181

THE ASYMPTOTIC APPROACH TO BEHRENS'S INTEGRAL, WITH FURTHER TABLES FOR THE d TEST OF SIGNIFICANCE

By R. A. FISHER, F.R.S.

1. The nature of fiducial inference

To the present generation of statisticians, familiar with 'Student's' distribution and with the process of 'Studentizing' unknown parameters, it has for some time appeared to be a somewhat puzzling historical fact that this advance in simple statistical procedure was not made long before, and was not made rather by a mathematician than by a research chemist. Light is perhaps thrown on this puzzle by the contrast, which has been striking during the last twenty years, between the facility, confidence and skill with which the new tests have been applied by practical men in research departments, and the embarrassment and confusion of many discussions, in journals devoted to mathematical statistics, by mathematically minded authors lacking contact with practical research.

I have at various times suggested that the advance reserved for 'Student' would, in all probability, have been made early in the nineteenth century, were it not for the preoccupation of writers on mathematical probability with discontinuous distributions, and their comparative neglect of variable quantities which, unlike frequencies, can take any value within their ranges of variation. It is certain that the analytic power of the mathematicians interested in Probability and the Theory of Errors was sufficient to master the particular problem broached by 'Student'. Helmert (1875), for example, determined the true distribution in samples of the sum of squares

$$S(x-\bar{x})^2$$

from which the estimate of error s^2 is calculated. He did not, however, make 'Student's' use of this knowledge. Again, W. Burnside (1923), in evident ignorance both of Helmert's and of 'Student's' work, gives the same solution, and arrives almost at 'Student's' distribution. His result differs from 'Student's' by one degree of freedom, through his feeling it necessary to introduce a Baysian a priori probability distribution for the precision constant,

$$h^2 = 1/2\sigma^2,$$

for which he assumes the frequency element to be proportional to dh. This discrepancy is instructive on the historical question, for it illustrates well how progress in this field in the nineteenth century was inhibited. Firstly, that the widespread teaching of inverse probability in mathematical textbooks had practically excluded the possibility of an approach involving different logical concepts from those of Bayes, while, secondly, the criticisms and doubts which in the latter half of the century Bayes's theorem had engendered, removed all confidence from conclusions in which it was implicated. Burnside, for example, leaves the subject with the words: 'It is a matter of individual judgment which form of second assumption is the more reasonable.'

Burnside gives a table of the quartiles (50 % points) of 'Student's' distribution. It evidently did not occur to him that a 5 or 1 % table would be more useful; or perhaps he was satisfied to use some arbitrary multiple of his quartile deviation, though this would at once throw away the value of having obtained the true distribution. The numerical values of this table of quartile deviations are also incorrect, which by itself may be taken to indicate that he regarded his solution rather as a matter of academic interest than as meeting a need for guidance in practical decisions.

This example, together with others arising from extensions of 'Student's' method, suggests that H. Jeffreys (1940), whose logical standpoint is very different from my own, may be right in proposing that 'Student's' method involves logical reasoning of so novel a type that a new postulate should be introduced to make its deductive basis rigorous. The postulate he frames is that, if, from a sample we obtain the mean

$$\bar{x} = \frac{1}{n} S(x),$$

and the estimated variance of the mean

$$s^2 = \frac{1}{(n-1)n} S(x-\bar{x})^2,$$

and set

$$t = \frac{\bar{x}-\mu}{s},$$

where μ is the hypothetical true mean of the population sampled, then the distribution of t, subject to the observational data $\bar{x}$ and s, and the non-observational data that the n values are drawn independently from a normal population, itself depends only on the size of the sample n.

While this statement will serve for the particular test of significance proposed by 'Student', the conditions on which analogous postulates should be introduced in analogous reasoning deserve consideration. The properties on which we rely are clearly: (*a*) that the distribution of t is independent of the true mean (μ) and variance (σ^2) of the population sampled, (*b*) that there is nothing about our particular statistics, $\bar{x}$ and s^2, or the way they were obtained, to bias the test. If, for example, there were reason to think that the mean $\bar{x}$ were higher than usual from samples of the same population, the test would obviously be biased. The same is true if there were reason to think that s were higher or lower than usual. Since the distribution of s/σ is independent of the population sampled, and is known, this implies that we regard our sample with given value of s as one of a population of similar samples drawn from populations having values of σ distributed according to the standard distribution of the ratio $s:\sigma$. The fiducial distribution of σ (Fisher, 1933) and the simultaneous fiducial distribution of σ and μ which I have discussed elsewhere (Fisher, 1935) are thus inherent in the logic of 'Student's' approach.

The generalizations of 'Student's' method with which I was first concerned were the wider applications of the t distribution to more than one sample, and to regression coefficients

(Fisher, 1924, 1925, 1926*a*), and those of the z distribution for the logarithmic ratio of two estimates of a variance. It was obvious that these involved no logical principle beyond that of 'Student's' original test, but were the fruits of the exact solutions of problems of distribution which I had arrived at in the meanwhile. No misunderstandings were apparent at this period; perhaps because writers without sufficient logical penetration had not at that time undertaken the elaborate theories of 'testing hypotheses' which have appeared in recent years.

In 1931, in the introduction to the *Hh* function published by the British Association in the first volume of their *Mathematical Tables* (Fisher, 1931), I gave the solution of an allied problem closely akin to the more recent developments.

If α is the true deviation of a value in terms of the true standard deviation σ, and a is the apparent deviation in terms of the estimate s, then

$$\mu+\alpha\sigma = \bar{x}+as,$$

and for each value of α the quantity a will have a determinate sampling distribution depending only on the sample number n and the deviation α. The solution is of value in a number of practical problems, for, given α, the percentile values of a are calculable, and given a the corresponding percentile values of α. For an industrial product, α may determine the percentage of the total output which fails in a specific test, and if a is designed to set the limits of a test included in a specification in such a way that the probability of failure to meet the specification is controlled, so as not to exceed some known value, the corresponding value of a shows how far the specification can go, even on the basis of a limited number of routine tests.

The solution given in the British Association Tables (Fisher, 1931), and for which the fine table of the hyperbolic Hermite function *Hh* was specially calculated by Dr J.R. Airey, was adopted without acknowledgement by a pupil of Dr Neyman, a certain S. Kolodziejczyk, who published a note (Kolodziejczyk, 1933) in the *Comptes Rendus* of the Académie des Sciences. As I had previously rather pressed this solution on Neyman's notice, owing to its important industrial applications, I was led to inquire why no acknowledgement was given of the origin of the solution, but acknowledgements only to Neyman's writings. Dr Neyman assured me, however, that in the original form of the note, reference to my introduction had been inserted, but had been cut out by the editor of *Comptes Rendus*, in *shortening* Kolodziejczyk's note. So far as I know, neither Neyman nor his follower has done anything to rectify the invidious position in which they have been placed. It would appear, however, that in this case at least Neyman did not feel any general objection to the logic of fiducial inference.

The principle of fiducial inference is very conveniently illustrated with reference to the variance of a normal distribution. If σ^2 is the true variance of the distribution, and if from a sample of n' observations we calculate the estimate,

$$s^2 = \frac{1}{n'-1}S(x-\bar{x})^2,$$

(Fisher, 1924, 1925, 1926*a*), and those of the z distribution for the logarithmic ratio of two estimates of a variance. It was obvious that these involved no logical principle beyond that of 'Student's' original test, but were the fruits of the exact solutions of problems of distribution which I had arrived at in the meanwhile. No misunderstandings were apparent at this period; perhaps because writers without sufficient logical penetration had not at that time undertaken the elaborate theories of 'testing hypotheses' which have appeared in recent years.

In 1931, in the introduction to the *Hh* function published by the British Association in the first volume of their *Mathematical Tables* (Fisher, 1931), I gave the solution of an allied problem closely akin to the more recent developments.

If α is the true deviation of a value in terms of the true standard deviation σ, and a is the apparent deviation in terms of the estimate s, then

$$\mu+\alpha\sigma = \bar{x}+as,$$

and for each value of α the quantity a will have a determinate sampling distribution depending only on the sample number n and the deviation α. The solution is of value in a number of practical problems, for, given α, the percentile values of a are calculable, and given a the corresponding percentile values of α. For an industrial product, α may determine the percentage of the total output which fails in a specific test, and if a is designed to set the limits of a test included in a specification in such a way that the probability of failure to meet the specification is controlled, so as not to exceed some known value, the corresponding value of a shows how far the specification can go, even on the basis of a limited number of routine tests.

The solution given in the British Association Tables (Fisher, 1931), and for which the fine table of the hyperbolic Hermite function *Hh* was specially calculated by Dr J.R. Airey, was adopted without acknowledgement by a pupil of Dr Neyman, a certain S. Kolodziejczyk, who published a note (Kolodziejczyk, 1933) in the *Comptes Rendus* of the Académie des Sciences. As I had previously rather pressed this solution on Neyman's notice, owing to its important industrial applications, I was led to inquire why no acknowledgement was given of the origin of the solution, but acknowledgements only to Neyman's writings. Dr Neyman assured me, however, that in the original form of the note, reference to my introduction had been inserted, but had been cut out by the editor of *Comptes Rendus*, in *shortening* Kolodziejczyk's note. So far as I know, neither Neyman nor his follower has done anything to rectify the invidious position in which they have been placed. It would appear, however, that in this case at least Neyman did not feel any general objection to the logic of fiducial inference.

The principle of fiducial inference is very conveniently illustrated with reference to the variance of a normal distribution. If σ^2 is the true variance of the distribution, and if from a sample of n' observations we calculate the estimate,

$$s^2 = \frac{1}{n'-1}S(x-\bar{x})^2,$$

then it has been demonstrated by a variety of methods that

$$\frac{(n'-1)s^2}{\sigma^2} = \chi^2,$$

for $n'-1$ degrees of freedom, where χ^2 is distributed in random samples, from the same or from different populations, as in the sum of the squares of $n'-1$ independent quantities, each normally distributed about zero with unit variance.

This distribution is therefore capable of tabulation for different known values of the degrees of freedom, and has in fact been tabulated rather thoroughly. For any value of n' the values exceeded by χ^2 in 99, 98, ..., 1 % of trials, or, in general, with a probability p, are therefore determinate.

If, then, we accept s^2 as an unbiased or unprejudiced estimate of σ^2, we must recognize that the values of σ^2 in the populations which in fact supply samples providing the estimate s^2 must be distributed as is

$$\frac{(n-1)s^2}{\chi^2},$$

where χ^2 is given this distribution. I.e. that if p is the probability that any given value χ^2 will be exceeded by chance, then p is the probability that σ^2 will exceed

$$\frac{(n-1)s^2}{\chi^2},$$

where s^2 is the estimated variance observed. For distinctness the probabilities arrived at by this and analogous arguments are known as fiducial probabilities.

The condition that s^2 is unbiased is most clearly satisfied if the sample from which it was derived constitutes the whole of the available information concerning the variance in question. It was doubtless an apprehension of this point which led 'Student' in his original paper to entitle it 'On the probable error of a unique sample'. The uniqueness postulated for the sample precludes the application of this method to cases in which the estimated variance can be compared with other estimates obtained by a series of repeated samplings. The condition of uniqueness may in such more complicated cases be satisfied by combining all the available information to form a single estimate. The criterion of Sufficiency is relevant here, since only sufficient estimates contain the whole of the information available. Alternatively, as I have stressed in an earlier paper (1939), we may take the responsibility for assuring the validity of the method by deciding to examine the inferences to be drawn from a given body of data, *as if it were unique*, that is, to ignore *pro tempore* any additional information, often of a vague and uncertain character, which may be available about the unknowns involved.

To extend this form of reasoning to the simultaneous distribution of the two parameters of a normal distribution, the mean and the variance, one need only note that if

$$\chi_1 = (\bar{x}-\mu)\frac{\sqrt{n'}}{\sigma}, \tag{1}$$

$$\chi^2 = \frac{(n'-1)s^2}{\sigma^2}, \tag{2}$$

μ being the mean of the population, then χ_1 will be distributed about zero with unit variance, while χ^2 will have the same distribution as before, these two distributions being independent when simultaneous variation of both variates is considered. Since jointly they determine the values of both μ and σ^2 in terms of their unprejudiced estimates $\bar{x}$ and s^2, the simultaneous fiducial distribution is determined.

It is possible to resolve the distribution into factors in several ways. Instead of equations (1) and (2), we might equally have written

$$t = (\bar{x}-\mu)\frac{\sqrt{n'}}{s}, \tag{3}$$

$$\chi^2_{n'} = \frac{1}{\sigma^2}\{(n'-1)s^2 + n'(\bar{x}-\mu)^2\}, \tag{4}$$

in which $\chi^2_{n'}$ has n' degrees of freedom, and t is distributed in 'Student's' distribution for $n'-1$ degrees of freedom. The simultaneous distributions defined in these two ways are, of course, strictly equivalent. We may note that in (2) μ has been eliminated so as to give the marginal distribution of σ; in (3) likewise σ is eliminated so that we have the marginal distribution of μ. These equations give the fiducial distributions of the two statistics singly.

Expression (1), complementary to (2), shows that if σ is known, the fiducial distribution of μ is normal, and not of 'Student's' form, as is the general, or marginal, distribution. Similarly, expression (4), complementary to (3), shows that if μ is known, σ may be estimated from n' instead of $n'-1$ degrees of freedom, and its fiducial distribution depends on $\chi^2_{n'}$, whereas the marginal distribution is derived from $\chi^2_{n'-1}$. It is not in any way surprising or embarrassing that the distributions should be thus mutually dependent.

Bartlett (1937) has thrown doubt on the validity of the simultaneous distribution for reasons which appear to be obscure. I must therefore quote the passage:

'It has been noted (Bartlett, 1936*a*) that if our information on a population parameter θ can be confined to a single degree of freedom, a fiducial distribution for θ can be expected to follow, and possible sufficiency properties that would achieve this result have been enumerated. A corresponding classification of fiducial distributions is possible.

Since recently Fisher (1935) has put forward the idea of a simultaneous fiducial distribution, it is important to notice that the sufficient set of statistics $\bar{x}$ and s^2 obtained from a sample drawn from a normal population (usual notation) do not at once determine fiducial distributions for the mean m and variance σ^2. That for σ^2 follows at once from the relation

$$P(\bar{x}, s^2 \mid m, \sigma^2) = p(\bar{x} \mid m, \sigma^2)\, p(s^2 \mid \sigma^2), \tag{1}$$

but that for $\bar{x}$ depends on the possibility of the alternative division

$$p\{\Sigma(x-m)^2/\sigma^2\}\, p(t), \tag{2}$$

where t depends only on the unknown quantity m. No justification has yet been given that because the above relations are equivalent respectively to fiducial distributions denoted by $fp(m/\sigma^2)fp(\sigma^2)$ and $fp(\sigma^2/m)fp(m)$, and hence symbolically to $fp(m,\sigma^2)$, that the idea of a simultaneous fiducial distribution, and hence by integration the fiducial distribution of either of the two parameters, is valid when *both* relations of form (1) and (2) do not exist (Bartlett, 1936*b*). Moreover, even in the above example, the simultaneous distribution is only to be regarded as a symbolic one, for there is no reason to suppose that from it we may infer the fiducial distribution of, say, $m+\sigma$.'

In spite of this opinion I can see no more than analytical difficulty in obtaining the fiducial distribution of any function $f(\mu,\sigma)$ of the two parameters, by integration of the simultaneous distribution. In particular, the fiducial distribution of $\mu+\sigma$, chosen by Bartlett, follows from the special case $\alpha = 1$, of the problem of α and a mentioned above, which was obtained by just such an integration. The known distribution of a when $\alpha = 1$, i.e.

$$\frac{(n'-1)!}{2^{\frac{1}{2}(n'-3)}\{\frac{1}{2}(n'-3)\}!}(1+a^2)^{-\frac{1}{2}n'}e^{-n'/2(1+a^2)}I_{n'-1}\left(\frac{-a\sqrt{n'}}{\sqrt{(1+a^2)}}\right)da,$$

where $$I_n(z) = \int_z^\infty I_{n-1}(t)\,dt, \quad I_0(z) = \frac{1}{\sqrt{(2\pi)}}\int_z^\infty e^{-\frac{1}{2}t^2}dt,$$

supplies the fiducial distribution, which he seems to seek, merely by substituting

$$\mu+\sigma = \bar{x}+as\sqrt{\frac{n-1}{n}}.$$

The representation is, of course, symbolical, and the concepts are abstract, like other mathematical concepts, but the numerical probabilities supplied by the distribution are none the less definite, and are as capable as any other of experimental verification and of practical application.

2. The significance of the difference between the means of samples from populations having variances in an unknown ratio

It has been seen that the difficulties which stood in the way of an early discovery of 'Student's' test lay largely in the logical concepts involved, and that in the modern period both Burnside and Bartlett have failed to grasp the essential reasoning needed to argue from the data presented, without the introduction of prior knowledge. It is not surprising that Bartlett also found difficulties with the application of the same methods to the more intricate problem of the comparison of the means of samples having unequal variances, or more correctly from populations, of which the variance ratio is unknown, and itself constitutes one of the parameters which require to be 'Studentized'. For it is obvious that if we use this method to eliminate the sampling errors of the estimates of both of two variances, this process is equivalent to the similar elimination of the unknown variance ratio, and of an estimate of either of them based on the two samples and on a known variance ratio.

This solution, first given in a different form by W.-V. Behrens (1929), may be developed very simply; for, if μ is the mean common to the two populations, we have from 'Student's' work

$$\bar{x}_1-\mu = s_1t_1,$$
$$\bar{x}_2-\mu = s_2t_2,$$

where s_1^2 and s_2^2 are the variances of the means, each estimated from its own sample, and t_1

and t_2 are distributed in repeated sampling in 'Student's' distribution for n_1 and n_2 degrees of freedom respectively. Moreover, t_1 and t_2 are distributed independently. Consequently, the difference between the observed means

$$\bar{x}_1 - \bar{x}_2$$

will be distributed as is

$$s_1 t_1 - s_2 t_2,$$

where s_1 and s_2 are known, and t_1 and t_2 have a known simultaneous distribution.

Sukhatmé (1938) has tabulated the 5 and 1 % levels of significance in terms of

$$d = \frac{s_1 t_1 - s_2 t_2}{\sqrt{(s_1^2 + s_2^2)}}, \quad \theta = \tan^{-1}(s_1/s_2).$$

He was thus able to utilize the fact that the values of d to be tabulated do not change very rapidly with change of angle, so that intervals of 15° are sufficiently close; a consideration of great importance, seeing that even at a single level of significance the table is one of triple entry.

The objections raised by Bartlett to this solution do not seem to be the same that weighed with him in the fiducial distribution of the parameters of a single normal distribution, though doubtless they arise from a similar underlying misapprehension of the argument.

(*a*) It was for a time claimed that an alternative test of significance was available giving different numerical values from that obtained above (Bartlett, 1936). An examination of the particular case adduced showed that the proposed test was defective, and as it has now been abandoned it need not be further discussed.

(*b*) It was argued that in repeated sampling from populations having variances in a fixed ratio, the value of d tabulated would not be exceeded by chance with the frequency indicated by the test of significance. One or other of two misapprehensions seems to be involved here. In 'Student's' test the quantity t appears in two roles. First, it is the pivotal quantity the distribution of which is independent of the population sampled, and the distribution of which is therefore accepted for the particular sample under consideration, if this sample is unprejudiced, by selection, or in any other way. Secondly, it is the quantity tabulated. In our present problem it is the pair of quantities t_1 and t_2 the simultaneous distribution of which is independent of the parameters of the population sampled. It was never imagined that d also preserved this property. The tabular value d is only introduced as a means of testing the significance of

$$s_1 t_1 - s_2 t_2$$

by integrating the frequency of simultaneous values of t_1 and t_2 over the region for which this exceeds

$$d\sqrt{(s_1^2 + s_2^2)}.$$

By a parallel but less suitable convention the same quantity might have been equated to, for example,

$$d'(s_1 + s_2),$$

and the tabulated values of d' would then have supplied the test of significance. The distribution of d or of d', or whatever might be used in place of them, in successive samples from a fixed population is entirely irrelevant.

The other misapprehension which may be involved is that Bartlett does not perhaps perceive that the unknown ratio of the variances of the two populations sampled has been 'Studentized', i.e. has been eliminated in a manner analogous to 'Student's' elimination of the unknown variance of a single sample.

In 'Student's' case no one imagines that for any fixed value of s

$$\mu + st$$

will be exceeded by the mean of a sample, $\bar{x}$, in $2\frac{1}{2}$ % of trials from a population with constant variance, where t is the tabulated value at the 5 % point, which itself is exceeded in just $2\frac{1}{2}$ % of trials. In such a case $\bar{x}$ is, and has long been known to be, distributed normally, and not in 'Student's' distribution. It is to allow for the fact that the variance of this normal distribution is unknown, that 'Student's' distribution is introduced. The possibility of eliminating the unknown variance is due to our knowledge of its fiducial distribution. The notion of repeated sampling from a fixed population has served its sole purpose when the distribution of t has been established.

Similarly, it is not to be supposed that Behrens's test of significance will represent the frequency with which

$$s_1 t_1 + s_2 t_2$$

exceeds any chosen value in a population of samples drawn from populations having a fixed variance ratio. The notion of repeated sampling from a fixed population has completed its usefulness when the simultaneous distribution of t_1 and t_2 has been obtained.

(c) Finally, Bartlett developed a numerical argument in the following words (1936, p. 565): 'An examination of Behrens' complete table ($n_1 = n_2$) might be sufficient to make us suspect its validity, for in all cases the fiducial probability given is less for $s_1/s_2 = 1$ than $s_1/s_2 = 0$ or ∞, whereas, given T, we should expect to be more sure that the observed difference is significant than if $s_1/s_2 = 1$, since in that case there is evidence that $\sigma_1^2 + \sigma_2^2$ is more efficiently estimated.'

This comment seems to show no consciousness that the ratio of the true standard deviations σ_1/σ_2 is not necessarily equal to that of the estimated values s_1/s_2. Recognizing that it is the essential business of the test to allow for the sampling error in the estimate of the variance ratio, we should not be inclined quickly to generalize on the contrast between the cases $\theta = 0°$ and $\theta = 45°$, even when $n_1 = n_2$.

In publishing his tables Sukhatmé (1938) has pointed out that at the 5 % point ($2\frac{1}{2}$ % in each tail) d is less at 45° than at 0° in accordance with Bartlett's expectation, when n_1 and n_2 exceed 5, but that for the smaller values the reverse is true. Even if Bartlett's expectation were justified for large samples, there is thus no ground for suspecting the accuracy of the slight table given by Behrens. It will appear in the present paper that even the first correction from the normal value, in the asymptotic expansion appropriate to large samples, is not of constant sign in this respect. At the 5 % point d for large samples is, as Sukhatmé states, less at 45° than at 0°, but at the 10 % point (5 % in each tail) the reverse is the case. The contribution of the second approximation term is moreover greater at 45° than at

0 or 90°, as far as the 2 % point. In the neighbourhood of the zero of the first term, the second term will of course govern the sign of the difference.

The first impression derived from Bartlett's criticism was that Behrens's solution, which I had confirmed later (1935) using a purely fiducial argument, was affected by an analytic error of some sort. The careful examination of the problem, which his strictures provoked, leaves no doubt that Bartlett's difficulties were entirely of a logical character. There is no error in the analysis. In the intervening period both Jeffreys (1940) and Yates (1939), from entirely different standpoints, have explained the logic of the argument, and its close analogy, or more properly identity, with that required for 'Student's' original test.

3. The treatment of the problem by asymptotic approximation

In 1926, in one of a series of papers written in collaboration with 'Student' for *Metron* I developed the ordinate and integral of 'Student's' distribution in a series in powers of n^{-1}, giving the polynomial coefficients so far as the fifth adjustment (Fisher, 1926*b*). The purpose of this expansion was to supply sufficiently accurate values of the probabilities corresponding to any values of t for values of n beyond the range, which it was proposed to tabulate. The series was also used for computing and checking the values tabulated.

In the case of Behrens's extension of 'Student's' test there are even stronger reasons for using a similar method. The direct calculations carried out by Sukhatmé are very much more laborious than those needed for 'Student's' integral. At a single level of significance values are needed, not as in 'Student's' case, for a single series of numbers of degrees of freedom, but for various values of three parameters provided by the two numbers of degrees of freedom of the two samples, and the estimated ratio of the variances of the two means. For functions of many variables there is a great advantage in the use of explicit formulae in which the several variables may be substituted, and there is much to be gained by extending the use of such formulae over regions too extensive for complete tabulation. Thirdly, it should be noted that the logical situation in which we would prefer to rely on the separate estimates of variances from the two samples, rather than on any process of pooling these estimates, is of more frequent occurrence with large samples than with small, and is particularly applicable to cases, such as arise in Physics and in Astronomy, in which we wish to compare estimates of the value of the same quantity (*a*) from relatively ample data of low intrinsic accuracy, and (*b*) from a small series of observations of relatively high precision. When, as often happens, the estimates of precision of the means obtained in these two ways are of the same order of magnitude, the only satisfactory test of significance is that based on Behrens's solution. For the discrepancy between the two means will be interpreted as the sum or difference of two errors, one distributed normally with a well-determined variance, while the other is of 'Student's' type. Tables 5 and 6 have been constructed from the asymptotic expansion with this special application in view.

The work of making Behrens's test available in practical use has been admirably initiated in Sukhatmé's two excellently arranged tables for the 5 and 1 % values (1938), for all combinations of n_1 and n_2 in the harmonic series 6, 8, 12, 24, ∞. We may expect the asymptotic expansion to provide (i) A check on Sukhatmé's values, obtained by a completely independent method, and applicable at least for the higher values of n_1 and n_2. (ii) Values of higher accuracy than could be obtained from Sukhatmé's table for values of n_1 and n_2 above 12. (iii) A wider range of levels of significance in the region to which the asymptotic expansion is applicable. (iv) The theoretical guidance offered by the algebraic form of the leading terms of the expansion.

The expansion of the ordinates of 'Student's' distribution is of the form

$$\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t^2} \sum_0^\infty P_r n^{-r},$$

where P_r is a polynomial in t^2 of degree $2r$. In particular

$$P_0 = 1,$$
$$P_1 = (t^4 - 2t^2 - 1) \div 4,$$
$$P_2 = (3t^8 - 28t^6 + 30t^4 + 12t^2 + 3) \div 96,$$
$$P_3 = (t^{12} - 22t^{10} + 113t^8 - 92t^6 - 33t^4 - 6t^2 + 15) \div 384,$$
$$P_4 = (15t^{16} - 600t^{14} + 7100t^{12} - 26616t^{10} + 18330t^8 + 6360t^6 + 1980t^4 - 1800t^2 - 945) \div 92160,$$
$$P_5 = (3t^{20} - 190t^{18} + 4025t^{16} - 33976t^{14} + 103702t^{12} - 63444t^{10} - 21270t^8 - 7800t^6 + 4455t^4 + 1890t^2 - 17955) \div 368640.$$

Integrating the expansion for the ordinate from t to ∞ we have the expansion for the probability, which may be written

$$q + \frac{1}{\sqrt{(2\pi)}} te^{-\frac{1}{2}t^2} \sum_1^\infty Q_r n^{-r},$$

in which

$$q = \frac{1}{\sqrt{(2\pi)}} \int_t^\infty e^{-\frac{1}{2}t^2} dt,$$

$$Q_1 = (t^2 + 1) \div 4,$$
$$Q_2 = (3t^6 - 7t^4 - 5t^2 - 3) \div 6(4^2),$$
$$Q_3 = (t^{10} - 11t^8 + 14t^6 + 6t^4 - 3t^2 - 15) \div 6(4^3),$$
$$Q_4 = (15t^{14} - 375t^{12} + 2225t^{10} - 2141t^8 - 939t^6 - 213t^4 + 915t^2 + 945) \div 360(4^4),$$
$$Q_5 = (3t^{18} - 133t^{16} + 1764t^{14} - 7516t^{12} + 5994t^{10} + 2498t^8 + 1140t^6 + 180t^4 + 5355t^2 + 17955) \div 360(4^5).$$

By calculating the maxima of the fifth correction, it was shown that this never exceeds 10^{-5} when n is greater than 18. For our present purpose it was therefore supposed that a useful level of accuracy would be obtained by using the expansion up to P_4.

In 1937 Cornish & Fisher discussed the problem of using a probability integral expressed as a series of corrective terms to the probability integral of the normal distribution to obtain explicit expressions for the percentile points at chosen levels of significance. If x is the normal deviate at any such level, and t the corresponding deviate of the distribution to be tabulated, then writing

$$t-x = f(t)$$

we may equate the expression for the probability of falling beyond any value t to

$$q+\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}t^2}\{f+\tfrac{1}{2}tf^2+\tfrac{1}{6}(t^2-1)f^3+\tfrac{1}{24}(t^3-3t)f^4+\ldots\},$$

where the coefficients involve the Hermite polynomials, and whence, substituting for each order of magnitude in succession, as, for example,

$$\begin{aligned} f(t) &= tQ_1+ \\ &\quad tQ_2-\tfrac{1}{2}t(tQ_1)^2+, \end{aligned}$$

we develop an expansion expressing f in terms of t. Since it is more convenient to use an expansion in terms of x, the values of which are available in advance, we convert the expansion by means of the very useful formula

$$f(t) = f(x)+\frac{1}{2}\frac{d}{dx}f^2(x)+\frac{1}{6}\frac{d^2}{dx^2}f^3(x)+\frac{1}{24}\frac{d^3}{dx^3}f^4(x)+\ldots,$$

so obtaining the percentile deviate explicitly in terms of the corresponding deviate of the normal distribution.

As a by-product of applying this process to Sukhatmé's d the following expansion has been obtained for 'Student's' t:

$$t = x+x\sum_1^\infty R_r n^{-r},$$

where R is a polynomial in x^2 of degree r, the first five being

$$\begin{aligned} R_1 &= (x^2+1)\div 4, \\ R_2 &= (5x^4+16x^2+3)\div 6(4^2), \\ R_3 &= (3x^6+19x^4+17x^2-15)\div 6(4^3), \\ R_4 &= (79x^8+776x^6+1482x^4-1920x^2-945)\div 360(4^4), \\ R_5 &= (27x^{10}+339x^8+930x^6-1782x^4-765x^2+17955)\div 360(4^5). \end{aligned}$$

In this expansion we may substitute, for example, $x = 1{\cdot}959963985$ for a 5 % point, or $x = 2{\cdot}575829303$ for the 1 % point, to obtain the percentile values for any number of degrees of freedom, beyond those for which tables are available.

4. The evaluation of Behrens's integral

We require to find the total frequency for which

$$s_1t_1+s_2t_2$$

exceeds

$$d\sqrt{(s_1^2+s_2^2)}$$

when t_1 and t_2 are independently distributed, with frequency elements

$$\frac{dt_1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}t_1^2}\sum_0^\infty P_r n_1^{-r}$$

and

$$\frac{dt_2}{\sqrt{(2\pi)}}e^{-\frac{1}{2}t_2^2}\sum_0^\infty P_r n_2^{-r}$$

respectively.

If $$t_1 = x\cos\theta + y\sin\theta, \quad t_2 = x\sin\theta - y\cos\theta, \quad s_1/s_2 = \tan\theta,$$
then the limit of integration is simply $y = d$. Moreover, the exponential terms may be at once transformed, since
$$t_1^2 + t_2^2 = x^2 + y^2,$$
so that the integral may be written
$$\int_d^\infty \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y^2}\,dy \int_{-\infty}^\infty \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\,dx\, \Sigma P_r(xc+ys)\,n_1^{-r}\, \Sigma P_r(xs-yc)\,n_2^{-r}.$$

Neither stage of the integration offers difficulty. With respect to x, the integrals of all odd powers vanish, while any even power x^{2s} is replaced by
$$(2s-1)(2s-3)\dots 3.1.$$

Since the polynomials are even in t_1 and t_2, integration with respect to x removes all odd powers not only of x but of y. For integration with respect to y we may use
$$\frac{1}{\sqrt{(2\pi)}}\int_d^\infty y^{2s}e^{-\frac{1}{2}y^2}\,dy = -\frac{1}{\sqrt{(2\pi)}}\left[y^{2s-1}e^{-\frac{1}{2}y^2}\right] + (2s-1)\frac{1}{\sqrt{(2\pi)}}\int_d^\infty y^{2s-2}e^{-\frac{1}{2}y^2}\,dy$$
$$= \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}d^2}\{d^{2s-1} + (2s-1)\,d^{2s-3} + (2s-1)(2s-3)\,d^{2s-5} + \dots\}$$
$$+ (2s-1)(2s-3)\dots 3q(d).$$

Now, for all cases, the whole integral must take the limiting values zero when d is infinite, and unity when d is negatively infinite, as does the function $q(d)$. Moreover, at these limits the expression in the upper line tends to zero. Consequently, the polynomials must be such that the term in $q(d)$ vanishes for all cases except the leading term
$$P_0(t_1)\,P_0(t_2).$$

It follows that the integral is obtained in the form
$$q(d) + \frac{d}{\sqrt{(2\pi)}} e^{-\frac{1}{2}d^2} S,$$
in which the term in S involving $$n_1^{-r} n_2^{-s}$$
has even powers of d only up to $4(r+s)-2$, and powers of $\cos\theta$ and $\sin\theta$ up to $4(r+s)$.

For example, if, in the polynomial
$$(t_1^4 - 2t_1^2 - 1),$$
we substitute $$t_1 = xc + ys,$$
and integrate with respect to x, we obtain
$$(3c^4 + 6c^2s^2y^2 + s^4y^4),$$
$$-2(c^2 + s^2y^2),$$
$$-1.$$

Making this homogeneous in c and s and arranging in powers of y, we have
$$s^4y^4,$$
$$(-2s^4 + 4s^2c^2)\,y^2,$$
$$-s^4 - 4s^2c^2.$$

Click for pt. 2 of paper