frac{\partial \log p_0}{\partial \alpha^{(i)}} = \frac{1}{p_0}\frac{\partial p_0}{\partial \alpha^{(i)}}; \quad \phi'_i = \frac{\partial \phi_i}{\partial \alpha^{(i)}}, \quad \ldots \ldots \quad (138)$$

it will be assumed that for every $i = 1, 2, \ldots c$.

$$\phi'_i = A_i + B_i \phi_i, \quad \ldots \ldots \quad (139)$$

A_i and B_i being independent of the x's.

(C) Further there will need to be conditions concerning the hypersurfaces $\phi_i = \text{const.}$ Denote by $S(\alpha^{(i)}, C_1, C_2, \ldots C_{i-1})$ the intersection of the hypersurfaces

$$\phi_j = C_j, \qquad j = 1, 2, \ldots i - 1, \quad \ldots \ldots \quad (140)$$

corresponding to fixed values of the α's and C's. $F(\alpha^{(i)})$ will denote the family of hypersurfaces $S(\alpha^{(i)}, C_1, \ldots C_{i-1})$ corresponding to fixed values of the α's and to different systems of values of the C's.

We shall assume that any family $F(\alpha^{(i)})$ is independent of $\alpha^{(i)}$ for $i = 2, 3, \ldots c$.

It will be noted that the order in which the parameters α are numbered is of no importance and therefore that the above condition means simply that it is possible to find an order of the parameters α, such that each family $F(\alpha^{(i)})$ is independent of $\alpha^{(i)}$. An illustration of these points will be given in examples (10) and (11) below.

We shall now prove that if the above conditions are satisfied,* then regions similar to W with regard to the set of parameters $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$, and of any given size ε, do exist. In doing so, we shall show the actual process of construction of the most general region similar to W for any value of ε.

Assume that the function p_0 satisfies the above conditions and that w is a region similar to W with regard to $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$ and of the size ε, so that

$$P(\{\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}\} w) = \varepsilon, \quad \ldots \ldots \quad (141)$$

* We are aware that these conditions are more stringent than is necessary for the existence of similar regions; this is a point requiring further investigation.

whatever be the values of the α's. As we have seen above, the similarity with regard to the set of parameters means the similarity with regard to any one of them separately, as for example $\alpha^{(1)}$, for any set of values of the others. It follows from the results proved in the case $c = 1$, that w is built up of parts, $w(\phi)$ of hypersurfaces $\phi_1 = C_1$ corresponding to fixed values of parameters α and to different values of C_1. Each region $w(\phi_1)$ must satisfy the condition

$$P(w(\phi_1)) = \varepsilon P(W(\phi_1)), \quad \ldots \ldots \ldots \ldots \quad (142)$$

being otherwise unrestricted. We have assumed that the family $F(\alpha^{(2)})$ of hypersurfaces $S(\alpha^{(2)}, C_1)$ corresponding to the equations $\phi_1 = C_1$ is independent of $\alpha^{(2)}$. This of course does not mean that a particular hypersurface $S(\alpha^{(2)}, C_1)$ is independent of $\alpha^{(2)}$, but $W(\phi_1)$ and consequently $w(\phi_1)$ will mean a specified member of $F(\alpha^{(2)})$ fixed in the space W and independent of $\alpha^{(2)}$. It will correspond to some definite complex of values of α's and of C_1; if $\alpha^{(2)}$ is changed, then $W(\phi_1)$ will remain unchanged, though it will correspond to some other value of C_1. The possibility of decomposition of W and w into such regions $W(\phi_1)$, and $w(\phi_1)$ which do not change when $\alpha^{(2)}$ varies, is guaranteed by the condition that $F(\alpha^{(2)})$ is independent of $\alpha^{(2)}$.

We shall now use the condition that w is similar to W with regard to $\alpha^{(1)}$ whatever be the values of other parameters, and thus of $\alpha^{(2)}$. This means that the variation of $\alpha^{(2)}$ does not destroy the equation (142).

It follows that

$$\frac{\partial^k P(w(\phi_1))}{\partial(\alpha^{(2)})^k} = \varepsilon \frac{\partial^k P(W(\phi_1))}{\partial(\alpha^{(2)})^k}, \; k = 1, 2, 3. \ldots \ldots \ldots \quad (143)$$

As the regions $W(\phi_1)$ and $w(\phi_1)$ are independent of $\alpha^{(2)}$, these conditions may be written in the form

$$\int \ldots \int_{w(\phi_1)} \frac{\partial^k p}{\partial(\alpha^{(2)})^k} dw(\phi_1) = \varepsilon \int \ldots \int_{W(\phi_1)} \frac{\partial^k p}{\partial(\alpha^{(2)})^k} dW(\phi_1). \quad \ldots \quad (144)$$

We may now use the condition (139) for $i = 2$, and applying the method used when dealing with the case $c = 1$, show that (144) is equivalent to

$$\frac{1}{\varepsilon} \int \ldots \int_{w(\phi_1)} p\phi_2{}^k dw(\phi_1) = \int \ldots \int_{W(\phi_1)} p\phi_2{}^k dW(\phi_1). \quad \ldots \ldots \quad (145)$$

Following the same method of argument we find immediately that the necessary (and clearly also sufficient) condition for w being similar to W with regard to $\alpha^{(1)}$ and $\alpha^{(2)}$ is that

$$\int \ldots \int_{w(\phi_1, \phi_2)} p dw(\phi_1, \phi_2) = \varepsilon \int \ldots \int_{W(\phi_1, \phi_2)} p dW(\phi_1, \phi_2). \quad \ldots \quad (146)$$

where $W(\phi_1, \phi_2)$ means the intersection in the sample space W of the hypersurfaces $\phi_1 = C_1$ and $\phi_2 = C_2$ for any values of C_1 and C_2, and $w(\phi_1, \phi_2)$—the part of the same, contained in w.

It is easily seen that the same argument may be repeated $c-1$ times and that finally we shall find that the necessary and sufficient condition for w being similar to W with regard to the whole set of parameters $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$ is that

$$\left.\begin{array}{c}\int \cdots \int_{w(\phi_1, \phi_2, \ldots \phi_c)} p\,dw\,(\phi_1, \phi_2, \ldots \phi_c) = \varepsilon \int \cdots \int_{\mathrm{W}(\phi_1, \phi_2, \ldots \phi_c)} p\,d\mathrm{W}\,(\phi_1, \phi_2, \ldots \phi_c) \\ \text{or} \qquad \mathrm{P}\,(w\,(\phi_1, \phi_2, \ldots \phi_c)) = \varepsilon\,\mathrm{P}\,(\mathrm{W}\,(\phi_1, \phi_2, \ldots \phi_c))\end{array}\right\} . \quad (147)$$

Here W $(\phi_1, \phi_2, \ldots \phi_c)$ means the intersection in W of hypersurfaces $\phi_i = \mathrm{C}_i$, $(i = 1, 2, \ldots c)$ for fixed values of the α's and for any system of values of the C's. The symbol $w\,(\phi_1, \phi_2, \ldots \phi_c)$ means the part of W $(\phi_1, \phi_2, \ldots \phi_c)$ included in w.

Having established this result it is now easy to construct the most general region w, similar to W with regard to $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$, provided that the function p satisfies the above conditions.

We fix any system of values of parameters $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$ and consider the hypersurface W $(\phi_1, \phi_2, \ldots \phi_c)$ corresponding to some system of values of $\mathrm{C}_1, \ldots \mathrm{C}_c$. From this hypersurface we take an arbitrary part $w\,(\phi_1, \phi_2, \ldots \phi_c)$ satisfying only the condition (147). The aggregate of $w\,(\phi_1, \phi_2, \ldots \phi_c)$, corresponding to all different systems of values of $\mathrm{C}_1, \mathrm{C}_2, \ldots \mathrm{C}_c$ will be the region w required, similar to W and of the size ε.

In the section which follows it will be assumed that the function p satisfies the conditions (A), (B) and (C) under which we are able to construct the most general similar region. Though these conditions seem to be very limiting, there are many important cases in which they are satisfied, and in which it is therefore possible to treat the problem of best critical regions by this method.

(*b*) *The Determination of the Best Critical Region.*

The set Ω of admissible hypotheses will be defined by the probability law (66), depending on the $c + d$ parameters (67), each simple hypothesis specifying the values of all parameters. For the composite hypothesis H'_0 with c degrees of freedom the law, p_0, will be given by (69) and the parameters will fall into two groups as in (68). The best critical region w_0 of size ε with regard to a simple alternative H_t defined by (70) and (71), must satisfy the following conditions :—

(1) w_0 must be similar to W with regard to the c parameters $\alpha^{(1)}, \alpha^{(2)} \ldots \alpha^{(c)}$; that is to say

$$\mathrm{P}_0\,(w_0) = \varepsilon, \quad \ldots \ldots \ldots \quad (148)$$

must be independent of the values of the α's. This we have shown in the preceding section is equivalent under certain assumptions to the condition (147).

(2) If v be any other region of size ε similar to W with regard to the same parameters,

$$\mathrm{P}_t\,(w_0) \geq \mathrm{P}_t\,(v). \quad \ldots \ldots \ldots \quad (149)$$

As in the case where the probability law p_o depended only upon the value of one unspecified parameter $\alpha^{(1)}$, we can prove that if w_0 is a region maximising $P_t(w)$, then except perhaps for a set of values of $\phi_1, \phi_2, \ldots \phi_c$ of measure zero, the region $w_0(\phi_1, \phi_2, \ldots \phi_c)$ will have the property of maximising $P_t(w(\phi_1, \phi_2, \ldots \phi_c))$. That is to say, $P_t(w_0(\phi_1, \phi_2, \ldots \phi_c))$ will be greater than or at least equal to the integral of p_t taken over any other part of the region $W(\phi_1, \phi_2, \ldots \phi_c)$, satisfying (147). The proof is identical to that given in Section IV (*c*) and will not be repeated.

In this way the problem of finding the best critical region for testing H'_0 is reduced to that of maximising

$$P_t(w(\phi_1, \phi_2, \ldots \phi_c)) \quad \ldots \ldots \ldots \ldots \quad (150)$$

under the condition (147) for every set of values,

$$\phi_1 = C_1, \ \phi_2 = C_2, \ldots \phi_c = C_c, \quad \ldots \ldots \ldots \ldots \quad (151)$$

The problem does not differ essentially from that dealt with when considering a simple hypothesis (Section III (*a*)), and the resulting solution is as follows. The necessary and sufficient condition that $w_0(\phi_1, \phi_2, \ldots \phi_c)$ must satisfy in order to maximise (150) is that inside the region we should have

$$p_t \geq k(\phi_1, \phi_2, \ldots \phi_c)\, p_0, \quad \ldots \ldots \ldots \ldots \quad (152)$$

k being possibly a function of the values of ϕ's, which must be determined to satisfy (147).

Finally, therefore, the method at present advanced of finding a best critical region for testing a composite hypothesis H'_0 with c degrees of freedom may be summed up as follows.

We start by examining whether the limiting conditions assumed under the above theory are satisfied :

(A) The first condition concerns the indefinite differentiability of p_0 with regard to any parameter $\alpha^{(1)}, \alpha^{(2)}, \ldots \alpha^{(c)}$.

(B) Next we calculate

$$\phi_i = \frac{\partial \log p_0}{\partial (\alpha^{(i)})}, \quad \ldots \ldots \ldots \ldots \quad (153)$$

and see whether

$$\phi'_i = \frac{\partial(\phi_i)}{\partial(\alpha^{(i)})} = A_i + B_i \phi_i \quad \ldots \ldots \ldots \ldots \quad (154)$$

the coefficients A_i and B_i being independent of the sample variates $x_1, x_2, \ldots x_n$.

(C) Up to this stage the order in which the c parameters $\alpha^{(i)}$ are numbered is indifferent. Now we must consider whether it is possible to arrange this order in such a way, that (1) the family $F(\alpha^{(2)})$ of hypersurfaces $S(\alpha^{(2)}, C_1)$, corresponding to the equation

$$\phi_1 = C_1 \quad \ldots \ldots \ldots \ldots \quad (155)$$

is independent of $\alpha^{(2)}$; (2) that the family $F(\alpha^{(3)})$ of hypersurfaces $S(\alpha^{(3)}, C_1, C_2)$, each of which is an intersection of two hypersurfaces

$$\phi_1 = C_1 \, ; \ \phi_2 = C_2, \quad \ldots \ldots \ldots \ldots \quad (156)$$

is independent of $\alpha^{(3)}$; and so on in general for $\alpha^{(i)}$ until lastly we find that the family $F(\alpha^{(c)})$ of hypersurfaces $S(\alpha^{(c)}, C_1, C_2, \ldots C_{c-1})$, formed of points satisfying $c-1$ equations

$$\phi_1 = C_1, \ \phi_2 = C_2, \ \ldots \ \phi_{c-1} = C_{c-1}, \quad \ldots \ldots \ldots \quad (157)$$

is independent of $\alpha^{(c)}$.

If all these conditions are satisfied, then the best critical region w_0 of size ε with regard to a simple alternative, H_t, determining a frequency law p_t, must be built up of pieces, $w_0(\phi_1, \phi_2, \ldots \phi_c)$, of the hypersurfaces $W(\phi_1, \phi_2, \ldots \phi_c)$ on which the inequality (152) holds, the coefficient $k(\phi_1, \phi_2, \ldots \phi_c)$ being determined to satisfy (147).

We note that if the boundaries of the regions $w_0(\phi_1, \phi_2, \ldots \phi_c)$ are independent of the d additional parameters,

$$\alpha^{(c+1)}, \alpha^{(c+2)}, \ldots \alpha^{(c+d)}, \quad \ldots \ldots \ldots \ldots \quad (158)$$

specified in (67), then w_0 will be a common best critical region with regard to every H_t of the set Ω.

(c) *Illustrative Examples.*

We shall give two illustrations in which we suppose that two samples,

(1) Σ_1 of size n_1, mean $= \bar{x}_1$, standard deviation $= s_1$.

(2) Σ_2 of size n_2, mean $= \bar{x}_2$, standard deviation $= s_2$.

have been drawn at random from *some* normal populations. If this is so, the most general probability law for the observed event may be written

$$p(x_1, x_2, \ldots x_{n_1}; \ x_{n_1+1}, \ldots x_N) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \frac{1}{\sigma_1^{n_1}\sigma_2^{n_2}} e^{-n_1\frac{(\bar{x}_1-a_1)^2+s_1^2}{2\sigma_1^2} - n_2\frac{(\bar{x}_2-a_2)^2+s_2^2}{2\sigma_2^2}} \quad (159)$$

where $n_1 + n_2 = N$, and a_1, σ_1 are the mean and standard deviation of the first, and a_2, σ_2 of the second sampled population.

(1) *Example* (10).—*The test for the significance of the difference between two variances.*—The admissible simple hypotheses include pairs of sampled populations for which $a_1, a_2, \sigma_1 > 0_1, \sigma_2 > 0$ may have any values whatever. H'_0 is the composite hypothesis that $\sigma_1 = \sigma_2$. This is the test for the significance of the difference between the variances in two independent samples. The parameters may be defined as follows:

For a simple alternative H_t:

$$\alpha_t^{(1)} = a_1 \, ; \ \alpha_t^{(2)} = a_2 - a_1 = b_t \, ; \ \alpha_t^{(3)} = \sigma_1 \, ; \ \alpha_t^{(4)} = \theta_t = \frac{\sigma_2}{\sigma_1}. \quad \ldots \quad (160)$$

For the hypothesis to be tested, H'_0:

$$\alpha^{(1)} = a\ ;\ \alpha^{(2)} = b\ ;\ \alpha^{(3)} = \sigma\ ;\ \alpha_0^{(4)} = 1. \quad \quad (161)$$

H'_0 it will be seen, is a composite hypothesis with 3 degrees of freedom, a, b and σ being unspecified.

We shall now consider whether the conditions (A), (B) and (C) of the above theory are satisfied.

(A) The condition of differentiability of p_0 with regard to all parameters, for all values of $\alpha^{(1)}$ and $\alpha^{(2)}$ and for $\alpha^{(3)} > 0$, is obviously satisfied.

(B) Making use of (159), we find that

$$\log p_0 = -\mathrm{N}\log\sqrt{2\pi} - \mathrm{N}\log\sigma - \frac{1}{2\sigma^2}\{n_1(\bar{x}_1 - a)^2 + n_2(\bar{x}_2 - a - b)^2 + n_1 s_1^2 + n_2 s_2^2\}, \quad (162)$$

$$\phi_1 = \frac{\partial \log p_0}{\partial a} = \frac{1}{\sigma^2}\{n_1(\bar{x}_1 - a) + n_2(\bar{x}_2 - a - b)\}, \quad (163)$$

$$\phi_2 = \frac{\partial \log p_0}{\partial b} = \frac{1}{\sigma^2}n_2(\bar{x}_2 - a - b), \quad (164)$$

$$\phi_3 = \frac{\partial \log p_0}{\partial \sigma} = -\frac{\mathrm{N}}{\sigma} + \frac{1}{\sigma^3}\{n_1(\bar{x}_1 - a)^2 + n_2(\bar{x}_2 - a - b)^2 + n_1 s_1^2 + n_2 s_2^2\}. \quad (165)$$

We see that

$$\phi'_1 = \mathrm{A}_1, \quad \phi'_2 = \mathrm{A}_2, \quad \phi'_3 = \mathrm{A}_3 + \mathrm{B}_3\phi_3, \quad (166)$$

where the A's and B's are independent of the sample variates, so that the condition (B) is satisfied.

(C) The equation of the hypersurface $\mathrm{S}(\alpha^{(2)}, \mathrm{C}_1)$, namely,

$$\phi_1 = \mathrm{C}_1 \quad (167)$$

where C_1 is an arbitrary constant, is obviously equivalent to the following,

$$n_1\bar{x}_1 + n_2\bar{x}_2 = \mathrm{C}'_1 \quad (168)$$

depending only upon one arbitrary parameter C'_1. Hence the family of these hypersurfaces, $\mathrm{F}(\alpha^{(2)})$, is independent of $\alpha^{(2)}$.

Similarly the equation $\phi_2 = \mathrm{C}_2$, in which C_2 is arbitrary, is equivalent to

$$\bar{x}_2 = \mathrm{C}'_2 \quad (169)$$

in which C'_2 is arbitrary. The intersections $\mathrm{S}(\alpha^{(3)}, \mathrm{C}_1, \mathrm{C}_2)$ of (168) and (169), which satisfy also the equations

$$\bar{x}_1 = \text{const.}; \quad \bar{x}_2 = \text{const.} \quad (170)$$

form a family, $\mathrm{F}(\alpha^{(3)})$, which is independent of $\alpha^{(3)}$.

Hence also the condition (C) is fulfilled, and we may now attempt to construct the best critical region w_0. Its elements $w_0(\phi_1, \phi_2, \phi_3)$ are parts of the hypersurfaces $W(\phi_1, \phi_2, \phi_3)$, satisfying the system of three equations $\phi_i = C_i$ ($i = 1, 2, 3$), containing certain fixed values of the $\alpha^{(i)}$ and arbitrary values of the constants C_i. Instead of this system of equations we may use the following which is equivalent

$$\bar{x}_1 = \text{const.} \qquad \ldots \ldots \ldots \ldots \ldots \quad (171)$$

$$\bar{x}_2 = \text{const.} \qquad \ldots \ldots \ldots \ldots \ldots \quad (172)$$

$$\frac{1}{N}(n_1 s_1^2 + n_2 s_2^2) = s_a^2 = \text{const.} \qquad \ldots \ldots \ldots \quad (173)$$

The element $w_0(\phi_1, \phi_2, \phi_3)$ is the part of $W(\phi_1, \phi_2, \phi_3)$, within which

$$p_t \geq k(\bar{x}_1, \bar{x}_2, s_a)\, p_0, \qquad \ldots \ldots \ldots \ldots \quad (174)$$

the value of k being determined for each system of values of $\bar{x}_1, \bar{x}_2, s_a$, so that

$$P_0(w_0(\phi_1, \phi_2, \phi_3)) = \varepsilon\, P_0(W(\phi_1, \phi_2, \phi_3)). \qquad \ldots \ldots \quad (175)$$

The condition (174) becomes

$$\frac{1}{\sigma^N} e^{-\frac{n_1[(\bar{x}_1-a)^2+s_1^2]+n_2[(\bar{x}_2-a-b)^2+s_2^2]}{2\sigma^2}} \leq \frac{k}{\sigma_1^N \theta^{n_2}} e^{-\frac{n_1[(\bar{x}_1-a_1)^2+s_1^2]+n_2\theta^{-2}[(\bar{x}_2-a_1-b_1)^2+s_2^2]}{2\sigma_1^2}} \ldots \quad (176)$$

Since the region determined by (175) will be similar to W with regard to a, b and σ we may put $a = a_1$, $b = b_1$, $\sigma = \sigma_1$, and the condition (176) will be found on taking logarithms to reduce to

$$n_2\{(\bar{x}_2 - a_1 - b_1)^2 + s_2^2\}(1 - \theta^2) \leq 2\sigma_1^2\theta^2(\log k - n_2 \log \theta). \quad \ldots \quad (177)$$

Since this inequality must hold good on the hypersurface on which $\bar{x}_2$ is constant, it contains only one variable, namely, s_2^2. Solving with regard to s_2^2 we find that the solution will depend upon the sign of the difference $1 - \theta^2$. Accordingly we shall have to consider separately the two classes of alternatives,

(a) $\theta = \frac{\sigma_2}{\sigma_1} > 1$; the B.C.R. will be defined by $s_2^2 \geq k'_1(\bar{x}_1, \bar{x}_2, s_a^2)$ . . (178)

(b) $\theta = \frac{\sigma_2}{\sigma_1} < 1$; the B.C.R. will be defined by $s_2^2 \leq k'_2(\bar{x}_1, \bar{x}_2, s_a^2)$, . . (179)

where k' stands for

$$\frac{2\sigma_1^2\theta^2(\log k - n_2 \log \theta)}{n_2(1 - \theta^2)} - (\bar{x}_2 - a_1 - b_1)^2. \qquad \ldots \ldots \quad (180)$$

The problem of finding the best critical region w_0 of size ε consists now in determining k' so as to satisfy (175). This condition may be expressed as follows: Since W (ϕ_1, ϕ_2, ϕ_3) is the locus of points in which $\bar{x}_1$, $\bar{x}_2$ and s_a^2 have certain fixed values, the right-hand side of (175) is the product of ε and the corresponding value of the frequency function of the three variates $\bar{x}_1$, $\bar{x}_2$ and s_a^2 or $p_0(\bar{x}_1, \bar{x}_2, s_a^2)$. The left-hand side of the same equation is the integral of p_0 over that part of W (ϕ_1, ϕ_2, ϕ_3) upon which s_2^2 satisfies either (178) or (179). To calculate this expression, we may start with the frequency function $p_0(\bar{x}_1, \bar{x}_2, s_a^2, s_2^2)$. Then

$$P_0(w(\phi_1, \phi_2, \phi_3)) = \int_{k'_1}^{k''} p_0(\bar{x}_1, \bar{x}_2, s_a^2, s_2^2)\, ds_2^2, \text{ in the case } (a), \quad . \quad . \quad (181)$$

or

$$P_0(w(\phi_1, \phi_2, \phi_3)) = \int_{k'''}^{k'_2} p_0(\bar{x}_1, \bar{x}_2, s_a^2, s_2^2)\, ds_2^2, \text{ in the case } (b). \quad . \quad . \quad (182)$$

Here k'' and k''' are the upper and the lower limits of variation of s_2^2 for fixed values of $\bar{x}_1$, $\bar{x}_2$ and s_a^2. Further, we shall have

$$P_0(W(\phi_1, \phi_2, \phi_3)) = p_0(\bar{x}_1, \bar{x}_2, s_a^2)) = \int_{k'''}^{k''} p_0(\bar{x}_1, \bar{x}_2, s_a^2, s_2^2)\, ds_2^2. \quad . \quad . \quad . \quad (183)$$

It is known that

$$p_0(\bar{x}_1, \bar{x}_2, s_1^2, s_2^2) = \frac{\text{const.}}{\sigma_1^{N}} s_1^{n_1-3}\ s_2^{n_2-3}\ e^{-\frac{n_1(\bar{x}_1-a_1)^2+n_2(\bar{x}_2-a_2)^2+Ns_a^2}{2\sigma_1^2}}. \quad . \quad (184)$$

Introducing s_a^2 instead of s_1^2 as a new variate, we have

$$s_1^2 = \frac{1}{n_1}(Ns_a^2 - n_2 s_2^2) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (185)$$

and

$$p_0(\bar{x}_1, \bar{x}_2, s_a^2, s_2^2) = \text{const.}\,(Ns_a^2 - n_2 s_2^2)^{\frac{n_1-3}{2}}\ s_2^{n_2-3}\ e^{-\frac{n_1(\bar{x}_1-a_1)^2+n_2(\bar{x}_2-a_2)^2+Ns_a^2}{2\sigma_1^2}}. \quad (186)$$

It is easily seen that

$$k'' = \frac{N}{n_2}s_a^2, \quad k''' = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (187)$$

Therefore, using (181), (183) and (186) we have from (175), after cancelling equal multipliers on both sides,

$$\int_{k'_1}^{k''} (Ns_a^2 - n_2 s_2^2)^{\frac{n_1-3}{2}}\ s_2^{n_2-3}\ ds_2^2 = \varepsilon \int_0^{k''} (Ns_a^2 - n_2 s_2^2)^{\frac{n_1-3}{2}}\ s_2^{n_2-3}\ ds_2^2 \qquad (188)$$

for case (a), and an analogous equation for case (b). Write

$$n_2 s_2^2 = Ns_a^2 u \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (189)$$

where u is the new variate. Then instead of (188) we shall have

$$\int_{u_0}^{1} (1-u)^{\frac{n_1-3}{2}} u^{\frac{n_2-3}{2}} du = \varepsilon \mathrm{B}\left(\frac{n_1-1}{2}, \frac{n_2-1}{2}\right) = \int_{0}^{u'_0} (1-u)^{\frac{n_1-3}{2}} u^{\frac{n_2-3}{2}} du, \quad . \quad (190)$$

where

$$u_0 = \frac{n_2 k'_1}{\mathrm{N}s_a^2}, \; u'_0 = \frac{n_2 k'_2}{\mathrm{N}s_a^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (191)$$

It follows from (190) that u_0 and u'_0 depend only upon n_1, n_2 and ε. Therefore, whatever be $\bar{x}_1$, $\bar{x}_2$ and s_a^2, the element $w(\phi_1, \phi_2, \phi_3)$ of the best critical region is defined by the inequality

$$s_2^2 \geq k'_1 = u_0 \frac{\mathrm{N}s_a^2}{n_2}, \qquad \text{in case } (a) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (192)$$

or

$$s_2^2 \leq k'_2 = u'_0 \frac{\mathrm{N}s_a^2}{n_2}, \qquad \text{in case } (b). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (193)$$

These two inequalities are equivalent to the following—

$$(a) \text{ For alternatives } \sigma_2 > \sigma_1 ; u = \frac{n_2 s_2^2}{n_1 s_1^2 + n_2 s_2^2} \geq u_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (194)$$

$$(b) \text{ For alternatives } \sigma_2 < \sigma_1 ; u = \frac{n_2 s_2^2}{n_1 s_1^2 + n_2 s_2^2} \leq u'_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (195)$$

which define the best critical regions in the two cases. We see that they are common for all the alternatives, included in each class (a) and (b). The constant u_0 depends only upon n_1 and n_2 and the value of ε chosen; it may be found from the incomplete beta-function integral (190), or from any suitable transformation, as for example, that to FISHER's z-function.*

Approaching the problem of testing whether the variances in two samples are significantly different, from the point of view of the best critical region, we have reached the criterion u, which is equivalent to that suggested on intuitive grounds by FISHER. This criterion is also that obtained by applying the principle of likelihood, but that method did not bring out clearly the need for distinction between the two classes of alternatives, since $\lambda < \lambda_0$ at both ends of the u-distribution.†

(2) *Example* (11).—*The test for the significance of the difference between two means.*—We have again two random and independent samples, Σ_1 and Σ_2, from normal populations, and the set Ω of admissible hypotheses includes pairs of populations in which the standard deviations σ_1 and σ_2 have the same (but unspecified) value $\sigma > 0$, while the means a_1 and $a_2 = a_1 + b$, may have any values whatever. H'_0 is the composite hypothesis with 2 degrees of freedom, that $b_0 = 0$.

* FISHER. "Statistical Methods for Research Workers," London, 1932. This contains tables giving z_0 (a function of u_0) for $\varepsilon = 0{\cdot}05$ and $0{\cdot}01$.

† PEARSON and NEYMAN. "On the Problem of Two Samples." 'Bull. Acad. Polon. Sci. Lettres' (1930).

The test of the hypothesis H'_0 is the test for significance of a difference between two means, in the case where it is known that the sampled populations have common variances.

Any simple alternative, H_t, of the set Ω specifies the parameters

$$\alpha_t^{(1)} = a_1^{(t)}\,;\ \alpha_t^{(2)} = \sigma_t\,;\ \alpha_t^{(3)} = a_2^{(t)} - a_1^{(t)} = b_t \quad \ldots\ldots \quad (196)$$

H'_0, the composite hypothesis to be tested, specifies only one parameter

$$\alpha_0^{(3)} = b_0 = 0, \quad \ldots\ldots \quad (197)$$

the others, $\alpha^{(1)} = a_1$ and $\alpha^{(2)} = \sigma$, being arbitrary.

Besides the symbols previously defined, $\bar{x}_1$, s_1; $\bar{x}_2$, s_2 (mean and standard deviation of each sample), we shall need the following: $\bar{x}_0$, the mean, and s_0, the standard deviation of the sample of $N = n_1 + n_2$ individuals formed by putting together the two samples Σ_1 and Σ_2.

The probability law p_0 will be given by

$$\log p_0 = -N \log \sqrt{2\pi} - N \log \sigma - \frac{N}{2\sigma^2}\{(\bar{x}_0 - a_1)^2 + s_0^2\}. \quad \ldots \quad (198)$$

The condition (A) is obviously satisfied by p_0, and following the same line of argument as in example (10) we find that the other conditions (B) and (C) are satisfied also. In fact,

$$\phi_1 = \frac{\partial \log p_0}{\partial \alpha^{(1)}} = \frac{N}{\sigma^2}(\bar{x}_0 - a_1) \quad \ldots\ldots \quad (199)$$

$$\phi_2 = \frac{\partial \log p_0}{\partial \alpha^{(2)}} = -\frac{N}{\sigma} + \frac{N}{\sigma^3}\{(\bar{x}_0 - a_1)^2 + s_0^2\}, \quad \ldots\ldots \quad (200)$$

and it is easy to see that ϕ'_1 and ϕ'_2 are linear functions of the corresponding ϕ's. Now the equation, $\phi_1 =$ constant, is equivalent to

$$\bar{x}_0 = C_1, \quad \ldots\ldots \quad (201)$$

C_1 being an arbitrary constant. Clearly the family $F(\alpha^{(2)})$ of hypersurfaces $W(\phi_1)$ corresponding to this last equation is independent of $\alpha^{(2)}$, and hence we conclude that the best critical region w_0 of size ε may be built up of elements $w_0(\phi_1, \phi_2)$. To obtain such an element, we have to find the hypersurface $W(\phi_1, \phi_2)$, which is the locus of points in which

$$\phi_1 = \text{const.}, \quad \phi_2 = \text{const.} \quad \ldots\ldots \quad (202)$$

and to determine its part, satisfying the conditions

$$p_t \geq k(\phi_1, \phi_2)\, p_0, \quad \ldots\ldots \quad (203)$$

$$\int \cdots \int_{w_0(\phi_1, \phi_2)} p_0\, dw_0(\phi_1, \phi_2) = \varepsilon \int \cdots \int_{W(\phi_1, \phi_2)} p_0\, dW(\phi_1, \phi_2). \quad \ldots \quad (204)$$

Now the system of equations (202) determining W (ϕ_1, ϕ_2) is equivalent to the following—

$$\bar{x}_0 = C_1 \quad \ldots \ldots \ldots \quad (205)$$

$$s_0^2 = C_2 \geq 0. \quad \ldots \ldots \ldots \quad (206)$$

As the best critical region is independent of a_1 and σ we may put into p_0

$$a_1 = a_1^{(t)}, \quad \sigma = \sigma_t, \quad \ldots \ldots \ldots \quad (207)$$

and taking into account the fact that (205) and (206) must hold good on the hyper-surface, W (ϕ_1, ϕ_2), the condition (203) may be transformed into the equivalent

$$b_t (\bar{x}_1 - \bar{x}_2) \leq k' (\bar{x}_0, s_0^2), \quad \ldots \ldots \ldots \quad (208)$$

k' being a new constant depending upon $\bar{x}_0$ and s_0^2 (that is to say upon ϕ_1 and ϕ_2), and upon ε.

Again two classes of alternatives H_t must be considered separately: (*a*) if $b_t = a_2^{(t)} - a_1^{(t)} > 0$, then the region $w_0 (\phi_1, \phi_2)$ will be defined by

$$v = \bar{x}_1 - \bar{x}_2 \leq k''_1 (\bar{x}_0, s_0^2). \quad \ldots \ldots \ldots \quad (209)$$

(*b*) if, however, $b_t = a_2^{(t)} - a_1^{(t)} < 0$, then instead of the above inequality, we shall have

$$v = \bar{x}_1 - \bar{x}_2 \geq k''_2 (\bar{x}_0, s_0^2). \quad \ldots \ldots \ldots \quad (210)$$

The solutions in both cases are analogous, so we shall consider only the case (*a*). The problem consists in determining $k''_1 (\bar{x}_0, s_0^2)$ so as to satisfy the condition (204). This is equivalent to the equation

$$\int_{k'''}^{k''_1} p_0 (\bar{x}_0, s_0^2, v)\, dv = \varepsilon \int_{k'''}^{k^{\text{iv}}} p_0 (\bar{x}_0, s_0^2, v)\, dv, \quad \ldots \ldots \quad (211)$$

where $p_0 (x_0, s_0^2, v)$ is the frequency function of the variates $\bar{x}_0, s_0^2$ and v, and k''' and k^{iv} are the lower and the upper limits, respectively, of the variation in v for fixed values of $\bar{x}_0$ and s_0^2.

Thus we have to find $p_0 (\bar{x}_0, s_0^2, v)$. We start with the frequency function of the variates $\bar{x}_1, \bar{x}_2, s_1^2, s_2^2$, namely:

$$p_0 (\bar{x}_1, \bar{x}_2, s_1^2, s_2^2) = C\, s_1^{n_1-3}\, s_2^{n_2-3}\, e^{-N \frac{(\bar{x}_0 - a_1)^2 + s_0^2}{2\sigma^2}}, \quad \ldots \ldots \quad (212)$$

C being a constant. Substituting in (212)

$$\bar{x}_1 = \bar{x}_0 + \frac{n_2}{N} v, \quad \ldots \ldots \ldots \quad (213)$$

$$\bar{x}_2 = \bar{x}_0 - \frac{n_1}{N} v, \quad \ldots \ldots \ldots \quad (214)$$

$$s_2^2 = \frac{1}{n_2}\left(N s_0^2 - n_1 s_1^2 - \frac{n_1 n_2}{N} v^2\right), \quad \ldots \ldots \ldots \quad (215)$$

and multiplying by the absolute value of the Jacobian

$$\left|\frac{\partial\,(\bar{x}_1, \bar{x}_2, s_2^2)}{\partial\,(\bar{x}_0, s_0^2, v)}\right| = \frac{N}{2} \quad \ldots \ldots \ldots \ldots \ldots \quad (216)$$

we obtain the frequency function of $\bar{x}_0$, s_0^2, v, s_1^2, namely,

$$p_0\,(\bar{x}_0, s_0^2, s_1^2, v) = C_1 s_1^{n_1-3}\left(Ns_0^2 - n_1 s_1^2 - \frac{n_1 n_2}{N}\,v^2\right)^{\frac{n_2-3}{2}} e^{-N\frac{(\bar{x}_0-a_1)^2+s_0^2}{2\sigma^2}}. \quad (217)$$

We note that for fixed values of s_0^2 and v, the variate s_1^2 may lie between limits zero and

$$s'^2_1 = \left(Ns_0^2 - \frac{n_1 n_2}{N}\,v^2\right)\frac{1}{n_1}. \quad \ldots \ldots \ldots \quad (218)$$

The frequency function $p_0\,(\bar{x}_0, s_0^2, v)$ is found from $p_0\,(\bar{x}_0, s_0^2, s_1^2, v)$ by integrating it with regard to s_1^2, between the limits zero and s'^2_1. We have thus

$$p_0\,(\bar{x}_0, s_0^2, v) = C_2\left(s_0^2 - \frac{n_1 n_2}{N^2}\,v^2\right)^{\frac{N-4}{2}} e^{-N\frac{(\bar{x}_0-a_1)^2+s_0^2}{2\sigma^2}}. \quad \ldots \quad (219)$$

Putting this into the equation (211), and cancelling on both sides equal constants, we find

$$\int_{k'''}^{k''_1}\left(s_0^2 - \frac{n_1 n_2}{N^2}\,v^2\right)^{\frac{N-4}{2}} dv = \varepsilon\int_{k'''}^{k^{\text{iv}}}\left(s_0^2 - \frac{n_1 n_2}{N^2}\,v^2\right)^{\frac{N-4}{2}} dv = 2\varepsilon\int_0^{k^{\text{iv}}}\left(s_0^2 - \frac{n_1 n_2}{N}\,v^2\right)^{\frac{N-4}{2}} dv, \quad (220)$$

where

$$k''' = -\frac{Ns_0}{\sqrt{n_1 n_2}}, \quad k^{\text{iv}} = +\frac{Ns_0}{\sqrt{n_1 n_2}}. \quad \ldots \ldots \quad (221)$$

Make now the transformation

$$v = \frac{Ns_0}{\sqrt{n_1 n_2}}\,\frac{z}{\sqrt{1+z^2}}, \quad \ldots \ldots \ldots \ldots \quad (222)$$

z being a new variable. Since s_0 is constant on W (ϕ_1, ϕ_2)

$$dv = \frac{Ns_0}{\sqrt{n_1 n_2}}\,\frac{dz}{(1+z^2)^{3/2}}, \quad \ldots \ldots \ldots \quad (223)$$

and the equation (220) becomes

$$\int_{-\infty}^{z'_0}(1+z^2)^{-\frac{N-1}{2}}\,dz = \varepsilon\sqrt{\pi}\,\frac{\Gamma\left(\frac{N-2}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)} \quad \text{for case } (a). \quad \ldots \quad (224)$$

Similarly

$$\int_{z''_0}^{\infty}(1+z^2)^{-\frac{N-1}{2}}\,dz = \varepsilon\sqrt{\pi}\,\frac{\Gamma\left(\frac{N-2}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)} \quad \text{for case } (b). \quad \ldots \quad (225)$$

z'_0 and z''_0 are functions of k''_1 and k''_2 defined by (209) and (222). It follows from (224) and (225) that they can depend only on N and ε. Thus the best critical region is defined as follows :

(*a*) For alternatives $b_t > 0$;

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{\mathrm{N}}} \leq z'_0. \quad \ldots \ldots \ldots \quad (226)$$

(*b*) For alternatives $b_t < 0$;

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{\mathrm{N}}} \geq z''_0 \quad \ldots \ldots \ldots \quad (227)$$

whatever be $\bar{x}_0$ and s_0, where z_0 may be found from published tables for any ε chosen.*

We thus reach the well-known extension of " STUDENT'S " test given by FISHER,† who, however, uses instead of z,

$$t = z\sqrt{\mathrm{N} - 2}. \quad \ldots \ldots \ldots \ldots \ldots \quad (228)$$

It also follows from the principle of likelihood. Again it has been shown that on the basis of the information available no better test could be devised for the hypothesis under consideration.

VI.—SUMMARY OF RESULTS.

1. A new basis has been introduced for choosing among criteria suitable for testing any given statistical hypothesis, H_0, with regard to an alternative H_t. If θ_1 and θ_2 are two such possible criteria and if in using them there is the same chance, ε, of rejecting H_0 when it is in fact true, we should choose that one of the two which assures the minimum chance of accepting H_0 when the true hypothesis is H_t.

2. Starting from this point of view, since the choice of a criterion is equivalent to the choice of a critical region in multiple space, it was possible to introduce the conception of the best critical region with regard to the alternative hypothesis H_t. This is the region, the use of which, for a fixed value of ε, assures the minimum chance of accepting H_0 when the true hypothesis is H_t. The criterion, based on the best critical region, may be referred to as to the most efficient criterion with regard to the alternative H_t.

3. It has been shown that the choice of the most efficient criterion, or of the best critical region, is equivalent to the solution of a problem in the Calculus of Variations. We give the solution of this problem for the case of testing a simple hypothesis.

To solve the same problem in the case where the hypothesis tested is composite, the solution of a further problem is required ; this consists in determining what has been called a region similar to the sample space with regard to a parameter.

* For z : " Tables for Statisticians and Biometricians," Part I, Table XXV ; Part II, Table XXV.

† For t : 'Metron,' vol. 5, p. 114 (1926) ; FISHER, " Statistical Methods for Research Workers," p. 139.

We have been able to solve this auxiliary problem only under certain limiting conditions; at present, therefore, these conditions also restrict the generality of the solution given to the problem of the best critical region for testing composite hypotheses.

4. An important case arises, when the best critical regions are identical with regard to a certain class of alternatives, which may be considered to include all admissible hypotheses. In this case—which, as has been shown by several examples, is not an uncommon one—unless we are in a position to assign precise measures of *a priori* probability to the simple hypotheses contained in the composite H_0, it appears that no more efficient test than that given by the best critical region can be devised.

5. The question of the choice of a "good critical region" for testing a hypothesis, when there is no common best critical region with regard to every alternative admissible hypothesis, remains open. It has, however, been shown that the critical region based on the principle of likelihood satisfies our intuitive requirements of a "good critical region."

6. The method of finding best critical regions for testing both simple and composite hypotheses has been illustrated for several important problems commonly met in statistical analysis. Owing to the considerable size which the paper has already reached, the solution of the same problem for other important types of hypotheses must be left for separate publication.