

123

THE MATHEMATICAL DISTRIBUTIONS USED IN THE COMMON TESTS OF SIGNIFICANCE

Author's Note (CMS 13.352a)

This paper was written to give mathematical teachers a concise account of these distributions, with demonstrations and the analysis of their probability integrals. These are identified in various forms with partial sums of the binomial and Poisson series. Papers 36 & 123 together supply a compact mathematical background for the common tests of significance.

THE MATHEMATICAL DISTRIBUTIONS USED IN IN THE COMMON TESTS OF SIGNIFICANCE

By R. A. FISHER

Introduction.—The three frequency distributions which provide the greatest number of tests of significance in common use are all closely related. The main types of application will be found illustrated arithmetically in the author's book *Statistical Methods for Research Workers* and in other publications in which extensive use is made of the arithmetical arrangement known as the Analysis of Variance. Some need has, however, been felt by mathematicians for a concise account of the algebraic properties and relationships of these distributions, and the following are essentially lecture notes designed to give a mathematical student a clear account of their properties.

1. *The frequency distribution of χ^2 .*—If x_1, x_2, \dots, x_n , are independent values of a variate distributed normally about zero, with unit variance, then the quantity

$$\chi^2 = S(x^2),$$

where S , as usual, stands for summation over the sample, has a distribution given by:—

$$df = \frac{1}{\frac{n-2}{2}!} \left(\frac{1}{2}\chi^2\right)^{\frac{1}{2}(n-2)} e^{-\frac{1}{2}\chi^2} d\left(\frac{1}{2}\chi^2\right).$$

This may be proved in several ways, two of which deserve notice.

(a) By induction, for $n=1$, the expression reduces to

$$\sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}x^2} dx,$$

which is clearly the distribution of x^2 for a single observation. If, now, $2u$ is the sum of the squares of n independent values of the variate, and has the distribution,

$$df = \frac{1}{\frac{n-2}{2}!} u^{\frac{1}{2}(n-2)} e^{-u} du,$$

and x is an additional observation independent of the others, then

$$\chi^2 = 2u + x^2,$$

and its distribution is to be inferred from the simultaneous distribution

$$df = \frac{1}{\frac{n-2}{2}!} \sqrt{\frac{2}{\pi}} u^{\frac{1}{2}(n-2)} e^{-u-\frac{1}{2}x^2} du dx.$$

If we now substitute

$$u = \frac{1}{2}(\chi^2 - x^2), \quad du = d(\frac{1}{2}\chi^2),$$

we have

$$df = \frac{1}{\frac{n-2}{2}!} \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\chi^2} d(\frac{1}{2}\chi^2) \cdot \left(\frac{\chi^2 - x^2}{2}\right)^{\frac{1}{2}(n-2)} dx,$$

in which x takes all values from 0 to χ . Integration with respect to x will, therefore, yield a factor χ^{n-1} or $(\frac{1}{2}\chi^2)^{\frac{1}{2}(n-1)}$ (with a constant which need not be determined, but which may be obtained from the Eulerian integral of the first kind), giving the distribution

$$df = \frac{1}{\frac{n-1}{2}!} (\frac{1}{2}\chi^2)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2} d(\frac{1}{2}\chi^2),$$

in accordance with the general formula.

Although the proof by induction is an attractive exercise in Eulerian integrals, many students find an alternative proof based on Euclidean hyperspace more simple and direct.

If $x_1 \cdots x_n$ are the co-ordinates of a point in such space, the frequency density at any point is proportional to $e^{-\frac{1}{2}x^2}$, and depends only on the distance of the sample point from the origin. The region in which this density exceeds any specified value is, therefore, a hypersphere in n dimensions having volume proportional to χ^n . The volume in which χ lies within any elementary range $d\chi$ is, therefore, proportional to

$$\chi^{n-1} d\chi,$$

and the element of frequency in this range is proportional to

$$\chi^{n-1} e^{-\frac{1}{2}\chi^2} d\chi.$$

The Eulerian integral of the second kind,

$$\int_0^\infty t^p e^{-t} dt = p!,$$

then supplies the required constant factor and establishes the distribu-

tion of χ or χ^2 .¹

2. *The distribution of Student's t .*—If we have a value of χ^2 derived from n independent values, and an additional value x independent of the others, "Student's t " may be defined as

$$t = \frac{x\sqrt{n}}{\chi},$$

for n degrees of freedom. Writing down the simultaneous distribution of χ and x , as above, and substituting for x in terms of t , we obtain

$$df = \frac{1}{\frac{n-2}{2}!} \sqrt{\frac{2}{\pi}} \cdot \left(\frac{1}{2}\chi^2\right)^{\frac{1}{2}(n-2)} e^{-\frac{1}{2}\chi^2(1+t^2/n)} d\left(\frac{1}{2}\chi^2\right) \frac{\chi}{\sqrt{n}} dt;$$

or, putting u for $\frac{1}{2}\chi^2\left(1 + \frac{t^2}{n}\right)$,

$$\frac{1}{\frac{n-2}{2}!} \frac{2}{\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} u^{\frac{1}{2}(n-1)} e^{-u} du dt.$$

Integration with respect to u from 0 to infinity, recollecting that t must be equally frequently positive or negative, yields the distribution found by "Student" in 1908:—

$$\frac{\frac{n-1}{2}!}{\frac{n-2}{2}! \sqrt{\pi n} \left(1 + \frac{t^2}{n}\right)^{\frac{1}{2}(n+1)}} dt.$$

3. *The distribution of z .*—In the most general case arising in the analysis of variance, we consider two quantities, χ_1^2 , and χ_2^2 , based respectively on n_1 and n_2 values of the variate, all of which are independent. We may then define z so that

$$e^{2z} = u = \frac{n_2\chi_1^2}{n_1\chi_2^2},$$

and proceed to find the distribution of z . This is carried out, as in the previous cases, by writing down a simultaneous distribution of χ_1 and χ_2 and making the substitution

¹ This distribution was first given by Helmert in 1875; it was later found independently by Pearson, "Student," and others, in the examination of various special problems belonging to the wide class in which it occurs.

$$\chi_1^2 = \frac{n_1}{n_2} \chi_2^2 u.$$

The integration proceeds as before, yielding the general distribution for z ,

$$df = 2 \frac{\frac{n_1 + n_2 - 2}{2}!}{\frac{n_1 - 2}{2}! \frac{n_2 - 2}{2}!} n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2} \frac{e^{n_1 z} dz}{(n_1 e^{2z} + n_2)^{\frac{1}{2}(n_1 + n_2)}},$$

which is evidently that of the natural logarithm of the ratio between two independent estimates of the same standard deviation based on n_1 and n_2 degrees of freedom, respectively. The wide class of problem for which z provides the appropriate test of significance is most easily recognized from this property.

4. *The Probability integral of χ^2 .* The probability integral of the χ^2 distribution,

$$df = \frac{1}{\frac{n-2}{2}!} (\frac{1}{2}\chi^2)^{\frac{1}{2}(n-2)} e^{-\frac{1}{2}\chi^2} d(\frac{1}{2}\chi^2),$$

is

$$P = \int_{\frac{1}{2}\chi^2}^{\infty} \frac{1}{\frac{n-2}{2}!} t^{\frac{1}{2}(n-2)} e^{-t} dt,$$

which represents the probability of exceeding a given value of χ^2 . Now, integrating by parts,

$$\begin{aligned} \int_{\frac{1}{2}\chi^2}^{\infty} \frac{1}{r!} t^r e^{-t} dt &= \left[-\frac{1}{r!} t^r e^{-t} \right]_{\frac{1}{2}\chi^2}^{\infty} + \int_{\frac{1}{2}\chi^2}^{\infty} \frac{1}{(r-1)!} t^{r-1} e^{-t} dt \\ &= \frac{1}{r!} (\frac{1}{2}\chi^2)^r e^{-\frac{1}{2}\chi^2} + \int_{\frac{1}{2}\chi^2}^{\infty} \frac{1}{(r-1)!} t^{r-1} e^{-t} dt. \end{aligned}$$

When n is even, this process terminates, yielding the formula

$$P = e^{-\frac{1}{2}\chi^2} \left\{ 1 + \frac{1}{2}\chi^2 + \frac{1}{2!} (\frac{1}{2}\chi^2)^2 + \cdots + \frac{1}{\frac{n-2}{2}!} (\frac{1}{2}\chi^2)^{\frac{1}{2}(n-2)} \right\}$$

Thus, for

$$\begin{aligned} n = 2, & \quad P = e^{-\frac{1}{2}\chi^2}, \\ n = 4, & \quad P = e^{-\frac{1}{2}\chi^2} (1 + \frac{1}{2}\chi^2), \end{aligned}$$

$$n = 6, \quad P = e^{-\frac{1}{2}\chi^2} \left\{ 1 + \frac{1}{2}\chi^2 + \frac{1}{2!} \left(\frac{1}{2}\chi^2\right)^2 \right\},$$

$$n = 8, \quad P = e^{-\frac{1}{2}\chi^2} \left\{ 1 + \frac{1}{2}\chi^2 + \frac{1}{2!} \left(\frac{1}{2}\chi^2\right)^2 + \frac{1}{3!} \left(\frac{1}{2}\chi^2\right)^3 \right\},$$

all of which are easily calculated for a given value of $\frac{1}{2}\chi^2$.

When n is odd, the same process may be applied, terminating at $r = \frac{1}{2}$; we then have the formula,

$$P = \int_{\frac{1}{2}\chi^2}^{\infty} \frac{1}{(-\frac{1}{2})!} t^{-\frac{1}{2}} e^{-t} dt + e^{-\frac{1}{2}\chi^2} \left\{ \frac{1}{\frac{1}{2}!} \left(\frac{1}{2}\chi^2\right)^{\frac{1}{2}} + \frac{1}{\frac{3}{2}!} \left(\frac{1}{2}\chi^2\right)^{\frac{3}{2}} + \dots + \frac{1}{\frac{n-2}{2}!} \left(\frac{1}{2}\chi^2\right)^{\frac{1}{2}(n-2)} \right\}.$$

In the integral, write $\frac{1}{2}\chi^2$ for t , and substitute for the fractional factorials using $(-\frac{1}{2})! = \sqrt{\pi}$, and we find

$$P = \sqrt{\frac{2}{\pi}} \int_{\chi}^{\infty} e^{-\frac{1}{2}x^2} dx + \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\chi^2} \left\{ \chi + \frac{1}{3}\chi^3 + \frac{1}{3 \cdot 5}\chi^5 + \dots + \frac{1}{3 \cdot 5 \dots (n-2)} \chi^{n-2} \right\}.$$

The integral is the familiar probability integral of the normal curve, the contribution to P being the total frequency outside the limits $\pm \chi$ times the standard deviation. The series is easily evaluated as before.

5. *Relation of the χ^2 distribution to the Poisson series.*—It will be noticed that, when n is even, the probability of the variate $\frac{1}{2}\chi^2$ exceeding any specified value m is

$$e^{-m} \left(1 + m + \frac{m^2}{2!} + \dots + \frac{m^{\frac{1}{2}(n-2)}}{\frac{n-2}{2}!} \right),$$

which is the sum of the first $\frac{1}{2}n$ terms of the Poisson series, having the parameter m , or, in other words, the probability that a variate distributed in such a series is less than $\frac{1}{2}n$. This identity is expressed in the formula,

$$\int_m^{\infty} \frac{1}{p!} t^p e^{-t} dt = \sum_{x=0}^p \frac{1}{x!} m^x e^{-m},$$

where p , which takes the place of $\frac{1}{2}(n-2)$, is a positive integer or zero.

Thus, a table of χ^2 can be used as a table of the partial sum of the Poisson series; in particular, the 5 per cent value of χ^2 , which is the value exceeded once in 20 trials, gives (on halving) the value of m , the "expectation" of the Poisson series of which the first $\frac{1}{2}n$ terms occupy 5 per cent of the frequency.

For example, if n is 8, the 5 per cent value of χ^2 is 15.507; consequently, we may infer that, if a rare event has been observed only $3 [= \frac{1}{2}(n-2)]$ times, the observation has departed significantly from any expectation exceeding 7.754 occurrences and, consequently, its real frequency of occurrence probably does not exceed that which would give this number in our body of observations. Again, if n is 6, the 95 per cent point is 1.635, so that, if 3 cases have certainly been observed, the expectation probably exceeds 0.817, since for this value 95 per cent of the observed numbers will be 0, 1, or 2. We may thus use the table very simply to show just how much information about the frequency of rare events is contained in a record of only a few such occurrences.

6. *The probability integral of "Student's" t distribution.*—It has already been shown that the ratio t of a deviation to its standard error as estimated from n degrees of freedom is

$$df = \frac{\frac{n-1}{2}!}{\frac{n-2}{2}! \sqrt{\pi n}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} dt;$$

or, if $\tan \theta$ is written for t/\sqrt{n} ,

$$df = \frac{\frac{n-1}{2}!}{\frac{n-2}{2}! \sqrt{\pi}} \cos^{n-1} \theta \cdot d\theta.$$

Then the probability of exceeding a given value of t is

$$\int_{\alpha}^{\frac{1}{2}\pi} \frac{\frac{n-1}{2}!}{\frac{n-2}{2}! \sqrt{\pi}} \cos^{n-1} \theta \cdot d\theta,$$

where $t = \sqrt{n} \tan \alpha$.

Now, integrating by parts, it appears that

$$\int_{\alpha}^{\frac{1}{2}\pi} \frac{(r \pm \frac{1}{2})!}{r! \sqrt{\pi}} \cos^{2r+1} \theta \cdot d\theta = \left[\frac{(r + \frac{1}{2})!}{r! \sqrt{\pi}} \cos^{2r} \theta \sin \theta \right]_{\alpha}^{\frac{1}{2}\pi} + \int_{\alpha}^{\frac{1}{2}\pi} 2 \frac{(r + \frac{1}{2})!}{(r - 1)! \sqrt{\pi}} \cos^{2r-1} \theta \sin^2 \theta d\theta.$$

which, so long as r is positive, is equal to

$$-\frac{(r + \frac{1}{2})!}{r! \sqrt{\pi}} \cos^{2r} \alpha \sin \alpha + (2r + 1) \int_{\alpha}^{\frac{1}{2}\pi} \frac{(r - \frac{1}{2})!}{(r - 1)! \sqrt{\pi}} \cos^{2r-1} \theta \cdot d\theta - 2r \int_{\alpha}^{\frac{1}{2}\pi} \frac{(r + \frac{1}{2})!}{r! \sqrt{\pi}} \cos^{r+1} \theta \cdot d\theta.$$

Hence,

$$\int_{\alpha}^{\frac{1}{2}\pi} \frac{(r + \frac{1}{2})!}{r! \sqrt{\pi}} \cos^{2r+1} \theta \cdot d\theta = -\frac{1}{2} \cdot \frac{(r - \frac{1}{2})!}{r! \sqrt{\pi}} \cos^{2r} \alpha \sin \alpha + \int_{\alpha}^{\frac{1}{2}\pi} \frac{(r - \frac{1}{2})!}{(r - 1)! \sqrt{\pi}} \cos^{2r-1} \theta \cdot d\theta$$

when r is positive; but, when $r=0$,

$$\int_{\alpha}^{\frac{1}{2}\pi} \frac{(r + \frac{1}{2})!}{r! \sqrt{\pi}} \cos^{2r+1} \theta \cdot d\theta = \frac{1}{2}(1 - \sin \alpha).$$

Hence,

$$P = \frac{1}{2} - \frac{1}{2} \sin \alpha \left\{ 1 + \frac{1}{2} \cos \alpha + \frac{1 \cdot 3}{2 \cdot 4} \cos^3 \alpha \dots + \frac{1 \cdot 3 \cdot 5 \dots (n - 3)}{2 \cdot 4 \cdot 6 \dots (n - 2)} \cos^{n-2} \alpha \right\},$$

when n is even. When n is odd, we proceed in the same way until $r = \frac{1}{2}$, and obtain

$$P = \frac{1}{2} - \frac{\alpha}{\pi} - \frac{\sin \alpha}{\pi} \left\{ \cos \alpha + \frac{2}{3} \cos^3 \alpha + \dots + \frac{2 \cdot 4 \dots (n - 3)}{3 \cdot 5 \dots (n - 2)} \cos^{n-2} \alpha \right\}.$$

As in the case of χ^2 , when n is odd a transcendental function is required, in this case an inverse circular function, whereas when n is even, P is expressed as a function algebraic in t .

"Student" has given (1) four-figure tables of P up to $n=20$; beyond this value a good asymptotic expansion is available (3). "Student's" tables are for $1-P$ in the notation used above, and represent the probability of a value less than any given positive value of t . Since the distribution is symmetrical about zero, this probability is never less than $\frac{1}{2}$. For tests of significance, we often require the probability, $2P$, that the observed ratio of a deviation to its estimated standard error shall lie outside the limits $\pm t$, or the complementary probability, $1-2P$, that it shall lie within these limits.

It will help to make clear the analogy with the more general test of significance given by z , of which the χ^2 and t tests are special cases, to observe that, when n is even, the expansion for $1-2P$ is

$$\sin \alpha \left\{ 1 + \frac{1}{2} \cos^2 \alpha + \frac{1 \cdot 3}{2 \cdot 4} \cos^4 \alpha + \dots \right. \\ \left. + \frac{1 \cdot 3 \cdot 5 \dots n-3}{2 \cdot 4 \cdot 6 \dots n-2} \cos^{n-2} \alpha \right\};$$

or, in terms of $\frac{n}{n+t^2} = \cos^2 \alpha = q$,

$$\sqrt{1-q} \left\{ 1 + \frac{1}{2}q + \frac{1 \cdot 3}{2 \cdot 4} q^2 + \dots + \frac{1 \cdot 3 \cdot 5 \dots n-3}{2 \cdot 4 \cdot 6 \dots n-2} q^{\frac{1}{2}(n-2)} \right\},$$

in which the expression within the bracket is the first $\frac{1}{2}n$ terms of the binomial expansion

$$(1-q)^{-\frac{1}{2}}.$$

Just as the probability integral of χ^2 gives the partial sum of a Poisson series, so, therefore, does the probability integral of t give the partial sum of a special type of binomial expansion; in each case the external factor is the inverse of the sum of the complete series, and the identity holds for all even values of n .

7. *The probability integral of the distribution of z .*—The distribution of z involves two whole numbers, n_1 and n_2 , which are the numbers of degrees of freedom in the two lines of the analysis of variance to be compared, and is given by the general formula,

$$df = 2 \frac{\frac{n_1 + n_2 - 2}{2}!}{\frac{n_1 - 2}{2}! \frac{n_2 - 2}{2}!} n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2} \frac{e^{n_1 z} dz}{(n_1 e^{2z} + n_2)^{\frac{1}{2}(n_1 + n_2)}}$$

Writing

$$q = \frac{n_1 e^{2z}}{n_1 e^{2z} + n_2},$$

this becomes

$$df = \frac{\frac{n_1 + n_2 - 2}{2}!}{\frac{n_1 - 2}{2}! \frac{n_2 - 2}{2}!} q^{\frac{1}{2}(n_1 - 2)} (1 - q)^{\frac{1}{2}(n_2 - 2)} dq.$$

Now,

$$\begin{aligned} \int_q^1 \frac{(r + s + 1)!}{r!s!} x^r (1 - x)^s dx &= \left[-\frac{(r + s + 1)!}{r!(s + 1)!} x^r (1 - x)^{s+1} \right]_q \\ &+ \int_q^1 \frac{(r + s + 1)!}{(r - 1)!(s + 1)!} x^{r-1} (1 - x)^{s+1} dx \\ &= \frac{(r + s + 1)!}{r!(r + 1)!} q^r (1 - q)^{s+1} \\ &+ \int_q^1 \frac{(r + s + 1)!}{(r - 1)!(s + 1)!} x^{r-1} (1 - x)^s dx \\ &- \frac{r}{s + 1} \int_q^1 \frac{(r + s + 1)!}{r!s!} x^r (1 - x)^s dx. \end{aligned}$$

This establishes the recurrence relation

$$\begin{aligned} \int_q^1 \frac{(r + s + 1)!}{r!s!} x^r (1 - x)^s dx &= \frac{(r + s)!}{r!s!} q^r (1 - q)^{s+1} \\ &+ \int_q^1 \frac{(r + s)!}{(r - 1)!s!} x^{r-1} (1 - x)^s dx. \end{aligned}$$

Hence, when n_1 is even, the probability of

$$\frac{n_1 e^{2z}}{n_1 e^{2z} + n_2}$$

exceeding any fractional value q is

$$\begin{aligned} P &= (1 - q)^{\frac{1}{2}n_2} \left\{ 1 + \frac{n_2}{2} q + \frac{n_2(n_2 + 2)}{2 \cdot 4} q^2 \right. \\ &\quad \left. \dots \frac{n_2(n_2 + 2) \dots (n_2 + n_1 - 4)}{2 \cdot 4 \dots (n_1 - 2)} q^{\frac{1}{2}(n_1 - 2)} \right\}, \end{aligned} \tag{A}$$

of which the terms in the second component are the first $\frac{1}{2}n_1$ terms of the binomial expansion

$$(1 - q)^{-\frac{1}{2}n_1}.$$

From this expansion, when n_1 is even and not too large, the probability may be easily calculated.

Alternatively, using the direct result of integration by parts, we shall find the alternative expression

$$\begin{aligned} (1 - q)^{\frac{1}{2}(n_1+n_2-2)} & \left\{ 1 + \frac{n_1 + n_2 - 2}{2} \cdot \frac{q}{1 - q} \right. \\ & + \frac{(n_1 + n_2 - 2)(n_1 + n_2 - 4)}{2 \cdot 4} \left(\frac{q}{1 - q} \right)^2 + \dots \\ & \left. + \frac{(n_1 + n_2 - 2) \cdot \dots \cdot (n_2 + 2)}{2 \cdot 4 \cdot \dots \cdot (n_1 - 2)} \left(\frac{q}{1 - q} \right)^{\frac{1}{2}(n_1-2)} \right\} \quad (B) \end{aligned}$$

involving the first $\frac{1}{2}n_1$ terms of the expansion of the positive binomial

$$\left(1 + \frac{q}{1 - q} \right)^{\frac{1}{2}(n_1+n_2-2)}.$$

The probability integral of z , when n_1 is even, is thus equivalent to the sum of $\frac{1}{2}n_1$ terms of a negative binomial in form (A), or of a positive binomial in form (B). It is of some historical interest that the probability integral of the normal distribution was first introduced by De Moivre as an approximation to the sum of a terminating series of binomial terms. Indeed, had the eighteenth century mathematicians possessed greater analytic power, the distribution of z , which was unknown to statisticians up to about 10 years ago, might have been studied before the normal distribution.

If n_2 tends to infinity and n_2q to the limiting value χ^2 , both the forms (A) and (B) tend to the form

$$e^{-\frac{1}{2}\chi^2} \left\{ 1 + \frac{1}{2}\chi^2 + \frac{1}{2!} \left(\frac{1}{2}\chi^2 \right)^2 + \dots + \frac{1}{\frac{n_1 - 2}{2}!} \left(\frac{1}{2}\chi^2 \right)^{\frac{1}{2}(n_1-2)} \right\},$$

which we obtained for the distribution of χ^2 , if we identify n_1 of the general case with n of the χ^2 distribution. The distribution of χ^2 is, thus, as is obvious from its statistical derivation, the limiting case, when n_2 is infinite, of the general distribution, the substitution being

$$\frac{\chi^2}{n} = e^{2z}, \quad n = n_1.$$

Again, if $n_2 = 1$, the expression is evidently equivalent to that obtained for the probability that "Student's" test of significance t shall lie within the limits

$$\pm \sqrt{n_1 \frac{1-q}{q}},$$

but

$$\frac{1-q}{q} = \frac{n_2}{n_1} e^{-2z} = \frac{1}{n_1} e^{-2z},$$

hence the probability that z shall exceed a given value is the probability that t shall lie within the limits $\pm e^{-z}$, when $n_2 = 1$, $n_1 = n$.

Since z is the logarithm of the ratio of the estimates of the standard deviation derived respectively from n_1 and n_2 degrees of freedom, it follows that, if we interchange n_1 and n_2 and change the sign of z , the expression for the distribution is the same as before. Consequently, when n_2 is even, the probability integral may be expressed as the sum of $\frac{1}{2}n_2$ terms of a binomial expansion. The expression corresponding to (A) is, writing p for $1-q$,

$$1 - P = (1-p)^{\frac{1}{2}n_1} \left\{ 1 + \frac{n_1}{2} p + \frac{n_1(n_1+2)}{2 \cdot 4} p^2 + \dots + \frac{n_1(n_1+2) \dots (n_1+n_2-4)}{2 \cdot 4 \dots (n_2-2)} p^{\frac{1}{2}(n_2-2)} \right\}, \quad (A')$$

illustrating that, for $n_1 = 1$, $n_2 = n$, the probability that z exceeds any given value is the probability that t will fall *outside* the limits $\pm e^z$, and that, when n_1 is infinite, the χ^2 distribution is given by the transformation

$$\frac{\chi^2}{n} = e^{-2z}, \quad n = n_2.$$

Corresponding to expression (B), we have

$$1 - P = q^{\frac{1}{2}(n_1+n_2-2)} \left\{ 1 + \frac{n_1+n_2-2}{2} \frac{1-q}{q} + \dots + \frac{(n_1+n_2-2) \dots (n_1+2)}{2 \dots (n_2-2)} \left(\frac{1-q}{q} \right)^{\frac{1}{2}(n_2-2)} \right\}. \quad (B')$$

When both n_1 and n_2 are even, it appears that P and $1-P$ are the first $\frac{1}{2}n_1$ terms and the remaining $\frac{1}{2}n_2$ terms of the expansion of

$$(p+q)^{\frac{1}{2}(n_1+n_2-2)},$$

where

$$\frac{q}{p} = \frac{n_1}{n_2} e^{2z},$$

the ratio of the sums of squares in the analysis of variances.

In cases where either n_1 or n_2 is even, the probability integral can be expressed as the incomplete sum of a binomial series, either with positive or negative exponent, the exponent being the half of an odd integer if either n_1 or n_2 is odd. The case remains to be considered in which both n_1 and n_2 are odd.

For this case we apply the recurrence solution as far as $r = \frac{1}{2}$, obtaining

$$\begin{aligned} P &= \int_q^1 \frac{\frac{n_2 - 1}{2}!}{\frac{n_2 - 2}{2}! \sqrt{\pi}} x^{-\frac{1}{2}} (1 - x)^{\frac{1}{2}(n_2 - 2)} dx \\ &+ 2 \frac{\frac{n_2 - 1}{2}!}{\frac{n_2 - 2}{2}! \sqrt{\pi}} (1 - q)^{\frac{1}{2}n_2} q^{\frac{1}{2}} \left\{ 1 + \frac{n_2 + 1}{3} q + \dots \right. \\ &\left. + \frac{(n_2 + 1) \dots (n_2 + n_1 - 4)}{3 \dots (n_1 - 2)} q^{\frac{1}{2}(n_1 - 3)} \right\}. \end{aligned}$$

The numerical coefficient of the second term, when n_2 is odd, is

$$\frac{2 \cdot 2 \cdot 4 \dots (n_2 - 1)}{\pi \cdot 1 \cdot 3 \dots (n_2 - 2)},$$

and the integral remaining at this stage is just double the one which has been already evaluated for the t distribution; so, putting

$$x = \sin^2 \theta, \quad q = \sin^2 \alpha = \frac{t^2}{t^2 + n_2},$$

we find, since n_2 is odd,

$$\begin{aligned} P &= 1 - \frac{2\alpha}{\pi} - \frac{2 \sin \alpha}{\pi} \left\{ \cos \alpha + \frac{2}{3} \cos^3 \alpha + \dots \right. \\ &\left. + \frac{2 \cdot 4 \dots (n_2 - 3)}{3 \cdot 5 \dots (n_2 - 2)} \cos^{n_2 - 2} \alpha \right\} \end{aligned}$$

$$\begin{aligned}
& + \frac{2 \cdot 2 \cdot 4 \cdots (n_2 - 1)}{\pi \cdot 1 \cdot 3 \cdots (n_2 - 2)} \sin \alpha \cos^{n_2} \alpha \left\{ 1 + \frac{n_2 + 1}{3} \sin^2 \alpha + \cdots \right. \\
& \left. + \frac{(n_2 + 1) \cdots (n_1 + n_2 - 4)}{3 \cdot 5 \cdots (n_1 - 2)} \sin^{n_1 - 3} \alpha \right\}
\end{aligned}$$

in terms of α , where α is connected with z by the equation

$$\tan \alpha = \sqrt{\frac{n_1}{n_2}} \cdot e^z.$$

Tables for z for the 5 per cent and the 1 per cent points of the distribution have been given for $n_1 = 1, 2, 3, 4, 5, 6, 8, 12, 24, \infty$; the last five values are in harmonic progression and enable the table to be interpolated in the manner which I have called asymptotic interpolation. For n_2 , I have given values from 1 to 30, together with 60 and ∞ ; in this case again the series of values for 20, 30, 60, and ∞ , may be used for asymptotic interpolation and the table thus gives four-figure values of z , an accuracy fully sufficient for all common purposes for all combinations of n_1 and n_2 except the region in which n_1 exceeds 24 and n_2 exceeds 30.

R. A. FISHER

University College, London

REFERENCES

1. Helmert (1875), "Über die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler," *Zeitschrift für Mathematik und Physik*. xx, 300-303.
2. "Student," *New Tables for Testing Significance of Observations* (1925), *Metron* v. 105-120.
3. R. A. Fisher, *Expansion of "Student's" Integral in Power of n^{-1}* (1925).
4. A. de Moivre, *The Doctrine of Chances* (1756).