THE BEHAVIOR OF SOME STANDARD STATISTICAL TESTS UNDER NONSTANDARD CONDITIONS

HAROLD HOTELLING
UNIVERSITY OF NORTH CAROLINA

1. Introduction and summary

In the burst of new statistical developments that has followed the work of Sir Ronald Fisher in introducing and popularizing methods involving exact probability distributions in place of the old approximations, questions as to the effects of departures from the assumed normality, independence, and uniform variance have often been subordinated. It is true that much recognition has been given to the existence of serial correlation in time series, with the resultant vitiation of statistical tests carried out in the absence of due precautions, and to some other special situations, such as manifestly unequal variances in least-square problems. Wassily Hoeffding has established some general considerations on the role of hypotheses in statistical decisions [21]. Also, there have been many studies of distributions of the Student ratio, the sample variance, variance ratio, and correlation coefficient in samples from nonnormal populations. (Some are cited at the end of this paper.) These efforts have encountered formidable mathematical difficulties, since the distribution functions sought cannot usually, except in trivial cases, be easily specified or calculated in terms of familiar or tabulated functions. Because of these difficulties, mathematics has in some such studies been supplemented or replaced by experiment ([20], [38], [54], and others); or, as in some important work, approximations for which definite error bounds are not at hand have been used.

Practical statisticians have tended to disregard nonnormality, partly for lack of an adequate body of mathematical theory to which an appeal can be made, partly because they think it is too much trouble, and partly because of a hazy tradition that all mathematical ills arising from nonnormality will be cured by sufficiently large numbers. This last idea presumably stems from central limit theorems, or rumors or inaccurate recollections of them.

Central limit theorems have usually dealt only with linear functions of a large number of variates, and under various assumptions have proved convergence to normality as the number increases. For a large but fixed number the approxima-

This research was carried out in part under Office of Naval Research Project NR 042 031; Contract Nonr-855(06) and its predecessor contract.

tion of the distribution to normality is typically close within a restricted portion of its range, but bad in the tails. Yet it is the tails that are used in tests of significance, and the statistic used for a test is seldom a linear function of the observations. The nonlinear statistics t, F, and r in common use can be shown ([9], p. 266) under certain wide sets of assumptions to have distributions approaching the normal form, though slowly in the tails. But in the absence of a normal distribution in the basic population, with independence, there is a lack of convincing evidence that the errors in these approximations are less than in the nineteenth-century methods now supplanted, on grounds of greater accuracy, by the new statistics.

In tails corresponding to small probabilities the behavior of t, r, and various related statistics may be examined with the help of a method of tubes and spheres proposed in [10], [11], and [12], and applied by Daisy Starkey in 1939 to periodogram analysis [57]; by Bradley [4], [5], [6] to t, F, and multivariate T; by E. S. Keeping [29] to testing the fit of exponential regressions; and by Siddiqui [55], [56] to the distribution of a serial correlation coefficient.

The distribution of the Student-Fisher t, as used for testing deviations of means in samples from various nonnormal populations, and in certain cases of observations of unequal variances and intercorrelations, will be examined in the next few sections, with special reference to large values of t. New exact distributions of t will be found for a few cases. A limit for increasing t, with wide applicability, will be obtained for the ratio of two probabilities of t being exceeded, one probability for the standard normal theory with independence on which the tables in use are based, the other pertaining to various cases of nonnormality, dependence, and heteroscedasticity. This limit depends on the sample size, which as it increases provides a double limit for the ratio. This double limit, for random samples from a nonnormal population, is for many commonly considered populations either zero or one. After considering the accuracy of the first limit and exact distributions of t in certain cases of small samples, we obtain the conditions on the population that the double limit be unity, a situation favorable to the use of the standard table. The nature of the conditions leads to certain remarks on proposals to modify the t-test for nonnormality.

The marked effects of nonnormality on the distributions of sample standard deviations and correlation coefficients will then be shown. The paper will conclude with a discussion of the dilemmas with which statisticians are confronted by the anomolous behavior in many cases of statistical methods that have become standard, and of possible means of escape from such difficulties.

2. Geometry of the Student ratio and projections on a sphere

To test the deviation of the mean $\bar{x}$ of observations $x_1, \cdots, x_N$ from a hypothetical value, which is here assumed without loss of essential generality to be zero, the statistic appropriate in case the observations constitute a random sample from a normal population, that is, are independently distributed with the same normal distribution, is

$$t = \frac{\bar{x}N^{1/2}}{s}, \tag{2.1}$$

where s is the positive square root of

$$s^2 = \frac{S(x - \bar{x})^2}{n} = \frac{Sx^2 - N\bar{x}^2}{n} = \frac{Sx^2 - \bar{x}Sx}{n}. \tag{2.2}$$

where S stands for summation over the sample, $\bar{x} = Sx/N$, and $n = N - 1$. We shall continue to adhere to Fisher's convention that S shall denote summation over a sample and $\sum$ other kinds of summation.

In space of Cartesian coordinates $x_1, \cdots, x_N$ let O be the origin and X a random point whose coordinates are the sample values. Let θ be the angle, between O and π inclusive, made by OX with the positive extension of the equiangular line, that is, the line on which every point has all its coordinates equal. It will soon be shown that

$$t = n^{1/2} \cot \theta. \tag{2.3}$$

We deal only with population distributions specified everywhere by a joint density function. This we call $f(x_1, x_2, \cdots, x_N)$. On changing to spherical polar coordinates by means of the N equations $x_i = \rho\xi_i$, where $\xi_1, \cdots, \xi_N$ are functions of angular coordinates $\phi_1, \cdots, \phi_{N-1}$ and satisfy identically

$$S\xi^2 = \xi_1^2 + \cdots + \xi_N^2 = 1, \tag{2.4}$$

while the radius vector ρ is the positive square root of

$$\rho^2 = x_1^2 + \cdots + x_N^2; \tag{2.5}$$

the Jacobian by which f is multiplied is of the form $J = \rho^{N-1}H(\phi_1, \cdots, \phi_{N-1})$. Thus the element of probability is

$$f(x_1, \cdots, x_N)\, dx_1 \cdots dx_N = f(\rho\xi_1, \cdots, \rho\xi_N)\rho^{N-1}\, d\rho\, dS, \tag{2.6}$$

where $dS = H(\phi_1, \cdots, \phi_N)\, d\phi_1, \cdots, d\phi_N$ is the element of area on the unit sphere.

Let the random point X be projected onto the unit sphere by the radius vector OX into the unique point ξ on the same side of O as X. The Cartesian coordinates of ξ are $\xi_1, \cdots, \xi_N$, while

$$D_N(\xi) = \int_0^\infty f(\rho\xi_1, \cdots, \rho\xi_N)\rho^{N-1}\, d\rho, \tag{2.7}$$

in accordance with (2.6).

Any statistic that is a continuous function of the N observations, that is, of the coordinates throughout a region of the n-space, determines through each point of this region what Fisher has called an *isostatistical surface*, on all points of which the statistic takes the same constant value. If, as in the cases of t, F, and r, the statistic remains invariant when all observations are multiplied by any one positive constant, the isostatistical surface is a cone with vertex at the origin, or a nappe of such a cone. The probability that the statistic lies between specified values is the integral of (2.7) over the portion of the unit sphere between its intersections with the corresponding cones.

Since t is unaffected by the division of each x_i by ρ, and therefore by its replacement by the corresponding ξ_i, we may rewrite (2.1) and (2.2) in terms of the ξ_i and then simplify with the help of (2.4). We also use the notation

$$a = S\xi_i = \xi_1 + \cdots + \xi_N \tag{2.8}$$

and obtain

$$t = a(N - a^2)^{-1/2}(N - 1)^{1/2}. \tag{2.9}$$

Thus the locus on the unit sphere for which t is constant is also a locus on which $S\xi$ has the constant value a. This locus is therefore a subsphere of intrinsic dimensionality $N - 2 = n - 1$, that of the original unit sphere being $n = N - 1$.

The distance between O and the hyperplane (2.8) is $aN^{-1/2}$, and also equals $\cos\theta$. From these facts (2.3) easily follows.

The probability that t is greater than a constant T, which we shall write $P\{t > T\}$, is the integral of (2.7) over a spherical cap. In the case of independent observations all having the same central normal distribution, a case for which we attach a star to each of the symbols f, D, and P already introduced, we readily find

$$\begin{aligned} f^*(x_1, \cdots, x_N) &= (2\pi)^{-N/2}e^{-Sx^2/2}, \\ D_N^* = (2\pi)^{-N/2}\int_0^\infty e^{-\rho^2/2}\rho^{N-1}\,d\rho &= \frac{1}{2}\pi^{-N/2}\Gamma\left(\frac{N}{2}\right), \end{aligned} \tag{2.10}$$

independently of ξ.

The constancy of probability density over the sphere in this case, with the necessity that the integral over the whole sphere must in every case be unity, provides a ready derivation of the $(N - 1)$-dimensional "area" of a unit sphere, which is at once seen to be D_N^{*-1}, that is

$$\frac{2\pi^{N/2}}{\Gamma\left(\frac{N}{2}\right)}. \tag{2.11}$$

This is one of several contributions of statistics to geometry arising in connection with contributions of geometry to statistics.

The N-dimensional "volume" enclosed by such a sphere is found by integrating with respect to r from zero to one the product of (2.11) by r^{N-1}, and therefore equals

$$\frac{\pi^{N/2}}{\Gamma\left(\frac{N}{2} + 1\right)}. \tag{2.12}$$

Evaluation of the probability of t being exceeded in this standard normal case is equivalent to finding the area of a spherical cap, a problem solved by Archimedes for the case $N = 3$. A geometrical solution in any number of dimensions may be obtained by noticing that the subsphere of $N - 2 = n - 1$ dimensions considered above for a particular value of t has radius $\sin \theta$ and that its "area" is proportional to the $(n - 1)$st power of this radius. By considering a "zone" of breadth $d\theta$ obtained by differentiating (2.3), the Student distribution may be reached in the form obtained by Fisher [15] in 1926 by quite different reasoning and for a broader class of uses,

$$g^*(t)\,dt = (\pi n)^{-1/2}\Gamma\left[\frac{1}{2}(n+1)\right]\left[\Gamma\left(\frac{n}{2}\right)\right]^{-1}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} dt. \tag{2.13}$$

The integral of this from T to ∞ is the probability $P\{t > T\}$. This must of course be doubled to get the usual two-tailed probability.

3. Probabilities of large values of t in nonstandard cases

The probability $P\{t > T\}$ when $T > 0$ is the integral of the n-dimensional density $D_N(\xi)$ over a spherical cap. This cap is the locus of points on the unit sphere in N-space whose distance from the equiangular line is less than $\sin \theta$, where $T = n^{1/2} \cot \theta$ and $0 \leqq \theta < \pi/2$. Central to such a cap is the point A on the unit sphere at which all the Cartesian coordinates are equal and positive, and are therefore all equal to $N^{-1/2}$. The projected density (2.7) thus takes at A the value

$$D_N(A) = \int_0^\infty f(\rho N^{-1/2}, \cdots, \rho N^{-1/2})\rho^{N-1}\,d\rho. \tag{3.1}$$

A similar situation holds at the point A' diametrically opposite to A, whose coordinates are all $N^{-1/2}$.

We consider the approximation to $P\{t < T\}$ consisting of the product of the n-dimensional area of the cap centered at A by the density at A.

If corresponding to two different population distributions both the D_N functions are continuous, the limit of the ratio of these as T increases, if it exists, equals the limit of the ratios of the probabilities in the two cases of the same event $t > T$. This also equals the limit of the ratio of the two probability densities of t when t increases, provided that this last limit exists, as L'Hospital's rule shows.

We are particularly concerned with comparing $P\{t > T\}$ for various populations with the probability of the same event for a normally distributed population with zero mean and a fixed variance. Using a star attached to P or D to identify this case of normality, we introduce for each alternative population distribution for which the indicated limits exist the following notation:

$$R_N = \lim_{T\to\infty} \frac{P\{t > T\}}{P^*\{t > T\}} = \frac{D_N(A)}{D_N^*(A)} = \frac{2\pi^{N/2} D_N(A)}{\Gamma\left(\frac{N}{2}\right)}, \tag{3.2}$$

in accordance with (2.10). By L'Hospital's rule,

$$\lim_{t\to\infty} \frac{g(t)}{g^*(t)}, \tag{3.3}$$

where $g(t)$ is the probability density of t in the nonstandard case and $g^*(t)$ is given by the standard t density (2.13). We also put

$$R = \lim_{N\to\infty} R_N. \tag{3.4}$$

If R_N is less than unity, a standard one-tailed t-test of significance on a sufficiently stringent level, that is, corresponding to a sufficiently small probability of rejecting a null hypothesis because t exceeds a standard level given in tables of the Student distribution, will in samples of N overstate the probability of the Type I error of wrongfully rejecting the null hypothesis. If $R_N > 1$, the probability of such an error is understated. Such biases exist even for very large samples if $R \neq 1$.

Negative values of t may be treated just as are positive ones above, replacing A by the diametrically opposite point A', with only trivially different results. For symmetrical two-tailed t-tests, $D_N(A)$ is to be replaced by its average with $D_N(A')$.

In random samples of N from a population with density $f(x)$, that is, with independence, (3.1) takes the form

$$D_N(A) = \int_0^\infty [f(\rho N^{-1/2})]^N \rho^{N-1}\, d\rho = N^{N/2} \int_0^\infty [f(z)]^N z^{N-1}\, dz. \tag{3.5}$$

Substituting this in (3.2) yields

$$R_N = 2(\pi N)^{N/2} \left[\Gamma\left(\frac{N}{2}\right)\right]^{-1} \int_0^\infty [f(z)]^N z^{N-1}\, dz. \tag{3.6}$$

4. Samples from a Cauchy distribution

When the Cauchy density

$$f(x) = \pi^{-1}(1 + x^2)^{-1} \tag{4.1}$$

is substituted in (3.5) the integral may be reduced to a beta function by substitution of a new letter for $z^2/(1 + z^2)$, and we find

$$R_N = \frac{\left(\frac{N}{\pi}\right)^{N/2} \Gamma\left(\frac{N}{2}\right)}{\Gamma(N)}. \tag{4.2}$$

For large samples we find with the help of Stirling's formula that R_N approximates

$$2^{1/2}\left(\frac{e}{2\pi}\right)^{N/2}, \tag{4.3}$$

and thus approaches zero. For $N = 2, 3, 4$, the respective values of R_N are

$$(4.4) \qquad .6366, \quad .4134, \quad .2703.$$

The fact that these are markedly less than unity points to substantial overestimation of probabilities of getting large values of t when a population really of the Cauchy type is treated as if it were normal.

A desirable but difficult enterprise is to provide for these approximations, which are limits as t increases to infinity, definite and close upper and lower bounds for a fixed t. This can be done for samples of any size from a Cauchy distribution by a method that will be outlined below, but the algebra is lengthy except for very small samples. The case of the double exponential, or Laplace, distribution is more tractable, as will be seen later.

For $N = 2$ the exact distribution of t, which then equals $(x_1 + x_2)/|x_1 - x_2|$, is readily found directly by change of variables in $f(x_1)f(x_2)$ and integration. On the basis of the Cauchy population (4.1) this yields

$$(4.5) \qquad \frac{1}{2\pi^2 t} \log \left(\frac{t+1}{t-1}\right)^2 dt.$$

To prove this, we observe that $|t|$ has the same distribution as $|u|$, where $u = (x + y)/(x - y)$ and the joint distribution of x and y is specified by $\pi^{-2}(1 + x^2)^{-1}(1 + y^2)^{-1}\, dx\, dy$.

Substituting $x = y(u + 1)/(u - 1)$, $dx = -2y(u - 1)^{-2}\, du$, we have for the joint distribution of y and u,

$$(4.6) \qquad \frac{2|y|\, dy\, du}{\pi^2(1 + y^2)[(u - 1)^2 + (u + 1)^2 y^2]}.$$

In spite of a singularity at the point $u = 1$, $y = 0$, this may be integrated with respect to y over all real values. The result, when t is substituted for u, is (4.5).

The density given by (4.5) is infinite for $t = \pm 1$, but is elsewhere continuous, and is integrable over any interval. Upon expanding in a series of powers of t^{-1} and integrating, it is seen that the probability that any $t > 1$ should be exceeded in absolute value is given by the series

$$(4.7) \qquad \frac{4}{\pi^2}\left(\frac{1}{t} + \frac{1}{3^2 t^3} + \frac{1}{5^2 t^5} + \cdots\right),$$

which converges even for $t = 1$, and then takes the same value 1/2 as the corresponding probability for $t = 1$ based on a fundamentally normal population, for which the distribution (2.13) becomes

$$(4.8) \qquad \frac{dt}{\pi(1 + t^2)}.$$

The formula $\sum_1^\infty n^{-2} = \pi^2/8$ used to sum (4.7) may be derived by integrating from 0 to $\pi/2$ the well known Fourier series $\sin x + (\sin 3x) + (\sin 5x)/5 + \cdots = \pi/4$.

Expansion of (4.8) in powers of t^{-1} and integration gives for the probability corresponding to (4.7),

$$\frac{2}{\pi}\left(\frac{1}{t} - \frac{1}{3t^3} + \frac{1}{5t^5} - \cdots\right). \tag{4.9}$$

The ratio of (4.7) to (4.9), or of (4.5) to (4.8), approaches $2/\pi$ when t^2 increases, agreeing with the result for $N = 2$ in (4.4), but exceeds this limit for finite values of t.

The five per cent point for t based on normal theory, that is, the value of t for which the probability (4.9) equals .05, is 12.706. When this value of t is substituted in (4.7) the resulting exact probability is found to be .0218. The approximation obtained by multiplying R_2 from (4.4) by the originally chosen probability .05 is .0318.

The true five per cent point when the underlying population is of the Cauchy form is 8.119.

When we go farther out in the tail we may expect a better approximation in using the product of R_2 by the "normal theory" probability to estimate the true "Cauchy" theory probability, and indeed this is soon verified: the one per cent point from standard normal theory tables for $n = 1$, that is, $N = 2$ in the present case, is 63.657. The approximation to the true Cauchy probability obtained on multiplying the chosen "normal theory" probability .01 by R_2 is .006366. The true value given by (4.7) is .006364. Thus the approximations by the R_2 method are rough at the five per cent point, but quite satisfactory at the one per cent point or beyond.

5. Accuracy of approximation. Inequalities

Various inequalities may be obtained relevant to the accuracy of the approximations that must usually be used to estimate the probabilities associated with standard statistical tests under nonstandard conditions.

For the Cauchy distribution and others similarly expressible by rational functions, and some others, the theorem that the geometric mean of positive quantities cannot exceed the arithmetic mean can be brought into play. Thus when the n-dimensional density $D(\xi)$ on the unit sphere is specified by substituting in (2.7) for f the product of N Cauchy functions (4.1) with different arguments, the result,

$$D(\xi) = \pi^{-N} \int_0^\infty [(1 + \rho^2\xi_1^2) \cdots (1 + \rho^2\xi_N^2)]^{-1}\rho^{N-1}\, d\rho, \tag{5.1}$$

in which $S\xi^2 = 1$ and $S\xi = a$, can be shown to exceed $D(A)$, in which each ξ^2 in (5.1) is replaced by N^{-1} and which is given by (3.5). This readily follows from the fact that the product within the square brackets in (5.1) is the Nth power of the geometric mean of these binomials, and is therefore not greater than the Nth power of their mean $1 + \rho^2/N$. Since the density on the sphere thus takes a minimum value at A, the estimate of the tail probability obtained by multiplying the area of the polar cap by the density at its center A, equivalent to multiplying the tail probability given by "normal theory" by the factors R_N

given in (4.2) and (4.4), is an underestimate. A numerical example of this is at the end of section 4.

Another kind of inequality may be used to set upper bounds on the difference between the probability densities on the sphere at different points, and from these may be derived others setting lower bounds. Consider for example samples of three from a Cauchy population, and instead of (5.1) use the notation D_{000}. This will be changed to D_{100} if ξ_1 is changed to a new value ξ_1' while ξ_2 and ξ_3 remain unchanged. In relation to a new set of values ξ_1', ξ_2', ξ_3' and the old set ξ_1, ξ_2, ξ_3, we shall use D with three subscripts, each subscript equal to 0 if the corresponding argument is an old ξ, or to 1 if it is a new ξ'. Now

(5.2)

$$D_{100} - D_{000} = \pi^{-N}(\xi_1^2 - \xi_1'^2) \int_0^\infty [(1 + \rho^2\xi_1'^2)(1 + \rho^2\xi_1^2) \cdots (1 + \rho^2\xi_N^2)]^{-1}\rho^{N+1}\, d\rho.$$

This difference is of the same sign as $\xi_1^2 - \xi_2^2$. The integrand on the right is the product of that of D_{000} by $\rho^2/(1 + \rho^2\xi_1'^2)$. This last factor is less than $\xi_1'^{-2}$ for all positive values of ρ. Thus we find

$$|D_{100} - D_{000}| \leqq \left|\frac{\xi_1^2}{\xi_1'^2} - 1\right| D_{000}. \tag{5.3}$$

The equality holds only if $\xi_1^2 = \xi_1'^2$. Put $\gamma_k = |\xi_k^2/\xi_k'^2 - 1|$. Then

$$|D_{100} - D_{000}| \leqq \gamma_1 D_{000}, \tag{5.4}$$

whence

$$(1 - \gamma_1)D_{000} \leqq D_{100} \leqq (1 + \gamma_1)D_{000}. \tag{5.5}$$

In the same way,

$$\begin{aligned} &|D_{110} - D_{100}| \leqq \gamma_2 D_{100}, \qquad (1 - \gamma_2)D_{100} \leqq D_{110} \leqq (1 + \gamma_2)D_{100}, \\ &|D_{111} - D_{110}| \leqq \gamma_3 D_{110}, \qquad (1 - \gamma_3)D_{110} \leqq D_{111} \leqq (1 + \gamma_3)D_{110}. \end{aligned} \tag{5.6}$$

Combining these results gives

$$(1 - \gamma_1)(1 - \gamma_2)(1 - \gamma_3)D_{000} \leqq D_{111} \leqq (1 + \gamma_1)(1 + \gamma_2)(1 + \gamma_3)D_{000}, \tag{5.7}$$

and

$$\begin{aligned} |D_{111} - D_{000}| &\leqq \gamma_1 D_{000} + \gamma_2 D_{100} + \gamma_3 D_{110} \\ &\leqq [\gamma_1 + \gamma_2(1 + \gamma_1) + \gamma_3(1 + \gamma_2)(1 + \gamma_1)]D_{000} \\ &= [(1 + \gamma_1)(1 + \gamma_2)(1 + \gamma_3) - 1]D_{000}. \end{aligned} \tag{5.8}$$

6. Exact probabilities for $N = 3$: Cauchy distribution

For samples of three or more from nonnormal populations the direct method used for samples of two in section 4 leads usually to almost inextricable difficulties. Another method, particularly simple for finding the portion of the distribution of t pertaining to values of t greater than n, will be applied in this section to samples of three from the Cauchy distribution, and in the next section to samples from the double exponential. We continue to use what may seem an odd dualism of notation, with N for the sample value and $n = N - 1$, in dealing with the Student ratio, partly in deference to differing traditions, but primarily because some of the algebra is simpler in terms of the sample number and some in terms of the number of degrees of freedom.

It is easy to prove the following lemma, in which we continue to use the notation x_i for the ith observations, $\xi_i = \rho x_i$, $i = 1, \cdots, N$, $\rho > 0$, $S\xi^2 = 1$, $a = S\xi$.

LEMMA. *If* $a \geqq n^{1/2}$, *then all* $\xi_i \geqq 0$. *On the other hand, if all* $\xi_i \geqq 0$, *then* $a \geqq 1$.

If the first statement were not true, one or more of the ξ must be negative; then since the sum of the ξ is positive, there must be a subset of them, say $\xi_1, \cdots, \xi_r$, with $1 \leqq r < N$, such that $a < \xi_1 + \cdots + \xi_r$, $\xi_1^2 + \cdots + \xi_r^2 < 1$. The maximum possible value of this upper bound for a, subject to the last condition, is $r^{1/2}$. Since r is an integer less than N, this leads to $a < n^{1/2}$, which contradicts the hypothesis. The second part of the lemma follows from the relations $a^2 = (S\xi)^2 = S\xi^2 + 2S\xi_i\xi_j$, with $S\xi^2 = 1$, and, because all the ξ are nonnegative, $S\xi_i\xi_j \geqq 0$.

The Cauchy distribution (4.1) generates on the unit sphere in N-space the n-dimensional density

$$D_N(\xi) = \frac{1}{\pi^N} \int_0^\infty \frac{\rho^{N-1}\, d\rho}{(1 + \rho^2\xi_1^2) \cdots (1 + \rho^2\xi_N^2)}. \tag{6.1}$$

This may be decomposed into partial fractions and integrated by elementary methods. The process is slightly more troublesome for even than for odd values of N because of the infinite limit, but the integral converges to a finite positive value for every positive integral N. For $N = 3$ the expression under the integral sign equals $Sa_k/(1 + \rho^2\xi_k^2)$, where

$$a_1 = -\frac{\xi_1^2}{(\xi_1^2 - \xi_2^2)(\xi_1^2 - \xi_3^2)} \tag{6.2}$$

and a_2 and a_3 may be obtained from this by cyclic permutations. Since

$$\int_0^\infty (1 + \rho^2\xi^2)^{-1}\, d\rho = \frac{\pi}{2|\xi|}, \tag{6.3}$$

we find

$$D_3(\xi) = \frac{1}{2}\pi^{-2} S \frac{a_k}{|\xi_k|}. \tag{6.4}$$

When this sum is evaluated with the help of (6.2) and reduced to a single fraction, both numerator and denominator equal determinants of the form known as alternants, and their quotient is a symmetric function of the form called by Muir a bialternant; this gives a form of (6.1) for all odd values of N. In the present case we obtain after cancellation of the common factors,

$$D_3(\xi) = \frac{1}{2}\pi^{-2}[(|\xi_1| + |\xi_2|)(|\xi_2| + |\xi_3|)(|\xi_3| + |\xi_1|)]^{-1} \tag{6.5}$$

as the density on an ordinary sphere. This density is continuous excepting at the six singular points where two observations, and hence two of the ξ, are zero. At each of these six points, $t = \pm 1$.

We shall use on the unit sphere polar coordinates θ, ϕ with the equiangular line as polar axis, and θ the angle between this axis and a random point having the density (6.5). For samples of three, $n = 2$, and so we have from (2.3),

$$t = 2^{1/2} \cot \theta. \tag{6.6}$$

The element of area on the sphere is $\sin \theta \, d\theta \, d\phi$. When this is multiplied by the density (6.5), with the ξ's replaced by appropriate functions of the polar coordinates, and then integrated with respect to ϕ from 0 to 2π, the result will be the element of probability for θ, which will be transformed through (6.6) into that for t.

Let y_1, y_2, y_3 be rectangular coordinates of which the third is measured along the old equiangular line. These are related to the old coordinates ξ by an orthogonal transformation

$$\xi_i = \sum c_{ij} y_j, \qquad \text{with} \quad c_{i3} = 3^{-1/2}; \qquad i = 1, 2, 3. \tag{6.7}$$

The orthogonality of the transformation implies the following conditions, in which the sums are with respect to k from one to three:

$$\sum c_{k1} c_{k2} = 0, \qquad \sum c_{kj} = 0, \qquad \sum c_{kj}^2 = 1, \qquad j = 1,2. \tag{6.8}$$

The orthogonality also implies that

$$c_{11}c_{21} + c_{12}c_{22} = c_{21}c_{31} + c_{22}c_{32} = c_{11}c_{31} + c_{12}c_{32} = -\frac{1}{3}. \tag{6.9}$$

On the sphere, whose equation in terms of the y is of course $\sum y_k^2 = 1$, the transformation to the polar coordinates may be written

$$y_1 = \sin \phi \sin \theta, \qquad y_2 = \cos \phi \sin \theta, \qquad y_3 = \cos \theta. \tag{6.10}$$

The five conditions (6.8) on the six unspecified c_{ij} are sufficient for the orthogonality of the transformation (6.7), so one degree of freedom remains in choosing the c_{ij}. This degree of freedom permits adding an arbitrary constant to ϕ, thus changing at will the point on a circle $\theta =$ constant from which ϕ is measured. In going around one of these circles, ϕ always ranges over an interval of length 2π. This degree of freedom will be used later to simplify a complicated expression by taking $c_{32} = 0$.

We shall now derive as a single integral and as a series the probability $P\{t \geqq T\}$ that the Student t should in a random sample of three from the Cauchy distribution (4.1) exceed a number T which is itself greater than two. We thus confine attention to values of t greater than two, and these correspond to values of a greater than $2^{1/2}$, as is seen by putting $N = 3$ and $a = 2^{1/2}$ in (2.9), which specifies a monotonic increasing relation between a and t. Then from the lemma at the beginning of this section it follows that all the observations and all the ξ are positive in the samples with which we are now concerned, as well as in some others.

The positiveness of the ξ permits for the present purpose the discarding of the absolute value signs in (6.5). The resulting cubic expression may, since

$$\xi_1 + \xi_2 + \xi_3 = a, \tag{6.11}$$

be written

$$(a - \xi_1)(a - \xi_2)(a - \xi_3). \tag{6.12}$$

With the help of (6.11) and the relation, derived from (6.11) and $S\xi^2 = 1$,

$$\xi_1\xi_2 + \xi_2\xi_3 + \xi_3\xi_1 = \frac{1}{2}(a^2 - 1), \tag{6.13}$$

the cubic (6.12) reduces to

$$\frac{1}{2}a(a^2 - 1) - \xi_1\xi_2\xi_3. \tag{6.14}$$

With the new symbols

$$u_k = c_{k1}y_1 + c_{k2}y_2, \qquad k = 1, 2, 3, \tag{6.15}$$

the transformation (6.7) can be rewritten

$$\xi_k = u_k + 3^{-1/2}y_3 = u_k + \frac{a}{3}, \tag{6.16}$$

the last equation holding because solving (6.7) for y_3 gives

$$y_3 = 3^{-1/2}(\xi_1 + \xi_2 + \xi_3) = 3^{-1/2}a. \tag{6.17}$$

From (6.15), (6.8) and (6.10) it will be seen that

$$\sum u_k = 0, \qquad \sum u_k^2 = y_1^2 + y_2^2 = \sin^2\theta. \tag{6.18}$$

From (6.18), by reasoning analogous to that leading to (6.13), we find

$$u_1u_2 + u_2u_3 + u_3u_1 = -\frac{1}{2}\sin^2\theta. \tag{6.19}$$

From (6.15),

$$u_1u_2u_3 = k_0y_1^3 + k_1y_1^2y_2 + k_2y_1y_2^2 + k_3k_3^3, \tag{6.20}$$

where

$$\begin{aligned} k_0 &= c_{11}c_{21}c_{31}, \qquad k_3 = c_{12}c_{22}c_{32}, \\ k_1 &= c_{11}c_{21}c_{32} + c_{21}c_{31}c_{12} + c_{31}c_{11}c_{22}, \\ k_2 &= c_{11}c_{22}c_{32} + c_{21}c_{32}c_{12} + c_{31}c_{12}c_{22}. \end{aligned} \tag{6.21}$$

In each term of k_1, the first two factors may be eliminated with the help of (6.9), which may also be written $c_{i1}c_{j1} = -1/3 - c_{i2}c_{j2}$, where $i, j = 1, 2, 3; i \neq j$. The linear terms thus introduced cancel out in accordance with the second of

(6.8). The three remaining terms are each equal to k_3. Thus $k_1 = -3k_3$. Likewise, $k_2 = -3k_1$.

At this point a substantial simplification is introduced by using the one available degree of freedom to make $c_{32} = 0$, thereby making k_3 and k_1 vanish. The orthogonal matrix in which c_{ij} is the element in the ith row and jth column is now

$$(6.22)\qquad \begin{bmatrix} 6^{-1/2} & -2^{-1/2} & 3^{-1/2} \\ 6^{-1/2} & 2^{-1/2} & 3^{-1/2} \\ -2 \times 6^{-1/2} & 0 & 3^{-1/2} \end{bmatrix}.$$

Thus $k_0 = -2^{-1/2}3^{-3/2}$ and $k_3 = 0$. From (6.20) we now find, with the help of (6.10),

$$(6.23)\qquad \begin{aligned} u_1u_2u_3 &= k_0(y_1^3 - 3y_1y_2^2) = k_0 \sin^3\theta(\sin^3\phi - 3\sin\phi\cos^2\phi) \\ &= 2^{-1/2}3^{-3/2}\sin^3\theta\sin 3\phi. \end{aligned}$$

From (6.16), (6.18), (6.19), and (6.23) we find

$$(6.24)\qquad \begin{aligned} \xi_1\xi_2\xi_3 &= \left(\frac{a}{3} + u_1\right)\left(\frac{a}{3} + u_2\right)\left(\frac{a}{3} + u_3\right) \\ &= \frac{a^3}{27} - \frac{a}{6}\sin^2\theta + 2^{-1/2}3^{-3/2}\sin^3\theta\sin 3\phi. \end{aligned}$$

Substituting this in (6.14) and simplifying gives the expression that appears in the denominator of the density (6.5) in the form

$$(6.25)\qquad \frac{25a^3}{54} - \frac{a}{2} + \frac{a}{6}\sin^2\theta - 2^{-1/2}3^{-3/2}\sin^3\theta\sin 3\phi.$$

This may be put into terms of t and ϕ alone by substituting in it the expressions, derivable when $N = 3$ from (2.9) and (6.6),

(6.26)

$$a = 3^{1/2}t(2 + t^2)^{-1/2}, \qquad \cos\theta = t(2 + t^2)^{-1/2}, \qquad \sin\theta = 2^{1/2}(2 + t^2)^{-1/2}.$$

The result is

$$(6.27)\qquad 3^{-3/2}2(4t^3 - 3t - \sin 3\phi)(2 + t^2)^{-3/2},$$

and this equals the content of the square bracket in (6.5) when all the ξ are positive, as they are for $t > 2$.

In the probability element $D_3(\xi)\sin\theta\, d\theta\, d\phi$, the factor $\sin\theta\, d\theta$ may be replaced by $|d\cos\theta|$ from (6.26), that is, by $2(2 + t^2)^{-3/2}\,dt$. Substituting (6.27) in (6.5) and multiplying by the element of area as thus transformed we have as the element of probability

$$(6.28)\qquad 3^{3/2}(2\pi^2)^{-1}(4t^3 - 3t - \sin 3\phi)^{-1}\,d\phi\,dt.$$

The probability element for t when it is greater than two will be found by integrating (6.28) with respect to ϕ from 0 to 2π. We put

(6.29) $$w = 4t^3 - 3t,$$

and observe that

(6.30) $$w - 1 = (t - 1)(2t + 1)^2, \qquad w + 1 = (t + 1)(2t - 1)^2.$$

From the first of these relations it follows, since $t \geqq 2$, that $w \geqq 25$, and therefore that the integral

(6.31) $$J = \int_0^{2\pi} (w - \sin 3\phi)^{-1}\, d\phi$$

exists. If the substitution $\phi' = 3\phi$ is made in this integral, a factor 1/3 is introduced by the differential, but the range of integration is tripled, and we have

(6.32) $$J = \int_0^{2\pi} (w - \sin \phi')^{-1}\, d\phi'.$$

This is evaluated by a well known device. For the complex variable $z = \exp(i\phi')$, the path of integration runs counterclockwise along the unit circle C around O, and $d\phi' = dz/(iz)$. Since $\sin \phi' = (z - z^{-1})/(2i)$, we find

(6.33) $$J = -2 \int_C (z^2 - 2iwz - 1)^{-1}\, dz.$$

Poles are at

(6.34) $$z_1 = i[w + (w^2 - 1)^{1/2}], \qquad z_2 = i[w - (w^2 - 1)^{1/2}],$$

of which only the latter is within the unit circle. Hence

(6.35) $$J = -\frac{4\pi i}{z_2 - z_1} = 2\pi(w^2 - 1)^{-1/2}.$$

Substitution in this from (6.30) yields a function of t alone. When this is multiplied by the factors of (6.28) not included in the integrand of (6.31), the element of probability $g(t)\, dt$ is obtained, where

(6.36) $$g(t) = 3^{3/2}\pi^{-1}(4t^2 - 1)^{-1}(t^2 - 1)^{-1/2}, \qquad t \geqq 2.$$

A check on the reasoning of this whole section is provided by dividing this expression for $g(t)$ by the corresponding density of t as given by normal theory, that is, by $g^*(t)$ as given by (2.13), and letting t increase. The limit, $3^{3/2}(4\pi)^{-1}$, agrees with R_3 as given by (4.2), in accordance with (3.3).

The Laurent series obtained by expanding (6.36) in inverse powers of t is

(6.37) $$3^{3/2}(4\pi t^3)^{-1}\left(1 + \frac{3}{4t^2} + \frac{9}{16t^4} + \frac{29}{64t^6} + \cdots\right),$$

and converges uniformly for all values of $t \geqq 2$. For any such value it may therefore be integrated to infinity term by term. After doubling to include the like probability for negative values from the symmetric distribution this gives

(6.38) $$P\{|t| \geqq T\} = 3^{3/2}(4\pi T^2)^{-1}\left(1 + \frac{3}{8T^2} + \frac{3}{16T^4} + \frac{29}{256T^6} + \cdots\right), \qquad T \geqq 2,$$

a rapidly convergent series.

This probability is also available in closed form. Indeed, the integral of (6.36) may be evaluated after rationalizing the integrand by means of the substitution $t = (u^2 + 1)/(u^2 - 1)$. The indefinite integral then becomes

$$3^{1/2}(\tan^{-1} 3^{1/2}u - \tan^{-1} 3^{-1/2}u) + \text{constant}. \tag{6.39}$$

The two inverse tangents may be combined by the usual formula, and the probability may be expressed in various ways, of which the most compact appears to be

$$P\{|t| \geqq T\} = 1 - 6\pi^{-1} \arctan [3^{-1/2}T^{-1}(T^2 - 1)^{1/2}], \qquad T \geqq 2. \tag{6.40}$$

One consequence is that $P\{|t| \geqq 2\}$ is about .115.

This equation may be solved for T to obtain a "percentage point" with specified probability P of being exceeded in absolute value when the null hypothesis is true. We have thus the formula

$$T^{-2} = 1 - 3 \tan^2 \left[\frac{\pi(1 - P)}{6}\right], \tag{6.41}$$

which is true only if the value of $|T|$ obtained from it is at least two. Taking in turn P as .05 and .01, we find from (6.41):

$$T_{.05} = 2.95, \qquad T_{.01} = 3.69. \tag{6.42}$$

7. The double and one-sided exponential distributions

We now derive for random samples of $N = n + 1$ the probability that t should exceed any value $T \geqq n$, when the population sampled has a frequency function of either of the forms

$$f(x) = \begin{cases} ke^{-kx}, & x \geqq 0 \\ 0, & x < 0; \end{cases} \tag{7.1}$$

$$f(x) = \frac{1}{2} ke^{-k|x|}, \tag{7.2}$$

where k is in each case a positive scale factor. Since t is invariant under changes of scale, k does not enter into its distribution and we shall take $k = 1$ without any loss of generality. Thus in the following we shall use the simpler frequency functions

$$f_1(x) = \begin{cases} e^{-x}, & x \geqq 0 \\ 0, & x < 0; \end{cases} \tag{7.3}$$

$$f_2(x) = \frac{1}{2} e^{-|x|}, \qquad \text{all real } x; \tag{7.4}$$

and the results will be applicable without change to (7.1) and (7.2). We are free also to interchange the signs $\geqq$ and $>$, and to say "positive" for "nonnegative" without changing our probabilities. The limitation of T and therefore t to values not less than n implies, in accordance with the lemma at the beginning of section 6, that all observations are nonnegative in the samples with which we are dealing.

Such samples are represented by points in the "octant" of the sample space for which all coordinates are positive, and in this part of the space the density, which is the product of N values of the frequency function for independent variates, is $\exp(-Sx)$ for (7.3) and $2^{-N}\exp(-Sx)$ for (7.4). Both these densities are constant over each n-dimensional simplex (generalized triangle or tetrahedron) defined, for $c > 0$, by an equation of the form

$$x_1 + x_2 + \cdots + x_N = c, \tag{7.5}$$

and the N inequalities $x_i \geqq 0$. Since this definition is unchanged by permuting the x, the simplex is regular. Its n-dimensional volume, which we designate v_n, may be determined by noticing that this simplex is a face, which may be called the "base," of an N-dimensional simplex with vertices at the origin and at the N points where the coordinate axes meet the hyperplane (7.5). These last N points are all at distance c from the origin. A generalization, easily established inductively by integration, of the ordinary formulas for the area of a triangle and the volume of a pyramid, shows that the volume of an N-dimensional simplex is the product of the n-dimensional area of any face, which may be designated the "base," by one-Nth of the length of a perpendicular ("altitude") dropped upon this base from the opposite vertex. Since the perpendicular distance from the origin to the hyperplane (7.5) is $cN^{-1/2}$, the volume of the N-dimensional simplex enclosed between this hyperplane and the coordinate hyperplanes is $v_n cN^{-3/2}$. But this N-dimensional volume may also be computed by taking another face as base, with c as altitude, evaluating the n-dimensional area of this base by the same method, and so on through all lower dimensions. This gives $c^N/N!$ as the N-dimensional volume. Equating, we find

$$v_n = \frac{c^n N^{3/2}}{N!} = \frac{c^n N^{1/2}}{n!}. \tag{7.6}$$

The samples for which t exceeds T are represented by points within a right circular cone with vertex at the origin, axis on the equiangular line, and semivertical angle θ, where $T = n^{1/2}\cot\theta$. This cone intersects the hyperplane (7.5) in a sphere whose radius we may call r. Then, since the distance of the hyperplane from the origin is $cN^{-1/2}$, we find $r = cN^{-1/2}\tan\theta = cn^{1/2}N^{-1/2}T^{-1}$. The n-dimensional volume enclosed by this sphere is within the simplex and, in accordance with (2.12), equals

$$\frac{r^n \pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}. \tag{7.7}$$

When the expression for r is inserted here and the result is divided by the volume v_n of the simplex, the result is

$$\pi^{n/2} n^{n/2} n! \left[N^{N/2}\Gamma\left(\frac{n}{2}+1\right)T^n\right]^{-1}. \tag{7.8}$$

This is the ratio of the n-dimensional volume of the sphere to that of the simplex, and is independent of c. It is thus the same on all the hyperplanes specified by (7.5) with positive values of c. Because of the constancy of the density on each of these hyperplanes, (7.8) is the probability that $t \geqq T$ when it is given that the probability is unity that all observations are positive, as is the case when the frequency function (7.1) or (7.3) is the true one. But if (7.2) or (7.4) holds, (7.8) is only the conditional probability that $t \leqq T$, and to give the unconditional probability must be multiplied by the probability of the condition, that is, by 2^{-N}.

If a two-tailed t-test of the null hypothesis that the mean is zero is applied and the double exponential (7.2) is the actual population distribution, the probability of a verdict of significance because of t exceeding T is double the foregoing, and therefore 2^{-n} times (7.8). With a slight change in the latter through reducing the argument of the Gamma function, this gives

$$(7.9) \qquad P\{|t| \geqq T\} = (\pi n)^{n/2}(n-1)!\left[2^{n-1}N^{n/2}\Gamma\left(\frac{n}{2}\right)T^n\right]^{-1}, \qquad T \geqq n.$$

The obvious relation

$$(7.10) \qquad \Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{n-1} = (n-1)!\pi^{1/2},$$

which is a special case of the duplication formula for the Gamma function, makes it possible to write (7.9) in the simpler form

$$(7.11) \qquad P\{|t| \geqq T\} = \pi^{(n-1)/2}n^{n/2}\Gamma\left(\frac{N}{2}\right)N^{-N/2}T^{-n}, \qquad T \geqq n.$$

Such probabilities of errors of the first kind in a two-tailed t-test are illustrated in table I for the values of T found in tables based on normal theory

TABLE I

PROBABILITIES FOR t IN SAMPLES FROM THE DOUBLE EXPONENTIAL

Stars refer to normal theory.

Sample size N	$t^*_{.05}$	$P = .05$ $P\{\|t\| > t^*_{.05}\}$ Exact	Approx.	$t^*_{.01}$	$P = .01$ $P\{\|t\| > t^*_{.01}\}$ Exact	Approx.	$t^*_{.001}$	$P = .001$ $P\{\|t\| > t^*_{.001}\}$ Exact	Approx.
2	12.706	.039 35	.039 27	63.657	.007 855	.007 854	636.619	.000 799	.000 785
3	4.303	.032 65	.030 23	9.925	.006 138	.006 046	31.598	.000 606	.000 605
4	3.182	.031 051	.023 132	5.841	.005 120	.004 626	12.941	.000 471	.000 463
5	2.776	...	.017 656	4.604	.004 715	.003 531	8.610	.000 386	.000 353
6	2.571	...	.013 458	4.032	...	.002 692	6.859	.000 337	.000 269

corresponding to the frequently used .05, .01, and .001 probabilities. One of these probabilities appears at the top of each of the three main panels into which the table is divided, and is to be compared with the exact probabilities in the second column of the panel, computed as indicated above for the values t^* of T obtained from a table of the Student distribution and often referred to as "percentage points" or "levels." It will at once be evident that the normal-theory probabilities at the heads of the panels are materially greater than the respective probabilities of the same events when the basic distribution of the observed variate is of the double exponential type (7.2).

This table also illustrates the comparison of the exact probabilities with the approximations introduced in section 3 for large values of T. These approximations, presented in the third column of each panel, are obtained by multiplying the normal-theory probability at the head of the panel by the expression R_N of (3.6) as adapted to the particular case. For the double exponential (7.4) we find

$$R_N = \pi^{N/2} n! \left[2^n N^{N/2} \Gamma\left(\frac{N}{2}\right) \right]^{-1}. \tag{7.12}$$

This may, with the help of (7.10), be put in the simpler alternative form

$$R_N = \pi^{n/2} \Gamma\left(\frac{n}{2} + 1\right) N^{-N/2}. \tag{7.13}$$

The table illustrates the manner in which the approximation slowly improves as T increases for a fixed n but grows poorer when n increases and T remains fixed.

The blank spaces in table I correspond to cases in which $T < n$, to which our formulas do not apply. Expressions of different analytical forms could be, but have not been, derived for such cases.

All the probabilities in the body of table I are less than the normal-theory probabilities .05, .01, and .001 at the heads of the respective columns, illustrating the greater concentration of the new distribution about its center in comparison with the familiar Student distribution. This concentration is further illustrated by the circumstance that $R_N < 1$ for all values of N. Indeed, $R_2 = \pi/4$ and $R_3 = \pi/3^{3/2}$ are obviously less than unity, and for all N,

$$\frac{R_{N+2}}{R_N} = \frac{\pi}{2} \frac{N+1}{N+2} \left(\frac{N}{N+2}\right)^{N/2}, \tag{7.14}$$

may be shown to be less than unity by means of the sign of the error after two terms in the expansion of the logarithm of the last factor.

Percentage points, the values that t has assigned probabilities P of exceeding, can be found for the double exponential population by solving (7.9) or (7.11) for T, provided the value thus found is not less than n. For samples of three ($n = 2$) the five per cent point thus found is 3.48 and the one per cent point is 7.78. Each of these is between the corresponding points for the normal and the Cauchy parent distributions, the latter of which were found at the end of section 6. These results are summarized in table II, along with the values of R_3 for

TABLE II

PERCENTAGE POINTS OF t AND MULTIPLIERS OF SMALL EXTREME-TAIL NORMAL-THEORY PROBABILITIES FOR SAMPLES OF THREE

Population	$T_{.05}$	$T_{.01}$	R_3
Normal	4.30	9.92	1.000
Double exponential	3.48	7.78	.785
Cauchy	2.95	3.69	.413

the same populations, which provide good approximations when t is very large. The degree of concentration of t about zero is in the same order by all three of the measures in the columns of table II.

For a fixed basic distribution and a fixed probability of T being exceeded by t, T is a function of n alone. The behavior of this function as n increases depends on the basic distribution. If this is normal, a percentage point T approaches a fixed value, the corresponding percentage point of a standard normal distribution of unit variance. If the basic distribution is the double exponential, it is deducible from (7.11) with the help of Stirling's formula that $T = O(n^{1/2})$ when P is fixed. More exactly, $T \sim (\pi n)^{-1/2}(2e)^{-1/2}P^{-1/n}$.

8. Student ratios from certain other distributions

The distribution of the Student ratio in samples from the "rectangular" distribution with density constant between two limits symmetric about zero, and elsewhere zero, has been the object of considerable attention. The exact distribution for samples of two is easily found by a geometrical method. Perlo [41] in 1931 obtained the exact distribution for samples of three in terms of elementary functions—not the same analytic function over the whole range—again by geometry, but with such greatly increased complexity attendant on the extension from two to three dimensions as to discourage attempts to go on in this way to samples larger than three. Rider [45] gave the distribution for $N = 2$, and also, by enumeration of all possibilities, investigated the distribution in samples of two and of four from a discrete distribution with equal probabilities at points uniformly spaced along an interval; the result presumably resembles that for samples from the continuous rectangular distribution.

The distribution of t is independent of the scale of the original variate, and we take this to have density 1/2 from -1 to 1, and zero outside these limits. In an N-dimensional sample space the density is then 2^{-N} within a cube of vertices whose coordinates are all ± 1, and zero outside it. The projection of the volume of this cube from the center onto a surrounding sphere will obviously produce a concentration of density at the points A and A' on the sphere where the coordinates all have equal values, greater than at the points of minimum density in the ratio of the length of the diagonal of the cube to an edge, that is, of $N^{1/2}$ from -1 to 1. Substituting the frequency function

$$f(x) = \begin{cases} \frac{1}{2}, & -1 < x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{8.1}$$

in (3.6) gives

$$R_N = \frac{\left(\frac{\pi N}{4}\right)^{N/2}}{\Gamma\left(\frac{N}{2}+1\right)}. \tag{8.2}$$

This increases rapidly with N and without bound; indeed, R_{N+2}/R_N is nearly $\pi e/2$, so $R = \infty$. The lowest values are

$$R_2 = \frac{\pi}{2}, \qquad R_3 = \frac{\pi}{2} 3^{1/2}, \qquad R_4 = \frac{\pi^2}{2}. \tag{8.3}$$

This provides a sharp contrast with the cases of the Cauchy and the single and double exponential distributions, for which $R = 0$. There seems to be a real danger that a statistician applying the standard t-test to a sample of moderate size from a population which, unknown to him, is from a rectangular population, will mistakenly declare the deviation in the mean significant. The opposite was true for the populations considered earlier; that is, use of the standard normal-theory tables of t where the Cauchy or exponential distribution is the actual one, with central value zero the hypothesis to be tested, will not lead to rejection of this hypothesis as often as expected.

An interesting generalization of the rectangular distribution is the Pearson Type II frequency function

$$\begin{aligned} f(x) &= c_p(1-x^2)^{p-1}, \qquad p > 0, \qquad -1 < x < 1, \\ c_p &= \frac{\Gamma\left(p+\frac{1}{2}\right)\pi^{-1/2}}{\Gamma(p)}, \end{aligned} \tag{8.4}$$

which reduces to the rectangular for $p = 1$. Values of p greater than unity determine frequency curves looking something like normal curves, but the induced distribution of t is very different in its tails. These are given approximately for large t by the product of the corresponding normal-theory probability by

$$R_N = N^{N/2}\left[\frac{\Gamma\left(p+\frac{1}{2}\right)}{\Gamma(p)}\right]^N \frac{\Gamma(pN - N + 1)}{\Gamma\left(pN - \frac{N}{2} + 1\right)}. \tag{8.5}$$

This is a complicated function of which only an incomplete exploration has been made; it appears generally to increase rapidly with N, less rapidly with p; whether it has an upper bound has not been determined. If it approaches a definite limit as N increases, the tail distribution of t has a feature not yet encountered in other examples: the probabilities of the extreme tails on this Type II hypothesis, divided by their probabilities on a normal hypothesis, may have a definite positive limit as the sample size increases. Such limits of ratios have in previous examples been zero or infinity.

Values of p between zero and one (not inclusive) yield U-shaped frequency curves. The expression just given for R_N may then present difficulties because of zero or negative values of the Gamma functions. The usual nature of the function and the variety of possibilities connected with it appear particularly when studying large samples, as the application of Stirling's formula is not quite so straightforward when the arguments of the Gamma functions are not all positive; and the factor in square brackets with the exponent N may be either greater or less than unity.

Skew distributions and distributions symmetric about values different from zero generate distributions of t which can in many cases be studied by the methods used in this paper, particularly by means of R_N, which is often easy to compute and gives approximations good for sufficiently large values of t, even with skew distributions. In particular, power functions of one-sided or two-sided t-tests can readily be studied in this way for numerous nonnormal populations.

9. Conditions for approximation to Student distribution

We now derive conditions on probability density functions of a wide class, necessary and sufficient for the distribution of t in random samples to converge to the Student distribution in the special sense that $R = 1$; that is, that $R = \lim_{N=\infty} R_N = 1$; $R_N = \lim_{T=\infty} (P/P^*)$, where P is the probability that t exceeds T in a random sample of N from a distribution of density $f(x)$, and P^* is the probability of the same event on the hypothesis of a normal distribution with zero mean. A like result will hold, with only trivial and obvious modifications, for the limiting form when t approaches $-\infty$. It will be seen from the results of this section that the conditions for this kind of approach to the Student distribution are of a different character, and are far more restrictive than those treated in central limit theorems for approach to a central normal distribution, which is also approached by a Student distribution, with increasing sample size within a fixed finite interval.

The probability density (frequency) function we suppose such that the probability that x is positive is itself positive. In the opposite case the trivial variation mentioned above comes into play. We rewrite (3.6) in the form

$$R_N = 2(\pi N)^{N/2} \frac{I_N}{\Gamma\left(\frac{N}{2}\right)}, \tag{9.1}$$

where

$$I_N = \int_0^\infty x^{N-1}[f(x)]^N \, dx, \tag{9.2}$$

and consider only cases in which this last integral exists and can be approximated asymptotically by a method of Laplace, of which many accounts are available, including those of D. V. Widder in *The Laplace Transform* and of A. Erdélyi in *Asymptotic Expansions*. Direct application of this method requires that $xf(x)$, and therefore

$$g(x) = \log x + \log f(x), \tag{9.3}$$

shall be twice differentiable and shall attain its maximum value for a single positive value x_0 of x; and that this be an *ordinary* maximum in the sense that the first derivative is zero and the second derivative is negative at this point. It is further required that, for all positive $x \neq x_0$, the strong inequality $g(x) < g(x_0)$ shall hold. Variations of the method may be applied when these conditions are somewhat relaxed; for instance, when there are several equal maxima, the interval of integration may be broken into subintervals each satisfying the conditions just stated; but we shall not in this paper consider any such variations.

Thus we deal with cases in which there is a uniquely determined $x_0 > 0$ for which $g(x_0) > g(x)$ when $x_0 \neq x > 0$, $g'(x_0) = 0$, $g''(x_0) < 0$, and

$$g(x) = g(x_0) + \frac{1}{2}(x - x_0)^2 g''(x_0) + o(x - x_0)^2. \tag{9.4}$$

Then (9.2) may be written

$$\begin{aligned} I_N &= \int_0^\infty \exp\left[Ng(x)\right] x^{-1}\, dx \\ &= \exp\left[Ng(x_0)\right] \int_0^\infty \exp\left[\frac{Ng''(x_0)}{2}(x - x_0)^2 + N\, o(x - x_0)^2\right] x^{-1}\, dx. \end{aligned} \tag{9.5}$$

When f, g, and their derivatives appear without explicit arguments the argument x_0 will be understood. With this convention, a new variable of integration $z = (-Ng'')^{1/2}(x - x_0)$ leads to the form

$$I_N = e^{Ng} \int_{-x_0(-Ng'')^{-1/2}}^\infty \exp\left[-\frac{z^2}{2} + o\left(\frac{z^2}{N}\right)\right] \left[x_0 + O(N^{-1/2})\right] dz (-Ng'')^{-1/2}. \tag{9.6}$$

An asymptotic approximation to I_N, which by definition has a ratio to I_N that approaches unity as N increases, is found here by letting the lower limit tend to $-\infty$ and dropping the terms $o(z^2/N)$ and $O(N^{-1/2})$, which are small in comparison with the terms to which they are added. This gives

$$I_N \sim (2\pi)^{1/2} x_0^{-1} (-Ng'')^{-1/2} e^{Ng}. \tag{9.7}$$

With the understanding that the argument is x_0 in (9.3) and its first and second derivatives, we find

$$g = \log x_0 + \log f, \qquad g' = x_0^{-1} + \frac{f'}{f} = 0, \qquad g'' = -x_0^{-2} + \frac{ff'' - f'^2}{f^2}, \tag{9.8}$$

whence

$$e^g = x_0 f, \qquad x_0 = -\frac{f}{f'}, \qquad g'' = \frac{ff'' - 2f'^2}{f^2}. \tag{9.9}$$

By Stirling's formula, $\Gamma(N/2)$ is asymptotically equivalent to

$$(2\pi)^{1/2}\left(\frac{N-2}{2}\right)^{(N-1)/2} e^{-(N-2)/2}. \tag{9.10}$$

This contains the factor $(1 - 2/N)^{N/2-1/2}$, which may be replaced by $1/e$. In this way we obtain

$$\Gamma\left(\frac{1}{2}N\right) \sim 2^{1-N/2}\pi^{1/2}N^{(N-1)/2}e^{-N/2}. \tag{9.11}$$

Making both this substitution and (9.7) in (9.1) and simplifying gives

$$R_N \sim 2^{1/2}(-g'')^{-1/2}x_0^{-1}(2\pi e^{2g+1})^{N/2}. \tag{9.12}$$

Since $g'' < 0$ by assumption, it is clear from (9.12) that the limit R of R_N as N increases is zero, a positive number, or infinite, according as $2\pi \exp(2g + 1)$ is less than, equal to, or greater than unity. Since this critical quantity is, by the first of (9.9), equal to $2\pi e(x_0 f)^2$, we have

THEOREM 1. *For a frequency function $f(x)$ satisfying the general conditions for application of the Laplace asymptotic approximation to $xf(x)$, a necessary and sufficient condition that the corresponding distribution of the Student ratio have the property that R is a positive constant is that $x_0 f = (2\pi e)^{-1/2}$ is the absolute maximum of $xf(x)$, taken only where $x = x_0$, and is an ordinary maximum.*

If this condition is satisfied, then it is obvious from (9.12) that a further necessary and sufficient condition for R to be unity is that $2^{1/2}(-g'')^{-1/2}x_0^{-1} = 1$, that is, $x_0^2 g'' + 2 = 0$. In this we replace x_0 and g'' by the expressions for them in the second and third equations respectively in (9.9) and simplify. The result is simply $f'' = 0$. Hence

THEOREM 2. *In order that $R = 1$ for the distribution of the Student ratio in random samples from a distribution of x of the kind just described, it is necessary and sufficient that the conditions of theorem 1 hold and that $f''(x_0) = 0$.*

A graphic version of these conditions is as follows. Under the portion of the smooth frequency curve $y = f(x)$ for which x is positive let a rectangle be inscribed, with two sides on the coordinate axes and the opposite vertex at a point L on the curve. Let L be so chosen that the rectangle has the greatest possible area. Let the tangent to the curve at L meet the x-axis at a point M. Then the triangle formed by L and M with the origin O is isosceles, since the first-order condition for maximization, the second equation of (9.8) or (9.9), states that the subtangent at L equals its abscissa and is measured in the opposite direction.

The area of the isosceles triangle OLM equals that of the maximum rectangle. The necessary and sufficient condition that R be a positive constant is that this area have the same value $(2\pi e)^{-1/2}$ as for a normal curve symmetric about the y-axis. If this is satisfied, the further necessary and sufficient condition that

$R = 1$ is that L be a point of inflection of the frequency curve for which x is negative.

It is particularly to be noticed that these conditions, which indicate the circumstances under which the familiar Student distribution can be trusted for large values of t and large samples, have nothing to do with the behavior of the distribution of x for extremely large or extremely small values of this, the observed variate, nor with its moments, but are solely concerned with certain relations of the frequency function and its first and second derivatives at points of inflection.

Proposals to "correct" t for nonnormality by means of higher moments of x (up to the seventh, according to one suggestion), encounter the difficulty that such moments reflect primarily the behavior of the distribution for very large values, whereas the mainly relevant criteria seem from the results of this section to concern rather the vicinity of the points of inflection, which for a normal distribution are at a distance of one standard deviation from the center. It is such intermediate or "shoulder" portions of a frequency curve that seem chiefly to call for exploration when the suitability of the t tables for a particular application is in question, especially for stringent tests with large samples. Moments of high order, even if known exactly, do not appear to be at all sensitive to variations of frequency curves in the shoulders. Moreover, if moments are estimated from samples, their standard errors tend to increase rapidly with the order of the moment. More efficient statistics for the purpose are evidently to be sought by other methods.

10. Correlated and heteroscedastic observations. Time series

Much was written in the nineteenth century, principally by astronomers and surveyors, about the problem of unequal variances of errors of observation, and hence different weights, which are inversely proportional to the variances. Much has been written in the twentieth century, largely by economists and others concerned with time series, about the problems raised by the lack of independence of observations, particularly by the tendency for observations close to each other in time or space to be more alike than observations remote from each other. The two problems are closely related, though this fact is seldom mentioned in the two separate literatures. They combine into one problem where the observations, or the errors in them, have jointly a multivariate normal distribution in which both correlations and unequal variances may be present. Problems of estimating or testing means and regression coefficients may in such cases easily be reduced by linear transformation to the standard forms in which the observations are independent with a uniform variance, provided only that the correlations and the ratios of the variances are known, since these parameters enter into the transformation.

Thus, if a multivariate normal distribution is assumed, the leading methodological problem is the determination of this set of parameters, which collectively are equivalent to the covariance matrix or its inverse, apart from a common factor. The number of independent parameters needed to reduce the problem to one of linear estimation of standard type is less by unity than the total number of variances and correlations of the N observations, and therefore equals $N + N(N - 1)/2 - 1 = (N - 1)(N + 2)/2$. Since this exceeds the number of observations, often greatly, there is no hope of obtaining reasonable estimates from these N observations alone. Numerous expedients, none of them universally satisfactory, are available for arriving at values for these parameters in various applications. Thus astronomers add estimated variances due to different sources, one of which, the "personal equation," involves a sum of squares of estimated errors committed by an individual observer. For observations made at different times, a maintained assumption, ostensibly a priori knowledge, that the observations or their errors constitute a *stationary* stochastic process of a particular sort often reduces the number of independent parameters enough so that, with large samples, estimates can be made with small calculated standard error, but valid only if the maintained assumption is close to the real truth. Another expedient, often in time series the most satisfactory of all, is to adjust the observations by means of concurrent variables, either supplementary observations, such as temperature and barometric pressure in biological experiments, or variables known a priori, such as the time as used in estimating seasonal variation and secular trend with the help of a number of estimated parameters sufficiently small to leave an adequate number of degrees of freedom for the effects principally to be examined. The simplest of all such expedients is merely to ignore any possible differences of variances or of intercorrelations that may exist.

If methods of this kind leave something to be desired, the statistician using them may be well advised at the next stage, when estimating the expected value common to all the observations, or that of a specified linear function of all the observations, or testing a hypothesis that such a linear function has a specified expected value, to modify or supplement methods based on the Student distribution so as to bring out as clearly as possible how the probability used is affected by variations in the covariance matrix. In addition, when as here the maintained hypotheses about the ancillary parameters of the exact model used are somewhat suspect, a general protective measure, in the nature of insurance against too-frequent assertions that particular effects are significant, is simply to require for such a judgment of significance an unusually small value of the probability of greater discrepancy than that actually found for the sample. Thus instead of such conventional values of this probability P as .05 and .01, it may be reasonable, where some small intercorrelations or differences in variances in the observations are suspected but are not well enough known to be used in the calculation, to disclaim any finding of significance unless P is below some smaller value such as .001 or .0001. Such circumstances point to use of the t-statistic, with a new distribution based on a distribution of observations which, though multivariate normal, differs from that usually assumed in that the several observations may be intercorrelated and may have unequal variances. Moreover, the tails of the new t-distribution beyond large values of $|t|$ are here of particular interest. The probabilities of such tails will now be approximated by methods similar to those already used in this paper when dealing with t in samples from nonnormal distributions.

Consider N observations, whose deviations from their respective expectations will be called $x_1, \cdots, x_N$, having a nonsingular multivariate normal distribution whose density throughout an N-dimensional Euclidean space is

$$(10.1) \qquad f(x_1 \cdots x_N) = (2\pi)^{-N/2}|\lambda|^{1/2} \exp\left(-\frac{1}{2}\sum\sum \lambda_{ij}x_ix_j\right),$$

where λ is the inverse of the covariance matrix and is symmetric, $|\lambda|$ is its determinant, and $\lambda_{ij} = \lambda_{ji}$ is the element in the ith row and jth column of λ. Let the Student t be computed from these x by the formula (2.1), just as if λ were a scalar matrix as in familiar usage, with the object of testing whether the expectations of the x, assumed equal to each other, were zero. Then, as before, $t = (N-1)^{1/2} \cot \theta$, where θ is the angle between the equiangular line $x_1 = \cdots = x_N$ and the line through the origin O and the point x whose coördinates are $x_1, \cdots, x_N$. When the points x are projected onto the $(N-1)$-dimensional unit sphere about O, the density of the projected points $\xi(\xi_1, \cdots, \xi_N)$ is given by (2.7), which for the particular distribution (10.1) becomes

$$(10.2) \qquad \begin{aligned} D_N(\xi) &= \int_0^\infty f(\rho\xi_1, \cdots, \rho\xi_N)\rho^{N-1}\, d\rho \\ &= (2\pi)^{-N/2}|\lambda|^{1/2} \int_0^\infty \rho^{N-1} \exp\left(-\frac{1}{2}\rho^2 \sum\sum \lambda_{ij}\xi_i\xi_j\right) d\rho \\ &= \frac{1}{2}\pi^{-N/2}|\lambda|^{1/2}\Gamma\left(\frac{N}{2}\right)\left(\sum\sum \lambda_{ij}\xi_i\xi_j\right)^{-N/2}. \end{aligned}$$

When $\lambda = 1$ (the identity matrix), the double sum in the last parenthesis reduces to unity, since $\xi_1^2 + \cdots + \xi_N^2 = 1$, and in this case

$$(10.3) \qquad D_N(\xi) = D_N^*(\xi) = \frac{1}{2}\pi^{-N/2}\Gamma\left(\frac{N}{2}\right).$$

The ratio of (10.2) to (10.3) is

$$(10.4) \qquad R_N(\xi) = \frac{|\lambda|^{1/2}}{\left(\sum\sum \lambda_{ij}\xi_i\xi_j\right)^{N/2}},$$

and is the ratio of the density in our case to that in the standard case at point ξ on the sphere. If ξ moves to either of the points A, where all $\xi_i = N^{-1/2}$, or A', where all $\xi_i = -N^{-1/2}$, (10.4) approaches

$$(10.5) \qquad R_N(A) = R_N(A') = \frac{N^{N/2}|\lambda|^{1/2}}{\left(\sum\sum \lambda_{ij}\right)^{N/2}};$$

this we call simply R_N. For large values of t, (10.5) approximates both the ratio of the probability densities and the ratio of the probabilities of a greater value of t in our case to that in the standard case.

A small bias may here be mentioned in our earlier method of approximating the distribution of t in random samples, a bias generally absent from the applications in the present section. The approximation is, as was seen earlier, equivalent to replacing the integral of the density over a spherical cap centered at A or A' by the product of $D(A)$ or $D(A')$ respectively by the $(N - 1)$-dimensional area of the cap. In the applications we have considered to the distribution of t in random samples from nonnormal populations, A and A' are likely to be maximum or minimum points of the density on the sphere, in which case the approximation will tend to be biased, with the product of R_N by the probability in the standard case of t taking a greater value tending in case of a maximum to overestimate slightly the probability of this event and in case of a minimum to underestimate it. In nonrandom sampling of the kind dealt with in this section, maxima and minima of the density on the sphere can occur only at points that are also on latent vectors of the quadratic form; A and A' will be such points only for a special subset of the matrices λ. Apart from these special cases, the approximation should usually be more accurate in a certain qualitative sense with correlated and unequally variable observations than with random samples from nonnormal distributions.

The simplest and oldest case is that of heteroscedasticity with independence. Here both the covariance matrix and its inverse λ are of diagonal form. Let $w_1, \cdots, w_N$ be the principal diagonal elements of λ, while all its other elements are zero. These w are the reciprocals of the variances, and are therefore the true weights of the observations. From (10.5),

$$R_N = \frac{N^{N/2}(w_1 w_2 \cdots w_N)^{1/2}}{(\sum w_i)^{N/2}} = \left(\frac{\text{geometric mean}}{\text{arithmetic mean}}\right)^{N/2} \leqq 1. \tag{10.6}$$

Hence, for sufficiently large T,

$$P\{t > T\} \leqq P^*\{t > T\}, \tag{10.7}$$

where the right-hand probability is that found in familiar tables. Moreover, the ratio of the true probability to the standard one is seen from (10.6) to approach zero if the observations are replicated more and more times with the same unequal but correct weights, and with concurrent appropriate decreases in the critical probability used. The equality signs in (10.6) and (10.7) hold only if the true weights are all equal.

In contrast to the foregoing cases are those of correlated normal variates with equal variances. The common variance may without affecting the distribution of t be taken as unity, and the covariances then become the correlations. We consider two special cases.

A simple model useful in time series analysis has as the correlation between the ith and jth observations $\rho^{|i-j|}$, with ρ of course the serial correlation of consecutive observations; and $|\rho| < 1$. It is easy to verify that the determinant of this matrix of correlations among N consecutive observations is $(1-\rho^2)^{N-1}$ and that its inverse is

$$\lambda = (1-\rho^2)^{-1}\begin{bmatrix} 1 & -\rho & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & 0 & & -\rho & 1 \end{bmatrix} \tag{10.8}$$

From this we find $\sum\sum \lambda_{ij} = (N - N\rho + 2\rho)/(1+\rho)$. The determinant of λ is of course the reciprocal of that of the correlation matrix, and therefore equals $(1-\rho^2)^{-N+1}$. Substituting in (10.5) gives

$$R_N = (1-\rho^2)^{1/2}(1-\rho)^{-N}[1 + 2\rho(1-\rho)^{-1}N^{-1}]^{-N/2}. \tag{10.9}$$

If N increases and $\rho > 0$, R_N diverges, in contrast with the case of independent observations of unequal variances, where the limit was zero. R_N also becomes infinite as ρ approaches one with N fixed. But when ρ tends to -1 with N fixed, R_N vanishes. If ρ is fixed with $-1 < \rho < 0$, then $\lim_{N\to\infty} R_N = 0$.

As another example, consider N observations, each with correlation ρ with each of the others, and all with equal variances. The correlation matrix has a determinant equal to $(1-\rho)^{N-1}[1 + (N-1)\rho]$, which must be positive, implying that $\rho > -(N-1)^{-1}$; and the inverse matrix may be written

$$\lambda = (1-\rho)^{-1}\begin{bmatrix} 1+g & g & g & \cdots & g \\ g & 1+g & g & \cdots & g \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ g & g & g & \cdots & 1+g \end{bmatrix}, \tag{10.10}$$

where $g = -\rho[1 + (N-1)\rho]^{-1}$. A short calculation now gives

$$R_N = [1 + N\rho/(1-\rho)]^{(N-1)/2}. \tag{10.11}$$

As in the previous example, R_N becomes infinite if N increases with ρ fixed, or if ρ tends to unity with N fixed. The possibility that ρ approaches -1 cannot here arise.

A further application of the methods of this section having some importance is to situations in which the mean is in question of a normally distributed set of observations whose correlations and relative variances are known approximately but not exactly. An exact knowledge of these parameters would provide transformations of the observations $x_1, \cdots, x_N$ into new variates $x_1^*, \cdots, x_N^*$, independently and normally distributed with equal variance. If now t is computed from the starred quantities, it will have exactly the same distribution as in the standard case. But errors in the estimates used of the correlations and relative variances will generate errors in the transformation obtained by means of them, and therefore in the distribution of t. The true and erroneous values of the parameters together determine a matrix λ which, with (10.4) or (10.5), gives indications regarding the distribution of t, from which approximate percentage points may be obtained, with an accuracy depending on that of the original estimates of the correlations and relative variances. In this way general theory and independent observations leading to knowledge or estimates of these parameters may be used to improve the accuracy of the probabilities employed in a t-test. If a confidence region of known probability is available for the parameters, an intersteing set of values of R_N corresponding to the boundary of the confidence region might be considered.

11. Sample variances and correlations

The sensitiveness to nonnormality of the Student t-test is, as we have seen, chiefly a matter of the very nonstandard forms of tails distant from the center of the true t-distribution to an extent increasing with the sample number. Between fixed limits, or for fixed probability levels, it is easy to deduce from known limit theorems of probability that the Student distribution is still often reasonably accurate for sufficiently large samples. However, these theorems do not apply in all cases; for instance, they do not hold if the basic population has the Cauchy distribution; and unequal variances and intercorrelations of observations are notorious sources of fallacies. Moreover, the whole point in using the Student distribution instead of the older and simpler procedure of ascribing the normal distribution to the Student ratio is to improve the accuracy in a way having importance only for small samples from a normal distribution with independence and uniform variance. Deviations from these standard conditions may well, for small samples with fixed probability levels such as .05, or for larger samples with more stringent levels, produce much greater errors than the introduction of the Student distribution was designed to avoid. The limit theorems are of little use for small samples. To substitute actual observations uncritically in the beautiful formulas produced by modern mathematical statistics, without any examination of the applicability of the basic assumptions in the particular circumstances, may be straining at a gnat and swallowing a camel.

Even the poor consolation provided by limit theorems in the central part of some of the distributions of t is further weakened when we pass to more complicated statistics such as sample variances and correlation coefficients. Here the moments of the statistic in samples from nonnormal populations betray the huge errors that arise so easily when nonnormality is neglected. Thus the familiar expression for the variance of the sample variance,

$$\sigma_{s^2}^2 = \frac{\mu_4 - \mu_2^2}{n}, \tag{11.1}$$

when divided by the squared population variance yields a quotient two and a half times as great for a normal as for a rectangular population, and has no general upper bound.

The asymptotic standard error of the correlation coefficient (see, for example, Kendall, vol. 1, p. 211) reduces for a bivariate normal population to the well known approximation which is the square root of

$$\frac{(1-\rho^2)^2}{n}, \tag{11.2}$$

where ρ is the correlation in the population. This may now be compared with the asymptotic variance of the correlation coefficient between x and y when the density of points representing these variates is uniform within an ellipse in the xy-plane and zero outside it. The population correlation ρ and the moments of the distribution are determined by the shape of the ellipse and the inclinations of its axes to the coordinate axes. On substituting the moments in the general asymptotic formula for the standard error of the correlation coefficient r, the result reduces to the square root of

$$\frac{2(1-\rho^2)^2}{3n}, \tag{11.3}$$

only two-thirds of the standard expression (11.2). It is noteworthy that in the special case $\rho = 0$, in which the ellipse reduces to a circle, the variance is only $2/(3n)$, differing from $1/n$, the exact variance of r in samples from a bivariate normal distribution with $\rho = 0$, and the asymptotic variance of r in samples from any population in which there is independence between x and y, or even in which the moments of fourth and lower orders satisfy the conditions for independence.

Fisher's ingenious uses ([14], section 35) of the transformation

$$z = f(r) = \frac{1}{2}\log(1+r) - \frac{1}{2}\log(1-r), \qquad \zeta = f(\rho), \tag{11.4}$$

whose surprising accuracy has been verified in a study by Florence David [10], and which has since been studied and modified [26], owes its value primarily to the properties that, for large and even moderately large samples, z has a nearly normal distribution with mean close to ζ and variance n^{-1}, apart from terms of higher order. This transformation may be derived (though it was not by Fisher) from the criterion that z shall be such a function of r as to have the asymptotic variance n^{-1}, independently of the parameter ρ. This problem has been studied [26], and an asymptotic series for the solution has been found to have z as its leading term; two additional terms, multiples respectively of n^{-1} and n^{-2}, have been calculated, with resultant modifications of z that presumably improve somewhat the accuracy of the procedures using it.

All uses of z, however, depend for the accuracy of their probabilities on the applicability of the familiar formula (11.2) for the variance of r, which has been derived only from assumptions including a bivariate normal population, and even then, only as an approximation; see [26]. If the bivariate distribution is not normal, and in spite of this fact the variance (11.2) is used either directly, or indirectly through the z-transformation, the errors in the final conclusions may be very substantial. This transformation does not in any way atone for nonnormality or any other deviation from the standard conditions assumed in its derivation.

A different transformation having the same desirable asymptotic properties as Fisher's may be obtained for correlations in random samples from a nonnormal bivariate distribution of known form, or even one whose moments of order not exceeding four are all known, provided certain weak conditions are satisfied. In these cases the asymptotic standard error σ_r of r is of order $n^{-1/2}$ and may be found in the usual way as a function of ρ. If $z = f(r)$ and $\zeta = f(\rho)$, with f a sufficiently regular increasing function, a time-honored method of getting approximate standard errors may be applied: Expand $z - \zeta$ in a series of powers of $r - \rho$; square, and take the mathematical expectation; neglect terms of the resulting series that are of order higher than the first in n^{-1}, thus ordinarily retaining only the first term, which emerges as asymptotically equivalent to the variance of r. A theorem justifying this procedure under suitable conditions of regularity and boundedness is given by Cramér [9], but the technique has been in frequent use for more than a century without the benefit of a careful proof such as Cramér's. In the present case, after taking a square root, it yields $\sigma_z \sim \sigma_r(d\zeta/d\rho)$. If now we put $\sigma_z = n^{-1/2}$, integrate, and impose the additional condition that $f(O) = 0$, we find:

$$\zeta \sim n^{-1/2} \int_0^\rho \sigma_r^{-1}\, d\rho. \tag{11.5}$$

If we wish to obtain a transformation resembling Fisher's in being independent of n, as well as in other respects, we may replace the sign of asymptotic approximation in (11.5) by one of equality.

Substituting in (11.5) the reciprocal of the square root of the asymptotic variance (11.2) of r in samples from a normal distribution leads to Fisher's transformation (11.4). But for the bivariate distribution having uniform density in an ellipse we substitute from (11.3) instead of (11.2), thus redefining z and ζ by multiplying Fisher's values by $(3/2)^{1/2}$, while retaining the approximate variance n^{-1}.

12. Inequalities for variance distributions

Light may often be thrown on distributions of statistics by means of inequalities determined by maximum and minimum values, even where it is excessively difficult to calculate exact distributions. An example of this approach is provided by the following brief study of the sample variance in certain nonnormal populations. It will be seen that this method sometimes yields relevant information where none can be obtained from the asymptotic standard error formula because the fourth moment is infinite.

One family of populations presenting special interest for the study of sample variances is specified by the Pearson Type VII frequency curves

$$y = y_0\left(1 + \frac{x^2}{a^2}\right)^{-p}, \qquad p \geqq 1, \tag{12.1}$$

where

$$y_0 = \frac{\Gamma(p)}{a\pi^{1/2}\Gamma\left(p - \frac{1}{2}\right)}. \tag{12.2}$$

This family of symmetric distributions may be said to interpolate between the Cauchy and normal forms, which it takes respectively for $p = 1$ and $p = \infty$. In the last case we may take $a^2 = 2p$.

The even moments, so far as they exist, are given by

$$\mu_{2k} = \frac{a^{2k}\pi^{-1/2}\Gamma\left(k + \frac{1}{2}\right)\Gamma\left(p - k - \frac{1}{2}\right)}{\Gamma\left(p - \frac{1}{2}\right)}, \qquad k < p - \frac{1}{2}, \tag{12.3}$$

Putting $k = 1, 2$ we find

$$\begin{aligned} \sigma^2 = \mu_2 &= \frac{a^2}{2p - 3}, && \text{if } p > \frac{3}{2}, \\ \mu_4 &= 3a^4(2p - 3)^{-1}(2p - 5)^{-1}, && \text{if } p > \frac{5}{2}. \end{aligned} \tag{12.4}$$

We consider the case in which the population mean is known but methods appropriate to a normal distribution are used in connection with an estimate of σ and tests of hypotheses concerning σ. In a random sample let $x_1, \cdots, x_n$ be the deviations from the known population mean; then n is both the sample number and the number of degrees of freedom; and $s^2 = Sx^2/n$ is an unbiased estimate of the variance σ^2 in any population for which the latter exists. But despite this fact and the similarity of the Type VII and normal curves, the accuracy of s^2 as an estimate of σ^2 differs widely for the different populations, even if σ^2 has the same value.

When the formula (11.1) for the variance (that is, squared standard error) of the sample variance s^2 is applied to a normal distribution, the result is $2\sigma^4/n$ as is well known. But when the moments (12.4) of the Type VII distribution are inserted in the same formula, the result is

$$4a^4(p - 1)[(2p - 3)^2(2p - 5)n]^{-1}, \qquad p > \frac{5}{2}. \tag{12.5}$$

In this, a^2 may by (12.4) be replaced by $(2p - 3)\sigma^2$, yielding

$$\sigma^2_{s^2} = 4\sigma^4(p - 1)[(2p - 5)n]^{-1}. \tag{12.6}$$

The variance of s^2 is infinite when $p \leqq 5/2$, for instance when $p = 1$ and the Type VII takes the Cauchy form. When $p > 5/2$ the ratio of the variance of s^2 in samples from a Type VII to that in samples from a normal population of the same variance is equal to $1 + 3/(2p - 5)$; this may take any value greater than unity.

In studying the distributions of s let us put

$$(12.7) \qquad u^2 = ns^2 = Sx_i^2, \qquad u \geqq 0,$$
$$Q = S \log\left(1 + \frac{x_i^2}{a^2}\right).$$

Then in the sample space u is the radius of a sphere whose $(n-1)$-dimensional surface has, by (2.11), the "area" $2\pi^{n/2}u^{n-1}/\Gamma(n/2)$. Also, let us denote by $q_m(u)$, $q(u)$, and $q_M(u)$, or simply q_m, q, and q_M, the minimum, mean, and maximum respectively of Q over the spherical $(n-1)$-dimensional surface of radius u. Thus

$$(12.8) \qquad q_m(u) < q(u) < q_M(u).$$

The density in the sample space found by multiplying together n values of (12.1) with independent arguments is, in the new notation,

$$(12.9) \qquad y_0^n e^{-pQ}.$$

The probability element for u is seen, on considering a thin spherical shell, to be

$$(12.10) \qquad 2\pi^{n/2}\left[\Gamma\left(\frac{n}{2}\right)\right]^{-1} y_0^n e^{-pq(u)}u^{n-1}\,du.$$

Upper and lower bounds will be found by replacing $q(u)$ here by q_m and q_M respectively.

At the extremes of Q for a fixed value of u, differentiation yields

$$(12.11) \qquad x_i\left(1 + \frac{x_i^2}{a^2}\right)^{-1} = \lambda x_i, \qquad i = 1, 2, \cdots, n,$$

where λ is a Lagrange multiplier. This shows that all nonzero coordinates of an extreme point must have equal squares. Let K be the number of nonzero coordinates at a point satisfying the equations. Since the sum of the squares is u^2, each of these coordinates must be $\pm K^{-1/2}u$; and if $u \neq 0$, then K can have only the values $1, 2, \cdots, n$. At such a critical point,

$$(12.12) \qquad Q = K \log(1 + u^2K^{-1}a^{-2}).$$

This is a monotonic increasing function of K. Hence,

$$(12.13) \qquad q_m = \log(1 + u^2a^{-2}), \qquad q_M = \log(1 + u^2n^{-1}a^{-2})^n.$$

A slightly larger but simpler upper bound is also available; for the monotonic character of the function shows that $q_M < u^2/a^2$.

The well known and tabulated distribution of $\chi^2 = ns^2/\sigma^2$, based on normal theory, is used for various purposes, one of which is to test the hypothesis that σ has a certain value σ_0 against larger values of σ. Such a test might for example help to decide whether to launch an investigation of possible excessive errors of measurement. The simplest equivalent of the standard distribution is that of

$$w = \frac{1}{2}\chi^2 = ns^2(2\sigma^2)^{-1}, \tag{12.14}$$

of which the element of probability is

$$g_n(w)\,dw = \frac{w^{n/2-1}e^{-w}\,dw}{\Gamma\left(\frac{n}{2}\right)}. \tag{12.15}$$

We now inquire how this must be modified if the population is really of Type VII rather than normal.

Let us first consider the cases for which the Type VII distribution has a variance, that is, for which $p > 3/2$; then by (12.4), $a^2 = (2p - 3)\sigma^2$. Eliminating a between this and (12.2) gives

$$\frac{y_0(2\pi)^{-1/2}c_p}{\sigma} = 1, \tag{12.16}$$

where $c_p = (p - 3/2)^{-1/2}\Gamma(p)/\Gamma(p - 1/2)$. When p increases, c_p tends to unity through values greater than unity. Also, c_3 is approximately 1.23.

Since $ns^2 = Sx^2 = u^2$, we find from (12.14) that $u = \sigma(2w)^{1/2}$. This and (12.16) we substitute in the distribution (12.10), which may then be written with the help of the notation (12.15) in the form

$$h(w) = c_p^n \exp\{w - pq[\sigma(2w)^{1/2}]\}g_n(w)\,dw. \tag{12.17}$$

Making the substitutions for u and a also in (12.13), where u^2/a^2 becomes $2w/(2p - 3)$, and referring to (12.8), we observe that the q function in (12.17) is greater than $\log[1 + 2w(2p - 3)^{-1}]$ and less than $\log[1 + 2w\,n^{-1}(2p - 3)^{-1}]^n$. Consequently, for all $w > 0$,

$$h'(w) < h(w) < h''(w), \tag{12.18}$$

where

$$\begin{aligned} h'(w) &= c_p^n\left(\frac{2p-3}{2p-3+2nw}\right)^{np} e^w g_n(w), \\ h''(w) &= c_p^n\left(\frac{2p-3}{2p-3+2w}\right)^{p} e^w g_n(w). \end{aligned} \tag{12.19}$$

If p increases while n and w remain fixed, both $h'(w)$ and $h''(w)$ tend to $g_n(w)$, and their ratio tends to unity. Thus the ordinary normal-theory tables and usages of χ^2 give substantially correct results when applied to samples from a Type VII population if only p is large enough. This of course might have been expected since the Type VII distribution itself approaches the normal form when p grows large; but the bounds just determined for the error of assuming normality may often be useful in doubtful cases, or where p is not very large.

In all cases, $h'(w)$ and $h''(w)$ may be integrated by elementary methods.

13. What can be done about nonstandard conditions?

Standard statistical methods involving probabilities calculated on standard assumptions of normality, independence, and uniform variance of the observed variates lead to many fallacies when any of these assumed conditions are not fully present, and adequate corrections, often requiring additional knowledge, are not made. The object of this paper is to clarify the behavior of these aberrant probabilities with the help of some new mathematical techniques and to take some soundings in the vast ocean of possible cases.

We have concentrated for the most part on the effects of nonstandard conditions on the distribution of Student's t, and especially on the portions of this distribution for which $|t|$ is rather large. It is such probabilities (or their complements) that appear to supply the main channel for passing from numerical values of t to decisions and actions. A knowledge of the "tail" areas of the frequency distribution of t (as also of F, r, and other statistics) seems therefore of greater value than information about relative probabilities of different regions all so close to the center that no one would ever be likely to notice them, to say nothing of distinguishing among them. Thus instead of moments, which describe a distribution in an exceedingly diffuse manner and are often not sufficiently accurate when only conventional numbers of them are used, it seems better to utilize for the present purpose a functional only of outer portions of the distribution of t. Such a measure, here called R_N for samples of N and R in the limit for very large samples, is the limit as T increases of the ratio of two probabilities of the same event, such as $t > T$ or $|t| > T$, with the denominator probability based on standard normal central random sampling and the numerator on some other condition which the investigator wishes to study. Additional methods are here developed for evaluating the exact probabilities of tail areas for which T is at least as great as the number of degrees of freedom n, in such leading cases as the Cauchy and the single and double simple exponential distributions. These findings could be pushed further inward toward the center, but with increasing labor and, at least in some cases, diminishing utility.

To heal the ills that result from attempts to absorb into the smooth routine of standard normal theory a mass of misbehaving data of doubtful antecedents, an inquiry into the nature and circumstances of the trouble is a natural part of the preparations for a prescription. Such an inquiry may in difficult cases require the combined efforts of a mathematical statistician, a specialist in the field of application, and a psychiatrist. But many relatively simple situations yield to knowledge obtainable with a few simple instruments. One of these instruments is R_N, which in case of erratically high or low values of t will eliminate many possible explanations of the peculiarity. Positive serial correlations in time series, one of Markov type and the other with equal correlations, tend to make R_N and t too large, whereas independence without accurate weights makes them too small. The Cauchy distribution makes them too low, some others too high. Seldom will R_N be close to unity, especially in small samples, unless the standard conditions come close to being satisfied.

The conditions that R, which is one for random sampling from a normal distribution, be also one with independent random sampling, but without actual normality, are very special, and are developed in section 9. These are rather picturesque, and show that this kind of approximate normality has absolutely nothing to do with moments, or with the behavior of the frequency curve at infinity, or very near the center of symmetry if there is one, but are solely concerned with what happens in a small neighborhood of each point of inflection. Any correction of t for nonnormality should evidently be responsive to deviations from the conditions of section 9.

In some cases the trouble can be dispelled as an illusion, generated perhaps by too few or too ill-controlled observations.

When a remedy is required for the errors resulting from nonnormality, a first procedure sometimes available is to transform the variate to a normal form, or some approximation thereto. The transformation must be based on some real or assumed form of the population distribution, and this should be obtained from some positive evidence, which may be in an empirical good fit or in reasoning grounded on an acceptable body of theory regarding the nature of the variates observed. One form of the first, practiced by at least some psychologists, consists of new scores constituting a monotonic function of the raw scores but adjusted by means of a large sample so as to have a nearly normal distribution, thus bringing the battery of tests closer to the domain of normal correlation analysis. This is satisfactory from the standpoint of univariate normal distributions, though for consistency of multivariate normal theory additional conditions are necessary, which may or may not be contradicted by the facts.

The replacement of price relatives by their logarithms appears to be a sound practice. Indeed, a positive factor, the general level, persisting through all prices, has a meaning and movement of its own; and with it are independent fluctuations of a multiplicative character. All this points to the logarithms of the prices as having the right general properties of approximate symmetry and normality instead of the very skew distributions of price relatives themselves, taken in some definite sense such as a price divided by the price of the same good at a definite earlier time. There is also some empirical evidence that logarithms of such price relatives have normal, or at least symmetric, distributions. In preparing a program of logarithmic transformations it is desirable to provide for suitable steps in the rare instances of a price becoming zero, infinite, or negative. The advantages of replacing each price, with these rare exceptions, by its logarithm (perhaps four-place logarithms are best) at the very beginning of a study are considerable, and extend also to some other time variables in economics.

Nonparametric methods, including rank correlation and the use of contingency tables and of order statistics such as quantiles, provide a means of escape from such errors as applying normal theory, for example, through the Student ratio, to distributions that are not normal. Getting suitable exact probabilities based purely on ranks is a combinatorial problem after a decision has been reached in detail as to the alternative hypotheses to be compared, and the steps to be taken after each relevant set of observations. These however are often difficult decisions, lacking such definite criteria as occur naturally in parametric situations like those involving correlation coefficients in normal distributions. Changing from measures to ranks often implies some sacrifice of information. For example, in a sample from a bivariate normal distribution, it is possible to estimate the correlation parameter either by the sample product-moment correlation or by a function of the rank correlation coefficient; but the latter has a higher standard error.

In spite of such strictures, nonparametric methods may suitably be preferred in some situations. The median is a definitely useful statistic, and so, for some purposes, is M. G. Kendall's rank correlation coefficient.

One caution about nonparametric methods should be kept in mind. Their use escapes the possible errors due to nonnormality, but does nothing to avoid those of falsely assumed independence, or of heteroscedasticity, or of bias; and all these may be even greater threats to sound statistical inference than is nonnormality. These prevalent sources of trouble, and the relative inefficiency of nonparametric methods manifested in their higher standard errors where they compete with parametric methods based on correct models, must be weighed against their advantage of not assuming any particular type of population.

We have not until the last paragraph even mentioned biased errors of observation. Such errors, though troublesome and common enough, are in great part best discussed in connection with specific applications rather than general theory. However, the statistical theory of design of experiments does provide very material help in combating the age-old menace of bias, and at the same time contributes partial answers to the question at the head of this section.

Objective randomization in the allocation of treatments to experimental units, for example with the help of mechanisms of games of chance, serves an important function that has been described in at least two different ways. It may be called elimination of bias, since in a balanced and randomized experiment any bias tends to be distributed equally in opposite directions, and thereby to be transformed into an increment of variance. Randomization may also be regarded as a restoration of independence impaired by special identifiable features of individual units, as in shuffling cards to destroy any relations between their positions that may be known or suspected. This aspect of randomization offers a substantial opportunity to get rid of one of the three chief kinds of nonstandard conditions that weaken the applicability of standard procedures of mathematical statistics.

Normality, independence, and uniformity of variance can all be promoted by attention to their desirability during careful design and execution of experiments and analogous investigations, such as sample surveys. Emphasis during such research planning should be placed on features tending to normality, including use of composite observations, such as mental test scores, formed by addition of individual item scores; and the removal of major individual causes of variance leaving as dominant a usually larger number of minor causes, which by their convolutions produce a tendency to normality.

After factorial and other modern balanced experiments, the unknown quantities are commonly estimated by linear functions of a considerable number of observations, frequently with equal coefficients. In addition to the other benefits of such designs, this feature is conducive to close approach to normality in the distributions of these estimates. A case in point is in weighing several light objects, which is much better done together in various combinations than separately one at a time, especially if both pans of the scale can be used, weighing some combinations against others. In the development of this subject (Yates [62], Hotelling [25], Kishen [31], Mood [33]) the primary motives were cancellation of bias and reduction of variance, followed by advancement of the combinatorial theory. But from the point of view of the present inquiry, an important part of the gain is the increase in the number of independent observations combined linearly to estimate each unknown weight. Errors in weights arrived at through such experiments must have not only sharply reduced bias and variance, but much closer adherence to the normal form.

One other protection is available from the dangerous fallacies that have been the main subject of this paper. It lies in continual scrutiny of the sources of the observations in order to understand as fully as possible the nature of the random elements in them, as well as of the biases, to the end that the most accurate and reasonable models possible be employed. To employ the new models most effectively will also call frequently for dealing in an informed and imaginative manner with new problems of mathematical statistics.

REFERENCES

The references below have been selected mainly because of their bearing on the problems of distributions of t and F in nonstandard cases, with a few dealing instead with distributions of correlation and serial correlation coefficients and other statistics. The many papers on distributions of means alone or of standard deviations alone have been barred, save in a few cases such as Rietz' derivation of the distribution of s in samples of three from a rectangular distribution, which throws light on the nature of the difficulty of the problem for larger samples. A few papers on weighing designs have been included for a reason arising in the final paragraphs of the paper.

An effort has been made to list all studies giving particular attention to extreme tail probabilities of the several statistics; as to this approach, which is also much used in the present paper, it is possible the list is complete. But in the general field indicated by the title, completeness has not been attained, and could not be without a truly enormous expansion. It is with great regret that many meritorious papers have had to be omitted. But references are abundant, and well-selected lists can be found in many of the publications listed below. The short bibliographies of Bradley and Rietz are extremely informative regarding the condition of our subject when they wrote. Kendall's is virtually all-inclusive, and when taken with appropriate parts of his text gives the most nearly adequate conspectus of the great and devoted efforts that have gone into the field since the First World War.

A relatively new nonstandard situation for application of standard methods is that in which different members of a sample are drawn from normal populations with different means. This is not treated in the present paper, but is in papers by Hyrenius, Quensel, Robbins, and Zachrisson, listed below.

[1] G. A. Baker, "Distribution of the mean divided by the standard deviations of samples from non-homogeneous populations," *Ann. Math. Statist.*, Vol. 3 (1932), pp. 1–9.

[2] M. S. Bartlett, "The effects of non-normality on the t-distribution," *Proc. Cambridge Philos. Soc.*, Vol. 31 (1935), pp. 223–231.

[3] G. E. P. Box, "Non-normality and tests on variances," *Biometrika*, Vol. 40 (1953), pp. 318–335.

[4] R. A. Bradley, "Sampling distributions for non-normal universes," unpublished Ph.D. thesis, Chapel Hill, University of North Carolina, 1949.

[5] ———, "The distribution of the t and F statistics for a class of non-normal populations," *Virginia J. Sci.*, Vol. 3 (NS) (1952), pp. 1–32.

[6] ———, "Corrections for non-normality in the use of the two-sample t and F tests at high significance levels," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 103–113.

[7] K. L. Chung, "The approximate distribution of Student's statistic," *Ann. Math. Statist.*, Vol. 17 (1946), pp. 447–465.

[8] A. T. Craig, "The simultaneous distribution of mean and standard deviation in small samples," *Ann. Math. Statist.*, Vol. 3 (1932), pp. 126–140.

[9] H. Cramér, *Mathematical Methods of Statistics*, Princeton, Princeton University Press, 1946.

[10] F. N. David, *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples*, London, *Biometrika*, 1938.

[11] ———, "The moments of the z and F distributions," *Biometrika*, Vol. 36 (1949), pp. 394–403.

[12] F. N. David and N. L. Johnson, "The effect of non-normality on the power function of the F-test in the analysis of variance," *Biometrika*, Vol. 38 (1951), pp. 43–57.

[13] A. Erdélyi, *Asymptotic Expansions*, New York, Dover, 1956.

[14] R. A. Fisher, *Statistical Methods for Research Workers*, Edinburgh, Oliver and Boyd, 1925 (1st ed.).

[15] ———, "Applications of Student's distribution," *Metron*, Vol. 5 (1925), pp. 90–104.

[16] A. K. Gayen, "The distribution of 'Student's' t in random samples of any size drawn from non-normal universes," *Biometrika*, Vol. 36 (1949), pp. 353–369.

[17] ———, "The distribution of the variance ratio in random samples of any size drawn from non-normal universes," *Biometrika*, Vol. 37 (1950), pp. 236–255.

[18] ———, "The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes," *Biometrika*, Vol. 38 (1951), pp. 219–247.

[19] R. C. Geary, "The distribution of Student's ratio for non-normal samples," *J. Roy. Statist. Soc., Suppl.*, Vol. 3 (1946), pp. 178–184.

[20] G. B. Hey, "A new method of experimental sampling illustrated in certain non-normal populations," *Biometrika*, Vol. 30 (1958), pp. 8–80. (Contains an important bibliography.)

[21] W. Hoeffding, "The role of hypotheses in statistical decisions," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California, 1956, Vol. 1, pp. 105–116.

[22] H. Hotelling, "An application of analysis situs to statistics," *Bull. Amer. Math. Soc.*, Vol. 33 (1927), pp. 467–476.

[23] ———, "Tubes and spheres in n-spaces and a class of statistical problems," *Amer. J. Math.*, Vol. 61 (1939), pp. 440–460.

[24] ———, "Effects of non-normality at high significance levels" (abstract), *Ann. Math. Statist.*, Vol. 18 (1947), p. 608.

[25] ———, "Some improvements in weighing and other experimental techniques," *Ann. Math. Statist.*, Vol. 15 (1944), pp. 297–306.

[26] ———, "New light on the correlation coefficient and its transforms," *J. Roy. Statist. Soc., Ser. B*, Vol. 15 (1953), pp. 193–232.

[27] H. HYRENIUS, "Sampling distributions from a compound normal parent population," *Skand. Aktuarietidskr.*, Vol. 32 (1949), pp. 180–187.

[28] ———, "Distribution of 'Student'-Fisher's t in samples from compound normal functions," *Biometrika*, Vol. 37 (1950), pp. 429–442.

[29] E. S. KEEPING, "A significance test for exponential regression," *Ann. Math. Statist.*, Vol. 22 (1950), pp. 180–198.

[30] M. G. KENDALL, *The Advanced Theory of Statistics*, 2 vols., London, Griffin, 1948.

[31] K. KISHEN, "On the design of experiments for weighing and making other types of measurements," *Ann. Math. Statist.*, Vol. 16 (1945), pp. 294–300.

[32] J. LADERMAN, "The distribution of 'Student's' ratio for samples of two items drawn from non-normal universes," *Ann. Math. Statist.*, Vol. 10 (1939), pp. 376–379.

[33] A. M. MOOD, "On Hotelling's weighing problem," *Ann. Math. Statist.*, Vol. 17 (1946), pp. 432–446.

[34] J. NEYMAN, "On the correlation of the mean and variance in samples drawn from an infinite population," *Biometrika*, Vol. 18 (1926), pp. 401–413.

[35] J. NEYMAN and E. S. PEARSON, "On the use and interpretation of certain test criteria for purposes of statistical inference," *Biometrika*, Vol. 20A (1928), pp. 175–240.

[36] E. S. PEARSON, assisted by N. K. Adayantha and others; "The distribution of frequency constants in small samples from non-normal symmetrical and skew populations," *Biometrika*, Vol. 21 (1929), pp. 259–286.

[37] E. S. PEARSON, "The analysis of variance in cases of non-normal variation," *Biometrika*, Vol. 23 (1931), pp. 114–133.

[38] E. S. PEARSON, L. CHESIRE and E. OLDIS, "Further experiments on the sampling distribution of the correlation coefficient," *J. Amer. Statist. Assoc.*, Vol. 27 (1932), pp. 121–128.

[39] E. S. PEARSON, "The test of significance for the correlation coefficient: some further results," *J. Amer. Statist. Assoc.*, Vol. 27 (1932), pp. 424–426.

[40] ———, "The test of significance for the correlation coefficient," *J. Amer. Statist. Assoc.* Vol. 26 (1931), pp. 128–134.

[41] V. PERLO, "On the distribution of Student's ratio for samples of three drawn from a rectangular distribution," *Biometrika*, Vol. 25 (1933), pp. 203–204.

[42] C. E. QUENSEL, "An extension of the validity of Student-Fisher's law of distribution," *Skand. Aktuarietidskr.*, Vol. 26 (1943), pp. 210–219. The distribution is examined of t for normally distributed observations having a common mean but different variances, and is found to be more closely concentrated about zero than in the standard case.)

[43] ———, "The validity of the z-criterion when the variates are taken from different normal populations," *Skand. Aktuarietidskr.*, Vol. 30 (1947), pp. 44–45. (Generalizes the foregoing to analysis of variance, with similar results.)

[44] ———, "The distribution of the second-order moments in random samples from non-normal multivariate universes," *Lunds Univ. Årsskrift. N. F. Avd*, 2. No. 4, 1952.

[45] P. R. RIDER, "On the distribution of the ratio of mean to standard deviation in small samples from certain non-normal universes," *Biometrika*, Vol. 21 (1929), pp. 124–143.

[46] ———, "On small samples from certain non-normal universes," *Ann. Math. Statist.*, Vol. 2 (1931), pp. 48–65.

[47] ———, "A note on small sample theory," *J. Amer. Statist. Assoc.*, Vol. 26 (1931), pp. 172–174.

[48] ———, "On the distribution of the correlation coefficients in small samples," *Biometrika*, Vol. 24 (1932), pp. 382–403.

[49] H. L. Rietz, "Note on the distribution of the standard deviation of sets of three variates drawn at random from a rectangular distribution," *Biometrika*, Vol. 23 (1931), pp. 424–426.
[50] ———, "Comments on applications of recently developed theory of small samples," *J. Amer. Statist. Assoc.*, Vol. 26 (1931), pp. 36–44.
[51] ———, "Some topics in sampling theory," *Bull. Amer. Math. Soc.*, Vol. 43 (1937), pp. 209–230.
[52] ———, "On the distribution of the 'Student' ratio for small samples from certain non-normal populations," *Ann. Math. Statist.*, Vol. 10 (1939), pp. 265–274.
[53] H. Robbins, "The distribution of Student's t when the population means are unequal," *Ann. Math. Statist.*, Vol. 19 (1948), pp. 406–410.
[54] W. A. Shewhart and F. W. Winters, "Small samples—new experimental results," *J. Amer. Statist. Assoc.*, Vol. 23 (1928), pp. 144–153. (This article combines experimental with other approaches, especially that of Neyman [34].)
[55] M. M. Siddiqui, "Distribution of a serial correlation coefficient near the ends of the range," *Ann. Math. Statist.*, Vol. 29 (1958), pp. 852–861.
[56] ———, "Distributions of some serial correlation coefficients," unpublished Ph.D. thesis, Chapel Hill, University of North Carolina, 1957.
[57] D. M. Starkey, "The distribution of the multiple correlation coefficient in periodogram analysis," *Ann. Math. Statist.*, Vol. 10 (1939), pp. 327–336.
[58] M. Weibull, "The distribution of the t and z variables in the case of stratified sample with individuals taken from normal parent populations with varying means," *Skand. Aktuarietidskr.*, Vol. 33 (1950), pp. 137–167.
[59] ———, "The regression problem involving non-random variates in the case of stratified sample from normal parent populations with varying regression coefficients," *Skand. Aktuarietidskr.*, Vol. 34 (1951), pp. 53–71.
[60] ———, "The distribution of t- and F-statistics and of correlation and regression coefficients in stratified samples from normal populations with different means," *Skand. Aktuarietidskr.*, Vol. 36 (1953), Suppl. 1–2, 106 pp.
[61] D. V. Widder, *The Laplace Transform*, Princeton, Princeton University Press, 1941.
[62] F. Yates, "Complex experiments," *J. Roy. Statist. Soc., Suppl.*, Vol. 2 (1935), pp. 181–247.
[63] U. Zackrisson, *The Distribution of "Student's" t in Samples from Individual Non-normal Populations*, Statistical Institute, University of Gothenburg, Publ. No. 6, 1959.