A TEST OF THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS OF SAMPLES FROM TWO NORMAL POPULATIONS WITHOUT ASSUMING EQUAL VARIANCES[1]

By Daisy M. Starkey

1. **History of the problem.** If the only available evidence about two normally distributed populations is contained in two samples, one from each, it has hitherto been the custom to the test the hypothesis that the means are equal by assuming that the quantity $\frac{\bar{x} - \bar{x}'}{\sqrt{ks^2 + k's'^2}}$ is distributed in Student's distribution, with $N + N' - 2$ degrees of freedom, where $s^2 = \frac{\Sigma(x - \bar{x})^2}{N(N - 1)}$ and $k = \frac{(N - 1)(N + N')}{N'(N + N' - 2)}$, the other notation being that used by R. A. Fisher.[2] The hypothesis underlying this test, however, is that the variances are equal. Although in many cases this may seem a reasonable assumption to adopt concurrently with that of equal means, it is undoubtedly not a necessary one, and it is, therefore, desirable that the test should be adapted to meet this difficulty.

The first advance on the problem was made by W. V. Behrens[3] who suggested that the distribution of the difference of the means could be expressed in terms of the observations in the samples from the two populations, the argument being entirely independent of the variances. R. A. Fisher[4] obtained substantially the same result, but expressed the argument in terms of fiducial probability. M. S. Bartlett[5] was of the opinion that Behrens' reasoning was incorrect, as he obtained some results which were apparently inconsistent with those tabulated in Behrens' paper, but R. A. Fisher[6] showed that Bartlett's argument was open to criticism. In the latter work, he actually obtained distributions for the case of two samples of two observations, and in the following we shall indicate some extensions of this more detailed work of Fisher, firstly, to the case

---

[1] Presented at the joint meeting of the American Mathematical Society and the Institute of Mathematical Statistics, Indianapolis, December 30, 1937.

Research done under a grant-in-aid from the Carnegie Corporation of New York City.

[2] *Statistical Methods for Research Workers*, 1925–1936.

[3] "Ein Beitrag zur Fehlerberechnung bei wenige Beobachtungen," *Landw. Jb.* 68, 807–37 (1929).

[4] "The Fiducial Argument in Statistical Inference," *Annals of Eugenics*, 6 (1935) pp. 391–8.

[5] "The Information Available in Small Samples," *Proc. Camb. Phil. Soc.* 32, pp. 560–6 (1936).

[6] "On a Point Raised by M. S. Bartlett on Fiducial Probability," *Annals of Eugenics* 7 Part IV, 370–5 (1937).

of other small samples of even numbers of observations, and, secondly, to samples of very large numbers.

2. **The case of small samples.** We recapitulate, briefly, the preliminary argument of R. A. Fisher[4], in which he denotes the unknown population means by μ and μ'. Since $\frac{\bar{x} - \mu}{s} = t$, where t is distributed in Student's distribution, we may write $\mu = \bar{x} - st$, and obtain the fiducial distribution of the population parameter μ in the form $G_1(\mu)\, d\mu$, where

$$G_1(\mu)\, d\mu = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \frac{d\mu}{s\left\{1 + \left(\frac{\bar{x} - \mu}{s}\right)^2 \Big/ n\right\}^{\frac{n+1}{2}}},$$

and a similar result for the fiducial distribution of μ'. The simultaneous fiducial distribution of μ and μ' is thus $G_1(\mu)\, G_2(\mu') d\mu d\mu'$ from which the fiducial distribution of $\mu - \mu'$ may be found. We may note that the characteristic function of $-(\mu - \mu') + (\bar{x} - \bar{x}')$ is $M(x)$, where

$$M(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ix[-(\mu-\mu')+(\bar{x}-\bar{x}')]}\, G_1(\mu) G_2(\mu')\, d\mu\, d\mu'$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ix(ts-t's')}\, H_1(t) H_2(t')\, dt\, dt',$$

where

$$H_1(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \frac{dt}{(1 + t^2/n)^{\frac{n+1}{2}}},$$

with a similar expression for $H_2(t)$. Thus from the fiducial point of view, the problem is essentially that of formally determining the distribution of the variate $ts - t's'$, or $at + bt'$, where $a = s$, $b = -s'$ are regarded as constants, t and t' being distributed in "Student's" distribution. The hypothesis $\mu = \mu'$ may then be examined by testing whether $\bar{x} - \bar{x}'$ is a significantly large value of this variate. We shall approach this distribution problem through the use of characteristic functions.

By definition, the characteristic function of "Student's" distribution is represented by the integral

$$(1) \qquad \frac{1}{\sqrt{\pi n}} \frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)} \int_{-\infty}^{\infty} \frac{e^{itx}}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}\, dt,$$

and may be evaluated by three methods which will be briefly considered.

First, by integrating the function

$$\frac{1}{\sqrt{\pi n}} \frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)} \frac{e^{iz|x|}}{\left(1+\frac{z^2}{n}\right)^{\frac{n+1}{2}}}$$

around a standard semicircular contour in the upper half of the z-plane, the value of the characteristic function is at once proved to be $2\pi i$ times the sum of the residues of the integrand within the contour when the radius of the semicircle becomes infinite. Within the contour there is one pole only at $z = i\sqrt{n}$. The residue at this pole is the coefficient of $1/h$ in the expansion of

$$\frac{1}{\sqrt{\pi n}} \frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)} \frac{e^{-|x|\sqrt{n}+i|x|h}}{\left[\left(\frac{h}{\sqrt{n}}\right)\left(2i+\frac{h}{\sqrt{n}}\right)\right]^{\frac{n+1}{1}}}$$

in ascending powers of h, which may readily be evaluated when n is odd.

Second, by using the result that

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{ixt}}{1+t^2}\,dt = e^{-|x|},$$

from which we deduce that

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{ixt}}{a+t^2}\,dt = \frac{1}{\sqrt{a}}\, e^{-\sqrt{a}|x|}.$$

Differentiating this result $(n - 1)$ times with respect to a, again considering odd values of n, we have that

$$(-1)^{\frac{1}{2}(n-1)} \frac{\left(\frac{n-1}{2}\right)!}{\pi} \int_{-\infty}^{\infty} \frac{e^{ixt}}{(a+t^2)^{\frac{n+1}{2}}}\,dt = \frac{d^{\frac{n-1}{2}}}{da^{\frac{n-1}{2}}}\left[\frac{1}{\sqrt{a}}\, e^{-\sqrt{a}|x|}\right].$$

By forming the first order differential equation in $y = \frac{1}{\sqrt{a}}\, e^{-\sqrt{a}|x|}$, and differentiating it $\frac{1}{2}(n - 3)$ times using Leibnitz[5] theorem, we may obtain a linear relation between the derivatives of order $\frac{1}{2}(n - 1)$ and lower; similarly, by differentiating $\frac{1}{2}(n - 5)$ times, we may obtain a linear relation between the derivatives of all orders up to and including $\frac{1}{2}(n - 3)$, and continuing in this way, we obtain a set of $\frac{1}{2}(n - 1)$ linear equations in the $\frac{1}{2}(n - 1)$ unknown derivatives. These equations may be solved for the derivative of order $\frac{1}{2}(n - 1)$ by the determinant rule. The denominator determinant is independent of x, and the numerator is $e^{-|x|\sqrt{a}}$ multiplied by a polynomial of degree $\frac{1}{2}(n - 1)$ in x. Using this fact, we may specify undetermined values for the coefficients in this polynomial, and obtain relations between these values for two consecutive values of n by differentiating once. The recurrence relations thus obtained may be used to verify by mathematical induction the following relation, after substituting $a = \sqrt{n}$,

$$\frac{1}{\sqrt{\pi n}} \frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)} \int_{-\infty}^{\infty} \frac{e^{itx}}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} dt \tag{2}$$

$$= e^{-|x|\sqrt{n}} \left[1 + |x|\sqrt{n} + \frac{(|x|\sqrt{n})^2}{2!}\frac{(n-3)}{(n-2)} + \frac{(|x|\sqrt{n})^3}{3!}\frac{(n-5)}{(n-2)} \cdots \right],$$

the coefficient of $(|x|\sqrt{u})^{2k}$ being

$$\frac{1}{(2k)!} \frac{(n-4k+1)(n-4k+3)(n-4k+5)\cdots(n-2k-1)}{(n-2)(n-4)(n-6)\cdots(n-2k)},$$

and the coefficient of $(|x|\sqrt{n})^{2k+1}$ being

$$\frac{1}{(2k+1)!} \frac{(n-4k-1)(n-4k+1)\cdots(n-2k-3)}{(n-2)(n-4)\cdots(n-2k)}.$$

This is, therefore, the value of the characteristic function, and is the same in form as the result which may be obtained by the first method. There are evidently a finite number of terms, the degree of the polynomial being $\frac{1}{2}(n-1)$.

Third, the characteristic function may be shown to satisfy the second order differential equation.

$$x\frac{d^2y}{dx^2} - (n-1)\frac{dy}{dx} - nxy = 0.$$

By change of variables $y = e^{-x\sqrt{n}}v$ (we assume that x is positive, as it may be replaced by its absolute value in the integral) and $u = x\sqrt{n}$, we obtain

$$u\frac{d^2v}{du^2} - \frac{dv}{du}(n-1+2u) + (n-1)v = 0.$$

Using the Frobenius method of solution in series, we obtain as one solution when n is odd the expression

$$v = 1 + u + \frac{u^2}{2!}\frac{(n-3)}{(n-2)} + \frac{u^3}{3!}\frac{(n-5)}{(n-2)} + \cdots$$

and the corresponding value of y has already been proved to be the value of the characteristic function. It is probable that the corresponding solution of the differential equation would also be the value of the characteristic function when n is even. In this case, however, the indicial equation has roots differing by an integer, and the solution of the differential equation is much more complicated in form. Nevertheless, it seems possible to find a series expansion for the characteristic function of "Student's" distribution in this way whatever be the value of n.

The characteristic functions of the distributions of at and bt' may now be readily obtained by replacing x by ax and bx in the above expression. Multiplying the characteristic functions of these two independent distributions, we obtain the characteristic function of the distribution of $at + bt'$, which is of the form

$$M(x) = e^{-|x|(|a|\sqrt{n}+|b|\sqrt{n})}\,[1 + |x|(|a|\sqrt{n} + |b|\sqrt{n'}) + \cdots],$$

the term in brackets being a polynomial of degree $\dfrac{(n + n' - 2)}{2}$. We may now use the result that the distribution is given by the integral

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\, M(x)\, dx,$$

and so obtain the distribution of $u = at + bt'$.

A distribution so obtained would involve four constants, a, b, n and n', and a derived probability table would thus be very complicated. It may, however, be simplified firstly by considering the case of equal sample numbers, and, secondly, making the transformation

$$v = \frac{(at + bt')}{|a| + |b|}, \tag{3}$$

whence the resulting distribution involves only two constants, n, and the ratio a/b. In this case the form of the characteristic function is

$$e^{-|x|\sqrt{n}}\left[1 + |x|\sqrt{n} + \frac{(|x|\sqrt{n})^2}{2!}\cdot\frac{(a^2 + b^2)\dfrac{n-3}{n-2} + 2|ab|}{(|a| + |b|)^2} + \cdots\right], \tag{4}$$

In determining the form of the distribution, we shall encounter integrals of the form

$$(n)^{\frac{1}{2}p}\int_{-\infty}^{\infty} e^{-|x|\sqrt{n}-ixv}\,|x|^p\,dx.$$

This can be reduced to

$$n^{\frac{1}{2}p}\int_0^{\infty} e^{-x\sqrt{n}-ixv}x^p\,dx + n^{\frac{1}{2}p}\int_0^{\infty} e^{-x\sqrt{n}+ixv}x^p\,dx,$$

and, integrating by parts, or using the Gamma Function integral, we obtain as the value of this integral

$$n^{\frac{1}{2}p}\,p!\left[\frac{1}{(\sqrt{n} + iv)^{p+1}} + \frac{1}{(\sqrt{n} - iv)^{p+1}}\right].$$

Writing $v = \sqrt{n}\tan\theta$, this reduces to

$$\frac{p!}{\sqrt{n}}\,2\cos(p + 1)\,\theta\,\cos^{p+1}\theta,$$

The distribution is thus seen to be:—

(5) $$\frac{1}{\pi}\,[p_0 + p_1 \cos 2\theta + p_2 \cos\theta \operatorname{cor} 3\theta + \cdots + p_{n-1} \cos^{n-2}\theta \cos n\theta]\, d\theta,$$

where

$$p_0 = 1,\ p_1 = 1,\ p_2 = \frac{(a^2 + b^2)\,\dfrac{n-3}{n-2} + 2\,|\,ab\,|}{(|\,a\,| + |\,b\,|)^2}, \ \ldots .$$

It is obvious that the values of the coefficients p may all be expressed in terms of the ratio $\left|\dfrac{a}{b}\right|$. Denoting this ratio by r,

$$p_2 = \frac{(r^2 + 1)\,\dfrac{n-3}{n-2} + 2r}{(r + 1)^2},$$

and thus we could construct a table for the probability integral involving n, r and v only.

The process of evaluating the probability integral may be simplified by considering the term already evaluated,

$$n^{\frac{1}{2}p} \int e^{-|x|\sqrt{n}-ixv}|\,x\,|^p\,dx.$$

Integrating this expression under the integral sign with respect to v, between the limits v and ∞, the contribution to the probability integral from this term is seen to be

$$\frac{n^{\frac{1}{2}p}}{i}\int_{-\infty}^{\infty} e^{-|x|\sqrt{n}-ixv}|\,x\,|^{p-1}\,dx,$$

which on the introduction of the same transformation as before, gives the value

$$-\,2\,(p-1)!\,\cos^p\theta \sin p\theta.$$

Thus, from (5), the total probability that θ should lie between $\frac{\pi}{2}$ and a given value, θ, is

(6) $$\frac{1}{\pi}\left[\frac{\pi}{2} - \theta - \cos\theta\sin\theta - \frac{p_2}{2}\cos^2\theta\sin 2\theta - \cdots - \frac{p_{n-1}}{n-1}\cos^{n-1}\theta\sin(n-1)\theta\right],$$

where $-\frac{\pi}{2} \leqq \theta \leqq \frac{\pi}{2}$.

The following summarises the results for small values of n.

1. $n = 1.\qquad \tan\theta = \dfrac{at + bt'}{|\,a\,| + |\,b\,|}.$

The results reduce to those already given by Fisher. The distribution is then simply $\frac{d\theta}{\pi}$, or Student's distribution, and is independent of a and b, and the probability integral is $\frac{1}{2} - \frac{\theta}{\pi}$.

2. $$n = 3. \qquad \tan\theta = \frac{at + bt'}{\sqrt{3}\,(|a| + |b|)}.$$

The distribution function is

$$\frac{d\theta}{\pi}\left[1 + \cos 2\theta + \frac{2r}{(1+r)^2}\cos\theta\cos 3\theta\right],$$

and the probability integral

$$\frac{1}{\pi}\left[\frac{\pi}{2} - \theta - \cos\theta\sin\theta - \frac{r}{(1+r)^2}\cos^2\theta\sin 2\theta\right].$$

3. $$n = 5. \qquad \tan\theta = \frac{at + bt'}{\sqrt{5}(|a| + |b|)}.$$

The distribution function is

$$\frac{d\theta}{\pi}\left[1 + \cos 2\theta + \frac{2}{3}\frac{(r^2 + 1 + 3r)}{(1+r)^2}\cos\theta\cos 3\theta + \frac{2r}{(1+r)^2}\cos^2\theta\cos 4\theta + \frac{8r^2}{3(1+r)^2}\cos^3\theta\cos 5\theta\right],$$

and the probability integral

$$\frac{1}{\pi}\left[\frac{\pi}{2} - \theta - \cos\theta\sin\theta - \frac{1}{3}\frac{(r^2 + 1 + 3r)}{(1+r)^2}\cos^2\theta\sin 2\theta + \frac{2r}{3(1+r)^2}\cos^3\theta\sin 3\theta - \frac{2r^2}{3(1+r)^4}\cos^4\theta\sin 4\theta\right].$$

3. **Samples of large numbers.** The foregoing method is not suitable when n and n' are large. In this case we use the asymptotic expansion of "Student's" distribution which has been worked out by R. A. Fisher[7] and is of the form,

$$f(t)\,dt = \frac{1}{\sqrt{\pi n}}\frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)}\frac{dt}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}} \tag{7}$$
$$\sim \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\,dt\left(1 + \frac{P_1}{n} + \frac{P_2}{n^2} + \cdots + \frac{P_k}{n^k} + \cdots\right),$$

[7] "The Expansion of 'Student's' Distribution in Powers of n^{-1}," *Metron*, Vol. 5, no. 3 (1925), pp. 22–25.

where P_k is a polynomial of degree $4k$ in t, such that

$$P_1 = \frac{t^4 - 2t^2 - 1}{4}, \qquad P_2 = \frac{3t^8 - 28t^6 + 30t^4 + 12t^2 + 3}{96}, \quad \text{etc.}$$

The development of an asymptotic expansion for the distribution of $at + bt'$ is obtained by combining the asymptotic expansions of t and t'. The theoretical justification of the process used makes use of the following lemma:—

If $R_k(t)$ is the remainder after the first $(k + 1)$ terms in the asymptotic expansion of "Student's" distribution in descending powers of n, then $\lim_{n \to \infty} n_k \int_0^\infty | R_k(t) | \, dt = 0.$

In the proof, the symbol "lim" will be used to denote the limit as n tends to infinity of the quantity in question. Let $S_k(t)$ represent the sum of the first $(k + 1)$ terms of the above expansion. It may readily be shown that if $0 < \delta < \frac{1}{2}$,

$$\lim n^k \int_{n^\delta}^\infty f(t) \, dt = 0,$$

and hence that

$$\lim n^k \int_{n^\delta}^\infty | R_k(t) | \, dt = 0.$$

Using an expansion for the logarithm of $\frac{1}{\sqrt{\pi n}} \frac{1}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ and the asymptotic expansion for the logarithm of the Gamma function, the following asymptotic expansion may be obtained, $\log f(t) = -\frac{1}{2} \log 2\pi - \frac{1}{2}t^2 + w$, where

$$(8) \qquad w = \frac{1}{4n}(t^4 + 2t^2 - 1) + \frac{1}{12n^2}(-2t^6 + 3t^4) + \cdots + \frac{G_{2p+2}}{p(p+1)n^p} + \tfrac{1}{2}T_{p+1} - \frac{n+1}{2} R_p + R_p',$$

G_{2p+2} being a polynomial of degree $2p+2$, and

$$| R_p | < \frac{t^{2p+4}}{(p+2)n^{p+2}} = \frac{\alpha t^{2p+4}}{(p+2)n^{p+2}}, \qquad \text{where } 0 < \alpha < 1,$$

$$| T_{p+1} | = \frac{t^{2p+2}}{(p+1)n^{p+1}}$$

$$| R_p' | < \frac{2^{p+3}A}{(p+1)(p+2)n^{p+1}} = \frac{\beta 2^{p+3}A}{(p+1)(p+2)n^{p+1}}, \qquad \text{where } 0 < \beta < 1,$$

A being a constant independent of n.

Thus, using Taylor's expansion, we obtain

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left(1 + w + \frac{w^2}{2!} + \frac{w^3}{3!} + \cdots + \frac{w^k}{k!} + \frac{w^{k+1}}{(k+1)!} e^{\theta w}\right),$$

where $0 < \theta < 1$.

Evidently $R_k(t)$ is of the form

$$(9) \qquad \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{1}{2}t^2} \left(\frac{q_{k+1}}{n^{k+1}} + \frac{q_{k+2}}{n^{k+2}} + \cdots + \frac{q_{k(p+2)}}{n^{k(p+2)}} + \frac{w^{k+1}}{(k+1)!} e^{\theta w} \right) \right],$$

the quantities q being polynomials in t.

Using the moments of the normal distribution, it may readily be shown that

$$\lim n^k \int_0^{n^\delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left(\frac{q_{k+1}}{n^{k+1}} + \frac{q_{k+2}}{n^{k+2}} + \cdots + \frac{q_{k(p+2)}}{n^{k(p+2)}} \right) dt = 0.$$

In the range of integration, when n is sufficiently large, it is evident that

$$|w| < \frac{2n^{4\delta}}{n} + \frac{2n^{6\delta}}{n^2} + \cdots + \frac{2n^{(2p+2)\delta}}{n^p} + \frac{n+1}{2} |R_p| + |R_p'| + |T_{p+1}| = O(n^{4\delta-1}) \quad \text{if } 0 < \delta < \tfrac{1}{4}.$$

Thus

$$\left| \frac{w^{k+1}}{\sqrt{2\pi}(k+1)!} e^{\theta w} e^{-\frac{1}{2}t^2} \right| < \frac{Kn^\delta}{n^{1-4\delta-4K\delta+K}} e^{\frac{\theta K'}{n^{1-4\delta}}}, \quad \text{where } K \text{ and } K' \text{ are constants.}$$

and hence

$$\lim n^k \int_0^{n^\delta} |R_k(t)|\, dt < \lim \frac{K}{n^{1-5\delta-4K\delta}} e^{\frac{\theta K'}{n^{1-4\delta}}} = 0, \quad \text{if } \delta < \frac{1}{5+4k}.$$

We can also deduce the following results:

1. Since the value of the integrand is unaltered if t is replaced by $-t$, we have at once

$$\lim n^k \int_{-\infty}^0 |R_k(t)|\, dt = 0.$$

2. Using both of these results it follows that

$$\lim n^k \int_{-\infty}^{\infty} |R_k(t)|\, dt = 0.$$

Hence

$$3. \qquad \lim n^k \int_{t'}^{t} |R_k(t)|\, dt = 0,$$

where t and t' have any real values, and thus it is legitimate to integrate the asymptotic expansion of $f(t)$ term by term with respect to t between any given limits.

4. If $\phi(t)$ is a function independent of n which is bounded for all values of t, the asymptotic expansion of $f(t)\phi(t)$ in terms of n may be integrated term by term with respect to t. In particular, if $\phi(t) = e^{itx}$, an asymptotic expansion for the characteristic function of "Student's" distribution may be obtained.

We may now consider the form of the distribution of $at + bt'$, and in order to simplify the argument, the following reasoning applies to the case in which the sample numbers are equal, although a similar theory may be developed for sample numbers which are of equal orders of magnitude. We may write

$$f(t) = S_k(t) + R_k(t),$$
$$f(t') = S_k(t') + R_k(t'),$$
$$u = at + bt',$$

and hence $t' = \dfrac{u - at}{b}$. The joint distribution of u and t is therefore

$$\left[S_k(t)S_k\left(\frac{u-at}{b}\right) + R_k(t)S_k\left(\frac{u-at}{b}\right) + S_k(t)R_k\left(\frac{u-at}{b}\right) + R_k(t)R_k\left(\frac{u-at}{b}\right)\right]\frac{dt\,du}{b}.$$

The distribution of u is obtained by integrating this expression with respect to t over all the possible values of t between $-\infty$ and $+\infty$. The product $S_k(t)S_k\left(\dfrac{u-at}{b}\right)$ gives the first $k+1$ terms in the asymptotic expansion which is the product of the asymptotic expansions of $f(t)$ and $f\left(\dfrac{u-at}{b}\right)$, and a remainder $\phi(t)$, where

$$\phi(t) = e^{-\frac{1}{2}t^2 - \frac{1}{2}\left(\frac{u-at}{b}\right)^2}\left(\frac{v_1}{n^{k+1}} + \frac{v_2}{n^{k+2}} + \cdots + \frac{v_k}{n^{2k}}\right), \tag{10}$$

$v_1, v_2, \cdots v_k$ being polynomials in t. Let

$$R_k'(t) = \phi(t) + R_k(t)S_k\left(\frac{u-at}{b}\right) + R_k\left(\frac{u-at}{b}\right)S_k(t) + R_k(t)R_k\left(\frac{u-at}{b}\right).$$

Using the expressions for the moments of the normal distribution, it may be shown that $\int_{-\infty}^{\infty} |\phi(t)|\,dt = O\left(\dfrac{1}{n^{k+1}}\right)$. Let the upper bound of the bounded function $S_k(t)$ for all values of n and t be B. Then

$$\int_{-\infty}^{\infty}\left|S_k\left(\frac{u-at}{b}\right)R_k(t)\right|dt < B\int_{-\infty}^{\infty}|R_k(t)|\,dt$$
$$= o(n^{-k}).$$

Similarly

$$\int_{-\infty}^{\infty} \left| S_k(t) R_k\left(\frac{u - at}{b}\right) \right| dt = o(n^{-k}).$$

and

$$\int_{-\infty}^{\infty} \left| R_k(t) R_k\left(\frac{u - at}{b}\right) \right| dt = o(n^{-k}).$$

Thus

$$\lim n^k \int_{-\infty}^{\infty} | R_k'(t) | \, dt = 0$$

and hence the distribution of u may be obtained by integrating the asymptotic expansion which is the product of the asymptotic expansions of $f(t)$ and $f\left(\frac{u - at}{b}\right)$ term by term.

In practice, it is convenient to find the distribution of

$$w = \frac{at + bt'}{\sqrt{a^2 + b^2}}. \tag{11}$$

We substitute $y = \dfrac{bt - at'}{\sqrt{a^2 + b^2}}$ and, using the above result, it follows that the distribution of w is given by

$$\frac{dw}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \left\{ 1 + \frac{1}{4n}\left[\frac{(aw + by)^4}{(a^2 + b^2)^2} - 2\,\frac{(aw + by)^2}{a^2 + b^2} + \frac{(bw - ay)^4}{(a^2 + b^2)^2} - 2\,\frac{(bw - ay)^2}{(a^2 + b^2)} - 2 \right] + \cdots \right\} dy$$

which is equal to

$$\frac{dw}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} \left\{ 1 + \frac{1}{4n}\left[\frac{w^4(a^4 + b^4) + 12w^2a^2b^2 + 3(a^4 + b^4)}{(a^2 + b^2)^2} - 4 - 2w^2 \right] + \cdots \right\}. \tag{12}$$

It may be noticed that this distribution may be expressed in terms of the ratio a/b only. The probability integral may readily be obtained. There is no theoretical difficulty involved in obtaining any desired number of terms of this expansion, but they rapidly become too complicated to handle with any ease. Moreover, it is difficult to find a limit of the error committed in using any given number of terms of the series for the probability integral as an approximation to the value of this integral, as the somewhat complicated method of obtaining the series masks the form of the remainder. While it is undoubtedly true that when n is very large the distribution approaches normality, and for a somewhat lower range of values of n the first two terms of the expansion should be taken, etc., it is difficult to forecast the number of terms which should be retained for any given value of n. In fact the same difficulty seems to exist when using the original asymptotic expansion of "Student's" distribution for the calculation of probabilities. For instance, the coefficients of the powers of t which occur in the sixth term of the asymptotic expansion of the probability integral are larger than those occurring in the fifth term, and, in consequence, in spite of the greater power of n in the denominator, for certain values of n these may contribute more to the probability integral than the previous term. We are unable to say anything about the aggregate of succeeding terms in general, and, therefore, it does not seem legitimate to drop all the terms following a term which yields a contribution beyond the limit of accuracy desired. This difficulty is even more apparent in the case in which the coefficients of the various powers of t occurring in the terms beyond the first involve also the ratio a/b, and it is probable that the different values of this ratio which are possible would lead to different numbers of terms being taken for the same value of n in order to gain the same degree of precision in the probability integral.

4. **The distributions of the test quantities which correspond to (3) and (11) for equal means, when the ratio of the variances is a known quantity.** When the ratio ϕ of the variances is given, the foregoing arguments, which are independent of the parameters specifying the distribution, may no longer be applied, for this would be information not supplied by the sample. In this case, the distributions of the test quantities which have been used take forms which depend only on the ratio of the variances, and are independent of the sample estimates of the variances.

The quantity (3), used in §2, when n was a small odd number, was $\dfrac{\bar{x} - \bar{x}'}{s + s'} = v$, where $s^2 = \dfrac{\Sigma(x - \bar{x})^2}{N(N - 1)}$, $s'^2 = \dfrac{\Sigma(x' - \bar{x}')^2}{N(N-1)}$ and $n = N - 1$. On the assumption of equal population means, the distribution of this quantity takes the form

$$\frac{2\Gamma(n + \frac{1}{2})\, dv}{\Gamma\left(\frac{n}{2}\right)^2 \sqrt{n\pi}\,(1 + \phi)^{n-1}} \int_{-\frac{1}{\sqrt{\phi}}}^{\sqrt{\phi}} \frac{(\sqrt{\phi}z + 1)^{n-1}(\sqrt{\phi} - z)^{n-1}}{\left(z^2 + 1 + \frac{v^2}{n}\right)^{n+\frac{1}{2}}}\, dz. \tag{13}$$

Thus in the case $n = 1$, we obtain

$$\frac{dv}{\pi(1 + v)^2}\left[\frac{\sqrt{\phi}}{(v^2 + 1 + \phi)^{\frac{1}{2}}} + \frac{1}{(\phi v^2 + \phi + 1)^{\frac{1}{2}}}\right],$$

which is the result given by R. A. Fisher.[6] The integral may be evaluated in terms of elementary functions for small odd integral values of n.

In §3, (11), the distribution of the statistic $w = \dfrac{\bar{x} - \bar{x}'}{\sqrt{s^2 + s'^2}}$ was considered when N was large. The exact distribution may be proved to be

$$(14)\quad \phi^{\frac{n}{2}+\frac{1}{2}} \frac{\Gamma(n+\frac{1}{2})n^n(1+\phi)^n}{\sqrt{\pi}\,(w^2\phi+n(1+\phi))^{n+\frac{1}{2}}\Gamma(n)}\,dw\cdot F\left(n+\tfrac{1}{2}, \tfrac{1}{2}n, n, \frac{n(1-\phi^2)}{vw^2\phi+n(1+\phi)}\right)$$

where F is the hypergeometric function. If $\phi = 1$, we have the limiting case in which the argument of the hypergeometric function is zero, and obtain "Student's" distribution, which is to be expected in view of the evidence stated in §1, the numbers in the samples being equal.

COLUMBIA UNIVERSITY.