On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I

Author(s): J. Neyman and E. S. Pearson

Source: *Biometrika*, Jul., 1928, Vol. 20A, No. 1/2 (Jul., 1928), pp. 175-240

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2331945

ON THE USE AND INTERPRETATION OF CERTAIN TEST CRITERIA FOR PURPOSES OF STATISTICAL INFERENCE. PART I.

By J. NEYMAN, Ph.D. and E. S. PEARSON, D.Sc.

CONTENTS.

I. Introductory.

One of the most common as well as most important problems which arise in the interpretation of statistical results, is that of deciding whether or not a particular sample may be judged as likely to have been randomly drawn from a certain population, whose form may be either completely or only partially specified. We may term Hypothesis A the hypothesis that the population from which the sample Σ has been randomly drawn is that specified, namely Π. In general the method of procedure is to apply certain tests or criteria, the results of which will enable the investigator to decide with a greater or less degree of confidence whether to accept or reject Hypothesis A, or, as is often the case, will show him that further data are required before a decision can be reached. At first sight the problem may be thought to be a simple one, but upon fuller examination one is forced to the conclusion that in many cases there is probably no single "best" method of solution. The sum total of the reasons which will weigh with the investigator in accepting or rejecting the hypothesis can very rarely be expressed in numerical terms. All that is possible for him is to balance the results of a mathematical summary, formed upon certain assumptions, against other less precise impressions based upon *à priori* or *à posteriori* considerations. The tests themselves give no final verdict, but as tools help the worker who is using them to form his final decision; one man may prefer to use one method, a second another, and yet in the long run there may be little to choose between the value of their conclusions. What is of chief importance in order that a sound judgment may be formed is that the method adopted, its scope and its limitations, should be clearly understood, and it is because we believe this often not to be the case that it has seemed worth while to us to discuss the principles involved in some detail and to illustrate their application to certain important sampling tests.

There are two distinct methods of approach, one to start from the population Π, and to ask what is the probability that a sample such as Σ should have been drawn from it, and the other the inverse method of starting from Σ and seeking the probability that Π is the population sampled. The first is the more customary method of approach, partly because it seems natural to take Π as the point of departure since in practice there are often strong *à priori* grounds for believing that this is the population sampled, and partly because there is a common tendency to view with suspicion any method involving the use of inverse probability. But in fact, however strong may be the *à priori* evidence in favour of Π, there would be no problem at all to answer if we were not prepared to consider the possibility of alternative hypotheses as to the population sampled; and we shall find that it is impossible to follow the first method very far without introducing certain ideas of inverse probability—that is to say, arguing from the sample to the population. If on the other hand we start boldly with assumptions regarding *à priori* and *à posteriori* probability, we reach by an almost simpler method sampling tests very nearly equivalent to those obtained from the first starting-point. Indeed the inverse method may be considered by some the more logical of the two; we shall consider first however the other solution.

Perhaps the most suggestive method of description is to represent Σ by a point in a hyperspace whose dimensions will depend upon the particular problem considered; and to associate the criteria for acceptance or rejection with a system of contours in this space, so chosen that in moving out from contour to contour Hypothesis A becomes less and less probable*. The frequency with which the sample corresponding to a particular point will occur in random sampling from Π may be represented by giving to the space an appropriate "point-density." Thus the chance of drawing a sample whose representative point lies within a certain

* Here and later the term "probability" used in connection with Hypothesis A must be taken in a very wide sense. It cannot necessarily be described by a single numerical measure of inverse probability; as the hypothesis becomes "less probable," our confidence in it decreases, and the reason for this lies in the meaning of the particular contour system that has been chosen.

region in the space is proportional to the density integrated throughout that region, or in other words to the "weight" of the region. The contours are not necessarily contours of equal density, but may be surfaces or regions throughout which some statistical measure such as the mean or standard deviation remains at a constant "level." Although it is impossible to visualise a space of high dimensions, it will be found that with the help of the analogy of a three-dimensional density space this method of description is often a very considerable aid to purely algebraic discussion.

Setting aside the possibility that the sampling has not been random or that the population has changed during its course*, Σ must either have been drawn randomly from Π or from Π', where the latter is some other population which may have any one of an infinite variety of forms differing only slightly or very greatly from Π. The nature of the problem is such that it is impossible to find criteria which will distinguish exactly between these alternatives, and whatever method we adopt two sources of error must arise:

(1) Sometimes, when Hypothesis A is rejected, Σ will in fact have been drawn from Π.

(2) More often, in accepting Hypothesis A, Σ will really have been drawn from Π'.

In the long run of statistical experience the frequency of the first source of error (or in a single instance its probability) can be controlled by choosing as a discriminating contour, one outside which the frequency of occurrence of samples from Π is very small—say, 5 in 100 or 5 in 1000. In the density space such a contour will include almost the whole weight of the field. Clearly there will be an infinite variety of systems from which it is possible to choose a contour satisfying such a condition. For example there will be the system of contours upon any one of which (*a*) the mean, or (*b*) the standard deviation, or (*c*) the ratio of mean to standard deviation, of Σ is constant.

The second source of error is more difficult to control, but if wrong judgments cannot be avoided, their seriousness will at any rate be diminished if on the whole Hypothesis A is wrongly accepted only in cases where the true sampled population, Π', differs but slightly from Π. It is not of course possible to determine Π', but making use of some clearly defined conception of probability we may determine a "probable" or "likely" form of it, and hence fix the contours so that in moving "inwards" across them the difference between Π and the population from which it is "most likely" that Σ has been sampled should become less and less. This choice also implies that on moving "outwards" across the contours, other hypotheses as to the population sampled become more and more likely than Hypothesis A.

* A non-random or biased sample from Π may appear to be a random sample from Π'; in practice it may be very important to be able to attribute Σ to Π rather than Π'. But clearly no general rules to cover such cases can be given, for the position must depend in each case on the possible forms of bias that may have occurred.

Both these aspects of the problem must be taken into account. Using only the first control, any given sample could always be found to lie outside an extremely divergent contour of *some* system. That is to say a criterion of any desired degree of stringency could always be found for a sample. But regarding the position from the second point of view, it is seen that there will only be certain systems which are of any value as criteria.

The application of these principles will become clearer when illustrated in the cases of particular sampling tests, but it seems well to emphasise at the outset the importance of careful thinking in these matters. For example it might readily be supposed that if, in sampling from Π, a sample of form Σ_1 occurs more frequently than one of form Σ_2, then greater confidence in Hypothesis A would be justified if Σ_1 were drawn rather than Σ_2. But we shall see that under certain conditions this may not be the case, because in the first event alternative hypotheses are relatively far more probable than A than they are in the second. It is indeed obvious, upon a little consideration, that the mere fact that a particular sample may be expected to occur very rarely in sampling from Π would not in itself justify the rejection of the hypothesis that it had been so drawn, if there were no other more probable hypotheses conceivable.

II. Sampling from a Normal Population.

(1) *Description of the Fundamental Space.*

We shall discuss first the case of sampling from an "infinite" population in which the single variable under consideration follows a Normal Distribution, and we shall suppose that this population, Π, is completely specified, its mean a and standard deviation σ being known. We have a sample, Σ, of n observations, $X_1, X_2, \dots X_n$, and the hypothesis whose probability we wish to test is that Σ is a random sample from Π; this is Hypothesis A. The methods to be discussed are perfectly general, but we have particularly in mind the case of small samples where the data are insufficient for the application of the ordinary (P, χ^2) test for goodness of fit.

Making use of the geometrical method first introduced into this problem by R. A. Fisher*, we shall imagine an n-dimensioned space in which we take an origin at the point $O\ (a, a, \dots a)$ and rectangular axes $Ox_1, Ox_2, \dots Ox_n$. Referred to these axes, Σ is the point $x_1 = X_1 - a,\ x_2 = X_2 - a, \dots x_n = X_n - a$. We now fill this space with the density field appropriate for samples of n drawn from Π, making the point-density at $(x_1, x_2, \dots x_n)$ equal

$$D = 1/(\sqrt{2\pi}\sigma)^n \cdot e^{-\frac{1}{2\sigma^2}(x_1^2 + x_2^2 + \dots + x_n^2)} \qquad \text{(i)},$$

so that the chance of drawing a sample with variates lying within the range $x_1 \pm \frac{1}{2}dx_1, \dots x_n \pm \frac{1}{2}dx_n$ is $D dx_1, \dots dx_n$.

* *Biometrika*, Vol. x. p. 507.

In the great majority of problems we are neither able nor do we wish to distinguish between x_1, x_2, ... x_n as differentiated variables, but in the space as defined there will be n ! points corresponding to a given unordered set of n values of the variable. Each of these points will lie in a different region of the general hyperspace; one such region will be that for which

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n \quad \text{.................................(ii),}$$

which may be described as a wedge-shaped region lying between $n-1$ primes passing through the line

$$x_1 = x_2 = \ldots = x_n \text{................................(iii).}$$

These n ! regions are identical in shape and density distribution and one may be superposed upon another or all combined together by a rotation about the axis (iii)*. The contour surfaces of equal density and the other contours that will be considered are all figures of revolution about this axis, and it will follow that any conclusions which may be drawn regarding the integral of the density taken throughout regions lying between such contours will apply to a single

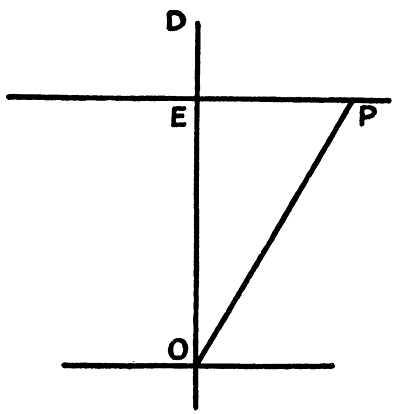

FIG. 1

segmental region as well as to the complete space. For simplicity in treatment we shall therefore consider the latter space, which may be termed the *fundamental space*, bearing in mind that in practice it is generally only possible to locate Σ in a single region such as that defined by the conditions (ii). This region may be supposed filled with a density obtained by superposing the n! similar regions.

Denote the mean of Σ by $a + m$ and its standard deviation by s. The section of the fundamental space by the two-dimensioned plane which passes through the line (iii) and the point P representing Σ is shown in Figure 1. O is the origin, OD the line (iii), and E the point where this line is cut by

$$x_1 + x_2 + \ldots + x_n = nm \text{.............................(iv).}$$

* In the case $n=2$, the regions are the halves of the x_1, x_2 plane divided by the line $x_1 = x_2$. For $n=3$, there are six regions, each being a "wedge" or "slice" in three-dimensioned space lying between two planes which cut at an angle of 60° in the line $x_1 = x_2 = x_3$. We shall use the term "prime" to describe any linear function of the n variables. Such a locus is a "flat" space of $n-1$ dimensions lying in n dimensions.

The prime (iv) passes through P and lies at right angles to OD, so that the angle PEO is a right angle. Further P is the point $(X_1 - a, X_2 - a, \ldots X_n - a)$ and E the point $(m, m, \ldots m)$, so that

$$OP^2 = \mathop{S}_{t=1}^{n} (X_t - a)^2 = n(m^2 + s^2),$$

$$OE^2 = nm^2 \text{ and consequently } EP^2 = ns^2.$$

Hence if we reduce the scale of this section of the hyperspace in the linear ratio of $1/\sqrt{n}$, we shall obtain a plane in which the position of the point corresponding to P is given by rectangular coordinates (m, s). This plane will be termed the (m, s)-plane. Any contour in the hyperspace, whose equation can be expressed in terms of the two variables m and s only, may therefore be obtained by a rotation of the corresponding curve in the (m, s)-plane about the axis (iii), and a linear enlargement in the ratio $\sqrt{n} : 1$.

Consider now the following:

(a) The point density D in the hyperspace may be written

$$D = \text{constant} \times e^{-\frac{n(m^2+s^2)}{2\sigma^2}} = \text{constant} \times e^{-\frac{1}{2}\chi^2},$$

where*
$$\chi^2\sigma^2 = \mathop{S}_{t=1}^{n} (X_t - a)^2 = n(m^2 + s^2).$$

(b) The contours of equal density are $(n-1)$-fold hyperspheres centred at the origin O.

(c) If the two-dimensioned (m, s)-plane be rotated in m dimensions about the axis (iii), the point P will trace out an $(n-2)$-fold hypersphere lying in a prime perpendicular to the axis of rotation; the boundary surface of this hypersphere is proportional to s^{n-2}.

Consequently the integral of D taken between any contour surfaces in the fundamental space which can be obtained (after enlargement) by a rotation of curves lying in the (m, s)-plane, will be the same as the integral of d between the corresponding curves in that plane, where d is a point-density assigned to the plane measured by

$$d = \text{constant} \times s^{n-2}\, e^{-\frac{n(m^2+s^2)}{2\sigma^2}},$$

or choosing the constant so that

$$\int_{-\infty}^{+\infty}\int_0^{\infty} d \,.\, dm \,.\, ds = 1,$$

$$d = \frac{n^{\frac{n}{2}}}{\sqrt{\pi}\, 2^{\frac{n-2}{2}}\, \Gamma\left(\frac{n-1}{2}\right)} \frac{s^{n-2}}{\sigma^n} e^{-\frac{n(m^2+s^2)}{2\sigma^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(v)}.$$

* This χ^2 must not be confused with the χ^2 of the Goodness of Fit Test.

The equations of the five contour systems defined below can all be expressed in terms of m and s. In order to grasp their bearing on the problem of testing Hypothesis A, it is necessary to consider their form and position in the fundamental space, but their reduction to the (m, s)-plane makes clearer their relation to one another and enables a ready integration of D to be obtained in certain regions of the hyperspace.

These contours, which are shown diagrammatically in the (m, s)-plane in Figure 2, are as follows:

(1) The contours of equal density D (contours of χ). These are semicircles, centred at O, of radius $\chi\sigma/\sqrt{n} = \sqrt{m^2 + s^2}$.

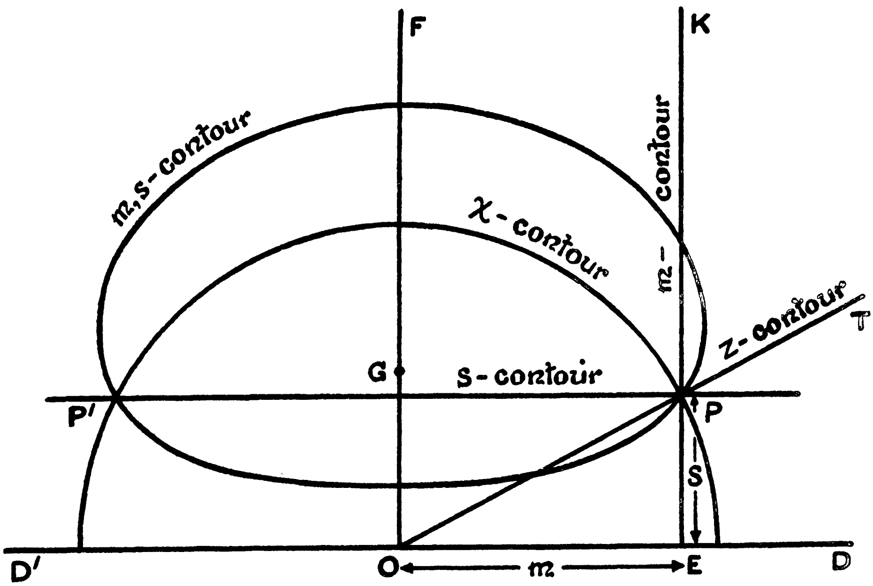

FIG. 2

(2) The contours of constant m; a series of lines parallel to the axis of s.

(3) The contours of constant s; a series of lines parallel to the axis of m.

(4) The contours upon which each of which the ratio of $z = m/s$ is constant (contours of z). These are a series of straight lines radiating from O.

(5) The contours of equal density d, or of equi-probable doublets (m, s). They form a series of oval curves in the (m, s)-plane with equations

$$(n-2)\log_e s - \tfrac{1}{2}n(m^2 + s^2)/\sigma^2 = K \quad \text{.....................(vi)},$$

and surround the point of maximum density G, or $m = 0$, $s = \sigma\sqrt{(n-2)/n}$.

By integrating d outside the first four contours in the (m, s)-plane we obtain the four standard integrals for the chance, P, of drawing a sample with χ, m, s, or z, equal to or greater than the value observed:

(1) $$P_\chi = c_1 \int_\chi^\infty \chi^{n-1} e^{-\frac{1}{2}\chi^2} d\chi \quad \text{(vii)},$$

(2) $$P_m = c_2 \int_m^\infty e^{-\frac{nm^2}{2\sigma^2}} dm \quad \text{(viii)}.$$

(3) The distribution of s being skew, we may take

$$P_s = c_3 \int_s^\infty s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds \quad \text{(ix)},$$

or

$$P_s' = c_3 \int_0^s s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds \quad \text{(x)},$$

according as s is greater or less than the modal value $s_0 = \sigma \sqrt{(n-2)/n}$.

(4) $$P_z = c_4 \int_z^\infty (1 + z^2)^{-\frac{n}{2}} dz \quad \text{(xi)},$$

and finally, for the chance of obtaining a less probable doublet (m, s),

(5) $$P_{m,s} = c_5 \int_\Omega s^{n-2} e^{-\frac{n(m^2+s^2)}{2\sigma^2}} dm\, ds \quad \text{(xii)},$$

where $\int_\Omega$ is to be taken outside the contour of system (vi), which passes through the sample point. This integral has not so far been computed.

The form of the contours in the fundamental space obtained by rotating Figure 2 about $D'OD$ in n dimensions can be easily visualised in the case of $n = 3$. The χ-contour becomes a sphere, the m-contour a plane at right angles to the axis, the s-contour a cylinder, the z-contour a cone, and the (m, s)-contour a hollow ring.

The foregoing results are not of course new, but this restatement has seemed necessary in order to make clear the discussion which follows*.

(2) *The Use of the Contour Systems in testing Hypothesis A.*

In the first place it is clear that any of the five contour systems described above will enable us to control the first source of error referred to in the introductory section. By rejecting Hypothesis A only when the point representing Σ falls upon a contour for which P (whether P_χ, P_m, ...) is, say, $\leqslant \cdot 01$, we shall be certain in the long run, in random sampling, of only rejecting the hypothesis when Σ has in fact been drawn from Π, in 1 case out of 100. But the samples rejected and those included will vary according to the system chosen. Consider the systems individually.

* See, for example, Student, *Biometrika*, Vol. VI. p. 1, 1908; R. A. Fisher, *Biometrika*, Vol. x. p. 507, 1914; *Metron*, Vol. v. No. 3, p. 3, 1925.

(2), (3) and (4), contours of m, s and z.

If we wish to judge whether a certain deviation in the mean, or a certain standard deviation, or a ratio of the two, is likely to be found in sampling from Π, these systems are entirely satisfactory. For instance, the smaller be P_m, the more divergent is the sample mean at $a+m$ from the population mean at a, the less frequently will a sample with mean in the limits $a+m \pm \frac{1}{2}dm$ be drawn in random sampling, and other things being equal the more likely become alternative hypotheses as to the value of the population mean.

But if our purpose is to make full use of the information supplied by Σ, and to ask whether, taken as a whole, it is likely to have been drawn from Π, the contours of m, s, and z are inadequate. In the first place they are not closed contours, for the space bounded by any one of them stretches away to infinity; this is true of the fundamental space as well as of the (m, s)-plane. Consequently the space on the inner side of any contour, however near to unity be the P associated with it, will contain points representing samples which are infinitely divergent from the population type. Therefore in using one of these systems alone as a criterion we should sometimes find ourselves being led to accept samples for which our common sense tells us that Hypothesis A is extraordinarily improbable.

It will also be seen that, while in the fundamental space the contours of the criteria of (viii) and (ix) for m and s contain the central region of the density field, the z-contours are orthogonal to the surfaces of equal density, and select out for rejection in a quite arbitrary fashion, the sample points lying at right angles to the axis (iii)*.

(1), contours of χ†.

These contours alone have the property of coinciding in the fundamental space with the surfaces of equal density, so that any contour divides off samples of more from those of less frequent occurrence. P_χ will provide a satisfactory measure in any problem where the judgment of the statistician is aided by a scale which expresses in numerical terms the fact that, in drawing from a given population, a sample is of relatively common, rather rare, exceptionally rare occurrence, etc. But in the great majority of problems we cannot so isolate the relation of Σ to Π; we reject Hypothesis A not merely because Σ is of rare occurrence, but because there are other hypotheses as to the origin of Σ which it seems more reasonable to accept. This attitude may be illustrated by the following example.

Suppose that Π is "normal" with mean at the origin of reference and $\sigma = 100{\cdot}0$, and that Σ is a sample of 8 for which the variate values are as follows:

$$+1{\cdot}0,\ +0{\cdot}4,\ -0{\cdot}6,\ +0{\cdot}1,\ -1{\cdot}1,\ -0{\cdot}3,\ -0{\cdot}1,\ +0{\cdot}2.$$

* This criticism of the z-contours does not of course apply to their use in Student's Problem, where the standard deviation of Π is unknown, which is discussed below. In the case where σ is known, however, they provide a good illustration of a system which gives for any sample a measure P_z, which is completely useless as a criterion to apply in testing Hypothesis A.

† The χ-contours here considered are of course quite different from those occurring in the (P, χ^2) test for goodness of fit, or what may be termed the "χ^2 group test." Here the variables are $x_1, x_2, \ldots x_n$; there they are the group frequencies $n_1, n_2, \ldots n_k$.

Here $$\chi^2 = \mathop{S}_{t=1}^{t=8} (X_t - a)^2/\sigma^2 = \cdot 000288.$$

P_χ is therefore very close to unity, the sample point lying almost at the centre of the density field. Yet unless the investigator had very strong *à priori* grounds for supposing that the standard deviation of the sampled population was 100·0, he would instinctively reject Hypothesis A, feeling that some other hypothesis, that gave σ in the neighbourhood of 0·5, was far more probable. This would appear to be a decision of common sense; we shall seek later to express it in more exact terms.

(5), the (m, s)-contours or contours of constant d.

These, just as the χ-contours, have the advantage of being closed, and if we take as the criterion of rejection $P_{m,s} \leqslant \cdot 01$, let us say, we shall not accept samples for which Hypothesis A seems exceedingly improbable on any common-sense grounds. As only the most divergent of the contours approach the origin O, we should reject Hypothesis A in the case of the numerical example given above. Further, since in sampling from a Normal population the mean and standard deviation may be considered as the two most important descriptive constants, the (m, s)-plane has a certain fundamental importance and it seems rational to adopt the contours of equal density in that plane. At any rate in making use of them we are applying what may be termed a "two constant" test, as distinct from the less adequate "one constant" test of either m or s alone*. But it must be remembered that it is difficult to find any logical reason for accepting the contours of equi-probable doublets (m, s), rather than those of the doublets (m, s^2), or (m, s^3), etc. For a given sample P_{m,s^2} will differ from $P_{m,s}$; the contours of the former in the fundamental space are obtained by rotating the oval curves in the (m, s)-plane for which

$$(n-3)\log_e s - \tfrac{1}{2}n(m^2+s^2)/\sigma^2 = K \quad\text{.....................(xiii)},$$

an equation differing from (vi) in the factor $n-3$ instead of $n-2$.

We must in fact consider whether it is possible to find a contour system which will take into account the probability of alternative hypotheses, and will not depend for its validity on the particular statistical constants chosen to describe the sample.

(3) *The Criterion of Likelihood.*

Suppose there to be two hypotheses regarding the population from which a given sample has been drawn; there will be one density field for Hypothesis A corresponding to Π, and another for Hypothesis A′ corresponding to Π'. By surrounding the sample point in the fundamental space with an element of volume $dX_1, dX_2, \ldots dX_n$ we have in the ratio of the "weights" of the elements a measure of the relative frequency of occurrence, on the two hypotheses, of samples lying

* It will be seen that if both the m-test and the s-test are applied to a sample, and the Hypothesis A is rejected if, let us say, $|m| > 3\sigma_m$ and $|s-\sigma| > 3\sigma_s$, the criterion of rejection will form a rectangle in the (m, s)-plane corresponding very roughly to the oval of the (m, s)-contours.

within the prescribed limits. These elements may seem a little artificial since we know in fact that the n variates of the sample have each unique values, even if these cannot be determined exactly. It will therefore be well to consider the position more carefully.

The possible range of the variable $x = X - a$ may be broken up into small subranges of equal length h. The n variates of the sample will each fall into one of these subranges, and the corresponding point P in the fundamental space will fall into one of the n-dimensioned hypercubes of side h which make up the framework into which the space will now be divided. As pointed out above, on p. 179, for a given set of unordered values $x_1, x_2, \dots x_n$ there will be a large number of hypercubes in the complete space into which P may fall. If n_1 of the variates fall into one subrange, n_2 into another and so on, where $n_1 + n_2 + \dots + n_k = n$, then the chance of drawing a sample with its representative point in one of these cubes is

$$C_p = \frac{n!}{n_1!\, n_2! \dots n_k!}\, p \qquad \text{(xiv)},$$

where p is the "weight" of one of the cubes. If none of the variates falls into the same subrange this expression reduces to $n!\,p$.

If we confine our attention to a single one of the $n!$ segmental regions such as that defined by (ii), we may consider the form of (xiv) as due to a boundary effect. When several of the x's fall into a single subrange, only a portion of the corresponding hypercube or $1/(n_1!\, n_2! \dots n_k!)$ lies within the region (ii), part being cut off by one or more of the $n-1$ primes bounding the region.

As h/σ is decreased so as to make the block surrounding P a differential element of volume, this boundary effect becomes negligible, none of the n variates falling into the same differential subrange. On the other hand as h/σ is increased the effect becomes more and more important, until all the hypercubes within the field of significant density cut a considerable number of the bounding primes. In the first case we measure the chance of drawing a sample having the n variates in the range

$$x_1 \pm \tfrac{1}{2}h, \dots x_n \pm \tfrac{1}{2}h,$$

by

$$f(x_1), f(x_2), \dots f(x_n)\, h^n \qquad \text{(xv)},$$

where for the Normal population

$$f(x) = 1/\sqrt{2\pi}\sigma \,.\, e^{-\frac{1}{2}\frac{x^2}{\sigma^2}};$$

in the second case, by a term of the multinomial expansion, namely

$$\frac{n!}{n_1!\, n_2! \dots n_k!}\, p_1^{n_1} \dots p_k^{n_k} \qquad \text{(xvi)},$$

where

$$p_s = \int_{x_s - \frac{1}{2}h}^{x_s + \frac{1}{2}h} f(x)\, dx.$$

Now it can be shown that when h is a finite subrange, a comparison of the probability of two hypotheses obtained from two expressions such as (xvi) leads directly to the ordinary (P, χ^2) test for goodness of fit*. It would seem therefore that by supposing that each observation lies somewhere within a differential element of fixed length, and comparing expressions of form (xv), we employ a method which differs only in degree from that used in cases of broad grouping. The use of a grouping unit of fixed length† whatever be the hypothetical population is common to both methods; the difference consists mainly in the fact that in one case the grouping unit is so large that many variates will generally fall into each group, while in the other the great majority of groups are vacant, and those filled contain only one observation.

We shall therefore take the ratio of two expressions such as (xv), with a common h, that is to say the ratio of the point-densities at P in the two fields, as a measure for the comparison of Hypotheses A and A′. Probability is a ratio of frequencies and this relative measure cannot be termed the ratio of the probabilities of the hypotheses, unless we speak of probability *à posteriori* and postulate some *à priori* frequency distribution of sampled populations. Fisher has therefore introduced the term *likelihood*, and calls this comparative measure the ratio of the likelihoods of the two hypotheses‡. For a given sample and two given hypotheses, the ratio of the likelihoods is completely independent of the coordinate space in which the sample point is represented. For example the ratio of the densities of the two fields at P is the same in the space of $x_1, x_2, \dots x_n$ as in that of $x_1^2, x_2^2, \dots x_n^2$. This is a point of considerable importance.

As an illustration, return to the example of p. 183. We may suppose that there are two hypotheses as to the population: (A_1) Mean at O, $\sigma = 100{\cdot}0$, and (A_2) Mean at O, $\sigma = 0{\cdot}5$. Then the two corresponding densities at the sample point are (A_1) $6{\cdot}42 \times 10^{-20}$, ($A_2$) $5{\cdot}18 \times 10^{-4}$ and the ratio of the likelihood of Hypothesis A_1 to that of A_2 is $1{\cdot}24 \times 10^{-16}$. This ratio confirms the common-sense judgment that A_2 is far more plausible than A_1.

There is little doubt that the criterion of likelihood is one which will assist the investigator in reaching his final judgment; the greater be the likelihood of some alternative Hypothesis A′ (not ruled out by other considerations), compared with that of A, the greater will become his hesitation in accepting the latter. It is true that in practice when asking whether Σ can have come from Π, we have usually certain *à priori* grounds for believing that this may be true, or if not so, for expecting that Π' differs from Π in certain directions only. But such expectations can rarely be expressed in numerical terms. The statistician can balance the numerical verdict of likelihood against the vaguer expectations derived from *à priori* considerations, and it must be left to his judgment to decide

* This point will be discussed later in connection with the (P, χ^2) test. R. A. Fisher has dealt with it in considering the relation between Maximum Likelihood and Minimum χ^2, *Phil. Trans.* A, Vol. 222, p. 357; *Journ. of Roy. Stat. Soc.* Vol. LXXXVII. p. 446.

† It might be supposed that the element should be made to depend upon the parameters of the hypothetical population. But if, for example, h were taken to be proportional to σ, it is seen that the Normal population, for which the chance of drawing a sample with observations in the prescribed limits is a maximum, will have $\sigma = \infty$. Such an interpretation of the element would be clearly of little value.

‡ *Phil. Trans.* A, Vol. 222, p. 326.

at what point the evidence in favour of alternative hypotheses becomes so convincing that Hypothesis A must be rejected.

As for all criteria, there is a limitation to its value in the case of very small samples. If all hypotheses were equally probable *à priori*, the criterion provides a means of picking out the most likely ones *à posteriori*, whatever be the size of n. But if we have as is usual some *à priori* grounds for believing in A, we shall not feel disposed to give up this belief unless the alternative suggested by the method of likelihood is what may be termed a stable alternative. For example, in the previous illustration, while it seemed reasonable to reject Hypothesis A upon the evidence of a sample of 8, we should not do so if we had only a sample of 2 with values $+0{\cdot}4$ and $-0{\cdot}6$, because we should be quite ready to find on adding a third observation to the sample that it had a value of $+50{\cdot}0$ or $-100{\cdot}0$, or of almost any quantity whatsoever. Here the alternative of a population with $\sigma = 0{\cdot}5$ does not appear sufficiently stable to be very seriously considered.

(4) *The Application of the Criterion of Likelihood in testing Hypothesis A.*

We shall suppose that while there is uncertainty as to the values of the mean and standard deviation, a and σ of the population sampled, there are good reasons for believing that Σ has been drawn from *some* normally distributed population. This is not a necessary limitation of the general theory of likelihood, for it is possible, in theory at any rate, to compare the point-densities of any types of field whatsoever, but we consider here only the simplest case*. Taking $\bar{X}$ as the mean of Σ referred to a fixed origin and s as before, its standard deviation, the density at P of the field associated with some population Π', mean at a', and standard deviation σ', is given by

$$D = \frac{1}{(\sqrt{2\pi}\,\sigma')^n}\, e^{-\frac{n}{2\sigma'^2}\{(\bar{X}-a')^2+s^2\}} \text{ ....(xvii).}$$

To find the Π' for which D is a maximum we solve

$$\frac{\partial D}{\partial a'} = 0 \text{ or } \bar{X} = a',$$

$$\frac{\partial D}{\partial \sigma'} = 0 \text{ or } -\frac{n}{\sigma'^{n+1}} + \frac{1}{\sigma'^n}\, \frac{n\{(\bar{X}-a')^2+s^2\}}{\sigma'^3} = 0,$$

giving

$$\sigma'^2 = (\bar{X}-a')^2 + s^2.$$

The population of maximum likelihood† is therefore that for which $a' = \bar{X}$ $\sigma' = s$, and the corresponding field gives as the density at the sample point P,

$$D_M = 1/(\sqrt{2\pi}\,s)^n \,.\, e^{-\frac{n}{2}} \text{ ....(xviii).}$$

From this we obtain the ratio

$$\lambda = \frac{\text{Likelihood of } \Pi}{\text{Likelihood of } \Pi' \text{ (max.)}} = \frac{D}{D_M} = \left(\frac{s}{\sigma}\right)^n e^{-\frac{n}{2}\left(\frac{m^2+s^2}{\sigma^2}-1\right)} \text{ ......(xix).}$$

* The effect of a moderate divergence from normality is considered experimentally below.

† It can be shown on a further differentiation that this is a maximum solution.

This ratio remains constant upon a certain contour surface in the density space of Π. That is to say, the ratio of the likelihood of Π to the maximum likelihood that can be associated with an alternative normal population is constant upon this contour, but steadily diminishes on passing outwards from one contour to another of the system. The surfaces are obtained by the rotation about the axis (iii) of the following curves in the (m, s)-plane:

$$(m^2+s^2)/\sigma^2 - \log_e (s^2/\sigma^2) = 1-(2/n)\log_e \lambda = k \log_e 10 \quad \text{.........(xx)}.$$

This equation represents a series of oval curves surrounding the point $m=0$, $s=\sigma$; the particular contour is determined by the constant k, and as n appears only on the right-hand side of the equation it will be seen that the form of the system as a whole is independent of the size of the sample. For large values of n the curves will differ only slightly from those of the contours of equi-probable doublets (m, s) given in (vi). In the fundamental space the surfaces will be closed contours which will not contain the origin. For the case $n=3$ it is possible to visualise the system of hollow rings obtained by rotating about DOD' curves much like the (m, s)-contour drawn in Figure 2.

Without claiming that this method is necessarily the "best" to adopt, we suggest that the use of this contour system in association with the density field of Π provides at any rate one clearly defined method of discriminating between samples for which Hypothesis A is more probable and those for which it is less probable. It is a method which takes into account the likelihood of alternative hypotheses, and while its justification does not rest on the choice of any particular statistical constants, the close agreement as n becomes large of its contour system with that of the equi-probable doublets (m, s) provides another form of interpretation.

To employ the method, it is necessary to know

$$P_\lambda = c_5 \int_\omega s^{n-2}\, e^{-\frac{n(m^2+s^2)}{2\sigma^2}}\, dm\, ds \quad \text{....................(xxi)},$$

an expression corresponding to (xii), except that here the integral is to be taken outside the contour of (xx) which passes through the point (m, s) corresponding in the (m, s)-plane to the observed sample, Σ. If then we only reject Hypothesis A when P_λ is, let us say, $\leqslant \cdot 01$, we can control the error of form (1)*, while the use of the contours of likelihood will minimise as far as possible the effect of errors of form (2). It would be possible to use the ratio λ of (xix) as a criterion, but this without a knowledge of P_λ does not enable us to estimate the extent of the form (1) error.

Tables of P_λ for $n=3$ to 50 are given in the Appendix. These are entered with k, which may either be calculated for the case of a given sample exactly from (xx), or found readily with sufficient accuracy for most practical purposes from the diagrams of Figures 13 and 14. The method by which the Tables were

* See p. 177.

computed and also the manner in which P_λ can be obtained for higher values of n are described in the Appendix. Illustrations of the use of the Tables are given in Section (9) below.

(5) *Student's Test and Hypothesis B.*

This is the problem to which Student's z-test has been applied*. The sample, Σ, of n observations $X_1, X_2, \ldots X_n$ has a mean $\bar{X}$ and a standard deviation s. There is every reason to believe that Σ has been drawn from *some* normally distributed population, and we wish to test Hypothesis B—namely, that Σ has been randomly drawn from such a population with mean a, but unspecified standard deviation.

In the fundamental space the axes and the position of the point P representing Σ are as before, but the density field appropriate to Π depends upon the unspecified σ. Of the contour systems considered above only those of z are independent of σ, and it might therefore appear at first sight that here was sufficient justification for the use of Student's integral, P_z, to test the probability of Hypothesis B. But B is really a multiple hypothesis concerning the sub-universe of normal populations, $M(\Pi)$, with means at a and with varying standard deviations. It only becomes precise upon definition of the manner in which σ is distributed within this sub-universe, that is to say, upon defining the *à priori* probability distribution of σ. Although it is true that whatever this may be, the distribution of z in sampling from $M(\Pi)$ will be the same with the probability integral given by (xi), it does not follow that the z-contours will therefore be the appropriate contours to use in testing the hypothesis as a whole. If it were possible to determine an *à priori* distribution for σ,

$$y = \phi(\sigma), \text{ where } \int_0^\infty \phi(\sigma) = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xxii)},$$

then we might fill the fundamental space with a density field appropriate to $M(\Pi)$ given by

$$\Delta = \frac{1}{(2\pi)^{n/2}} \int_0^\infty \frac{1}{\sigma^n} \phi(\sigma) \, e^{-\frac{1}{2\sigma^2} S(X_t - a)^2} d\sigma \ldots\ldots\ldots\ldots\text{(xxiii)},$$

and consider what in this were the appropriate contours to use. In practice, however, it is doubtful if $\phi(\sigma)$ could ever be determined, and if some arbitrary form is to be assigned, it is simpler to approach the problem from the point of view of inverse probability, and think of the density space as associated with Σ rather than $M(\Pi)$. This will be done in a later section.

In a problem where Hypothesis B is to be tested, the investigator will usually be quite certain that σ cannot have all values between 0 and ∞. He may often be able to define broad limits outside which it is impossible that it should lie, and to suggest narrower but less clearly defined limits within which it most probably lies. But although he might say, "As far as I know, σ may lie equally well anywhere between b_1 and b_2," he would probably find it difficult on reflection to

* *Biometrika*, Vol. VI. p. 1 *et seq.*

justify this statement, rather than, "As far as I know, σ^2 may lie equally well anywhere between b_1^2 and b_2^2." In fact, his *à priori* information as to σ is generally not sufficiently precise to bear exact numerical expression. Under such conditions the criterion of likelihood will probably be of service.

Corresponding to any given Σ, we can find the population Π out of $M(\Pi)$ for which the likelihood is a maximum. This is obtained from (xvii) by equating $\partial D/\partial\sigma'$ to zero, considering $a = a'$ as fixed. We find $\sigma'^2 = m^2 + s^2$, and for this population the density at P is

$$D_0 = \frac{1}{(2\pi)^{n/2}} \frac{1}{(m^2+s^2)^{n/2}} e^{-\frac{n}{2}} \quad \text{....................(xxiv).}$$

We can now obtain the ratio of (a) maximum likelihood for a member of $M(\Pi)$, to (b) maximum likelihood for any normal population (the D_M of (xviii)), or

$$\lambda' = \frac{D_0}{D_M} = \left(\frac{s^2}{m^2+s^2}\right)^{\frac{n}{2}} = (1+z^2)^{-\frac{n}{2}} \quad \text{..................(xxv).}$$

This ratio is constant along the contours of z. It follows that if the sample point is surrounded by an element of volume, the ratio of the maximum frequency with which a sample lying within these limits will be drawn from a normal population with mean at a to the maximum frequency with which it could be drawn from any unrestricted normal population whatever, is constant upon one of these contours. Speaking therefore in somewhat loose terms, we may say that, other things being equal, we shall be more ready to accept Hypothesis B for a sample lying inside or on a certain z-contour, than for one lying outside it. And we may measure our confidence in a decision to reject or accept the hypothesis by P_z, which is the integral of the density outside the contour whatever be σ. Student's test was originally devised to allow for the use of s instead of σ in testing the significance of a deviation in the sample mean. We have above another interpretation of it as a method of testing the sample as a whole, based upon the criterion of likelihood, which will be valid as long as that criterion can be employed.

Whilst the *à priori* knowledge regarding the value of σ remains loosely defined, the open z-contours are probably the best that can be used. But if there are grounds for believing that σ must lie within a limited range, the contours which it would be natural to adopt become deflected until, when σ can be almost exactly located, they will form a closed system tending to the λ-contours. Suppose, for illustration, that in a sample of 10 there is a mean of 20·4 and a standard deviation of 3·6. We may ask whether it is likely that this sample has been drawn from a population with mean of 18·0. As long as there is no *à priori* information which enables the value of σ to be closely determined, or which is contradicted *à posteriori* by the sample value of $s = 3{\cdot}6$, it seems reasonable to apply Student's test. But if there were grounds for believing that σ could not exceed 1·5, the knowledge that for $z = \frac{2}{3}$, $P_z = {\cdot}26$ would be of little value in solving the problem in hand. This point is illustrated more fully in the first example of Section (9) below.

(6) *Alternative Method of examining Hypothesis B.*

Take as variables $M = \frac{m}{\sigma}$ and $S = \frac{s}{\sigma}$, and suppose Figure 3 (p. 192) to represent the (M, S)-plane. The λ contours of (xx) become

$$M^2 + S^2 - \log_e S^2 = k \log_e 10 \quad \text{.....................(xxvi)},$$

and surround the point $C\,(M = 0, S = 1)$. Since the point-density in the (m, s)-plane is given by (v), that in the (M, S)-plane will be

$$d' = \text{constant} \times S^{n-2} e^{-\frac{n}{2}(M^2+S^2)} \quad \text{..................(xxvii)};$$

consequently this plane with density d', and the contour system (xxvi), can be looked upon as a fixed field of reference whatever be σ. In testing Hypothesis B, we cannot locate in the plane the point P corresponding to Σ, but only the radial line OT somewhere on which it must lie, where

$$M/S = m/s = z = \cot TOD.$$

P_z is the integral of the density throughout the sector TOD of the plane. If σ were known, the point P could be identified, falling let us suppose at Q. We could then find P_λ, the integral of d' over the region lying outside the λ-contour drawn through Q. P_λ will clearly be a maximum if P were found to fall at R, the point where OT touches a member of the λ-system. This has been seen from a different method of approach* to occur when $\sigma^2 = m^2 + s^2$, or $M^2 + S^2 = 1$. Consequently in Figure 3, $OR^2 = 1 = OC^2$.

We may now argue as follows. For a given Σ, the most favourable case giving a maximum value to P_λ arises when we suppose the sampled population from $M(\Pi)$ to be Π_1 with $\sigma = \sqrt{m^2 + s^2}$. If, on giving this value to σ, Hypothesis A judged by P_λ and the criterion of likelihood is improbable, then it will be more improbable for every other normal population with mean at a. Hence we reject Hypothesis B. If on the other hand P_λ is such that Hypothesis A appears reasonable for population Π_1, then Hypothesis B can hardly be rejected, because while it may not be possible to assert that it is most probably true, we know that for values of σ in the neighbourhood of $\sqrt{m^2 + s^2}$ the sample Σ is not exceptional. To the criticism that there is no justification for assuming that Π_1 is the population sampled, it can be answered that unless some *à priori* distribution of σ be postulated, there is no justification whatever for assuming that Π_1 or a population

* On p. 190 above. The result may also be proved directly as follows:

The contour equation is

$$\phi(M, S) \equiv M^2 + S^2 - \log_e S^2 = k \log_e 10,$$

and the tangent to this at the point (M_1, S_1) will be

$$S - S_1 = -(M - M_1)\,\frac{\partial \phi}{\partial M_1}\Big/\frac{\partial \phi}{\partial S_1}.$$

But

$$\frac{\partial \phi}{\partial M_1} = 2M_1; \quad \frac{\partial \phi}{\partial S_1} = 2S_1 - \frac{2}{S_1},$$

so that the tangent becomes

$$S - S_1 = (M - M_1)\, M_1 S_1/(1 - S_1^2).$$

As this is to be the line OT passing through the origin,

$$S_1 = M_1^2 S_1/(1 - S_1^2) \text{ or } M_1^2 + S_1^2 = 1,\ m^2 + s^2 = \sigma^2.$$

differing slightly from it is improbable. The greater the range of values of σ above and below $\sqrt{m^2+s^2}$, which give reasonable values to P_λ, the more probable becomes Hypothesis B, but no attempt is made to give this probability a unique numerical measure.

It will be seen from Figure 3 that if Student's test be applied, $2P_z$ is the integral of the field between the axis of M and the tangents OT, OT'. Clearly $2P_z < P_\lambda$ (max), and the one will be a single-valued function of the other. It follows that if we were to reject Hypothesis B whenever $P_\lambda\,(\max) \leqslant \alpha$, we should in fact always reject it for samples for which $z \geqslant \gamma$ and $2P_z \leqslant \beta$, where $\beta < \alpha$. If then a statistician thoughtlessly decides, whatever be the test, to reject an hypothesis when $P \leqslant \cdot 01$, say, and accept it when $P > \cdot 01$, it will make a considerable difference to his conclusions whether he uses P_z or P_λ (max). But as the ultimate

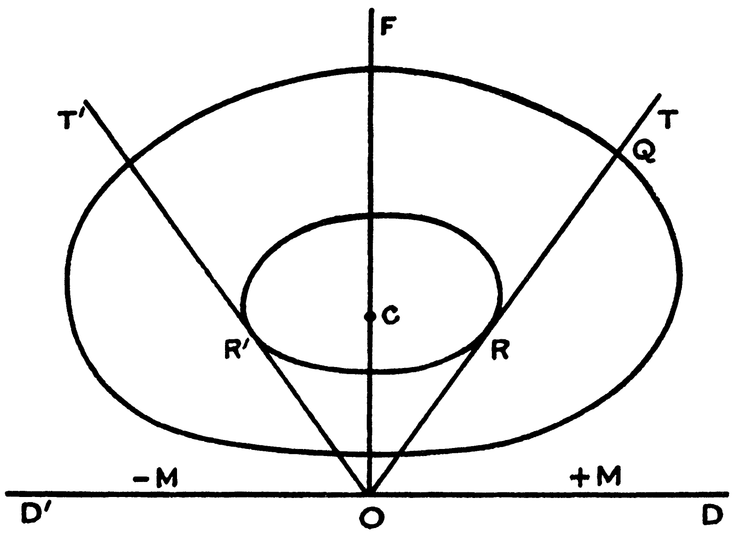

FIG. 3

value of statistical judgment depends upon a clear understanding of the meaning of the statistical tests applied, the difference between the values of the two P's should present no difficulty. The difference in the two scales of probability simply corresponds to the difference in attitude of mind with which the problem is approached; the use of P_λ (max) perhaps corresponds to the more cautious attitude, but it would seem impossible to claim that one approach is the correct one and the other is erroneous.

(7) *Solutions obtained by the Inverse Method.*

The inverse method of approach is to start from Σ, the observed sample, as the one certain fact in the problem. Points in the space no longer represent samples but populations from which Σ may have been drawn; that is to say, the density field is made appropriate to Σ and not to Π, and we shall reject an hypothesis concerning the population sampled on entering regions of the field in which the density falls below a certain level. This alternative point of view is of some interest, and although it leads to tests which are almost identical with those reached above, we shall discuss it in some detail in the present case of sampling from a normal population.

We must again assume that Σ has been drawn from *some* normally distributed population, and as this can be described completely by its mean, a, and standard deviation, σ, the whole discussion may be confined to the two-dimensioned (a, σ)-plane. Referred to a fixed origin in this plane and rectangular axes, Σ is the point $G(\bar{X}, s)$ and any population Π a point $P(a, \sigma)$. The position is illustrated in Figure 4.

The ordinary method of inverse probability consists in postulating some function $\phi(a, \sigma)$ to represent the probability *à priori* that the sampled population is Π, and taking for the point-density of the field

$$D = \text{constant} \times \phi(a, \sigma) \times \sigma^{-n} e^{-\frac{n}{2\sigma^2}\{(\bar{X}-a)^2+s^2\}} \quad \text{.........(xxviii).}$$

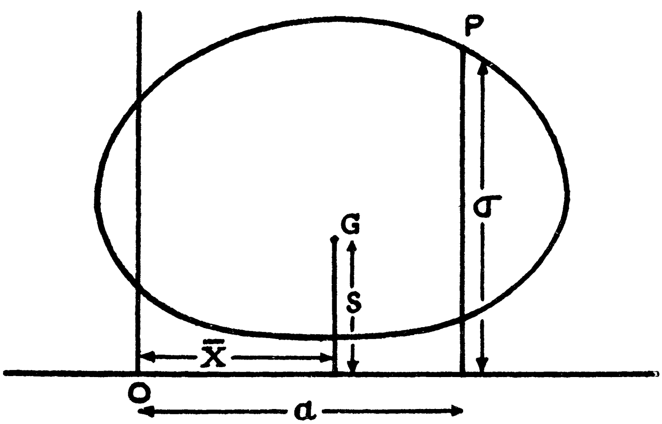

FIG. 4

$D . da . d\sigma$ would then be termed the probability *à posteriori* that the population sampled has a mean and standard deviation in the range, $a \pm \frac{1}{2}da$, $\sigma \pm \frac{1}{2}d\sigma$. The difficulty of this procedure in any practical problem lies in the fact that it is almost impossible to express ϕ in exact terms. We prefer therefore to follow a line of argument which while really equivalent to the above with ϕ assumed constant makes use of the principle of likelihood rather than the somewhat vaguer conception of *à posteriori* probability.

The likelihood of Π may be written

$$L \propto D = (1/\sqrt{2\pi})^n \sigma^{-n} e^{-\frac{n}{2\sigma^2}\{(\bar{X}-a)^2+s^2\}} \quad \text{...............(xxix).}$$

Population points such as P of Figure 4, for which L is constant will lie on closed curves,

$$\log_e \sigma^2 + \{(\bar{X}-a)^2 + s^2\}/\sigma^2 = h/n \quad \text{.....................(xxx),}$$

which surround the sample point G. Give to the field a point-density D. If we are right in believing that the value of the likelihood is a useful criterion, then L (or D) itself would provide us with a measure of the confidence to be placed in an hypothesis as to Π. This quantity as it stands is not however measured on a very suitable scale, for it is difficult to know how to interpret the very small values that L will generally take. Let I be the integral of D taken throughout the whole (a, σ)-field*, and I' throughout that portion of it which lies outside the member of system (xxx) passing through P. Then the ratio I'/I appears to be a suitable measure to use. It is the same for all populations of a given likelihood and it varies between 0 and 1; when it is very small we know that the likelihood of Π stands very low in the scale, and when large that L approaches the maximum value. Likelihood as defined by Fisher is a quantity which cannot be integrated†. With this in the strict sense we agree, but look upon the relation $I'/I = f(L)$ as a form of transformation providing as a criterion a function of L placed on a scale more easily understood in common terms.

We can now write

$$P_{a,\sigma} = I'/I = c_6 \int_{\omega} \sigma^{-n} e^{-\frac{n}{2\sigma^2}\{(\bar{X}-a)^2+s^2\}}\, da\, d\sigma \quad \text{...........(xxxi)},$$

where the integral is taken outside the member of (xxx) passing through (a, σ), and c_6 is a quantity depending only on n and s, which is chosen so that $P_{a,\sigma} = 1$ when $a = \bar{X}$, $\sigma = s$.

It is of interest to compare the values of

$$P_{m,s},\ \text{(xii)};\ P_{\lambda},\ \text{(xxi)};\ \text{and}\ P_{a,\sigma},\ \text{(xxxi)}.$$

Each is a function of n and the four quantities, $\bar{X}_0$, s_0, a_0 and σ_0, which define the particular Σ and Π with which Hypothesis A is concerned. Further, they are each integrals taken in a two-dimensioned density field over regions outside an oval contour. They may be brought into comparable form by means of the following transformations.

(*a*) $P_{m,s}$.

Transformation $\quad m/\sigma_0 = u, \quad s/\sigma_0 = v, \quad dm\,ds = \sigma_0^2\, du\, dv.$

Take $\quad k_0 = n\{\log(s_0^2/\sigma_0^2) - [(\bar{X}_0 - a_0)^2 + s_0^2]/\sigma_0^2\}.$

The density in the (u, v)-plane becomes, on transforming (v),

$$D' = cv^{n-2} e^{-\frac{n}{2}(u^2+v^2)} \quad \text{........................(xxxii)}.$$

$P_{m,s}$ is obtained by integrating D' outside the contour

$$(n-2)\log v^2 - n(u^2+v^2) = k_0 - 2\log(s_0^2/\sigma_0^2) \quad \text{.........(xxxii } bis\text{)}.$$

* It is not necessary to suppose this to range from $a = +\infty$ to $-\infty$, $s = 0$ to $+\infty$. The conclusions will not be modified if we suppose that the field is bounded by one of the contours of (xxx) which contains, let us say, ·999 of the density field.

† See *Phil. Trans.* A, Vol. 222, p. 327.

The integral I does not equal unity, but $\Gamma\left(\frac{n-2}{2}\right)\Big/\{2^{\bar{2}}\,\Pi^{\frac{n-1}{2}}\, n^{\frac{n-1}{2}}\, s^{n-2}\}$.

(b) P_λ.

The transformation is as for $P_{m,s}$, and the density in the (u, v)-plane is as in (xxxii). The integral is to be taken outside the contour

$$n \log v^2 - n(u^2+v^2) = k_0 \quad \text{.......................(xxxiii).}$$

This corresponds to (xxvi).

(c) $P_{a,\sigma}$.

The transformation is $$(\bar{X}_0 - a)/\sigma = u, \quad s_0/\sigma = v,$$

whence $$a = X_0 - s_0 u/v, \quad \sigma = s_0/v,$$

and $$da\,d\sigma = J\,.\,du\,dv = s_0^2/v^3\,du\,dv.$$

Then the density becomes, using (xxix),

$$D'' = c' v^{n-3} e^{-\frac{n}{2}(u^2+v^2)} \quad \text{.......................(xxxiv).}$$

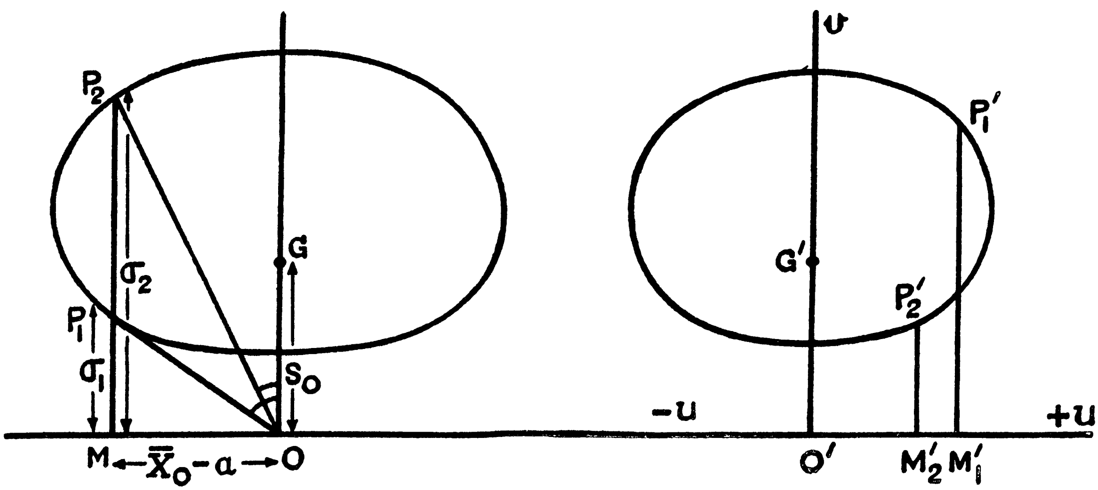

FIG. 5

And the contour, using (xxx), is

$$\log(s_0^2/v^2) + u^2 + v^2 = \log \sigma_0^2 + (\bar{X}_0 - a_0)^2/\sigma_0^2 + s_0^2/\sigma_0^2,$$

or $$n \log v^2 - n(u^2+v^2) = k_0 \quad \text{..................(xxxiii } bis\text{).}$$

The nature of this transformation is shown in Figure 5. The sample point G transforms into G', $u = 0$, $v = 1$, and two points P_1, P_2 on a contour become P_1', P_2', where

$$O'M_1' = u_1 = \tan GOP_1, \quad P_1'M_1' = v_1 = OG/P_1M,$$

$$O'M_2' = u_2 = \tan GOP_2, \quad P_2'M_2' = v_2 = OG/P_2M.$$

The field outside the contour in the (a, σ)-plane corresponds to that outside the contour in the (u, v)-plane and it is over this that D'' is to be integrated.

13—2

On comparing these results in the (u, v)-plane we see that

(1) the density fields for $P_{m,s}$ and P_λ are the same, namely that given by (xxxii), but the contours (xxxii *bis*) and (xxxiii) differ both in the factor $(n-2)$ for n, and in the additional constant term $-2\log(s_0^2/\sigma_0^2)$ contained in the former;

(2) the density fields for P_λ and $P_{a,\sigma}$ differ in the term v^{n-2} of (xxxii) and v^{n-3} of (xxxiv), but the contours are identical (xxxiii).

In large samples, for a given Σ and Π, the three tests become practically equivalent since the values of the three P's will scarcely differ. They each result however from different methods of approaching the problem, and in small samples will lead to somewhat different results. It is the quantity P_λ which has been tabled.

We shall now consider whether it is possible to obtain from the (a, σ)-field, measures of the probability that the population, from which Σ has been drawn, has

(1) a mean a, such that $a_1 \leqslant a \leqslant a_2$, irrespective of the standard deviation;

(2) a standard deviation σ, such that $\sigma_1 \leqslant \sigma \leqslant \sigma_2$, irrespective of the value of a.

Again, as likelihood can only be associated with a single population with unique values of a and σ, we cannot speak of the likelihood of populations for which (a, σ) falls within the regions of the field defined above, although we might speak of the probability *à posteriori* of such an event, assuming $\phi(a, \sigma)$ to be constant. Whatever be its exact interpretation, however, the integral of D in (xxix) taken over these regions and divided by I (the integral over the whole field) will provide us with some form of measure of the probability of the hypotheses we are considering, and will be a quantity varying between 0 and 1.

First integrate D, therefore, along a strip of breadth da at a distance a from the axis of σ:

$$da\int_0^\infty D\,.\,d\sigma = \text{constant}\times da\int_0^\infty \sigma^{-n}\,e^{-\frac{n}{2\sigma^2}\{(\bar{X}-a)^2+s^2\}}\,d\sigma.$$

Write $u^2 = n\{(\bar{X}-a)^2+s^2\}$ and $u^2/2\sigma^2 = v$, so that $-d\sigma\, u^2/\sigma^3 = dv$ and the integral becomes

$$= \text{constant}\times u^{-n+1}\,da\int_0^\infty e^{-v}\,v^{\frac{n-3}{2}}\,dv$$

$$= \text{constant}\times u^{-n+1}\,da,$$

the constant being a function of n only.

Dividing by I, and writing $z = (\bar{X}-a)/s$, we have finally for the measure of probability that the mean of Π is such that $a_1 \leqslant a \leqslant a_2$,

$$P(a_1, a_2) = c_7\int_{z_1}^{z_2}(1+z^2)^{-\frac{n-1}{2}}\,dz \quad \text{..................(xxxv)},$$

where $\quad z_1 = (\bar{X}-a_1)/s, \quad z_2 = (\bar{X}-a_2)/s.$

Except that we have now $n-1$ instead of n this expression corresponds to (xi), and its value may be obtained from the tables of Student's z or t function, using $n-1$ instead of n*.

Now integrate D along a strip of breadth $d\sigma$ at a distance σ from the axis of a:

$$d\sigma \int_{-\infty}^{+\infty} D \,.\, da = \text{constant} \times \sigma^{-n+1} e^{-\frac{ns^2}{2\sigma^2}} d\sigma.$$

On dividing by I we have for the measure of probability that the standard deviation of Π is such that $\sigma_1 \leqslant \sigma \leqslant \sigma_2$,

$$P(\sigma_1, \sigma_2) = c_8 \int_{\sigma_1}^{\sigma_2} \sigma^{-n+1} e^{-\frac{ns^2}{2\sigma^2}} d\sigma \quad \text{.................(xxxvi).}$$

By making the transformation $\sqrt{n}s/\sigma = \psi$, this becomes

$$P(\psi_1, \psi_2) = \int_{\psi_1}^{\psi_2} \psi^{n-3} e^{-\frac{1}{2}\psi^2} d\psi \Big/ \int_0^{\infty} \psi^{n-3} e^{-\frac{1}{2}\psi^2} d\psi \quad \text{...(xxxvi } \textit{bis}\text{),}$$

which is the difference of two ordinary χ^2 integrals and can be obtained from the χ^2 Tables by taking $\psi^2 = \chi^2$, $n' = n-1$.

If in equation (ix), where the variable is s, we write $\psi = \sqrt{n}s/\sigma$, we obtain a probability integral P_s of exactly the same type as (xxxvi), except that there will be a term ψ^{n-2} instead of ψ^{n-3}, that is to say, the χ^2 Tables must be entered with $n' = n$.

We have reached, then, three pairs of corresponding results, (xxi) and (xxxi), (xi) and (xxxv), (ix) and (xxxvi); in each case there is this difference of a unit in n which appears to result invariably on proceeding by the inverse method. Probably most statisticians will prefer to use the criteria (xxi), (xi) and (ix) of the direct method, which with the addition of the present P_λ Tables can be readily expressed in numerical terms. The difference of the two is however of no great importance except in very small samples, where final conclusions will be drawn in any case with some hesitation.

(8) *Analysis of Church's Samples from a Skew Population.*

Dr A. E. R. Church very kindly placed at our disposal one of his sets of sampling results previously obtained for another purpose†. These were the 1000 samples of 10 drawn from a smooth distribution with the following frequency constants:

$$\sigma = 3{\cdot}569{,}207, \quad \beta_1 = {\cdot}219{,}333, \quad \beta_2 = 3{\cdot}157{,}676.$$

This distribution not differing very greatly from a Type III form may be taken as typical of the moderately skew distributions frequently met with in

* Student's original table of $z = m/s$ for $n=2$ to 10 is given in *Tables for Statisticians and Biometricians*, p. 36 (second edition). A complete table for $n=2$ to 21 with auxiliary tables for use with higher values has been published in *Metron*, Vol. v. p. 3, 1925. These latter tables must be entered with $t = z\sqrt{n'-1}$, where n' is now taken as the size of sample.

† *Biometrika*, Vol. XVIII. pp. 321—324. This is Church's Population "B" and group of samples "IV."

practice. It is therefore of considerable importance to discover how far the sampling variation in the frequency constants can be represented by the appropriate "normal" formulae. In practice we may often not have available sufficient information to test whether the population Π be normally distributed or no, and even if this be available may be unable to deduce the appropriate formulae for

TABLE I.

Distribution of Frequency Constants for Church's 1000 *Samples of* 10.

Frequencies for S			Frequencies for z		
Central values of S	Observed	Normal Theory	Central values of z	Observed	Normal Theory
—	—	—	Less than −1·20	—	2·9
—	—	—	−1·15	2	1·7
—	—		−1·05	3	2·9
1·3	1·5		− ·95	3	4·7
1·5	0·5	20·3	− ·85	8	7·7
1·7	11		− ·75	10	12·6
1·9	21	21·2	− ·65	6	20·2
2·1	32·5	34·3	− ·55	38	31·2
2·3	71	50·3	− ·45	46	46·4
2·5	68	67·0	− ·35	59	65·5
2·7	81	82·5	− ·25	92	85·7
2·9	81·5	94·0	− ·15	119	103·9
3·1	101·5	100·1	− ·05	134	114·6
3·3	89·5	99·8	+ ·05	103	114·6
3·5	80	93·9	+ ·15	104	103·9
3·7	95	83·2	+ ·25	78	85·7
3·9	67	69·8	+ ·35	63	65·5
4·1	54	55·7	+ ·45	36	46·4
4·3	53	42·2	+ ·55	39	31·2
4·5	33	30·5	+ ·65	21	20·2
4·7	19·5	21·0	+ ·75	9	12·6
4·9	11·5	13·9	+ ·85	11	7·7
5·1	9	8·7	+ ·95	6	4·7
5·3	6	5·2	+1·05	4	2·9
5·5	6		+1·15	2	1·7
5·7	4		+1·25	—	1·1
5·9	2	6·4	+1·35	1	0·7
6·1	—		+1·45	1	0·4
6·3	1		+1·55	1	0·2
—	—	—	+2·05	1	0·5
—	—	—	Greater than +2·10	—	
Total	1000·0	1000·0	Total	1000	1000·0

Frequency Constants for s	Observed	Normal Theory	Standard Error
Mean	3·3181	3·2933	·0249
Standard Deviation	·8086	·7863	·0176 (approx.)
β_1	·1467	·0634	—
$\sqrt{\beta_1}$	+·3830	+·2518	·0761
β_2	2·9782	3·0106	·176

Frequency Constants for z	Observed	Normal Theory	Standard Error
Mean	+·0120	0	·0120
Standard Deviation	·3709	·3780	·0107
β_1	·1740	0	—
$\sqrt{\beta_1}$	+·4171	0	·181
β_2	4·5405	4·2000	1·363

Frequencies outside P_λ Contours in the (m, s)-field

k of Contour	Observed	Normal Theory
·45	847	849
·50	500	504
·55	321	299
·60	181	177
·65	102	105
·70	59	63
·80	19	22
·90	8	8

samples from skew distributions. To what extent, then, shall we be likely to draw faulty conclusions if we apply the normal criteria in these cases?

Church has fitted a normal curve with standard deviation of $\sigma/\sqrt{n}$ to the distribution of means, and finds on applying the test for goodness of fit, a value for P of ·5758 *. By taking the square roots of his values of s^2, we have obtained the distribution of s given in Table I. This is compared in the next column with the theoretical distribution appropriate to samples from a normal distribution with the same standard deviation, viz. $\sigma = 3{\cdot}569{,}207$. The distribution of s^2 for a normal population is of course a Type III curve, and the theoretical frequencies for s were found by taking the areas between the corresponding ordinates of the s^2 curve from the Tables of the Incomplete Gamma Function. Testing for goodness of fit, $\chi^2 = 26{\cdot}12$ and with $n' = 19$, $P = {\cdot}097$. The larger of the differences in frequency are rather irregularly scattered, but there appear on the whole to be too few low and too many high values of s among the observations. This means that the observational distribution is more skew than that of normal theory, a point brought out from a comparison of the frequency constants given in the same table. The constants for normal theory are taken from the table given in *Biometrika*, Vol. x. p. 529. The standard error for σ_s was obtained on the assumption that the theoretical distribution of s was normal †. The ratios of the differences between the observed and theoretical values of the four frequency constants to their standard errors are $+1{\cdot}0$, $+1{\cdot}2$, $+1{\cdot}69$ and $-0{\cdot}2$ respectively; the disagreement except perhaps in the case of $\sqrt{\beta_1}$ cannot be considered very serious, although we are probably just reaching a degree of skewness in the population for which the normal theory begins to fail.

Compare now the observed distribution and that of normal theory for $z = m/s$. They are given in Table I; here $\chi^2 = 29{\cdot}19$, $n' = 18$ and $P = {\cdot}033$. The most serious discrepancy, contributing 9·98 to χ^2, occurs in the group for $z = -{\cdot}6$ to $-{\cdot}7$, and there is no doubt that among the observations there are too few with large negative and too many with large positive values of z. The observed z-distribution is in fact significantly skew, whereas Student's distribution is symmetrical. A comparison of the frequency constants is given in the same table. Judged by the standard errors, it is only $\sqrt{\beta_1}$ that differs significantly from normal theory ‡.

* *Biometrika*, Vol. XVIII. p. 334.

† For low values of β_1 the table (*Tables for Statisticians and Biometricians*, Table XXXVII) giving the standard error of this quantity cannot be used, as in this region the second order term is as great or greater than the first order term which vanishes when $\beta_1 = 0$. It is the first order term which is tabled. We have therefore compared the values of $\sqrt{\beta_1}$ for which the standard error to first order is equal to

$$\{\beta_4 - 3\beta_3 - 6\beta_2 + \tfrac{9}{4}\beta_1\beta_2 + \tfrac{35}{4}\beta_1 + 9\}^{\frac{1}{2}}/\sqrt{N}$$

(where N is number of samples), and have inserted values for β_3 and β_4 obtained from the theoretical β_1 and β_2 on the assumption that the distribution can be represented by a Pearson curve.

‡ The standard error of $\sqrt{\beta_1}$ was calculated from the formula given in the preceding footnote; this gives to the first order for samples from symmetrical distributions,

$$(\text{S.E. } \sqrt{\beta_1})^2 = (\beta_4 - 6\beta_2 + 9)/N.$$

In the case of Student's distribution for $n = 10$, $\beta_4 = 49{\cdot}0$.

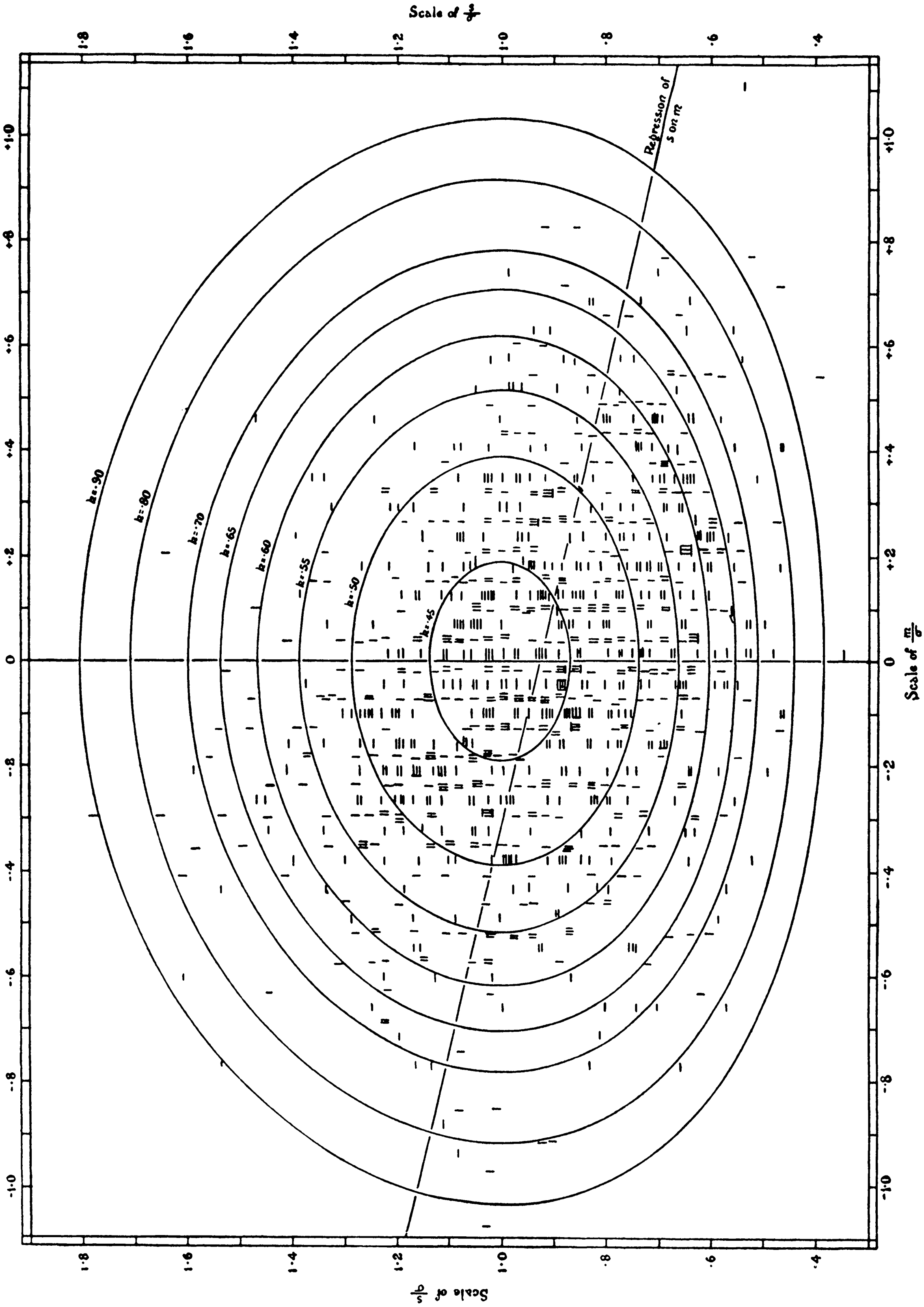

Fig. 6. Distribution of m and s for Church's 1000 Samples of 10 from a Skew Population

Church's population distribution was negatively skew, and this results in positive skewness in the distribution of z.

In practice the method of applying the test consists, when the observed value of z is large, in rejecting Hypothesis B with a degree of confidence which increases as P_z decreases. Taking P_z equal (to the nearest figure) to (*a*) ·05, and (*b*) ·005, we find from the observations,

(*a*)	for positive values of z:	57	cases out of	1000	instead of	53,
	„ negative „ „ :	32	„	„	„	53,
(*b*)	„ positive „ „ :	6	„	„	„	5,
	„ negative „ „ :	2	„	„	„	5.

This difference must not be passed over, but it is doubtful whether the knowledge that P_z was really ·03 (or ·06) rather than ·05, or again ·002 (or ·006) rather than ·005, would in fact ever modify our judgment when balancing the probabilities regarding the origin of a single sample.

Finally we may consider the distribution of the observations among the P_λ contours. Figure 6 is a reproduction of a rough working diagram containing entries for the 1000 doublets (m, s). The scale attached to the margins is that of the ratios m/σ and s/σ, where σ is as before the standard deviation of the population. Eight of the P_λ contours are drawn in the figure; the observed regression straight line of s on m has also been given. It is obvious from inspection that the condition which holds for samples from a normal distribution of no correlation between m and s and of homoscedasticity is now far from true. The coefficient of correlation calculated from the observations is $r = +\cdot 3206$ with standard error ·0191. The theoretical correlation between m and s^2, which gives a fair approximation to that between m and s, is ·3036*. The correlation ratio of s on m is $\eta = \cdot 3495$. In spite of this obvious disagreement with the distribution of the normal (m, s)-field, it is found that the frequencies of sample points lying outside the successive P_λ contours follow very closely the expected numbers. A comparison is shown at the end of Table I; the only disagreement of importance is that for the contour, $k = \cdot 55$, outside which 321 observations instead of 299 lie. The χ^2-test may be applied to the frequencies *between* contours, for which there will be 9 groups including that for observations inside $k = \cdot 45$ and for observations outside $k = \cdot 90$; here $\chi^2 = 7\cdot 36$ and $P = \cdot 50$.

The P_λ contours can no longer of course provide a completely satisfactory criterion, since they have ceased to divide on a logical basis the sample points for which Hypothesis A is more probable from those for which it is less probable. They appear however to enable us to control the first of the two sources of error described in the introductory section†, since they do enclose approximately the expected proportions of random samples drawn from Π. Whether the situation would be as satisfactory for samples of a different size, or for different population distributions, we cannot of course say without further experiment.

* Calculated from the formula (68) given in *Biometrika*, Vol. XVII. p. 479.

† See p. 177 above.

(9) ***Illustrations of the use of the Tables of P_λ.***

The tables are given in the Appendix with a description of the method by which they were computed and of how they may be extended. They must be entered with n, the size of sample, and k; in the following examples k has been obtained from one or other of the diagrams, Figures 13 and 14, pp. 236, 237, in which the marginal scales are (1) M, the ratio of the deviation in the sample mean from the population mean to the population standard deviation σ, and (2) S, the ratio of the sample standard deviation, s, to σ.

Example 1.

Measurements on 884 Egyptian skulls (XXVI—XXX Dynasties)* gave for the distribution of cephalic index, Mean = 75·06, Standard Deviation = 2·68. Would it be justifiable to consider that the 10 skulls with cephalic indices as follows, 74·1, 77·7, 74·4, 74·0, 73·8, 72·2, 75·2, 78·2, 77·1, 78·4, were a random sample from this population?

(a) The distribution of cephalic indices is in general found to be symmetrical and leptokurtic; we shall assume, pending the completion of further inquiries, that if $\beta_2 = 3{\cdot}4$ to $3{\cdot}5$, the P_λ contours are still adequate. It is found that the mean cephalic index is 75·51 and $s = 2{\cdot}059$, from which $M = +\,{\cdot}168$ and $S = {\cdot}768$. Using Figure 13 (the sign of M is immaterial) it is seen that the sample point lies just inside the contour $k = {\cdot}50$; entering the P_λ Table of p (see p. 238) with $n = 10$ and $k = {\cdot}50$ it is found that $P_\lambda = {\cdot}504$. It is therefore possible to say that *judged by cephalic index alone* there would be no reason to doubt that these 10 skulls were a random sample of XXVI—XXX Dynasty material†.

(b) Now apply the same test to the following 10 cephalic indices: 66·7, 69·4, 67·8, 73·2, 79·3, 80·7, 64·9, 82·2, 72·4, 78·1. Here $\bar{X} = 73{\cdot}47$, $s = 5{\cdot}942$, $M = -\,{\cdot}593$, $S = 2{\cdot}217$; the sample point falls near the contour $k = 1{\cdot}60$ in Figure 14, and for $n = 10$ this corresponds to P_λ less than ·0001. From the position of the point in the diagram it will be seen that the divergence is mainly due to a very large value of s compared with σ; the mean is not exceptionally divergent from the supposed population mean of 75·06. We should conclude that it was very improbable that the 10 skulls were a random sample from the Egyptian population, and that probably they consisted of heterogeneous material, the variability being greater than that generally found in a sample from a homogeneous race.

(c) Suppose that we had been given only the mean cephalic index and not the standard deviation of the population. We could ask whether it was likely having regard to the observed variability of $s = 5{\cdot}942$, that a random sample of 10 could have been drawn having a mean differing by 1·59 from the population mean. Applying Student's test‡ it is found that $z = -\,{\cdot}268$ or $t = z\sqrt{n-1} = -\,{\cdot}803$

* *Biometrika*, Vol. XVI. Tables II and III, pp. 337, 338.

† Actually the cephalic indices are those of the first 10 of the Farringdon Street skulls measured by Miss Hooke. *Biometrika*, Vol. XVIII. p. 54, Appendix A.

‡ Using either of the Tables referred to in the footnote to p. 197 above.

and $P_t = P_z = \cdot 221$. That is to say so large a deviation of z in excess or defect from zero would occur in 44 per cent. of random samples. Consequently the evidence does not contradict Hypothesis B*.

(*d*) If however we had known from other sources that the standard deviation of cephalic index within a single race lay within the range 2·5 to 4·0, we could have improved upon the z-test†. Although σ be unknown, we could say that the sample point in Figure 14 must lie somewhere on the line $z = \cdot 268$ which joins the origin, $M = 0 = S$, to the corresponding division of the radiating z-scale given in the right-hand margin of the diagram. The exact position will depend on the value given to σ, but if we might infer that this is unlikely to be greater than 4·0, we could say that it was very unlikely that the sample point lay farther to the left along the z-line than the point for which $S = s/\sigma = 5\cdot 942/4 = 1\cdot 485$. To the right of this point the contours cutting the line have $k > \cdot 68$ and therefore $P_\lambda < \cdot 08$. We should therefore argue that it was far less likely that the hypothesis as to mean cephalic index of population were true than the value of P_z obtained in (*c*) would suggest.

(*e*) If there had been no *à priori* information as to σ, it would still have been possible to locate the sample point on the line $z = \cdot 268$. By taking $\sigma = \sqrt{m^2 + s^2} = 6\cdot 13$, or $M^2 + S^2 = 1$, we should locate the sample at that point on the line where it touches a contour of the system. This will roughly be that with $k = \cdot 47$, giving $P_\lambda = \cdot 689$. Consequently were it quite possible for a homogeneous race to have a standard deviation of cephalic index as great as 6·0, we should have no reason for doubting the origin of the second sample of 10. This is an illustration of the line of reasoning followed in Section (6) above.

Example 2.

Records of weight in a large population of mice for males between 120 and 140 days of age show a mean of 23·823 grammes, $\sigma = 3\cdot 137$ grammes, and for the frequency distribution $\beta_1 = \cdot 086$, $\beta_2 = 2\cdot 687$.

Can the groups of 6, 4 and 5 mice with the following weights be considered random samples from this population?

(*a*) 22·5, 26·0, 20·5, 24·0, 18·0, 24·5.

(*b*) 21·0, 21·0, 20·0, 20·5.

(*c*) 22·5, 23·5, 23·5, 25·0, 24·5.

(*a*) $n = 6$, $M = -\cdot 395$, $S = \cdot 851$. Using Figure 13, we see that $k = \cdot 52$ approximately, giving $P_\lambda = \cdot 607$, so that there is no reason to doubt the origin of the sample.

* If the inverse argument discussed in Section (7) above is used, Student's Tables would be entered with $n = 9$, and this gives $P_z = \cdot 223$ against $\cdot 221$, a difference of no importance as far as interpretation is concerned.

† The following values for the standard deviation of cephalic index will be found in recent papers in this Journal: XXVI—XXX Dynasty Egyptian 2·68; Naqada 2·80; 17th Century English 3·26; Farringdon Street 3·48; Australian "A" 3·88.

(*b*) $n=4$, $M=-1{\cdot}020$, $S={\cdot}132$. The point is beyond the outermost contour of the diagram but this suggests a value for k of 2·1. Using the exact equation for k, or

$$k = {\cdot}434{,}294\,(M^2+S^2) - 2\log_{10} S,$$

it is found that $k=2{\cdot}218$ and $P_\lambda={\cdot}002$. This small value of P_λ is chiefly due to the low value of S.

(*c*) $n=5$, $M=-{\cdot}007$, $S={\cdot}278$. Here $k=1{\cdot}15$, $P_\lambda={\cdot}036$. Again the value of S is unusually small.

The values of P_λ in (*b*) and (*c*) are small enough to make us search for an alternative hypothesis to that of pure chance fluctuation. As a matter of fact while (*a*) was a randomly chosen sample, in both the cases (*b*) and (*c*) the mice belonged to a single litter, so that the individual weights were correlated partly from the effect of inheritance, and partly because weight is correlated with size of litter.

Example 3.

The tables may be used in cases where the normal curve gives an approximation to the binomial.

A count of over a thousand words taken at random from daily newspapers gave a proportion of ·6691 monosyllables. Out of 35 words chosen at random from the works of 12 English authors, the following numbers of monosyllables were found*: Kipling 26, Thackeray 26, Macaulay 24, Meredith 24, Carlyle 23, Scott 23, Stevenson 22, Conrad 22, Borrow 22, Ruskin 21, Bryce 19, Hazlitt 12. Do these representatives of literature either appear to use a lower proportion of monosyllables or show among themselves a greater diversity in proportion than would be expected to occur in random samples from a vocabulary in which the proportion of monosyllables was ·6691?

The appropriate binomial has $n=12$, $p={\cdot}6691$, $q={\cdot}3309$, and therefore a mean of 23·42 and standard deviation of 2·784 monosyllables. The $\beta_1={\cdot}015$ and $\beta_2=2{\cdot}957$, so that the fit by a normal curve should be satisfactory. The 12 samples of 35 have a mean of 22·00 and standard deviation of 3·559, giving $M=-{\cdot}510$, $S=1{\cdot}278$, and $k={\cdot}60$, $P_\lambda={\cdot}121$. The P_λ would be very much larger but for the sample from Hazlitt†, and there appears in this particular character to be no significant difference between the language of the 12 writers and that of the newspapers.

Example 4.

Another type of problem may be illustrated as follows. A number of laboratories are perhaps determining by chemical analysis the percentage of some constituent substance in a given material. Take as an example the hypothetical results

* The 35 words were the first on the last line of each of 35 consecutive pages.

† A further random sample of 100 words taken from Hazlitt's *Characters of Shakespear's Plays* contained 42 monosyllables, corresponding to 14·7 out of 35, and suggesting that there may really be some significant difference in Hazlitt's vocabulary!

given in the following table; for each laboratory there is a mean per cent. m_t based on n_t analyses; the weighted mean of the whole is 8·353 per cent. and the combined standard deviation of the 187 analyses ·329 per cent. Information is

TABLE II.

Laboratory	1	2	3	4	5	6	7	8	9	10	11	12
Mean per cent., m_t	8·432	8·127	8·483	8·002	8·317	8·582	8·239	8·531	8·163	8·405	8·391	7·976
No. of analyses, n_t	22	10	19	15	16	25	9	16	4	16	23	12
$(m_t - 8{\cdot}4)\sqrt{n_t}/\sigma$	+ ·46	− 2·63	+ 1·10	− 4·69	− 1·01	+ 2·77	− 1·47	+ 1·59	− 1·44	+ ·06	− ·13	− 4·47

required as to whether, (*a*) the results are consistent with a quoted trade percentage of 8·4 for the constituent substance, (*b*) the variation in means is no more than would be expected to arise from chance fluctuations, or (*c*) that variation points to systematic differences in the method of analysis or in the material distributed among the different laboratories. If the variation among the 187 observations is simply due to chance errors, then $\sigma = {\cdot}329$ may be taken as a fairly close measure of the standard deviation of the universe of possible variations in analysis which has been sampled; we may therefore test the hypothesis that the sampled population mean is 8·4 and that the grouping of the observations into 12 sets is of a purely random nature, in the following manner. The standard error of each laboratory mean will be $\sigma/\sqrt{n_t}$, and, if the hypothesis be true, the 12 quantities $(m_t - 8{\cdot}4)\sqrt{n_t}/\sigma$ given in the last row of the table should each be a chance fluctuation from zero reduced to a standard error of unity. On the assumption, which could be verified in a given case, that the distribution of individual analyses is not exceptionally anormal, the distribution of these deviations in mean may be taken as normal. Consequently if we calculate the mean and standard deviation of the figures in the last row of the table, we can test the hypothesis by taking these as M and S, and finding k and P_λ with $n = 12$.

It is found that $M = -{\cdot}821$, $S = 2{\cdot}192$, $k = 1{\cdot}70$, and P_λ is less than ·0001. The hypothesis is therefore extremely improbable and from the position of the point on Figure 14 it is clear that this is mainly due to the high value of S. Comparing the difference between the weighted mean of the 187 observations (8·353), and 8·4 with its standard error of $\sigma/\sqrt{187} = {\cdot}0240$, we find a ratio of $-1{\cdot}94$, and a deviation as great as this in excess or defect might be expected to occur in 5 per cent. of random samples. The variation among the means themselves is however quite inconsistent with the hypothesis of chance fluctuations from a general mean, and must probably be attributed to variation in technique or to lack of homogeneity in the material distributed for analysis.

Example 5.

It is of some interest to compare different methods of approaching the problem of determining whether two samples can have been drawn from the same population. Suppose that the following figures were the results of an experiment in Industrial Psychology. A piece of work has been carried out by one group of 30 workmen according to Method I, and by another group of 40 according to Method II, the groups being selected at random from a factory. Are the recorded distributions in times such as to justify the conclusion that Method I is the speedier of the two?

TABLE III.

Time in seconds		50	51	52	53	54	55	56	57	58	59	60	Totals
Frequency Distributions	Method I	1	3	5	4	7	5	3	1	1	—	—	30
	Method II	—	1	2	5	8	9	6	3	3	1	2	40

It is found on calculation that,

for I, mean, $\bar{x}_1 = 53{\cdot}700$ secs.; standard deviation, $s_1 = 1{\cdot}882$ secs.

for II, „ $\bar{x}_2 = 55{\cdot}175$ secs.; „ „ $s_2 = 2{\cdot}072$ secs.

The usual method of procedure is to compare the difference in means with the standard error of that difference, and to refer the ratio of the two to the appropriate probability tables. The value given to the standard error and the tables used will depend entirely on the line of reasoning employed in the final process of inference.

(*a*) We may compare $\bar{x}_1 - \bar{x}_2$ with $\sqrt{s_1^2/n_1 + s_2^2/n_2}$, and from the ratio, $-3{\cdot}107$, obtain $\frac{1}{2}(1+\alpha) = {\cdot}9991$ from Sheppard's Tables, and from this conclude that if the samples came from two populations with the same mean but standard deviations of s_1 and s_2, so great a difference in mean, positive or negative, would only occur in about two samples in 1000.

(*b*) Having regard to the mean values of s_1 and s_2 in sampling we may attempt to obtain estimates of σ_1 and σ_2 which in the long run are likely to be closer than s_1 and s_2, namely: $s_1\sqrt{n_1/(n_1-1)}$ and $s_2\sqrt{n_2/(n_2-1)}$, substitute these in the standard error, and hence obtain a value of the ratio $= -3{\cdot}060$ and a resulting $\frac{1}{2}(1+\alpha) = {\cdot}9989$.

(*c*) We may give to the result, (*b*), the inverse interpretation suggested by Rhodes*.

(*d*) We may approach the problem by making use of the principle of likelihood and reaching the test given by Fisher†. Suppose that we have reason to believe that two samples have been drawn from normal populations with the same standard deviation σ, but that it is necessary to compare the relative probability of two hypotheses, (1) that the samples come from identical populations with mean at a, and (2) that while identical as to variability there are two different means a_1 and a_2.

* *Journal of Royal Statistical Society*, 1926, Vol. LXXXIX. p. 544.

† *Metron*, 1925, Vol. v. p. 7. *Statistical Methods for Research Workers*, 1925, p. 109.

Then it is easy to show that the maximum likelihood of hypothesis (1) is associated with a population for which

$$a = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}; \ \sigma^2 = \frac{1}{n_1 + n_2}\left\{n_1 s_1^2 + n_2 s_2^2 + \frac{n_1 n_2}{n_1 + n_2}(\bar{x}_1 - \bar{x}_2)^2\right\},$$

giving

$$L(\max) \propto \frac{1}{\sigma^{n_1+n_2}} e^{-\frac{n_1+n_2}{2}}.$$

The maximum likelihood of hypothesis (2) is associated with a population for which

$$a_1 = \bar{x}_1; \ a_2 = \bar{x}_2; \ \sigma'^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2},$$

and

$$L'(\max) \propto \frac{1}{\sigma'^{n_1+n_2}} e^{-\frac{n_1+n_2}{2}}.$$

As a measure of comparison of the two hypotheses we may take the ratio of these two maximum likelihoods, or

$$L(\max)/L'(\max) = (\sigma'/\sigma)^{n_1+n_2} = (1 + z^2)^{-\frac{n_1+n_2}{2}},$$

where

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}.$$

The smaller be z, the less likely becomes the hypothesis of a single population compared to the alternative hypothesis of two populations differing in mean, and consequently we can take z or a function of z as a criterion to use in judging the probability that the two samples have come from the same population. It is most convenient to refer z to the sampling distribution that it would follow were hypothesis (1) true. This Fisher has shown to be

$$y = y_0(1 + z^2)^{-\frac{n_1+n_2-1}{2}},$$

so that in using the test we have only to enter Student's Tables with

$$t = \sqrt{n_1 + n_2 - 2} \, . \, z$$

and $n' = n_1 + n_2 - 1$ *.

In the present example $z = -\cdot 3662$, $t = -3\cdot 020$ and $P_z = P_t = \cdot 9982$.

This extension by Fisher of Student's test would seem to be of considerable value in dealing with small samples. Its limitation lies partly in the assumption of normality, but as the z-distribution is not very sensitive to changes in the population β's this will often not be serious; and partly in the fact that it does not take into account a possible difference in population standard deviations. It assumes without question that these are the same and that there is only uncertainty as to the means. In many problems however there will be *à priori* reasons for believing that there can be no significant difference in variability and it is in such cases that the test is applicable.

* See footnote to p. 197 above as to notation.

(*e*) Finally it is possible to make use of the P_λ Tables in the following manner. Having observed the two samples, the single normal population of maximum likelihood is that associated with hypothesis (1) of the preceding paragraph, (*d*), namely in the present case a population with $a = 54{\cdot}543$ secs., $\sigma = 2{\cdot}123$ secs. It may now be asked whether the two series of observations are likely to have arisen independently in random sampling *if this had been the population sampled.* It is found that

$$\text{for I, } n = 30,\ M = -{\cdot}397,\ S = {\cdot}887,\ k = {\cdot}51,\ P_\lambda = {\cdot}079,$$
$$\text{for II, } n = 40,\ M = -{\cdot}298,\ S = {\cdot}976,\ k = {\cdot}47,\ P_\lambda = {\cdot}201.$$

That is to say in the most favourable light possible, the combined probability of occurrence of the two samples may be measured by the product $\cdot079 \times \cdot201 = \cdot016$. For any other values assigned to a and σ the probability of the observed result measured in this manner will be less, and therefore we should feel considerable confidence in arguing that the hypothesis of a common origin for the samples is improbable. This method does take into account possible differences in variability, and used with discretion would seem to be of value. It would of course lead to some difficulty of interpretation should the resulting product of probabilities be, let us say, ·05 to ·20; if in the most favourable circumstances the hypothesis of a common origin is just possible, are we to discard it on the grounds that we have no right to postulate the most favourable conditions? A satisfactory answer to this question is probably impossible. At any rate the method is perhaps one which encourages an attitude of caution.

In the present example, however, whichever test be used, it is clear that we should conclude that process I does almost certainly lead to a really quicker timing than process II.

III. Sampling from a Rectangular Population.

(10) *Sampling Distributions of the Frequency Constants.*

The population, Π, may be defined by two quantities, the range w, and the distance of the centre of this range from a fixed origin, g. Then all values of the variable between $g - \frac{1}{2}w$ and $g + \frac{1}{2}w$ are of equal frequency of occurrence in Π. In a particular sample, Σ, consisting of n observations $X_1, X_2, \ldots X_n$, the range may be denoted by W and the distance of the mid-point or "centre" of this range from the fixed origin by G, that is to say the lowest and highest variates of the sample are $G - \frac{1}{2}W$ and $G + \frac{1}{2}W$ respectively. The fundamental space for samples of n drawn randomly from Π will be an n-dimensioned hypercube each of the sides of which is of length w, and having as a diagonal the line joining the points $(g - \frac{1}{2}w, \ldots g - \frac{1}{2}w)$ and $(g + \frac{1}{2}w, \ldots g + \frac{1}{2}w)$. As in the "normal" space, this diagonal forms the axis of the $n\,!$ similar regions of which one is that defined by the condition

$$x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_n \quad \text{.........................(ii } bis\text{)},$$

where $x_s = X_s - g + \frac{1}{2}w$, the origin for the x's being taken at one corner of the cube.

The density field within the hypercube will be uniform, for the chance of drawing a sample with observations in the limits $X_1 \pm \frac{1}{2}dX_1, \ldots X_n \pm \frac{1}{2}dX_n$ is $dX_1, \ldots dX_n/w^n$. Making use of this density field, Hall* has shown that the distribution of the means of samples of n is given by a curve consisting of n connected arcs of degree $n-1$, having $(n-1)$-point contact at their joins. The distribution of standard deviations seems likely also to consist of a number of connected arcs.

In testing Hypothesis A we cannot make use of contours of equal density as these do not exist. Nor, except from the established custom of regarding them as the most important frequency constants, is there any reason why the mean and standard deviation of the sample should be chosen as the most suitable estimates of the population parameters. Consider the problem from the point of view of the criterion of likelihood. Suppose that there are reasons for believing that Σ has been drawn from *some* population following a rectangular distribution, and that an alternative to Π is a population Π' with centre at g' and range of w'. Then

$$\text{Likelihood of } \Pi/\text{Likelihood of } \Pi' = w'^n/w^n.$$

For a given Σ, the population of maximum likelihood is that for which w' is a minimum. But w' cannot be less than the sample range, W, and if it has this value clearly $g' = G$. It follows that in the fundamental hypercube, the contours upon which

$$\lambda = \text{Likelihood of } \Pi/\text{Likelihood of } \Pi' \text{ (max)}$$

is constant are the contours of constant range.

If therefore we make use of the criterion of likelihood in forming a judgment on Hypothesis A, we need consider only the range of the sample, and our confidence in the hypothesis will increase as $W \to w$. If $W > w$ or any of the observations in the sample fall outside the range of Π, then of course the hypothesis becomes impossible. Remembering that we are only considering cases where the sample has come from *some* rectangular population, the use of range alone as a criterion is seen to be perfectly reasonable. This limitation in the form of the sampled population appears more serious than that involved in the previous part of this paper†, simply because populations of normal or approximately normal form are found in experience to be common, so that we can frequently assume normality without great risk of error, while it is more difficult to conceive of problems in which a rectangular form of population can be assumed *à priori*. There is little doubt however that useful conclusions would be reached by applying the tests developed below to samples from any nearly symmetrical population with sharply limited range—that is to say, a population for which β_1 and β_2 lay in the neighbourhood of 0 and 1·8.

* *Biometrika*, Vol. XIX. pp. 240—244.

† There it was assumed that Σ came from *some* normally distributed or nearly normally distributed population.

We will now consider the distribution of W and G in sampling from II. In Figure 7, OI is of length w, and A and B represent the lowest and highest observations of the sample.

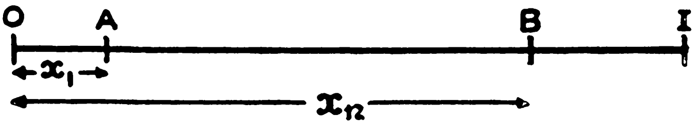

FIG. 7

The chance of obtaining one observation in an element dx centred at A, another in an element dx_n centred at B, and $n-2$ between these two values is

$$\frac{dx_1}{w} \times \left(\frac{x_n - x_1}{w}\right)^{n-2} \times \frac{dx_n}{w} = \left(\frac{x_n - x_1}{w^n}\right)^{n-2} dx_1\, dx_2.$$

Change the variables to $\begin{cases} G = \frac{1}{2}(x_1 + x_n) + g - \frac{1}{2}w, \\ W = x_n - x_1, \end{cases}$

so that $$dG\, dW = dx_1 dx_2.$$

Then the chance of obtaining a sample with centre and range in the limits $G \pm \frac{1}{2}dG$, $W \pm \frac{1}{2}dW$ is

$$\phi_1(G, W)\, dG\, dW \equiv k W^{n-2}/w^n \,.\, dG\, dW \quad \text{.............(xxxvii)}.$$

Choosing k so that the integral of ϕ taken throughout the total range of variation is unity, it is found that $k = n(n-1)$. Hence

$$\phi_1(G, W) \equiv n(n-1)\; W^{n-2}/w^n \quad \text{..............(xxxvii } bis\text{)}.$$

Integrating out for G between the limits $g - \frac{1}{2}(w - W)$ and $g + \frac{1}{2}(w - W)$ we obtain the chance, $\phi_2(W)\, dW$, of drawing a sample with range in the limits $W \pm \frac{1}{2}dW$ or

$$\phi_2(W) \equiv n(n-1)\frac{W^{n-2}}{w^{n-1}}\left(1 - \frac{W}{w}\right) \quad \text{..............(xxxviii)}.$$

This is a Type I curve (and also part of an $(n-1)$th order parabola) with the mean at $(n-1)w/(n+1)$, and a range from $W = 0$ to w*. For the reasons given above we therefore take as our criterion the probability integral

$$P_W = \int_0^W \phi_2(W)\, dW = \{nw - (n-1)W\}\, W^{n-1}/w^n \quad \text{.........(xxxix)}.$$

The chance of drawing a sample with its centre in the limits $G \pm \frac{1}{2}dG$ is obtained by integrating (xxxvii) between the limits for W of 0, and $w - 2\,|\,G - g\,|$† or

$$\phi_3(G) \equiv n\,(w - 2\,|\,G - g\,|)^{n-1}/w^n \quad \text{....................(xl)},$$

a symmetrical curve made up of two portions of $(n-1)$th order parabolae placed "back to back" on the ordinate through $G = g$. The form of distribution is shown roughly in Figure 8.

* This expression for the special case of $n=2$ has been given by Borel: *Traité du Calcul des Probabilités*, Tome I, Fasc. 1, p. 20, 1925.

† Throughout this paper $|\,q\,|$ denotes that the expression q within the rules is to be given a positive sign.

The chance of drawing a sample with centre at a distance as great or greater than $|G-g|$ from the centre of the population will be

$$P_G = 2\int_G^{g+\frac{1}{2}w} n\{w-2(G-g)\}^{n-1}/w^n \,.\, dG = \{w-2\,|\,G-g\,|\}^n/w^n \quad \ldots\text{(xli)}.$$

Finally we may find the distribution of $z' = (G-g)/\tfrac{1}{2}W$. Making the transformation $z' = 2(G-g)/W$, we have $\tfrac{1}{2}W dW dz' = dW dG$, and from (xxxvii),

$$\phi_4(z', W) \equiv \tfrac{1}{2}n(n-1)\,W^{n-1}/w^n \quad \ldots\ldots\ldots\ldots\ldots\text{(xlii)}.$$

For a given value of z', W can increase from 0 until $\tfrac{1}{2}W + |G-g| = \tfrac{1}{2}w$, or $W = w/(1+|z'|)$. Integrating out for W between these limits, we have

$$\phi_5(z') \equiv \tfrac{1}{2}(n-1)\{1+|z'|\}^{-n} \quad \ldots\ldots\ldots\ldots\ldots\text{(xliii)}.$$

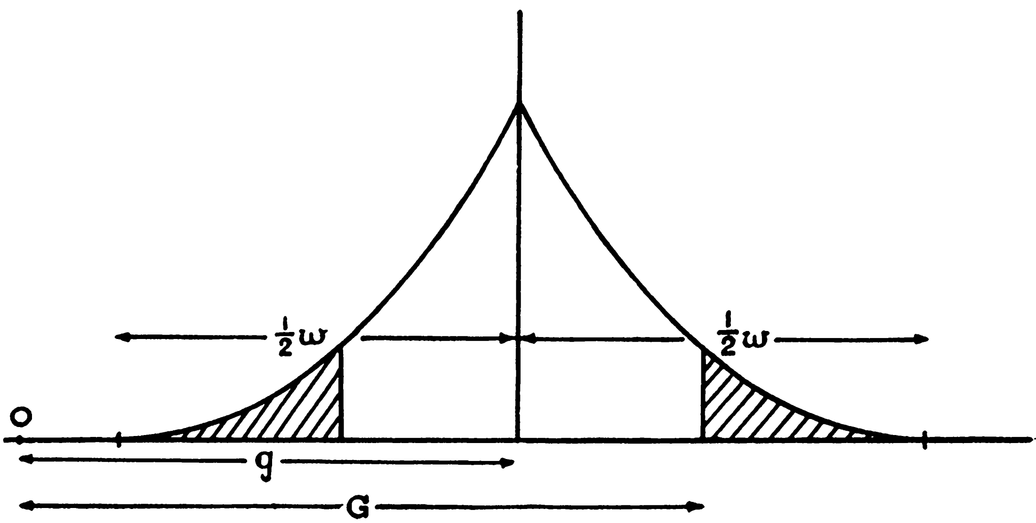

FIG. 8

This distribution of $z' = (G-g)/\tfrac{1}{2}W$ is of considerable interest. It corresponds to Student's distribution

$$f(z) = \text{constant} \times (1+z^2)^{-\frac{n}{2}}$$

for the ratio $z = m/s$ for samples from a normal population. z' is the ratio of the distance of the centre of the sample from the population centre to half the sample range, that is to say, like z it is the ratio of a measure of location to a measure of scale. The distribution of z' is independent of the population range w just as that of z is independent of the population standard deviation σ. It can be shown that z' is constant over parts of $n(n-1)$ primes passing through the centre of the fundamental hypercube and making an angle with its diagonal $x_1 = x_2 = \ldots = x_n$, which depends only on the value of z'. These primes correspond to the conical z contours in the normal space.

Suppose that we have a problem in which it is wished to know whether it is likely that a given sample $\Sigma\,(W, G)$ has been drawn from a rectangular population with centre at g, but for which the range, w, is unspecified. We are now testing Hypothesis B. Following the argument employed in dealing with normal populations, it is found, as shown in Figure 9, that the ratio of (1) Maximum

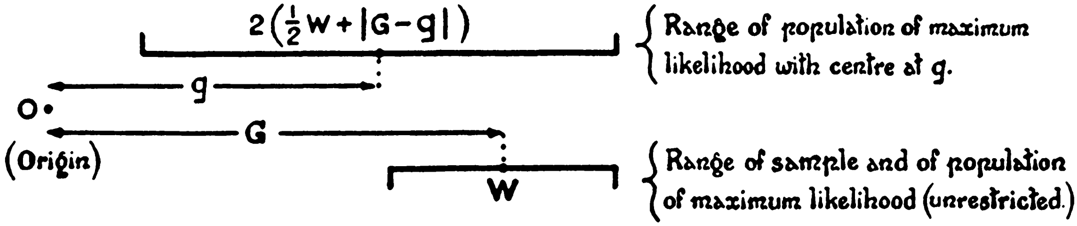

FIG. 9

likelihood for a rectangular population with centre at g, to (2) Maximum likelihood for any rectangular population, is

$$W^n/(W + 2\,|\,G - g\,|)^n = \{1 + |\,z'\,|\}^{-n}.$$

This expression is constant along the contours of z', and therefore leads to the probability integral of (xliii) as the criterion to employ in testing Hypothesis B, or

$$P_{z'} = (n-1)\int_{z'}^{\infty} (1 + z')^{-n}\,dz' = (1 + z')^{-n+1} \text{ ....(xliv).}$$

If the z' distribution is applicable only to exactly rectangular populations, its interest is mainly theoretical, but if it can be shown that it will represent adequately the distribution of the ratio (deviation in centre)/(half range) for small samples from a wider range of symmetrical populations, it will be of greater value. The test in this form or in its inverse form discussed below would enable an estimate to be made of the probable limits within which lies the mean of a symmetrical limited range population, from a knowledge of the two extreme individuals of the sample only. This it would do in the same way as Student's test does from a knowledge of the mean and standard deviation of the sample. Some further consideration of this point is being undertaken.

Fisher has given the limiting forms of the distributions of range (xxxviii) and centre (xl) when n is large*. He did not consider the case of small samples, but used his results as an illustration of the way in which the method of maximum likelihood picks out the "statistic" G as the estimate of the population mean which has the minimum standard error. If the estimate is to be based in this way on two observations only, it is of course necessary that the grouping should be fine, and that there should be no possibility that either of the extreme observations is subject to error of measurement or of record.

* *Phil. Trans.* A, Vol. 222, p. 348.

(11) *Solutions obtained by the Inverse Method.*

As a population is completely specified by the position of its centre g, and range w, the problem can be considered in the (g, w)-plane. This is represented in Figure 10.

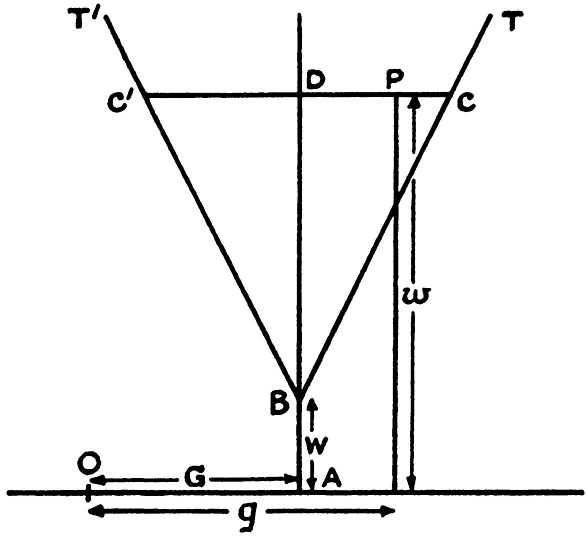

FIG. 10

B is the sample point (G, W), and P represents a population (g, w). For it to be possible that Σ has come from Π, it is necessary that

$$g - \tfrac{1}{2}w \leqslant G - \tfrac{1}{2}W \leqslant G + \tfrac{1}{2}W \leqslant g + \tfrac{1}{2}w,$$

or that
$$-\tfrac{1}{2}(w - W) \leqslant g - G \leqslant \tfrac{1}{2}(w - W),$$

or that
$$w \geqslant W + 2\,|\,g - G\,|.$$

Consequently the (g, w)-field is limited to the region inside the acute angle TBT', where $\angle\, TBD = \angle\, T'BD = \tan^{-1}\tfrac{1}{2}$. The likelihood of a rectangular population represented by P is $L \propto w^{-n}$, and we may give to the field a point-density $d = w^{-n}$. This is constant along lines such as CDC' parallel to the axis of g. L itself provides a criterion by which to judge the probability of Hypothesis A, but we obtain a more readily comprehensible scale by taking the ratio of (1) I', the integral of d throughout the portion of the field $TCC'T'$, to (2) I, the integral over the whole field TBT'. This gives, after simple integration,

$$P_w = I'/I = \{(n-1)\,w - (n-2)\,W\}\; W^{n-2}/w^{n-1} \quad \text{..........(xlv)},$$

and provides a measure not only of the probability that the range in the sampled population is $\geqslant w$, but also from the standpoint of likelihood it is the criterion to use as to the probability of Hypothesis A taken as a whole.

Finally we may consider the probability that the centre of the sampled population is such that $|\,g - G\,| \geqslant \alpha$, say. We shall obtain a measure of this by integrating d over the two corresponding regions of the field, and dividing the result by I. It is found that

$$P_g = \left\{1 + \frac{2\,|\,g - G\,|}{W}\right\}^{-n+2}, \text{ where } |\,g - G\,| = \alpha \text{ ....(xlvi).}$$

Compare now (xlv) with (xxxix) and (xlvi) with (xliv), remembering that $z' = 2\,|\,G - g\,|/W$. It will be seen that by changing $n - 1$ into n in the solutions of the inverse method, P_w becomes identical with P_W and P_g with $P_{z'}$. Consequently for a given sample (G, W) and hypothetical population (g, w), these pairs of criteria will give almost identical numerical values unless the sample is very small. The distributions of range and of the ratio z', (xxxviii) and (xliii), were in fact first obtained by one of us from the integrals P_w and P_g of the inverse method, making the change from $n - 1$ into n from analogy with the solutions in the normal field!

(12) *Illustrations from Experimental Sampling.*

In practical application the weakness of the methods discussed—and indeed of any method of dealing with small samples—lies in the fact that while questioning the position (as measured by m or g) and the scale (by σ or w) of the sampled population, they provide no test of the shape of the distribution. They assume that the β_1 and β_2 of the population are known to be 0 and 3, or 0 and 1·8, as the case may be. In Section (8), analysing Church's samples of 10 from a skew population, we have considered from one direction what is the risk of error involved in making these assumptions. We shall now consider this from another direction.

Two sets of random samples from a rectangular population were obtained*:

(*a*) 500 samples of 4. Tippett's Random Sampling Numbers† were used, the distribution of the population being divided into 21 groups. As a check on the random nature of the sampling, the distribution of the total 2000 observations was compared with the "expected" distribution, namely one containing 95·24 observations in each of the 21 groups. Applying the (P, χ^2) test for goodness of fit, it was found that $\chi^2 = 16{\cdot}08$ and $P = {\cdot}712$.

(*b*) 540 samples of 10. These were obtained by drawing 10 coloured beads from a box containing 50 beads of each of 11 different varieties. The beads were replaced and shuffled after each draw of 10. The population was not therefore strictly "infinite," but as $M = 55n$ the effect is hardly likely to have been of great importance. Unfortunately it was not realised until afterwards that the sampling was biased because the various types of beads were sufficiently different in size to make the chances of drawing beads of the different colours not always equal to 1/11. The distribution of frequencies among the 11 groups for 4200 beads was 409, 446, 362, 404, 355, 407, 342, 422, 363, 375, 315. The "expected" number in each group is 381·8. Applying the (P, χ^2) test, it is found that

* (*a*) were drawn by Miss M. Page, (*b*) by Miss Page and two students of University College, London, Miss F. E. James and Mr T. Uchida. The work of computation has been mainly Miss Page's.

† Published as No. XV of the *Tracts for Computers*, Cambridge University Press.

$\chi^2 = 39{\cdot}71$ and $P = {\cdot}00002$. The samples must therefore be considered as drawn from some irregular and slightly skew population, and no very detailed analysis of the results seemed to be justified.

Taking the experimental sampling (*a*), we may first test the formulae of the preceding section.

The Distribution of Range, W.

In sampling, an observation fell into one or other of the 21 equal groups into which the population rectangle was divided. An exact range for the sample was not therefore obtainable, but it seemed that an adequate test could be obtained by taking the population range as $w = 21$, and by supposing that samples whose highest and lowest observations fell into:

(1) neighbouring groups, (2) groups 2 apart, (3) groups 3 apart, ... (20) groups 20 apart,

had ranges, W, falling into groups

(1) 0–1·5, (2) 1·5–2·5, (3) 2·5–3·5, ... (20) 19·5–21·0.

The theoretical distribution of W obtained from (xxxviii) is

$$y = 12\,(21\,W^2 - W^3)/(21)^4 \quad\text{.......................(xlvii).}$$

Taking its integral over the corresponding grouping ranges for W, we have the set of observed and theoretical frequencies shown in Table IV. The (P, χ^2) test makes $\chi^2 = 13{\cdot}43$, which for $n' = 17$ gives $P = {\cdot}641$.

The Distribution of Centre, G.

The origin for G was taken at the centre of the middle or 11th population group; the values of G were either integral or contained half units.

A value of G, "0"	was taken as lying in a group ranging from	$-0{\cdot}25$	to $+0{\cdot}25$,
,, ,, "0·5"	,, ,, ,, ,,	$+0{\cdot}25$	,, $+0{\cdot}75$,
,, ,, "1·0"	,, ,, ,, ,,	$+0{\cdot}75$	,, $+1{\cdot}25$,
etc.			

The theoretical distribution of G obtained from (xl) is

$$y = 4\,\{1 - 2\,|\,G\,|/21\}^3/21 \quad\text{.......................(xlviii).}$$

Taking its integral for the required groups, we obtain the series of observed and theoretical frequencies for G given in Table IV. These results lead to $\chi^2 = 17{\cdot}31$, which for $n' = 21$ gives $P = {\cdot}633$.

For the mean and standard deviation of G, we have

	Observation	Theory	Standard Error
Mean	$-{\cdot}0950$	0	·1212
Standard deviation	2·6778	2·7111	·0857 *

* Approximately, on the assumption that the distribution of G is a normal curve.

TABLE IV. *Distribution of Frequency Constants in 500 Samples of 4 from a Rectangular Population.*

Frequencies for Range

W	Observed	Expected
0—1·5	1	0·7
1·5—2·5	1	2·3
2·5—	5	5·0
3·5—	3	8·4
4·5—	10	12·3
5·5—	25	16·7
6·5—	20	21·2
7·5—	25	25·7
8·5—	29	30·0
9·5—	40	33·9
10·5—	33	37·3
11·5—	45	39·9
12·5—	46	41·7
13·5—	47	42·3
14·5—	35	41·6
15·5—	32	39·4
16·5—	35	35·6
17·5—	28	29·9
18·5—19·5	24	22·2
19·5—21·0	16	13·9
Total	500	500·0

Frequencies for Centre

G	Observed	Expected
−*8·75* to −*9·25*	1	
−*8·25*—	—	
−*7·75*—	—	6·7*
−*7·25*—	3	
−*6·75*—	—	
−*6·25*—	5	
−*5·75*—	4	3·8
−*5·25*—	7	5·1
−*4·75*—	6	6·9
−*4·25*—	10	8·9
−*3·75*—	12	11·3
−*3·25*—	9	14·1
−*2·75*—	18	17·4
−*2·25*—	26	21·1
−*1·75*—	21	25·3
−*1·25*—	27	30·0
−*0·75*—	43	35·3
−*0·25*—	40	41·2
+*0·25* to −*0·25*	38	45·8
+*0·75* to +*0·25*	38	41·2
+*1·25*—	41	35·3
+*1·75*—	33	30·0
+*2·25*—	34	25·3
+*2·75*—	15	21·1
+*3·25*—	20	17·4
+*3·75*—	15	14·1
+*4·25*—	11	11·3
+*4·75*—	4	8·9
+*5·25*—	8	6·9
+*5·75*—	3	5·1
+*6·25*—	4	3·8
+*6·75*—	1	
+*7·25*—	2	6·7*
+*7·75* to +*7·25*	1	
Total	500	500·0

Frequencies for ratio, z'

z'	Observed	Expected
−*3·00* and less	6	3·9
−*2·75* to *3·00*	—	
−*2·50*—	1	5·4
−*2·25*—	—	
−*2·00*—	2	
−*1·75*—	5	6·7
−*1·50*—	8	
−*1·25*—	4	6·0
−*1·00*—	16	9·3
−*0·75*—	12	15·4
−*0·50*—	25	27·4
−*0·25*—	52	53·9
0 to −*0·25*	121	122·0
0 to +*0·25*	113	122·0
+*0·25*—	65	53·9
+*0·50*—	30	27·4
+*0·75*—	14	15·4
+*1·00*—	10	9·3
+*1·25*—	5	6·0
+*1·50*—	2	6·7
+*1·75*—	1	
+*2·00*—	3	
+*2·25*—	2	5·4
+*2·50*—	—	
+*2·75* to *3·00*	—	
+*3·00* and more	3	3·9
Total	500	500·0

Frequencies for ratio, z

z	Observed	Expected ("normal" case)
−*5·0* and less	3	
−*4·5* to *5·0*	1	1·6
−*4·0*—	2	
−*3·5*—	1	1·9
−*3·0*—	—	
−*2·5*—	7	2·2
−*2·0*—	3	4·4
−*1·5*—	12	10·0
−*1·0*—	20	25·3
−*0·5*—	60	67·1
0 to −*0·5*	139	137·5
0 to +*0·5*	145	137·5
+*0·5*—	72	67·1
+*1·0*—	18	25·3
+*1·5*—	7	10·0
+*2·0*—	3	4·4
+*2·5*—	3	2·2
+*3·0*—	2	1·9
+*3·5*—	1	
+*4·0*—	—	
+*4·5* to *5·0*	1	1·6
+*5·0* and more	—	
Total	500	500·0

* Theoretical frequencies from $G=6·25$ to $10·5$, or tail of curve.

The standard deviation of the sampled population is $21/(2\sqrt{3}) = 6{\cdot}0622$; consequently the standard deviation of the mean in samples of 4 is $\sigma/\sqrt{n} = 3{\cdot}0311$, illustrating Fisher's point referred to above that G provides a better estimate of the population centre or mean than $\bar{X}$, the sample mean.

The Distribution of $z' = 2(G-g)/W$.

Using (xliii), we see that the distribution of z' is

$$y = \tfrac{3}{2}(1+|z'|)^{-4} \qquad \text{(xlix)}.$$

Integrating this expression within the grouping limits shown, we obtain the observed and theoretical frequencies given in Table IV. Here $\chi^2 = 12{\cdot}64$ and $n' = 14$, so that $P = {\cdot}477$.

The agreement of observation with theory for these three distributions is therefore very satisfactory. We may consider now the further problem. There exist a great variety of limited range symmetrical (and unimodal) distributions with values of β_2 lying between $3{\cdot}0$ and $1{\cdot}8$; it is very unlikely that, however inadequate our information, we should suppose that the population from which a sample has been drawn to be normal when in fact it was rectangular. But we can obtain some idea of the degree to which the normal criteria will fail for samples from populations with intermediate forms—say with β_2 $2{\cdot}0$ to $2{\cdot}6$—by testing the adequacy of the normal theoretical distributions in this extreme case of $\beta_2 = 1{\cdot}8$.

Take first the distribution of range. For a rectangular distribution $w = 2\sqrt{3}\sigma$; it follows from (xxxviii) that the frequency constants of the sample range, W, are:

$$\left.\begin{aligned}
&\text{Mean} = \frac{n-1}{n+1}w = 2\sqrt{3}\,\frac{n-1}{n+1}\sigma\\
&\text{Standard Deviation} = \frac{w}{n+1}\sqrt{\frac{2(n-1)}{n+2}} = \frac{2\sigma}{n+1}\sqrt{\frac{6(n-1)}{n+2}}\\
&\beta_1 = \frac{2(n+2)(n-3)^2}{(n-1)(n+3)^2};\quad \beta_2 = \frac{6(n+2)(n^2-2n+3)}{(n-1)(n+3)(n+4)}
\end{aligned}\right\}.$$

The ratios of mean range and of standard deviation of range to the standard deviation of the population, and the β_1 and β_2 of the range distribution for samples from a normal population, are known*. These values are compared with the true values in Table V, for $n = 3$, 4, 6, 10 and 20. For the smaller samples there is a fair degree of correspondence in the mean range, but as n increases the ratio for the rectangular distribution tends to $2\sqrt{3} = 3{\cdot}4641$ while that for the normal distribution increases slowly but without limit. Even at $n = 3$, the standard deviation ratios have little correspondence, while the β_1 and β_2 values show no agreement, except in giving increasing skewness as n increases.

* L. H. C. Tippett, *Biometrika*, Vol. XVII. pp. 364—387. The values given in the table were actually taken from the summary given in *Biometrika*, Vol. XVIII. p. 192.

TABLE V.

Comparison of Frequency Constants of Range.

	Population	Size of Sample				
		3	*4*	*6*	*10*	*20*
Mean W/σ ...	Rectangular	1·732	2·078	2·474	2·834	3·134
	Normal	1·693	2·059	2·534	3·078	3·735
Standard Deviation of W/σ	Rectangular	·775	·693	·553	·386	·217
	Normal	·888	·880	·848	·797	·729
β_1 ...	Rectangular	0·000	·082	·356	·773	1·265
	Normal	·417	·273	·189	·156	·161
β_2 ...	Rectangular	2·143	2·357	2·880	3·648	4·569
	Normal	3·286	3·188	3·170	3·22	3·26

Consider next what would be the result of supposing the sampled population to have been normal, and applying Student's z-test to measure the probability of Hypothesis B. Does the distribution of $z = m/s$ for these samples follow approximately the law $f(z) = \text{constant} \times (1 + z^2)^{-2}$? A comparison of observation and normal theory is given in the fourth section of Table IV. It is found that $\chi^2 = 10{\cdot}77$, which for $n' = 10$ gives $P = {\cdot}293$. The main discrepancy, which is not serious, lies in some large negative values of the observed z's. The observed standard deviation is $\sigma_z = 1{\cdot}3030$ against the normal value of $1/\sqrt{n-3} = 1{\cdot}0000$. As the standard error of σ_z in the normal case is infinite, owing to the 4th moment of the z-distribution being infinite, it is not easy to estimate the significance of this difference. Judging only from the χ^2-test, however, Student's distribution appears quite adequate to describe the variation in z, as far as the evidence from these 500 samples goes.

It must be remembered, however, that even if z calculated for samples from a rectangular distribution follows Student's law, P_z has changed its meaning as a criterion; it no longer enables us to reject or accept Hypothesis B upon the same logical basis. As we have seen, a more suitable ratio to employ would now be $z' = 2(G - g)/W$. But is z highly correlated with z'? If this were so, in judging from z we should reach results not very different from those based on z'. We have calculated the coefficient of correlation between z and z' for the 500 samples; it has the very high value of $+ {\cdot}9756$. These results therefore suggest that for samples of 4 from a rectangular population, (a) the distribution of $z = m/s$ follows Student's law with sufficient approximation for ordinary purposes, and (b) the criterion P_z will be closely equivalent to $P_{z'}$. Consequently, if we assume erroneously that the sampled population is normal when in fact it has a β_1 and β_2 on or close to the stretch of the β_2 axis lying between $\beta_2 = 3{\cdot}0$ and $1{\cdot}8$, we shall not be led into serious errors of judgment by using Student's test. The results for samples of 10 considered below suggest that for this case also point (a) holds good, and since as n increases the distribution of m/s tends to that of m/σ or to a normal curve with standard deviation $\sigma/\sqrt{n}$ whatever be the form of the population sampled, the most crucial test of correspondence (except for the cases $n = 2$ and 3) would appear to have been made at $n = 4$. On the other hand, the correlation between z and z' which is perfect at $n = 2$ will decrease as n increases, and the mean and standard deviation of the sample are less and less influenced by the values of the two extreme observations. The interpretation of the z-test will therefore become more and more doubtful.

TABLE VI.

Distribution of z in 540 samples of 10.

Central values of z	Less than $-1{\cdot}80$	$-1{\cdot}75$	$-1{\cdot}65$	$-1{\cdot}55$	$-1{\cdot}45$	$-1{\cdot}35$	$-1{\cdot}25$	$-1{\cdot}15$	$-1{\cdot}05$	$-{\cdot}95$	$-{\cdot}85$	$-{\cdot}75$
Observation ...	—	1	—	—	1	1	1	1	3	1	5	5
Normal Theory	0·1	0·1	0·1	0·1	0·2	0·4	0·6	0·9	1·5	2·5	4·2	6·8
z	$-{\cdot}65$	$-{\cdot}55$	$-{\cdot}45$	$-{\cdot}35$	$-{\cdot}25$	$-{\cdot}15$	$-{\cdot}05$	$+{\cdot}05$	$+{\cdot}15$	$+{\cdot}25$	$+{\cdot}35$	$+{\cdot}45$
Observation ...	11	13	25	29	44	53	71	55	86	47	29	18
Normal Theory	10·9	16·8	25·1	35·4	46·3	56·1	61·9	61·9	56·1	46·3	35·4	25·1
z	$+{\cdot}55$	$+{\cdot}65$	$+{\cdot}75$	$+{\cdot}85$	$+{\cdot}95$	$+1{\cdot}05$	$+1{\cdot}15$	$+1{\cdot}25$	$+1{\cdot}35$	Greater than $+1{\cdot}40$		Total
Observation ...	19	11	1	2	2	3	1	—	1	—		540
Normal Theory	16·8	10·9	6·8	4·2	2·5	1·5	0·9	0·6	0·4	0·6		540·0

We shall now consider briefly the experimental sampling (b), with 540 samples of 10. As stated above, no very detailed analysis is justifiable. Since the population was only divided into 11 groups, it was not possible to obtain satisfactory distributions of W, G or z'. The mean, standard deviation and ratio z were, however, calculated. Owing to the biased sampling, the population could not be considered as exactly rectangular. In order to correct to some extent the error involved, the means, m, were measured not from the centre of the middle or 6th population group, but from a point $-{\cdot}19$ from this, corresponding more nearly to the mean as observed from the samples.

The observed distribution of z and the corresponding theoretical frequencies are shown in Table VI. The fit is bad, giving $\chi^2 = 27{\cdot}16$, which for $n' = 16$ makes $P = {\cdot}027$. A comparison of the frequencies shows that this is due to a tremendous contribution of 15·94 to χ^2 for the group $z = +\cdot 1$ to $+\cdot 2$. But beyond this, the observed distribution of z is somewhat skew, having $\beta_1 = \cdot 395$ and $\beta_2 = 4\cdot 988$. Now this cannot be due to the error involved in assuming that the z of samples from a rectangular population follows the distribution appropriate for a normal population, for in both cases the distribution of z must be symmetrical. It is probable that the asymmetry as well as the great discrepancy in the group $z = +\cdot 1$ to $+\cdot 2$ results from the irregularity of the bead sampling combined with some effect on the grouping of z introduced by the correction to m referred to above.

Using the only other measure of comparison, the standard deviation of z, we find:

$$\text{from observations } \sigma_z = \cdot 3707; \qquad \text{from theory } \sigma_z = \cdot 3780,$$

with a standard error which would, if the distribution of σ_z were normal, be ·0113. It appears therefore that as far as the scale goes, the observed z distribution is represented by the "normal" value of σ_z, or $1/\sqrt{n-3}$. Further sampling would be necessary to determine the form of the curve*, but as the probability integrals of symmetrical curves with the same standard deviation change very slowly with β_2, it seems very probable that for $n = 10$ Student's distribution will still represent adequately the distribution of z from a rectangular population. The interpretation of P_z in testing Hypothesis B is again of course doubtful.

While the distributions of z appear in both these series of samples to be represented approximately by Student's curve, the normal P_λ contours are found to be quite inadequate. This is partly because the regression of s on m is no longer linear, and partly because the distribution of the s-margin cannot be represented even approximately by the normal s-distribution.

The standard deviation of the variance, s^2, is given by

$$\sigma_{s^2} = \sigma^2 \frac{(n-1)}{n^{\frac{3}{2}}} \sqrt{\beta_2 - 3 + \frac{2n}{n-1}},$$

where β_2 refers to the sampled population. Further $\sigma_s = \sigma_{s^2}/2\sigma$ fairly closely for large samples, though with increasing approximation as n decreases. It follows

TABLE VII.

	Observations from Rectangular Population					Theory for Normal Population				
n	Mean s/σ	σ_s/σ	β_1	β_2	*Sk.*	Mean s/σ	σ_s/σ	β_1	β_2	*Sk.*
4	·822	·275	·006	2·379	−·074	·798	·337	·236	3·108	+·270
10	·932	·148	·070	2·761	−·174	·923	·220	·063	3·011	+·128

that σ_s is very sensitive to changes in β_2, and the drop from 3·0 to 1·8 in β_2 completely alters its value. Also, as the results in Table VII show, the observed

* In view of the complex nature of the distribution of m as found by Irwin and Hall (*Biometrika*, Vol. xix. pp. 225 and 240) it seems unlikely that the theoretical distribution of z would be easy to find.

distributions of s for the samples from the rectangular population are skew in the opposite sense to those from the normal population; the skewness increases with n.

In view of this complete lack of correspondence, except perhaps for the mean $\bar{s}$, it is not surprising to find that the P_λ contours are here of little value. Placing down upon each (m, s)-distribution three contours, the results of Table VIII below were obtained. The observed frequencies are in all cases less than those expected on normal theory, so that in using the P_λ contours we should be inclined to underestimate the significance of a divergence in the sample point. At $\beta_2 = 1{\cdot}8$ we have reached a position quite out of touch with the normal field; it is hoped shortly to provide information on the position at an intermediate point, say $\beta_2 = 2{\cdot}5$.

TABLE VIII.

$n=4$			$n=10$		
k	Observed frequency outside contour	$P_\lambda \times 500$	k	Observed frequency outside contour	$P_\lambda \times 540$
·70	138	196·5	·50	179	271·9
1·00	48	68·8	·70	19	33·8
1·50	11	12·1	·90	2	4·2

IV. SAMPLING FROM AN EXPONENTIAL POPULATION.

(13) *Sampling Distributions of Frequency Constants.*

The frequency distribution of the population may be written

$$y = \frac{1}{\sigma} e^{-\frac{X-b}{\sigma}} \qquad \text{(l)},$$

where b is the distance from the origin to the start of the curve, $b + \sigma$ the distance from the origin to the mean, and σ the standard deviation of the curve. In the fundamental space of Π, we take rectangular axes through the point $(b, b, \ldots b)$, and writing $x = X - b$, the field in which Σ may lie is that for which all the x's are positive. The appropriate density with which to fill this field will be

$$D = \sigma^{-n} e^{-\frac{S(x)}{\sigma}} = \sigma^{-n} e^{-\frac{n\bar{x}}{\sigma}} \qquad \text{(li)},$$

where $\bar{x}$ is the distance of the sample mean from the start of the population curve. The contours of equal density will be the m contours, which are primes in the n-fold space lying at right angles to the diagonal axis

$$x_1 = x_2 = \ldots = x_n \qquad \text{(iii } bis\text{)}.$$

We shall now find the contours in this space upon which the ratio

$$\lambda = \text{Likelihood of } \Pi/\text{Likelihood of } \Pi'\,(\max)$$

is constant. For an exponential population, Π', with start at b' and standard deviation σ', we have

$$L \propto \sigma'^{-n} e^{-\frac{n(\bar{X}-b')}{\sigma'}} \quad \text{.............................(lii).}$$

Whence $\dfrac{\partial L}{\partial b'} \propto n\sigma'^{-n-1} e^{-\frac{n(\bar{X}-b')}{\sigma'}}$, showing that for a given value of σ', L is a maximum when $\bar{X}-b'$ is as small as possible. Clearly this will occur when $b'=X_1$, the lowest of the sample variates, for it is impossible for Σ to have come from Π' if $X_1 < b'$.

Further $\dfrac{\partial L}{\partial \sigma'} \propto n\sigma'^{-n-2} e^{-\frac{n(\bar{X}-b')}{\sigma'}} \{\bar{X}-b-\sigma'\}$. This expression vanishes when

(a) $\sigma'=0$ making $L=0$.
(b) $\sigma'=\infty$ „ $L=0$.
(c) $\sigma'=\bar{X}-b'$ „ $L \propto (\bar{X}-b')^{-n} e^{-n}$.

The last is the maximum solution; it follows that for a given sample, the exponential population of maximum likelihood is that for which

$$b'=X_1=b+x_1; \quad \sigma'=\bar{X}-X_1=\bar{x}-x_1.$$

Hence the ratio λ is given by

$$\lambda = \left(\frac{\bar{x}-x_1}{\sigma}\right)^n e^{-\frac{n(\bar{x}-\sigma)}{\sigma}} \quad \text{..........................(liii).}$$

This result shows that the method of likelihood picks out $\bar{x}$ the mean, and x_1 the lowest observation, as the two descriptive measures of the sample; that is to say it suggests that we should take $X_1=b+x_1$ and $l=\bar{X}-X_1=\bar{x}-x_1$ as the sample estimates of the population parameters b and σ respectively (see Figure 12). In the fundamental space, λ is constant upon the contours

$$(x_1+l)/\sigma - \log_e(l/\sigma) = k = 1-(1/n)\log_e\lambda \text{.................(liv),}$$

and this will be the system to use in testing Hypothesis A according to the criterion of likelihood. The problem can be referred to the (x_1, l)-plane, and it is necessary to find the appropriate density, d, for this field. This can be obtained most readily by obtaining first the $(x_1, \bar{x})$-field.

The chance of drawing a sample from Π with the n observations in the limits $x_1 \pm \frac{1}{2}dx_1, \ldots x_n \pm \frac{1}{2}dx_n$ is

$$C\,dx_1 \ldots dx_n = \sigma^{-n} e^{-\frac{n\bar{x}}{\sigma}}\, dx_1 \ldots dx_n.$$

Take now as variables

$$\begin{cases} z_0 = x_1 + x_2 + \ldots + x_n, \\ z_1 = x_1, \text{ the lowest observation of sample,} \\ z_3 = x_3,\ z_4 = x_4, \ldots z_n = x_n. \end{cases}$$

Hence

$$C\,dx_1dx_2 \ldots dx_n = C\,dz_0dz_1dz_3 \ldots dz_n = \sigma^{-n} e^{-\frac{z_0}{\sigma}}\, dz_0dz_1dz_3 \ldots dz_n.$$

For a given set of values of $z_0, z_1, z_4, \dots z_n$, the variable z_3 cannot be less than z_1 and takes its highest value when $x_2 = x_1 = z_1$, and consequently

$$2x_1 + x_3 + x_4 + \dots + x_n = n\bar{x}, \text{ or } z_3 = z_0 - 2z_1 - z_4 \dots - z_n = u_1.$$

Hence

$$C\,dz_0\,dz_1\,dz_4 \dots dz_n = dz_0\,dz_1\,dz_4 \dots dz_n\,\sigma^{-n} e^{-\frac{z_0}{\sigma}} \int_{z_1}^{u_1} dz_3$$

$$= dz_0 \dots dz_n\,\sigma^{-n} e^{-\frac{z_0}{\sigma}} (z_0 - 3z_1 - z_4 \dots - z_n).$$

Now for given values of $z_0, z_1, z_5, \dots z_n$, the variable z_4 can vary from z_1 to $u_2 = z_0 - 3z_1 - z_5 \dots - z_n$, corresponding to the case where both x_2 and x_3 have their lowest possible values of x_1.

Hence

$$C\,dz_0\,dz_1\,dz_5 \dots dz_n = dz_0 \dots dz_n\,\sigma^{-n} e^{-\frac{z_0}{\sigma}} \int_{x_1}^{u_2} (z_0 - 3z_1 - z_4 \dots - z_n)\,dz_4$$

$$= dz_0 \dots dz_n\,\sigma^{-n} e^{-\frac{z_0}{\sigma}} \tfrac{1}{2} (z_0 - 4z_1 - z_5 \dots - z_n)^2.$$

Proceeding in this way we can integrate out successively for all the variables $z_5, z_6, \dots z_n$, and shall be left with

$$C\,dz_0\,dz_1 = dz_0\,dz_1\,\sigma^{-n} e^{-\frac{z_0}{\sigma}} (z_0 - nz_1)^{n-2}/(n-2)!.$$

This process of integration will have covered only $1/n$ of the complete fundamental space, namely that portion in which $x_1 = z_1$ is the lowest observation. There will be $n-1$ similar regions in which $x_2, x_3, \dots$ or x_n is the lowest. To obtain the chance that the mean of the sample lies in the limits $\bar{x} \pm \frac{1}{2} d\bar{x}$, and that the lowest value of n undifferentiated observations lies in the limits $x_1 \pm \frac{1}{2} dx_1$, we must multiply the above expression by n and write $z_0 = n\bar{x}$, $z_1 = x_1$, giving

$$\phi_1(\bar{x}, x_1)\,d\bar{x}\,dx_1 \equiv \sigma^{-n} (n^n/(n-2)!) (\bar{x} - x_1)^{n-2} e^{-\frac{n\bar{x}}{\sigma}} d\bar{x}\,dx_1 \quad \text{.........(lv).}$$

From (lv), integrating for x_1 between the limits 0 and $\bar{x}$, we find the distribution of means, $\bar{x}$, or

$$\phi_2(\bar{x}) \equiv \sigma^{-n} (n^n/(n-1)!)\,\bar{x}^{n-1} e^{-\frac{n\bar{x}}{\sigma}} \quad \text{....................(lvi),}$$

a Type III curve corresponding to a special case of the result obtained by Church*.

Next, upon integrating out for $\bar{x}$ between the limits x_1 and ∞, we obtain for the distribution of the lowest observation in the sample,

$$\phi_3(x_1) \equiv (n/\sigma) e^{-\frac{nx_1}{\sigma}} \quad \text{............................(lvii),}$$

which is another exponential curve, with standard deviation σ/n.

* *Biometrika*, Vol. XVIII. p. 336, giving the distribution of means in samples of n from a Type III population.

Putting in (lv), $\bar{x} = l + x_1$, we obtain for the frequency function of x_1 and l,

$$\phi_4(x_1, l) \equiv \sigma^{-n}(n^n/(n-2)!)\, l^{n-2} e^{-\frac{n(l+x_1)}{\sigma}} \quad \ldots\ldots\ldots\ldots\ldots \text{(lviii)}.$$

If then we know that the sample Σ has been drawn from *some* exponential population and we wish to judge the probability that it has been taken from Π, the suitable criterion to take is P_λ, or the integral of a density $d = \phi_4(x_1, l)$ taken over the portion of the (x_1, l)-plane lying outside the member of the system (liv) passing through the sample point, (x_1, l).

The (x_1, l)-plane is represented in Figure 11, OA, OB being the axes of l and x_1. G is the point $x_1 = 0$, $l = \sigma$, representing the sample for which Π would be the population of maximum likelihood. By differentiating (liv) it is seen that

$$\frac{dx_1}{dl} = \frac{\sigma - l}{l},$$

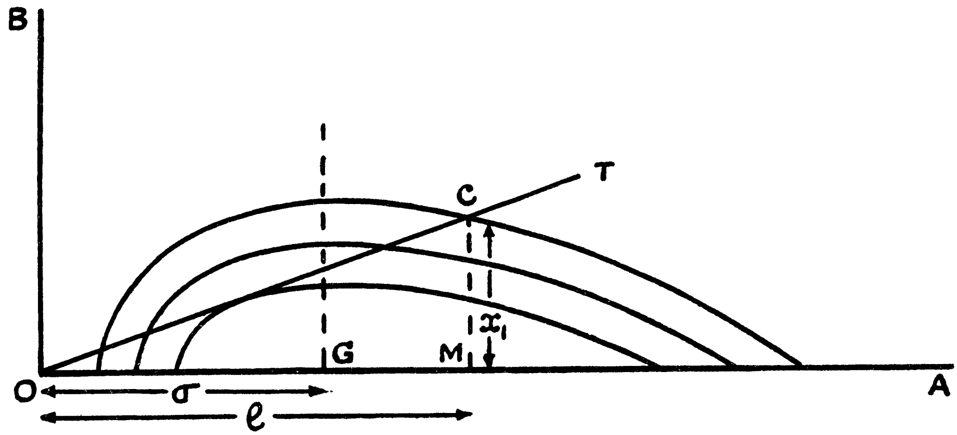

Fig. 11

and consequently all contours cut a line $l =$ constant at the same angle. The differential coefficient is zero along the line $l = \sigma$, and for very divergent contours tends to $+\infty$ and -1 at the two points of intersection with the axis of l. C represents a sample Σ; P_λ is the integral of the field taken outside the curve through C. We have not attempted to compute this.

Consider now Hypothesis B. Having a sample Σ, with observations $X_1, X_2, \ldots X_n$, which we have reason to believe has been drawn from *some* exponential population, we wish to measure the probability that it has been drawn from a population whose "start" or lower limit is at a distance b from the origin. The parameter, σ, of the population is unspecified. Then, as before, we consider the ratio of

(1) the maximum likelihood for an exponential population starting at b,

to (2) the maximum likelihood for any exponential population.

The former corresponds to the population for which $b' = b$, $\sigma' = \bar{X} - b$, the latter for which $b' = X_1$, $\sigma' = \bar{X} - X_1$, as found from (lii).

The position is shown diagrammatically in Figure 12, where the observations of the sample are shown as black circles, and the exponential curves of the two hypothetical populations have been dotted in. The ratio of the two likelihoods becomes

$$(\bar{X}-b)^{-n}e^{-n}/(\bar{X}-X_1)^{-n}e^{-n}=\left(1+\frac{x_1}{l}\right)^{-n}.$$

This is constant along the radiating lines, such as OCT in Figure 11, upon which the ratio $z''=x_1/l=\dfrac{X_1-b}{l}$ is fixed. We may therefore take these as the appropriate contours to use in testing Hypothesis B. It remains to find the distribution of z''. Writing in (lviii) $x_1=lz''$, we have

$$\phi_5(z'',l)\,dz''dl\equiv\sigma^{-n}(n^n/(n-2)\,!)\,l^{n-1}e^{-\frac{nl(1+z'')}{\sigma}}\,dz''dl \quad \text{.........(lix).}$$

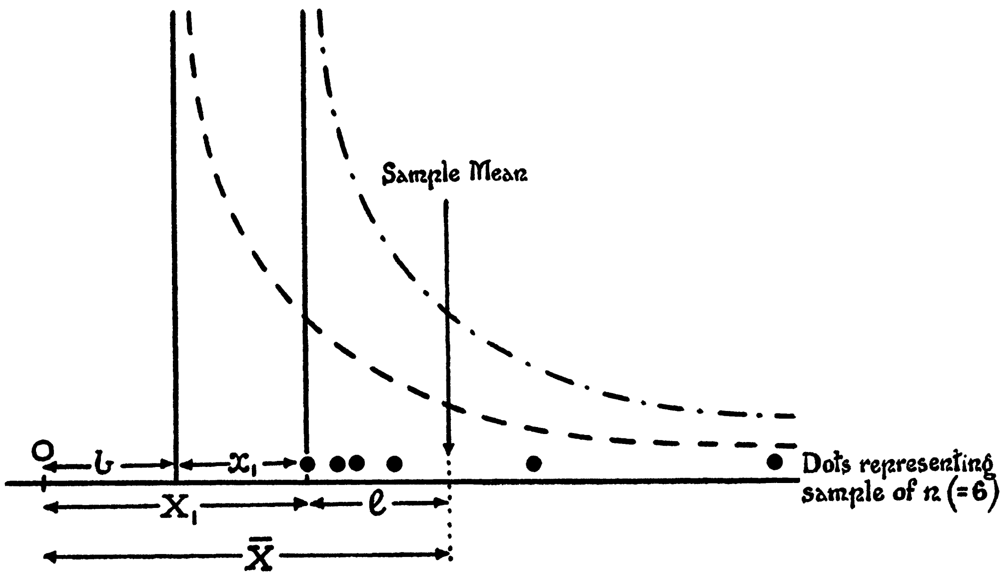

FIG. 12

This must be integrated for values of l between 0 and ∞. Writing

$$t=nl(1+z'')/\sigma,$$

we find

$$\phi_6(z'')\,dz''=(1+z'')^{-n}dz''\int_0^\infty t^{n-1}e^{-t}dt/(n-2)\,!,$$

or

$$\phi_6(z'')\equiv(n-1)(1+z'')^{-n} \quad \text{..........................(lx).}$$

It is easily shown that

$$\text{Mean } z''=1/(n-2).$$

$$\text{Standard Deviation of } z''=\sqrt{(n-1)/(n-2)^2(n-3)}.$$

Further we obtain

$$P_{z''}=\int_{z''}^\infty(n-1)(1+z'')^{-n}dz''=(1+z'')^{-n+1} \quad \text{..............(lxi),}$$

which may be used as a criterion for testing Hypothesis B. It is comparable with and very similar to P_z, (xi), and $P_{z'}$, (xliv), corresponding to the normal and rectangular cases. z, z' and z'' are all ratios of a deviation in the measure of position to the measure of scale obtained from the sample.

These results which have been expressed in terms of x_1 and l are again what may be termed the "ideal" results for random sampling. Unfortunately in practice two difficulties often arise:

(*a*) There may be slight errors in observation, or variation in the character measured in an individual, which are not due to chance fluctuations.

(*b*) The unit of measurement or the grouping unit with which data have been tabled may be too coarse for the value of x_1 to be determined with adequate precision.

Under such conditions a method which provides a criterion depending upon the measure of a single variate, such as x_1, will be at a disadvantage compared with one which depends upon an average value, such as the mean or the standard deviation. In the latter case the effect of a few badly measured observations, a few disturbed variations or of coarse grouping is likely to cancel out. There are, however, we believe, problems to which the methods of the present section can be applied, and one of these in connection with random intervals we shall take as an illustration in the section that follows.

It has not seemed necessary to develop the inverse solution considered in dealing with the normal and rectangular populations, but the working would be quite straightforward.

(14) *Test for Random Intervals.*

Whitworth, in *Choice and Chance**, has shown that if the intervals of time between events which happen once on the average in time c may be considered as random, then the chance after an event has happened that another will not occur in an interval t is given by $e^{-\frac{t}{c}}$. If $y = f(t)$ represents the frequency distribution of the random intervals between the events, then

$$e^{-\frac{t}{c}} = \int_t^\infty f(t)\,dt, \text{ or } f(t) = \frac{1}{c}e^{-\frac{t}{c}}.$$

Morant† has extended this idea to the case where there is necessarily a closed interval of length b after an event before another can occur. The frequency distribution of intervals will now be

$$y = \frac{1}{c}e^{-\frac{t-b}{c}} \quad \text{.................................(lxii).}$$

Suppose that we have $v = n + 1$ consecutive observations of the occurrence of an event, recorded at times $t_1, t_2, t_3, \ldots t_v$, and that the n intervals or differences between the t's are $i_1, i_2, \ldots i_n$. Let i_s be the shortest of these intervals. Let us

* See p. 200, Prop. LI.

† *Biometrika*, Vol. XIII. pp. 309—337.

imagine in the first place that we have reason to believe that the intervals are random, but that there is between each a closed period of length b. We wish to test from the observations whether this hypothesis as to b is reasonable, but are not interested in the average time c. The problem is exactly that of testing Hypothesis B. With the previous notation

$$\begin{cases} x_1 = i_s - b, \\ l = (i_1 + i_2 + \ldots + i_n)/n - i_s = (t_v - t_1)/n - i_s, \\ z'' = x_1/l, \\ P_{z''} = (1 + z'')^{-n+1}. \end{cases}$$

In following this method it is important to be clear exactly what interpretation may be placed on the value found for $P_{z''}$. A large value (say $1 > P_{z''} > 0{\cdot}5$) will not imply that the intervals (less b) are random. This result will follow whatever be the form of distribution of intervals, provided only that $i_s - b$ is small compared to the average interval. A large $P_{z''}$ merely suggests that *if* the intervals (less the closed period) are random, then it is reasonable to suppose that the closed period is b. On the other hand, a very small value of $P_{z''}$ suggests that either the supposed closed period b is incorrect, or that the intervals (less b) are not random. In cases where it is clear *à priori* that there can be no closed period, then a small $P_{z''}$ certainly suggests that the intervals are not random.

If however it is possible to observe a number, say N, of series of n consecutive intervals, we may calculate z'' in each case and compare the distribution of these N values of z'' with the theoretical frequency law (lx). If there is agreement, then not only is it probable that the hypothesis as to b is correct, but also that the intervals (less b) are random. As z'' is independent of c, the average "rate of happenings" of the events, it follows that if b is unchanged—and in particular is zero—we may combine together in the test, the z'''s of different series of n observations, even though this rate may change from series to series. The test is also applicable, however small n may be. The practical importance of this will be illustrated below.

Through the kindness of Miss E. M. Newbold and Professor Greenwood, we have been able to make use of some of the original data regarding Factory Accidents used by the former in Report No. 34 of the Industrial Fatigue Research Board*. We may ask what answer the z''-test gives to the following question: Are the lengths of interval between consecutive accidents incurred by the same individual worker distributed randomly? For this purpose use has been made of the cards of male workers in a Chocolate Factory during the year 1923†. The time of accident was recorded to the day and even hour, and by making allowance

* "A Contribution to the Study of the Human Factor in the Causation of Accidents," 1925. For a later paper by the same writer, see *Journal of Roy. Stat. Soc.* 1927, p. 487.

† These are the cards of the series M 1 of page 11 of the I.F.R.B. Report, No. 34. The injuries were mainly of minor character, i.e. "cut finger," "abrasion," "splinter," etc. attended to in the Ambulance Room of the Factory, and did not lead to absence from work.

for Sundays, Bank Holidays, and as far as possible extra days of absence, sickness, etc.*, it was possible to obtain fairly accurately the number of days which each worker was exposed to risk between accidents. As Miss Newbold has shown, the liability to accident varies considerably among the workers, varying from none to fifteen accidents in the year for the cases examined, but the ratios z'', which are independent of this liability or rate, can be combined together.

The cards have been analysed for 50 workers; of these, 25 incurred only three accidents in the year, giving two intervals or the minimum to which the test can be applied.

TABLE IX.

Intervals between Accidents. Distributions of z''.

$n=2$			$n=4$			$n=6$		
z''	Observation	Theory	z''	Observation	Theory	z''	Observation	Theory
0— ·5	20	16·7	*0— ·1*	4	6·2	*0—·1*	4	7·6
·5— 1·0	7	8·3	*·1— ·2*	8	4·3	*·1—·2*	10	4·4
1·0— 1·5	4	5·0	*·2— ·3*	4	3·1	*·2—·3*	5	2·6
1·5— 2·0	2	3·3	*·3— ·4*	1	2·3	*·3—·4*	—	1·7
2·0— 2·5	2	2·4	*·4— ·5*	1	1·7	*·4— ·5*	—	1·1
2·5— 3·0	2·5	1·8	*·5—1·0*	3	4·3	Greater than *·5*	1	2·6
3·0— 4·0	5	2·5	*1·0—1·5*	3	1·5			
4·0— 5·0	1·5	1·7	*1·5—2·0*	1	0·7	—	—	—
5·0— 6·0	—	1·2	*2·0—5·0*	—	0·8	—	—	—
6·0— 7·0	1	0·9	*5·0—8·0*	—	0·1	—	—	—
7·0— 8·0	—	0·7	—	—	—	—	—	—
8·0— 9·0	—	0·6	—	—	—	—	—	—
9·0—10·0	—	0·4	—	—	—	—	—	—
Greater than *10·0*	5	4·5	—	—	—	—	—	—
			—	—	—	—	—	—
Total	50	50·0	Total	25	25·0	Total	20	20·0
Mean z''	—	—		·460	·500		·179	·250
Difference in terms of standard error				$-$·23			$-$·99	

In the above table are given the calculated values of z'' for (*a*) the 50 cases, taking the first three accidents or $n = 2$, (*b*) the 25 of these for which there were at least five accidents ($n = 4$), and (*c*) the 20 of these for which there were at least seven accidents ($n = 6$). The intervals chosen were the first two, first four and first six of the year. It is clear that as the records were restricted to a year, they do not contain a completely random selection of intervals, at any rate for the workers

* As the dates of absence were not specified exactly, but grouped in thirteen 4-week periods, there was often some uncertainty as to the accident interval to which they belonged. But it was always possible to locate the longer periods of absence such as the 10—14 days annual holiday.

of small liability. The table compares the observed frequencies with the theoretical obtained by integrating (lx) between appropriate limits; also the mean z'', and the ratio of the difference between mean (observed) and mean (theory) to the standard error. In the case of $n=2$, the theoretical mean and standard deviation of z'' are infinite so that the latter comparison cannot be made, but as far as can be judged from a sample of 50, the correspondence in the frequencies appears very close. For $n=4$ and $n=6$ the suggestion of too few observed values of z'' in the first group would need confirmation from fuller data*; the mean values of z'' are

TABLE X.

Distribution of z''; Omnibuses and Pedestrians.

Intervals between Omnibuses						Intervals between Pedestrians		
$n=2$			$n=4$			$n=5$		
z''	Observation	Theory	z''	Observation	Theory	z''	Observation	Theory
0— ·5	4	9·3	0— ·1	3	7·0	0—·1	4	3·2
·5— 1·0	7·5	4·7	·1— ·2	1	4·8	·1—·2	3	2·0
1·0— 1·5	1·5	2·8	·2— ·3	2	3·5	·2—·3	—	1·3
1·5— 2·0	3	1·9	·3— ·4	5	2·5	·3—·4	—	0·9
2·0— 2·5	2	1·3	·4— ·5	2	1·9	·4—·5	1	0·6
2·5— 3·0	—	1·0	·5—1·0	6	4·8	·5—·6	1	0·4
3·0— 4·0	2	1·4	1·0—1·5	3	1·7	·6—·7	1	0·3
4·0— 5·0	4	0·9	1·5—2·0	2	0·8	Greater than ·7	—	1·3
5·0— 6·0	2	0·7	2·0—5·0	2	0·9			
6·0— 7·0	—	0·5	5·0—8·0	1	0·1	—	—	—
7·0— 8·0	—	0·4	Greater than 8·0	1	—	—	—	—
8·0— 9·0	—	0·3						
9·0—10·0	—	0·3	—	—	—	—	—	—
Greater than 10·0	2	2·5	—	—	—	—	—	—
Total	28	28·0	Total	28	28·0	Total	10	10·0
Mean z''	—	—		1·258	·500		·210	·333
Difference in terms of standard error				+4·6			−·82	

however quite in agreement with theory as far as can be judged with such large standard errors.

To obtain some further idea of the sensitiveness of the test, we have applied it to the time intervals observed between the passing of consecutive omnibuses on a London bus route. These intervals are not of course random, the omnibuses

* In general there could be no closed interval, b, the times between accidents varying from 0 to over 180 days, but it is possible that after an accident the worker is for a day or two more cautious, and this reduces the chance of very small intervals and therefore small values of z''.

having started at fixed intervals (apparently of 3 minutes) from three-quarters of an hour to an hour before the observations were made. During this time traffic delays, etc. would have superposed chance positive and negative variations on to the scheduled intervals, so that in some cases these had been reduced to a few seconds and in others had been doubled. Table X gives the values of z'' calculated for 28 sets of $n=2$ and 28 of $n=4$. The fit for $n=2$ is far less satisfactory than in the case of the accident intervals, and appears even worse for $n=4$, where the mean z'' exceeds the theoretical value by 4·6 times its standard error. Finally in the same table are given 10 values of z'' calculated from sets of 5 intervals of time observed between pedestrians passing a fixed point on a roadway; here, as far as can be judged, agreement is satisfactory, suggesting that the intervals were random.

There has not been time before going to press to extend these results, but inadequate as the numbers are, they certainly suggest that the method may be of value in testing the randomness of intervals in many problems, and further that the test will be sensitive even for pairs of consecutive intervals ($n = 2$) provided that the number of such pairs available is large enough.

V. Conclusion.

The main problem that we have considered is that of determining whether it is probable that a given sample, taken as a whole, has been drawn from a specified population. From the idea of the representation of a sample of n as a point in multiple space, there follows at once the conception that the discriminating criteria may be regarded as associated with surfaces in this space which divide regions containing points for which the hypothesis to be tested as to the sample's origin is less probable from those for which it is more probable. Regarded from this point of view it is seen that in any given problem there need be no single system of surfaces which it is "best" to make use of. The system adopted will provide a numerical measure, and this must be coordinated in the mind of the statistician with a clear understanding of the process of reasoning on which the test is based. We have endeavoured to connect in a logical sequence several of the most simple tests, and in so doing have found it essential to make use of what R. A. Fisher has termed "the principle of likelihood." The process of reasoning, however, is necessarily an individual matter, and we do not claim that the method which has been most helpful to ourselves will be of greatest assistance to others. It would seem to be a case where each individual must reason out for himself his own philosophy.

The differences in method of approach have been illustrated in the two problems that have been described as the testing of Hypothesis A and Hypothesis B. Both problems suppose a knowledge of the form or shape of the population sampled, A supposes its standard deviation known and B does not. In both cases there appear to be four or five methods open to the statistician; if properly interpreted we should not describe one as more *accurate* than another, but according to the problem in hand should recommend this one or that as providing information which is more *relevant* to the purpose. The principles involved have been illustrated in the case of samples from normal, rectangular and exponential populations, for each of which analogous tests exist. Most attention has been given to the normal case, simply because the β's of the populations of experience do tend to cluster round the Gaussian point. New tables have been computed for what is termed the P_λ test of Hypothesis A, and illustrations of some of the various types of problems in which it could be employed have been given.

Now these tests, whether applicable to the normal, the rectangular or the exponential population, have all assumed both that there is perfect random sampling and that while there is uncertainty as to position and scale there is none as to the shape of the population curve. In following what would appear a more efficient system of reasoning, it must be remembered that the ideal situation is not quite that of practical reality, and we must be careful that what is gained on the swings is not being lost on the roundabouts. It would be useless to emphasise the gain in relevance of a $P_z = \cdot08$ obtained from Student's Tables over a $P = \cdot15$ obtained from Sheppard's Tables, if in fact the former were in error by an amount of the order of ·05 owing to a faulty assumption that the population sampled was normal. Owing to the rapidity with which the distribution of means tends to normality, as n increases, the P obtained by entering Sheppard's Tables with $z = m/s$ has a definite meaning if properly interpreted, even if the population sampled be skew. But by using P_z we may be obtaining a false impression of the accuracy of the criterion upon which our judgment is to be based. In practice in dealing with small samples it is generally not possible to be certain of the form of the population distribution from internal evidence. But from other considerations, from previous experience in similar problems and so forth, we may have good grounds for believing that the distribution is roughly of a certain type, and if skew for knowing the direction in which it will be more sharply limited. There will undoubtedly be regions in the β_1, β_2 population field surrounding the normal, the rectangular, and the exponential points and any other points for which the problem can be completely solved, within which the criteria appropriate for these spots will be adequate. It is therefore a problem of first importance to ascertain what is the extent of these regions of adequacy, and failing a fresh advance in theory* the most straightforward method of doing this is to take experimental soundings in the β_1, β_2 field. In this process we have not at present gone very far, but have shown that

(*a*) At $\beta_1 = 0{\cdot}2$, $\beta_2 = 3{\cdot}2$, the normal theory for samples of 10 does provide criteria which for practical purposes are probably adequate in the case of the distribution of m, of s, of z and for the P_λ contours.

* It is of course a serious limitation to the practical value of the method of likelihood, that it appears only to provide readily soluble equations in certain simple cases. It is not until the samples are large enough for the application of the (P, χ^2) group test that the problem can be solved in a more general manner.

(*b*) At $\beta_1 = 0$, $\beta_2 = 1{\cdot}8$, the z-distribution for samples of 4 and of 10 appears adequate, but the normal theory quite fails for s and P_λ. In this direction an intermediate sounding is required.

(*c*) Some experimental work at present in progress suggests that the normal theory will also be adequate for s and z for samples of 5 and 10 from a symmetrical leptokurtic curve with β_2 as great as 4·2.

(*d*) The results obtained in connection with the exponential population are quite out of touch with the symmetrical cases, but the distributions reached in that case will probably be adequate for a fairly wide range of skew populations with β_1 and β_2 in the region surrounding the point (4, 9).

The limitation implied by the assumption of perfect random sampling must not of course be overlooked; this is of particular importance in the case of very small samples. It has been shown that there may be some latitude as to the exact shape of the population curve associated with each hypothesis, but it has been assumed that in any event the sample is a random one. It is true that the causes underlying the bias or selection which occurs in the drawing of a sample from Π may be such that the sample can be regarded as randomly drawn from Π'; but if a biased sample from an approximately normal population can only be considered as a random sample from a J-shaped curve, it is clear that our methods of comparison will be inadequate. The bias may consist only of a few errors of observation or of crudeness of grouping, but the smaller the sample the more important will be the effect.

The answer to criticism from this source is that the tests should only be regarded as tools which must be used with discretion and understanding, and not as instruments which in themselves give the final verdict. It is impossible to escape from the difficulties; they make it desirable to avoid the use of very small samples whenever possible, but do not prevent some conclusions of value being drawn when fuller data are not available. If the criteria that have been discussed indicate that the supposed population of origin is improbable in comparison with some other alternative, that is a sign which is not to be disregarded. But we must not discard the original hypothesis until we have examined the alternative suggested, and have satisfied ourselves that it does involve a change in the real underlying factors in which we are interested; until we have assured ourselves that the alternative hypothesis is not error in observation, error in record, variation due to some outside factor that it was believed had been controlled, or to any one of many causes which when they chance to modify even one observation in a sample of five may be of serious consequence.

We must not conclude without thanking Miss Ida McLearn for her diagrams, Miss M. Page, without whose assistance in computing the completion of the paper in its present form must have been indefinitely delayed, and Mrs L. J. Comrie who carried out the final stage in the tabling of P_λ.

VI. APPENDIX: TABLES OF P_λ.

P_λ is given by the relation

$$P_\lambda = \int_\omega \frac{n^{\frac{n}{2}}}{\sqrt{\pi}\, 2^{\frac{n-2}{2}}\, \Gamma\left(\frac{n-1}{2}\right)} S^{n-2} e^{-\frac{n}{2}(M^2+S^2)} dM dS \text{ ......(xxi } bis\text{)},$$

where the integral is to be taken outside the curve in the (M, S)-plane upon which λ is constant, and

$$\lambda = S^n e^{-\frac{n}{2}(M^2+S^2-1)} \text{ ....(xix } bis\text{)}.$$

We may write

$$\log_{10}\lambda = \frac{n}{2}\{\log_{10} S^2 - (M^2+S^2-1)\log_{10} e\} = \frac{n}{2}\{\log_{10} e - k\},$$

and obtain for the equations of the contours

$$(M^2+S^2)\log_{10} e - \log_{10} S^2 = k \text{ ....(xx } bis\text{)}.$$

Values of M for equidistant values of S* were first computed for the curves corresponding to a number of values of k between ·45 and 2·40. Some of the contours have been drawn in Figures 13 and 14. The lowest possible value of k corresponding to the centre of the system, $M = 0$, $S = 1$, $\lambda = 1$, is

$$k = \log_{10} e = \cdot 434{,}2945.$$

If S_1 and S_2 be the points at which a contour cuts the axis of S, if M_s indicates the value of M at the point on the curve corresponding to a given S, and if we write $x_s = \sqrt{n}\, M_s$, then the integral (xxi *bis*) can be put into the form

$$\tfrac{1}{2}(1 - P_\lambda) = \int_{S_1}^{S_2} \frac{n^{\frac{n-1}{2}}}{2^{\frac{n-3}{2}}\, \Gamma\left(\frac{n-1}{2}\right)} S^{n-2} e^{-\frac{n}{2}S^2} (\tfrac{1}{2}\alpha_s)\, dS \text{ ....(lxiii)},$$

where

$$\tfrac{1}{2}(1+\alpha_s) = \int_{-\infty}^{x_s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx,$$

and is obtained by entering Sheppard's Tables with $x_s = \sqrt{n}\, M_s$. The values of the integral (lxiii) were then calculated by quadrature. The task of calculating P_λ for an adequate framework of values of k and for $n = 3$ to $n = 50$ would have been very great, but it was discovered that the ratios P_λ/λ for a given k are very nearly independent of n. Table XI gives the values of P_λ actually computed divided by the corresponding λ. It was estimated that the computed value of P_λ might be in error by several units in the fifth decimal place, and this limits the number of significant figures that can be given for the ratio as k increases. It also means that the final figures in the tabled ratios are not exact. The results however appeared to justify the following procedure:

(*a*) To obtain interpolated series of ratios for $n = 3$, 4 and 10 from the ten or eleven values computed in each case.

* The argument interval for S varied from ·005 to ·04.

TABLE XI.

Computed Values of P_λ/λ.

k	$n=3$	$n=4$	$n=5$	$n=10$	$n=20$	$n=50$
$\log e$	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000
·45	—	1·0169	—	1·0170	—	1·0170
·50	1·0743	—	1·0736	1·0731	1·0732	1·072
·55	—	—	—	1·1327	—	1·135⁻
·60	1·1991	1·1977	—	1·1959	1·195⁻	—
·65	—	—	—	1·2622	—	—
·70	—	1·3363	—	1·332	1·33	—
·75	1·4151	1·4117	—	—	—	—
·80	1·4957	1·4921	1·4899	1·485⁺	—	—
·90	1·6721	1·6672	—	1·66	—	—
1·00	—	1·864	—	1·85	—	—
1·20	2·3432	2·335⁻	2·330	—	—	—
1·50	3·292	3·286	—	—	—	—
1·80	4·638	4·62	—	—	—	—
2·40	9·42	—	—	—	—	—
3·00	18·53	—	—	—	—	—

(*b*) To extrapolate for high values of k for $n=10$ and 4, from $n=4$ and 3 respectively, in view of the fact that as P_λ decreases only a few significant figures are needed in the ratio.

(*c*) With the help of these "frames" and the isolated values for $n=5$, to interpolate appropriate ratios for $n=5$, 6, 7, 8 and 9.

(*d*) Since the ratios at $n=10$ give accurately the three values of P_λ computed for $n=20$ and the three for $n=50$, to assume that the ratios for $n=10$ will be adequate for all values of n in the range 10 to 50.

To obtain P_λ it was only necessary to compute $\log_{10}\lambda = \frac{n}{2}(\log_{10} e - k)$, add to it the logarithm of the ratio P_λ/λ and find the anti-logarithm. The P_λ tables given below were calculated in this way. They are given to four decimal places, but the final figure may often be in error by a single unit, although not, it is believed, by as much as two units. After the work was commenced it was realised that the variable k was not perhaps the best to have taken from the point of view of interpolation in the tables; k^2 would probably have been better. It has been necessary to change the interval from ·01 to ·05 at $k=$ ·65, and from ·05 to ·30 at $k=1·50$. For high values of n and low values of k the tables are not easy to use if four-figure accuracy is required, but for most practical purposes it is sufficient to obtain P_λ to two decimal places only, and in this there is no difficulty.

Method of entry.

Having obtained from the data of the problem under consideration $M=m/\sigma$ and $S=s/\sigma$, k may either be calculated from (xx *bis*), or generally obtained with sufficient accuracy from one or other of the Figures 13 and 14. P_λ is then found by entering the tables with n and k. Examples of the use of the tables have been given in Section (9) above.

Extension of the Tables.

On the assumption that the ratios P_λ/λ for $n=10$ will continue to hold good for any value of n greater than 50, we may proceed as follows:

(a) Find the appropriate k, (b) calculate λ from the relation

$$\log_{10}\lambda = \frac{n}{2}(\log_{10}e - k),$$

(c) multiply the ratio P_λ/λ given in Table XII below by λ, and so obtain P_λ.

TABLE XII.

Values of P_λ/λ for $n = 10$ and beyond.

k	P_λ/λ	k	P_λ/λ	k	P_λ/λ	k	P_λ/λ
·435	1·0008	·500	1·0731	·565	1·1514	·630	1·2353
·440	1·0062	·505	1·0789	·570	1·1576	·635	1·2420
·445	1·0116	·510	1·0847	·575	1·1639	·640	1·2488
·450	1·0170	·515	1·0906	·580	1·1702	·645	1·2555
·455	1·0224	·520	1·0965	·585	1·1766	·650	1·2623
·460	1·0279	·525	1·1024	·590	1·1830	·70	1·332
·465	1·0334	·530	1·1084	·595	1·1894	·75	1·407
·470	1·0390	·535	1·1144	·600	1·1959	·80	1·485+
·475	1·0446	·540	1·1205	·605	1·2024	·85	1·569
·480	1·0502	·545	1·1266	·610	1·2089	·90	1·658
·485	1·0559	·550	1·1328	·615	1·2155	·95	1·753
·490	1·0616	·555	1·1390	·620	1·2221	1·00	1·855
·495	1·0673	·560	1·1452	·625	1·2287	—	—

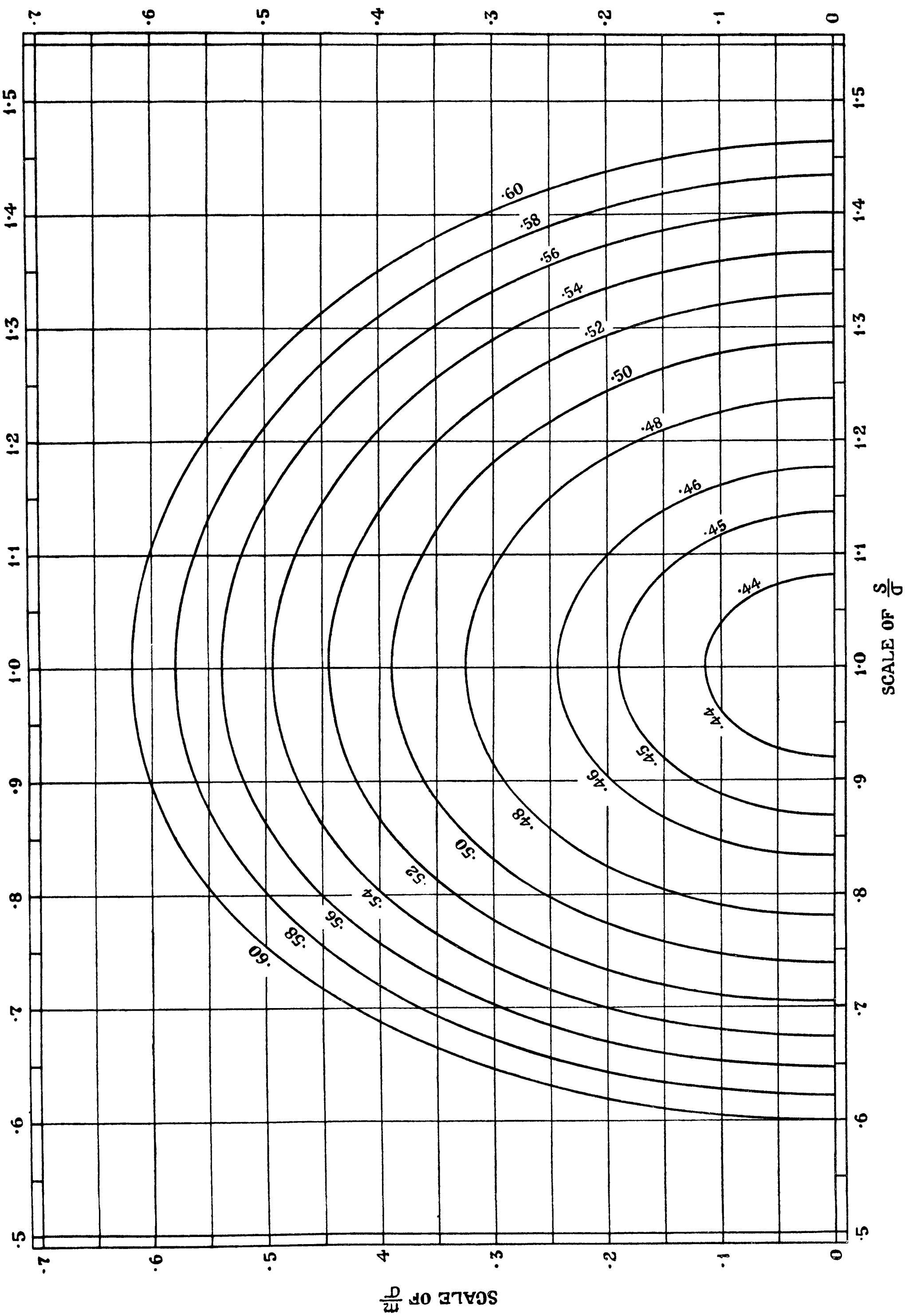

FIG. 13. k-CONTOURS FOR USE WITH P_λ TABLES

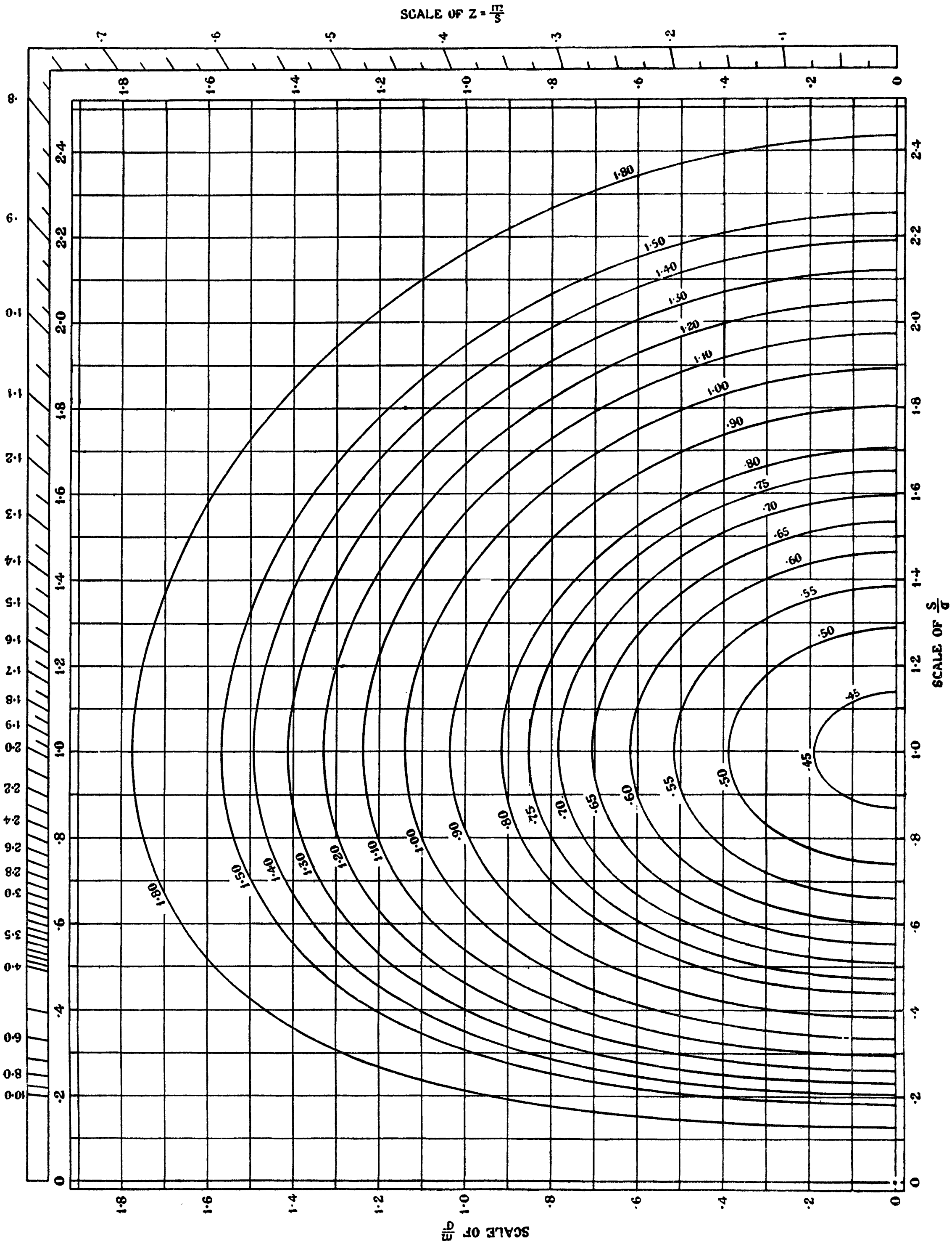

FIG 14. k-CONTOURS FOR USE WITH P_λ TABLES

TABLES OF P_λ.

Size of Sample, n.

k	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*k*
·44	·9865	·9800	·9736	·9672	·9609	·9546	·9484	·9421	·9360	·9298	·9238	·9177	*·44*
·45	·9634	·9460	·9291	·9124	·8961	·8800	·8642	·8487	·8335	·8186	·8039	·7895	*·45*
·46	·9408	·9132	·8866	·8607	·8356	·8113	·7876	·7646	·7423	·7207	·6996	·6792	*·46*
·47	·9189	·8817	·8461	·8120	·7792	·7478	·7177	·6888	·6611	·6344	·6089	·5844	*·47*
·48	·8975	·8513	·8075	·7660	·7267	·6894	·6541	·6205	·5887	·5585	·5299	·5027	*·48*
·49	·8766	·8219	·7707	·7227	·6778	·6356	·5961	·5590	·5243	·4917	·4612	·4325	*·49*
·50	·8562	·7935	·7355	·6818	·6321	·5859	·5432	·5036	·4669	·4329	·4014	·3721	*·50*
·51	·8363	·7661	·7019	·6432	·5895	·5402	·4951	·4537	·4159	·3811	·3493	·3202	*·51*
·52	·8168	·7396	·6699	·6068	·5497	·4980	·4512	·4088	·3704	·3356	·3040	·2755	*·52*
·53	·7978	·7140	·6393	·5724	·5126	·4591	·4112	·3683	·3299	·2954	·2646	·2370	*·53*
·54	·7792	·6894	·6101	·5401	·4781	·4233	·3748	·3318	·2938	·2601	·2303	·2039	*·54*
·55	·7611	·6656	·5823	·5096	·4459	·3903	·3416	·2990	·2617	·2290	·2005	·1755	*·55*
·56	·7434	·6426	·5558	·4808	·4159	·3598	·3113	·2694	·2331	·2017	·1745	·1510	*·56*
·57	·7261	·6204	·5304	·4536	·3879	·3318	·2837	·2427	·2076	·1776	·1519	·1299	*·57*
·58	·7092	·5990	·5063	·4280	·3618	·3059	·2586	·2187	·1849	·1563	·1322	·1118	*·58*
·59	·6927	·5784	·4832	·4038	·3374	·2820	·2357	·1970	·1647	·1376	·1151	·0962	*·59*
·60	·6766	·5584	·4611	·3810	·3147	·2600	·2148	·1775	·1467	·1212	·1001	·0827	*·60*
·61	·6608	·5391	·4401	·3594	·2935	·2397	·1958	·1599	·1306	·1067	·0872	·0712	*·61*
·62	·6454	·5205	·4201	·3391	·2737	·2210	·1784	·1441	·1163	·0939	·0759	·0613	*·62*
·63	·6305	·5026	·4009	·3199	·2553	·2037	·1626	·1298	·1036	·0827	·0660	·0527	*·63*
·64	·6158	·4852	·3826	·3018	·2381	·1878	·1482	·1169	·0923	·0728	·0575	·0453	*·64*
·65	·6015	·4685	·3652	·2847	·2220	·1732	·1350	·1053	·0822	·0641	·0500	·0390	*·65*
·70	·5348	·3931	·2892	·2129	·1567	·1153	·0849	·0625	·0461	·0339	·0250	·0184	*·70*
·75	·4756	·3299	·2291	·1591	·1106	·0768	·0534	·0371	·0258	·0180	·0125	·0087	*·75*
·80	·4229	·2769	·1815	·1190	·0781	·0512	·0336	·0221	·0145	·0095	·0062	·0041	*·80*
·85	·3762	·2325	·1439	·0891	·0551	·0341	·0211	·0131	·0081	·0050	·0031	·0019	*·85*
·90	·3347	·1952	·1140	·0666	·0389	·0228	·0133	·0078	·0046	·0027	·0016	·0009	*·90*
·95	·2979	·1640	·0904	·0499	·0275	·0152	·0084	·0046	·0026	·0014	·0008	·0004	*·95*
1·00	·2651	·1377	·0717	·0373	·0194	·0101	·0053	·0027	·0014	·0007	·0004	·0002	*1·00*
1·05	·2359	·1157	·0568	·0279	·0137	·0068	·0033	·0016	·0008	·0004	·0002	·0001	*1·05*
1·10	·2100	·0972	·0450	·0209	·0097	·0045	·0021	·0010	·0005	·0002	·0001	·0000	*1·10*
1·15	·1869	·0817	·0357	·0156	·0069	·0030	·0013	·0006	·0003	·0001	·0000	—	*1·15*
1·20	·1664	·0687	·0284	·0117	·0048	·0020	·0008	·0003	·0001	·0001	—	—	*1·20*
1·25	·1482	·0578	·0226	·0088	·0034	·0013	·0005	·0002	·0001	·0000	—	—	*1·25*
1·30	·1319	·0486	·0179	·0066	·0024	·0009	·0003	·0001	·0000	—	—	—	*1·30*
1·35	·1174	·0409	·0142	·0049	·0017	·0006	·0002	·0001	—	—	—	—	*1·35*
1·40	·1046	·0344	·0113	·0037	·0012	·0004	·0001	·0000	—	—	—	—	*1·40*
1·45	·0932	·0289	·0090	·0028	·0009	·0003	·0001	—	—	—	—	—	*1·45*
1·50	·0830	·0243	·0071	·0021	·0006	·0002	·0001	—	—	—	—	—	*1·50*
1·80	·0415	·0086	·0018	·0004	·0001	·0000	·0000	—	—	—	—	—	*1·80*
2·10	·0206	·0030	·0004	·0001	·0000	—	—	—	—	—	—	—	*2·10*
2·40	·0106	·0011	·0001	·0000	—	—	—	—	—	—	—	—	*2·40*
2·70	·0053	·0004	·0000	—	—	—	—	—	—	—	—	—	*2·70*
3·00	·0026	·0001	—	—	—	—	—	—	—	—	—	—	*3·00*

Size of Sample, n.

k	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*k*
·44	·9117	·9057	·8998	·8939	·8881	·8822	·8765	·8707	·8650	·8593	·8537	·8481	*·44*
·45	·7754	·7615	·7478	·7344	·7213	·7084	·6956	·6832	·6709	·6589	·6471	·6355	*·45*
·46	·6594	·6402	·6215	·6034	·5858	·5687	·5521	·5360	·5204	·5052	·4905	·4762	*·46*
·47	·5608	·5382	·5166	·4958	·4758	·4566	·4382	·4206	·4037	·3874	·3718	·3568	*·47*
·48	·4770	·4525	·4293	·4073	·3864	·3666	·3478	·3300	·3131	·2970	·2818	·2674	*·48*
·49	·4057	·3805	·3568	·3347	·3139	·2944	·2761	·2589	·2429	·2278	·2136	·2004	*·49*
·50	·3450	·3199	·2966	·2750	·2549	·2364	·2191	·2032	·1884	·1746	·1619	·1501	*·50*
·51	·2934	·2690	·2465	·2259	·2071	·1898	·1739	·1594	·1461	·1339	·1227	·1125	*·51*
·52	·2496	·2261	·2049	·1856	·1682	·1524	·1381	·1251	·1133	·1027	·0930	·0843	*·52*
·53	·2123	·1901	·1703	·1525	·1366	·1224	·1096	·0982	·0879	·0787	·0705	·0632	*·53*
·54	·1806	·1599	·1416	·1253	·1110	·0983	·0870	·0770	·0682	·0604	·0535	·0473	*·54*
·55	·1536	·1344	·1177	·1030	·0901	·0789	·0691	·0604	·0529	·0463	·0405	·0355	*·55*
·56	·1306	·1130	·0978	·0846	·0732	·0634	·0548	·0474	·0410	·0355	·0307	·0265	*·56*
·57	·1111	·0950	·0813	·0695	·0595	·0509	·0435	·0372	·0318	·0272	·0233	·0199	*·57*
·58	·0945	·0799	·0676	·0571	·0483	·0409	·0345	·0292	·0247	·0209	·0177	·0149	*·58*
·59	·0804	·0672	·0562	·0470	·0392	·0328	·0274	·0229	·0192	·0160	·0134	·0112	*·59*
·60	·0684	·0565	·0467	·0386	·0319	·0263	·0218	·0180	·0149	·0123	·0101	·0084	*·60*
·61	·0582	·0475	·0388	·0317	·0259	·0212	·0173	·0141	·0115	·0094	·0077	·0063	*·61*
·62	·0495	·0399	·0323	·0260	·0210	·0170	·0137	·0111	·0089	·0072	·0058	·0047	*·62*
·63	·0421	·0336	·0268	·0214	·0171	·0136	·0109	·0087	·0069	·0055	·0044	·0035	*·63*
·64	·0358	·0282	·0223	·0176	·0139	·0110	·0086	·0068	·0054	·0042	·0034	·0026	*·64*
·65	·0304	·0237	·0186	·0145	·0113	·0088	·0069	·0054	·0042	·0033	·0025	·0020	*·65*
·70	·0135	·0100	·0073	·0054	·0040	·0029	·0022	·0016	·0012	·0009	·0006	·0005	*·70*
·75	·0060	·0042	·0029	·0020	·0014	·0010	·0007	·0005	·0003	·0002	·0002	·0001	*·75*
·80	·0027	·0018	·0012	·0008	·0005	·0003	·0002	·0001	·0001	·0001	·0000	·0000	*·80*
·85	·0012	·0007	·0005	·0003	·0002	·0001	·0001	·0000	·0000	·0000	—	—	*·85*
·90	·0005	·0003	·0002	·0001	·0001	·0000	·0000	—	—	—	—	—	*·90*
·95	·0002	·0001	·0001	·0000	·0000	—	—	—	—	—	—	—	*·95*
1·00	·0001	·0001	·0000	—	—	—	—	—	—	—	—	—	*1·00*
1·05	·0000	·0000	—	—	—	—	—	—	—	—	—	—	*1·05*

Size of Sample, n.

k	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*k*
·44	·8426	·8371	·8316	·8262	·8207	·8154	·8100	·8047	·7994	·7942	·7890	·7839	*·44*
·45	·6241	·6129	·6020	·5912	·5806	·5702	·5600	·5499	·5401	·5304	·5209	·5116	*·45*
·46	·4623	·4488	·4357	·4230	·4107	·3987	·3871	·3758	·3648	·3542	·3439	·3339	*·46*
·47	·3425	·3287	·3154	·3027	·2905	·2788	·2676	·2568	·2465	·2365	·2270	·2179	*·47*
·48	·2537	·2407	·2283	·2166	·2055	·1950	·1850	·1755	·1665	·1580	·1499	·1422	*·48*
·49	·1879	·1762	·1653	·1550	·1454	·1364	·1279	·1199	·1125	·1055	·0989	·0928	*·49*
·50	·1392	·1290	·1196	·1109	·1028	·0954	·0884	·0820	·0760	·0705	·0653	·0606	*·50*
·51	·1031	·0945	·0866	·0794	·0728	·0667	·0611	·0560	·0513	·0471	·0431	·0395	*·51*
·52	·0764	·0692	·0627	·0568	·0515	·0466	·0423	·0383	·0347	·0314	·0285	·0258	*·52*
·53	·0566	·0507	·0454	·0407	·0364	·0326	·0292	·0262	·0234	·0210	·0188	·0168	*·53*
·54	·0419	·0371	·0329	·0291	·0258	·0228	·0202	·0179	·0158	·0140	·0125	·0110	*·54*
·55	·0311	·0272	·0238	·0208	·0182	·0160	·0140	·0122	·0107	·0094	·0082	·0072	*·55*
·56	·0230	·0199	·0172	·0149	·0129	·0112	·0097	·0084	·0072	·0063	·0054	·0047	*·56*
·57	·0170	·0146	·0125	·0107	·0091	·0078	·0067	·0057	·0049	·0042	·0036	·0031	*·57*
·58	·0126	·0107	·0090	·0076	·0065	·0055	·0046	·0039	·0033	·0028	·0024	·0020	*·58*
·59	·0094	·0078	·0065	·0055	·0046	·0038	·0032	·0027	·0022	·0019	·0016	·0013	*·59*
·60	·0069	·0057	·0047	·0039	·0032	·0027	·0022	·0018	·0015	·0012	·0010	·0008	*·60*
·61	·0051	·0042	·0034	·0028	·0023	·0019	·0015	·0012	·0010	·0008	·0007	·0006	*·61*
·62	·0038	·0031	·0025	·0020	·0016	·0013	·0011	·0009	·0007	·0006	·0004	·0004	*·62*
·63	·0028	·0022	·0018	·0014	·0011	·0009	·0007	·0006	·0005	·0004	·0003	·0002	*·63*
·64	·0021	·0016	·0013	·0010	·0008	·0006	·0005	·0004	·0003	·0002	·0002	·0002	*·64*
·65	·0015	·0012	·0009	·0007	·0006	·0004	·0003	·0003	·0002	·0002	·0001	·0001	*·65*
·70	·0003	·0003	·0002	·0001	·0001	·0001	·0001	·0000	·0000	·0000	·0000	·0000	*·70*
·75	·0001	·0001	·0000	·0000	·0000	·0000	·0000	—	—	—	—	—	*·75*
·80	·0000	·0000	—	—	—	—	—	—	—	—	—	—	*·80*

Size of Sample, n.

k	39	40	41	42	43	44	45	46	47	48	49	50	*k*
·44	·7787	·7736	·7686	·7635	·7585	·7536	·7486	·7437	·7389	·7340	·7292	·7244	·44
·45	·5024	·4934	·4846	·4759	·4673	·4590	·4507	·4427	·4347	·4269	·4193	·4118	·45
·46	·3241	·3147	·3055	·2966	·2879	·2795	·2714	·2635	·2558	·2483	·2411	·2341	·46
·47	·2091	·2007	·1926	·1848	·1774	·1703	·1634	·1568	·1505	·1444	·1386	·1330	·47
·48	·1349	·1280	·1214	·1152	·1093	·1037	·0984	·0933	·0886	·0840	·0797	·0756	·48
·49	·0870	·0816	·0766	·0718	·0673	·0632	·0592	·0556	·0521	·0489	·0458	·0430	·49
·50	·0562	·0521	·0483	·0448	·0415	·0385	·0357	·0331	·0307	·0284	·0264	·0244	·50
·51	·0362	·0332	·0304	·0279	·0256	·0234	·0215	·0197	·0180	·0165	·0152	·0139	·51
·52	·0234	·0212	·0192	·0174	·0158	·0143	·0129	·0117	·0106	·0096	·0087	·0079	·52
·53	·0151	·0135	·0121	·0108	·0097	·0087	·0078	·0070	·0062	·0056	·0050	·0045	·53
·54	·0097	·0086	·0076	·0068	·0060	·0053	·0047	·0042	·0037	·0033	·0029	·0026	·54
·55	·0063	·0055	·0048	·0042	·0037	·0032	·0028	·0025	·0022	·0019	·0017	·0015	·55
·56	·0041	·0035	·0030	·0026	·0023	·0020	·0017	·0015	·0013	·0011	·0010	·0008	·56
·57	·0026	·0022	·0019	·0016	·0014	·0012	·0010	·0009	·0007	·0006	·0005	·0005	·57
·58	·0017	·0014	·0012	·0010	·0009	·0007	·0006	·0005	·0004	·0004	·0003	·0003	·58
·59	·0011	·0009	·0008	·0006	·0005	·0004	·0004	·0003	·0003	·0002	·0002	·0002	·59
·60	·0007	·0006	·0005	·0004	·0003	·0003	·0002	·0002	·0002	·0001	·0001	·0001	·60
·61	·0005	·0004	·0003	·0002	·0002	·0002	·0001	·0001	·0001	·0001	·0001	·0000	·61
·62	·0003	·0002	·0002	·0002	·0001	·0001	·0001	·0001	·0001	·0000	·0000	—	·62
·63	·0002	·0002	·0001	·0001	·0001	·0001	·0000	·0000	·0000	—	—	—	·63
·64	·0001	·0001	·0001	·0001	·0000	·0000	—	—	—	—	—	—	·64
·65	·0001	·0001	·0000	·0000	—	—	—	—	—	—	—	—	·65

N.B. I feel it necessary to make a brief comment on the authorship of this paper. Its origin was a matter of close co-operation, both personal and by letter, and the ground covered included the general ideas and the illustration of these by sampling from a normal population. A part of the results reached in common are included in Chapters I, II and V. Later I was much occupied with other work, and therefore unable to co-operate. The experimental work, the calculation of tables and the developments of the theory of Chapters III and IV are due solely to Dr Egon S. Pearson.

J. Neyman.