AN APPROXIMATE FORMULA FOR THE CUMULATIVE z-DISTRIBUTION[1]

By John Wishart[2]

Statistical Laboratory, University of Cambridge and Princeton University

1. Summary. A straightforward expansion and integration of the frequency function for Fisher's z produces a formula for the probability that z is not exceeded, of which the successive terms decrease rapidly when n_1 and n_2 are large. It is given in terms of incomplete normal moment functions (or χ^2 probabilities), and as a polynomial in $zN^{1/2}$, where N is the harmonic mean of n_1 and n_2. This last form is identical with the inverted Cornish-Fisher expansion, originally deduced by quite different methods.

2. To obtain their well-known expansion for determining percentage points for the distribution of z (one-half of the natural logarithm of the ratio of two independent variance estimates from normal data) in cases where the degrees of freedom n_1 and n_2 are large, Cornish and Fisher (1937) used the method of the normalizing transformation. They developed a Gram-Charlier Type A series expansion which required knowledge of the cumulants of z. These they worked out in the approximate form for large n_1 and n_2, to a point sufficient for the order of approximation worked to. The method is rather complicated, but a final formula is given which enables chosen percentage points to be determined. Although it is possible by substitution to deduce the corresponding formula for determining the probability associated with a chosen value of z, the author does not recall having seen such a formula explicitly stated.[3]

3. The frequency function of z may be manipulated directly so as to give on integration this inverted formula. The method is direct and simple, requires no Gram-Charlier Type A series, and no cumulants.

Consider two independent variance estimates s_1^2 and s_2^2 from normal data, having degrees of freedom n_1 and n_2. z is then $\frac{1}{2}\ln(s_1^2/s_2^2)$. For the time being write $\frac{1}{2}n_1$ as c_1 and $\frac{1}{2}n_2$ as c_2. Then the frequency function of z is

$$\frac{2(c_1/c_2)^{c_1}}{\mathrm{B}(c_1, c_2)} \frac{e^{2c_1z}}{(1 + c_1e^{2z}/c_2)^{c_1+c_2}}, \tag{1}$$

where the range of z is from $-\infty$ to ∞, and $\mathrm{B}(c_1, c_2)$ is the Beta-Function, equal

Received February 22, 1956.

[1] Research partially supported by the Office of Naval Research.

[2] This paper was recommended for publication after the death of the author. It is published, with minor emendations, after consulting with a colleague of the author. The original title was "A new derivation of the inverted Cornish-Fisher expansion for the z-distribution." Ed.

[3] Reviewers note. Campbell [1] gave an expression for finding percentage points; he did not require a knowledge of cumulants or use the Gram-Charlier Type A series expansion. Student [4] used such an expansion for t to compute his original table, and Fisher [3] developed the expansion by methods similar to those used here.

to $\Gamma(c_1)\Gamma(c_2) / \Gamma(c_1 + c_2)$. We shall take n_1 as the smaller of the degrees of freedom, so that $c_1 / c_2 \leqq 1$. The frequency function may be written

$$\begin{aligned} &\frac{2(c_1/c_2)^{c_1}}{\mathrm{B}(c_1, c_2)} \exp \{2c_1 z - (c_1 + c_2) \ln (1 + c_1 e^{2z}/c_2)\} \\ &\qquad = \frac{2(c_1/c_2)^{c_1}}{\mathrm{B}(c_1, c_2)} \exp \left[2c_1 z - (c_1 + c_2) \ln \frac{c_1 + c_2}{c_2} \right. \\ &\qquad\qquad \left. - (c_1 + c_2) \ln \left\{1 + \frac{c_1(e^{2z} - 1)}{c_1 + c_2}\right\}\right]. \end{aligned} \tag{2}$$

The first logarithm can be put into the outside term, and the second may be expanded, noting that $c_1(e^{2z} - 1) / (c_1 + c_2)$ will lie between $+1$ and -1 except in the extreme tail of the distribution when n_1 and n_2 are nearly equal and of the order of 30 or less.

The frequency function then becomes

$$\begin{aligned} &\frac{2c_1^{c_1}c_2^{c_2}}{(c_1 + c_2)^{c_1+c_2}\mathrm{B}(c_1, c_2)} \exp \left[2c_1 z - c_1(2z) - \frac{C}{2} \cdot \frac{(2z)^2}{2!} - \frac{C}{2} \frac{c_2 - c_1}{c_1 + c_2} \frac{(2z)^3}{3!} \right. \\ &\qquad - \frac{C}{2}\left(1 - \frac{3C}{c_1 + c_2}\right)\frac{(2z)^4}{4!} - \frac{C}{2}\frac{c_2 - c_1}{c_1 + c_2}\left(1 - \frac{6C}{c_1 + c_2}\right)\frac{(2z)^5}{5!} \\ &\qquad \left. - \frac{C}{2}\left(1 - \frac{15C}{c_1 + c_2} + \frac{30C^2}{(c_1 + c_2)^2}\right)\frac{(2z)^6}{6!} - \cdots\right]. \end{aligned} \tag{3}$$

where C is the harmonic mean of c_1 and $c_2 = 2\, c_1 c_2 / (c_1 + c_2)$.

Now put $2z = x(2/C)^{1/2}$, whereupon the frequency function may be written

$$\begin{aligned} &\frac{\sqrt{(2\pi)}\, c_1^{c_1-1/2} c_2^{c_2-1/2}}{(c_1 + c_2)^{c_1+c_2-1/2}\mathrm{B}(c_1, c_2)} \cdot \frac{e^{-x^2/2}}{\sqrt{(2\pi)}} \\ &\qquad \cdot \exp\left\{-\left[\left(\frac{2}{C}\right)^{0.5} \cdot \frac{c_2 - c_1}{c_1 + c_2} \cdot \frac{x^3}{3!} + \frac{2}{C}\left(1 - \frac{3C}{c_1 + c_2}\right)\frac{x^4}{4!} + \left(\frac{2}{C}\right)^{1.5}\right.\right. \\ &\qquad \left.\left. \cdot \frac{c_2 - c_1}{c_1 + c_2}\left(1 - \frac{6C}{c_1 + c_2}\right)\frac{x^5}{5!} + \left(\frac{2}{C}\right)^2\left(1 - \frac{15C}{c_1 + c_2} + \frac{30C^2}{(c_1 + c_2)^2}\right)\frac{x^6}{6!} + \cdots\right]\right\}. \end{aligned} \tag{4}$$

On expanding the Γ-functions in $\mathrm{B}(c_1, c_2)$ by Stirling's formula, the first part of this expression becomes approximately

$$\begin{aligned} &\left(1 + \frac{1}{12(c_1 + c_2)} + \frac{1}{288(c_1 + c_2)^2}\right)\left(1 + \frac{1}{12c_1} + \frac{1}{288c_1^2}\right)^{-1} \\ &\qquad \cdot \left(1 + \frac{1}{12c_2} + \frac{1}{288c_2^2}\right)^{-1} = 1 - \frac{1}{6N}\left(2 - \frac{N}{n_1 + n_2}\right) \\ &\qquad\qquad + \frac{1}{72N^2}\left(2 - \frac{N}{n_1 + n_2}\right)^2 \end{aligned} \tag{5}$$

in terms of n_1 and n_2 and their harmonic mean N.

The second part of (4) is the normal frequency function, and the third part may be expanded into the following series in terms of n_1, n_2 and N:

$$\begin{aligned}
&1 - \frac{1}{3N^{0.5}} \cdot \frac{n_2 - n_1}{n_1 + n_2} x^3 + \frac{1}{18N}\left[\left(1 - \frac{2N}{n_1 + n_2}\right)x^6 - 3\left(1 - \frac{3N}{n_1 + n_2}\right)x^4\right] \\
&- \frac{1}{810N^{1.5}} \cdot \frac{n_2 - n_1}{n_1 + n_2}\left[5\left(1 - \frac{2N}{n_1 + n_2}\right)x^9 - 45\left(1 - \frac{3N}{n_1 + n_2}\right)x^7\right. \\
(6)\quad &\left. + 54\left(1 - \frac{6N}{n_1 + n_2}\right)x^5\right] + \frac{1}{9720N^2}\left[5\left(1 - \frac{2N}{n_1 + n_2}\right)^2 x^{12}\right. \\
&- 90\left(1 - \frac{2N}{n_1 + n_2}\right)\left(1 - \frac{3N}{n_1 + n_2}\right)x^{10} \\
&\left. + 27\left(13 - \frac{94N}{n_1 + n_2} + \frac{141N^2}{(n_1 + n_2)^2}\right)x^8 - 216\left(1 - \frac{15N}{n_1 + n_2} + \frac{30N^2}{(n_1 + n_2)^2}\right)x^6\right]
\end{aligned}$$

as far as terms in N^{-2}.

We now have a frequency function for the variable $x = zN^{1/2}$ $(-\infty \leqq x \leqq \infty)$ in terms of the normal frequency function multiplied by a polynomial in x. For a chosen X, the probability $P(x \leqq X)$ is given by the integral of the frequency function from $-\infty$ to X. Alternative forms can be found for the result of the integration. We may express it in terms of Pearson's incomplete normal moment functions

$$\mu_r(X) = \frac{1}{\sqrt{(2\pi)}} \int_0^X x^r e^{-x^2/2}\, dx$$

$$m_r(X) = \frac{\mu_r(X)}{(r-1)(r-3)\cdots 1 \text{ or } 2}$$

according as r is even or odd. Numerical values for $m_{12}(X)$ are given to seven decimals in *Tables for Statisticians and Biometricians* (Pearson, 1914, 1931), while $\mu_0(X) = P(X) - 0.5$, where

$$P(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^X e^{-x^2/2}\, dx$$

is given elsewhere in the same *Tables* (Table II), and also by Pearson and Hartley (1954), Table 1.

In this form the probability $P(0 \leqq x \leqq X)$ is

$$\begin{aligned}
&\left[1 - \frac{1}{6N}\left(2 - \frac{N}{n_1 + n_2}\right) + \frac{1}{72N^2}\left(2 - \frac{N}{n_1 + n_2}\right)^2\right] \\
&\quad \cdot \left\{\mu_0(X) - \frac{2}{3N^{0.5}} \cdot \frac{n_2 - n_1}{n_1 + n_2} m_3(X)\right. \\
&+ \frac{1}{6N}\left[5\left(1 - \frac{2N}{n_1 + n_2}\right)m_6(X) - 3\left(1 - \frac{3N}{n_1 + n_2}\right)m_4(X)\right]
\end{aligned}$$

$$
(7)\quad \begin{aligned}
&- \frac{8}{135N^{1.5}} \cdot \frac{n_2 - n_1}{n_1 + n_2}\left[40\left(1 - \frac{2N}{n_1 + n_2}\right) m_9(X)\right.\\
&\left. - 45\left(1 - \frac{3N}{n_1 + n_2}\right) m_7(X) + 9\left(1 - \frac{6N}{n_1 + n_2}\right) m_5(X)\right]\\
&+ \frac{1}{72N^2}\left[385\left(1 - \frac{2N}{n_1 + n_2}\right)^2 m_{12}(X) - 630\left(1 - \frac{2N}{n_1 + n_2}\right)\right.\\
&\cdot\left(1 - \frac{3N}{n_1 + n_2}\right) m_{10}(X) + 21\left(15 - \frac{94N}{n_1 + n_2} + \frac{141N^2}{(n_1 + n_2)^2}\right) m_8(X)\\
&\left.\left. - 24\left(1 - \frac{15N}{n_1 + n_2} + \frac{30N^2}{(n_1 + n_2)^2}\right) m_6(X)\right]\right\}.
\end{aligned}
$$

The probability $P(-\infty \leqq x \leqq 0)$ is got from (7) as a special case by putting $\mu_0(-\infty) = m_{2s}(-\infty) = 0.5$, and $m_{2s+1}(-\infty) = (-2\pi)^{-1/2}$. It then becomes[4]

$$
(8)\quad \begin{aligned}
&\left[1 - \frac{1}{6N}\left(2 - \frac{N}{n_1 + n_2}\right) + \frac{1}{72N^2}\left(2 - \frac{N}{n_1 + n_2}\right)^2\right]\\
&\quad\cdot\left\{\frac{1}{2} + \frac{1}{12N}\left(2 - \frac{N}{n_1 + n_2}\right) + \frac{1}{144N^2}\left(2 - \frac{N}{n_1 + n_2}\right)^2\right.\\
&\qquad\left. + \frac{2}{3\sqrt{(2\pi N)}} \cdot \frac{n_2 - n_1}{n_1 + n_2}\left[1 + \frac{4}{45N}\left(4 + \frac{N}{n_1 + n_2}\right)\right]\right\}\\
&\qquad = \frac{1}{2} + \frac{2}{3\sqrt{(2\pi N)}} \cdot \frac{n_2 - n_1}{n_1 + n_2}\left[1 + \frac{1}{90N}\left(2 + \frac{23N}{n_1 + n_2}\right)\right].
\end{aligned}
$$

The sum or difference of (7) and (8), according as X is positive or negative, gives the probability $P(-\infty \leqq x \leqq X)$.

Alternatively we may write in (7)

$$
\begin{aligned}
(2\pi)^{\frac{1}{2}} m_{2r+1}(X) &= P(X^2 \mid 2r + 2),\\
2m_{2r}(X) &= P(X^2 \mid 2r + 1),
\end{aligned}
$$

where $P(X^2 \mid \nu)$ denotes the probability that χ^2 does not exceed X^2, for ν degrees of freedom. These probabilities may be obtained to five decimals by subtracting from unity the χ^2 probabilities given in Pearson and Hartley (1954), Table 7.

A series expansion for the probability $P(-\infty \leqq x \leqq X)$ can be obtained in terms of $P(X)$ and $Z(X) = e^{-x^2/2} / \nu(2\pi)$, together with a polynomial in X, by associating (6) with $Z(x)$ and integrating term by term by parts. This gives the required probability as

$$
\left\{1 - \frac{1}{6N}\left(2 - \frac{N}{n_1 + n_2}\right) + \frac{1}{72N^2}\left(2 - \frac{N}{n_1 + n_2}\right)^2\right\}
$$

[4] Reviewer's note. The algebraic signs for $m_{2s}(-\infty)$ should be the opposites of those given here; when X is negative, (7) takes negative values.

$$\cdot \Bigg[P(X) + \frac{n_2 - n_1}{n_1 + n_2} \cdot \frac{X^2 + 2}{3N^{0.5}} Z(X) - \frac{1}{18N} \Bigg\{ \left(1 - \frac{2N}{n_1 + n_2}\right) X^5$$

$$+ \left(2 - \frac{N}{n_1 + n_2}\right) (X^3 + 3X) \Bigg\} Z(X) + \frac{1}{6N} \left(2 - \frac{N}{n_1 + n_2}\right) P(X)$$

$$+ \frac{1}{810N^{1.5}} \frac{n_2 - n_1}{n_1 + n_2} \Bigg\{ 5 \left(1 - \frac{2N}{n_1 + n_2}\right) X^8 - 5 \left(1 - \frac{11N}{n_1 + n_2}\right) X^6$$

$$+ 6 \left(4 + \frac{N}{n_1 + n_2}\right) (X^4 + 4X^2 + 8) \Bigg\} Z(X)$$

$$- \frac{1}{9720N^2} \Bigg\{ 5 \left(1 - \frac{2N}{n_1 + n_2}\right)^2 X^{11} - 5 \left(1 - \frac{2N}{n_1 + n_2}\right)$$

$$\cdot \left(7 - \frac{32N}{n_1 + n_2}\right) X^9 + 9 \left(4 - \frac{52N}{n_1 + n_2} + \frac{103N^2}{(n_1 + n_2)^2}\right) X^7$$

$$+ 9 \left(2 - \frac{N}{n_1 + n_2}\right)^2 (X^5 + 5X^3 + 15X) \Bigg\} Z(X)$$

$$+ \frac{1}{72N^2} \left(2 - \frac{N}{n_1 + n_2}\right)^2 P(X) \Bigg].$$

On multiplying in by the outside factor this becomes

$$P(X) + Z(X) \Bigg[\frac{n_2 - n_1}{n_1 + n_2} \cdot \frac{X^2 + 2}{3N^{0.5}} - \frac{1}{18N} \Bigg\{ \left(1 - \frac{2N}{n_1 + n_2}\right) X^5$$

$$+ \left(2 - \frac{N}{n_1 + n_2}\right) (X^3 + 3X) \Bigg\} + \frac{1}{810N^{1.5}} \frac{n_2 - n_1}{n_1 + n_2} \Bigg\{ 5 \left(1 - \frac{2N}{n_1 + n_2}\right) X^8$$

$$- 5 \left(1 - \frac{11N}{n_1 + n_2}\right) X^6 + 6 \left(4 + \frac{N}{n_1 + n_2}\right) X^4$$

$$+ 3 \left(2 + \frac{23N}{n_1 + n_2}\right) \left(X^2 + 2\right) \Bigg\} - \frac{1}{9720N^2} \Bigg\{ 5 \left(1 - \frac{2N}{n_1 + n_2}\right)^2 X^{11} \tag{9}$$

$$- 5 \left(1 - \frac{2N}{n_1 + n_2}\right) \left(7 - \frac{32N}{n_1 + n_2}\right) X^9 + 9 \left(4 - \frac{52N}{n_1 + n_2}\right.$$

$$\left. + \frac{103N^2}{(n_1 + n_2)^2}\right) X^7 - 9 \left(2 - \frac{N}{n_1 + n_2}\right) \left(8 - \frac{19N}{n_1 + n_2}\right) X^5$$

$$- 45 \left(2 - \frac{N}{n_1 + n_2}\right)^2 (X^3 + 3X) \Bigg\} \Bigg].$$

This is the expression which is the "direct" form of the Cornish-Fisher expansion, yielding, to terms in N^{-2}, the probability that z shall not exceed $xN^{-1/2}$. Additional terms could be worked out by noting that the terms of the exponential in (3) are equivalent to the binomial cumulants, but the terms in (9) should suffice for W of the order of 50 or above, and fewer terms will do if the probability is not required to a large number of significant figures.

For the benefit of those accustomed to the notation of Cornish and Fisher, (9) may be put into the form

$$
\begin{aligned}
(10)\quad & \mu + z\Bigg[\frac{\sqrt{(\frac{1}{2}\sigma)}}{3}\cdot\frac{\delta}{\sigma}(X^2+2) - \frac{\frac{1}{2}\sigma}{36}\left\{3(X^3+3X)+\frac{\delta^2}{\sigma^2}(2X^5+X^3+3X)\right\} \\
& + \frac{(\frac{1}{2}\sigma)^{1.5}}{1620}\cdot\frac{\delta}{\sigma}\left\{9(5X^6+6X^4+9X^2+18)+\frac{\delta^2}{\sigma^2}\right. \\
& \left.\cdot(10X^8-55X^6-6X^4-69X^2-138)\right\} - \frac{(\frac{1}{2}\sigma)^2}{38880} \\
& \cdot\left\{27(5X^7+3X^5-15X^5-45X)+18\frac{\delta^2}{\sigma^2}\right. \\
& \cdot(10X^9-51X^7-27X^5-15X^3-45X) \\
& \left.+\frac{\delta^4}{\sigma^4}(20X^{11}-320X^9+927X^7-171X^5-45X^3-135X)\right\}\Bigg],
\end{aligned}
$$

but note that z here is what we have hitherto written as $Z(X)$. In using (10) we take X as the chosen value of Fisher's z divided by $\sqrt{(\frac{1}{2}\sigma)}$, i.e. by its approximate standard derivation, $[\frac{1}{2}(1/n_1 + 1/n_2)]^{1/2}$; δ is, of course, $1/n_1 - 1/n_2$.

The order of the terms in (9) or (10) may be seen if we choose as an example $n_1 = 60, n_2 = 120, Z = (\sqrt{5})/20 = 0.1118034$. Then $N = 80$, while $\sqrt{(\frac{1}{2}\sigma)} = (\sqrt{5})/20$ and $\delta/\sigma = \frac{1}{3}$, and $X = 1$. Using Pearson and Hartley's Table 1 we find for the probability that this chosen value of Z is not exceeded

	0.8413	447
+	90	177
−	10	642
+		218
+		16
	0.8493	216

so that we are here close to the 15 per cent point of the z distribution.

When $n_1 = n_2 = n$ we have $N = n$, $N/(n_1 + n_2) = \frac{1}{2}$, also $\delta = 0$, $\frac{1}{2}\sigma = n^{-1}$. Then (7) and (8) give

$$(11)\quad \frac{1}{2} + \left(1 - \frac{1}{4n} + \frac{1}{32n^2}\right)\left(\mu_0(X) + \frac{m_4(X)}{4n} - \frac{32m_6(X) - 35m_8(X)}{96n^2}\right),$$

while (9) or (10) gives

$$(12)\quad P(X) - Z(X)\left\{\frac{X(X^2+3)}{12n} + \frac{5X^7 + 3X^5 - 15X(X^2+3)}{1440n^2}\right\}.$$

When $n_1 = n$, $n_2 = \infty$, we have an expansion from which can be calculated the probability that a chosen value of χ^2, for n degrees of freedom, is not exceeded.

If this value be denoted χ_0^2, then we take $X = \sqrt{(\frac{1}{2}n)}\cdot\ln(\chi_0^2/n)$, so that we are effectively transforming χ^2 by first forming the ratio of χ^2 to its mean, raised to the power of its standard deviation, and then taking one-half the natural logarithm of this quantity. The expansion for the probability may be obtained from (7) and (8), or from (9), by putting $N = 2n$, $(n_2 - n_1)/(n_1 + n_2) = 1$ and $N / (n_1 + n_2) = 0$, or from (10) with $\sigma = \delta = n^{-1}$. It has been developed from first principles by the author in [7].

REFERENCES

[1] G. A. Campbell, "Probability curves showing Poisson's exponential summation," *Bell System Technical Journal*, Vol. 2 (1923), pp. 95–113; and Collected Papers, N. Y. (1937), pp. 224–242.

[2] E. A. Cornish and R. A. Fisher, *Revue de l'Institut International de Statistique*, Vol. 4 (1937), p. 307.

[3] R. A. Fisher, "The asymptotic approach to Behrens' integral with further tables for the d-test of significance," *Ann. Eng. Lond.*, Vol. 11 (1941), p. 151.

[4] M. G. Kendall, "Advanced Theory of Statistics," Vol. II, Section 21.10, p. 101.

[5] K. Pearson, *Tables for Statisticians and Biometricians*, Part I (Table IX); Part II (Table XIII), Biometric Laboratory, London, 1914, 1931.

[6] E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. I, Cambridge University Press, 1954.

[7] J. Wishart, "χ^2 probabilities for large numbers of degrees of freedom" *Biometrika*, Vol. 43 (1956), pp. 92–95.

THE MIXTURE OF NORMAL DISTRIBUTIONS WITH DIFFERENT VARIANCES[1]

By D. Teichroew[2]

University of California, Los Angeles

1. Introduction. In some practical problems, the observed variable may have a normal distribution whose variance varies from one observation to the next. The purpose of this note is to give the formula for the marginal distribution when the variances are assumed to be distributed according to the Gamma distribution.

2. The distribution in the general case. We assume that the conditional density of X, given σ^2, is

$$f(x/\sigma^2) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} e^{-x^2/2\sigma^2} \qquad -\infty < x < \infty, \qquad \sigma^2 > 0,$$

Received May 25, 1956; revised July 18, 1956.

[1] The preparation of this paper was sponsored (in part) by the Office of Naval Research, USN.

[2] Present address: National Cash Register Company, Hawthorne, California.