NORMALIZING THE NONCENTRAL t AND F DISTRIBUTIONS

By Nico F. Laubscher

South African Council for Scientific and Industrial Research and Cornell University

1. Introduction and Summary. Let X be a random variable governed by one of a family of distributions which is conveniently parameterized by μ, the expectation of X, so that, in particular, the variance of X, σ^2, is a function of μ, which we denote by $\sigma^2(\mu)$. A transformation, $\psi(X)$, is sometimes sought so that the variance of $\psi(X)$, as μ sweeps over its domain, is independent of μ (or much more nearly constant than $\sigma^2(\mu)$).

A standard method of obtaining such a transformation for stabilization of the variance is to consider X as one of a sequence of random variables, the sequence converging asymptotically in distribution, usually to a normal distribution. One form of the basic theorem is stated and proved by C. R. Rao [8], pp. 207–8, as follows.

Theorem (Rao). *If X is asymptotically normally distributed about μ, with asymptotic variance $\sigma^2(\mu)$, then any function $\psi = \psi(X)$, with continuous first derivative in some neighborhood of μ, is asymptotically normally distributed with mean $\psi(\mu)$ and variance $\sigma^2(\mu)(d\psi/d\mu)^2$, where $(d\psi/d\mu)$ denotes the derivative of $\psi(X)$ with respect to X, evaluated at the point μ.*

From this we immediately have the following well-known

Corollary. *The random variable*

$$\psi(X) = c \int_K^X \frac{d\mu}{\sigma(\mu)}, \tag{1}$$

where $0 < x < \infty$, and where K is an arbitrary constant, has a variance which is stabilized asymptotically at c^2.

It is assumed, of course, that the integrand in (1) is integrable. If $\psi(X)$ is not a real-valued function on the domain of X, then the mapping is meaningless.

Transformations such as (1), perhaps slightly modified, not only often work well for stabilizing non-asymptotic variances, but also often serve as well to normalize non-normal distributions. In general, however, nothing is known about the relative closeness to normality of the distribution of a random variable before and after a variance-stabilizing transformation is applied. Nor can anything general be said about the relative rapidity of approach to asymptotic normality.

The study of concrete examples, however, suggests some connection between variance stabilization and normalization of non-normal distributions. A theoretical connection that may be relevant in certain cases has been put forward by N. L. Johnson [3], pp. 150–1. Johnson shows that, when the random variable of interest has a certain structure, then the differential equation for the normalizing

Received April 15, 1959; revised July 18, 1960.

transformation is similar to the differential equation for the variance-stabilizing transformation. The specified structure is that $X_n = Y_1 + Y_2G(X_1) + \cdots\cdots + Y_nG(X_{n-1})$, where the Y's are independent and small, and $G(\cdot)$ is some function.

In what follows, we obtain the variance-stabilizing transformation for the noncentral t distributions and consider its normalizing properties. We repeat the same procedure for the topside noncentral F distributions, although the variance-stabilizing transformation in this case is not well-defined. We then derive two other (well-defined) transformations for the approximate normalization of the topside noncentral F. Numerical comparisons of these approximations and the exact values are given.

2. Noncentral ***t***. If U and V are independent random variables and U is $N(0, 1)$, V is χ^2/n (where χ^2 denotes the central chi-square variable with n degrees of freedom), and δ is a real number, then the random variable defined by $t = (U + \delta)V^{-\frac{1}{2}}$ is known as the noncentral t variable with n degrees of freedom and with noncentrality parameter δ. We assume throughout that $n \geqq 4$. The first moment about zero, and the second and third central moments of noncentral t were obtained by Johnson and Welch [4]. These moments are, respectively,

$$\mu = (\tfrac{1}{2}n)^{\frac{1}{2}}\delta\Gamma(\tfrac{1}{2}n - \tfrac{1}{2})/\Gamma(\tfrac{1}{2}n),$$

$$\mu_2 = [n(1 + \delta^2)/(n - 2)] - \mu^2,$$

and

$$\mu_3 = \mu\{n(\delta^2 + 2n - 3)/[(n - 2)(n - 3)] - 2\mu_2\}.$$

Eliminating δ between μ and μ_2 we find that

$$\mu_2 = a^2 + b^2\mu^2$$

where

$$a = [n/(n - 2)]^{\frac{1}{2}}$$

and

$$b = \{2\Gamma^2(\tfrac{1}{2}n)/[(n - 2)\Gamma^2(\tfrac{1}{2}n - \tfrac{1}{2})] - 1\}^{\frac{1}{2}}$$

which is a positive real number for $n \geqq 4$.

Now, from (1), (with $K = 0$ and $c = 1$), the variance-stabilizing transformation, $\xi(t)$, of noncentral t is

$$\xi(t) = \int_0^t (a^2 + b^2\mu^2)^{-\frac{1}{2}}\,d\mu$$

$$= \alpha \sinh^{-1}(\beta t),$$

where $\alpha = b^{-1}$ and $\beta = b/a$.

The random variable

$$\xi_1(t) = \xi(t) - \alpha \sinh^{-1}(\beta\mu) \tag{2}$$

will have, approximately, mean value zero and unit variance.

Using the second- and third-derivative terms respectively in the Taylor series for expectations, the following transformations might be expected to eliminate more bias than (2):

$$\xi_2(t) = \xi_1(t) + \tfrac{1}{2}b^2\mu\mu_2^{-\frac{1}{2}}, \tag{3}$$

$$\xi_3(t) = \xi_2(t) - \tfrac{1}{6}b^4\mu_2^{-5/2}\mu_3[2\mu^2 - (a^2/b^2)]. \tag{4}$$

To find the degree of approximation to normality of these transformations, $P\{t \leqq n^{\frac{1}{2}}x\}$ is approximated by $P\{\xi_i(t) \leqq \xi_i(n^{\frac{1}{2}}x)\}$, where $\xi_i(t)$, $i = 1, 2, 3$, is the transformation used. If n and δ together with α, $0 < \alpha < 1$, are specified, then the equation $P\{t \leqq n^{\frac{1}{2}}x\} = 1 - \alpha$ determines x uniquely[1]. Let $\Phi(\cdot)$ denote the unit normal distribution. If $\xi_i(t)$ is exactly $N(0, 1)$, then $\Phi(\xi_i(n^{\frac{1}{2}}x)) = 1 - \alpha$ is an identity. From tables of the normal distribution, if $\alpha = 0.10$, it follows that $\xi_i(n^{\frac{1}{2}}x) = 1.282$, $i = 1, 2, 3$. Hence, given n, δ and $\alpha = 0.10$, we obtain x from the tables of Resnikoff and Lieberman [9], pp. 383–9[2]. We enter the values of $\xi_i(n^{\frac{1}{2}}x)$ in Table 1. To see how good the approximations are in terms of probabilities (rather than in terms of the deviates) it suffices to note a few values of $\Phi(\cdot)$ for quick reference: $\Phi(1.26) = 0.896$, $\Phi(1.27) = 0.898$, $\Phi(1.28) = 0.900$ and $\Phi(1.29) = 0.902$.

There is a considerable degree of skewness in the noncentral t distribution for large values of δ and small values of n (Johnson and Welch [4]). Thus we do not expect (2) to be a very good approximation for simultaneous small values of n and large values of δ. Table 1 confirms this suspicion. Also, for small n, the quality of approximation deteriorates as δ increases. Larger values of n improve this transformation. In this case (3) and (4) are seen to be very close to normality, even for large δ.

Other numerical work, not presented here, was done for the cases $a = 0.05$ and $\alpha = 0.01$ and for the same values of n and δ. These results show that (2) is the most suitable transformation when $\alpha = 0.05$. The probability integral is over-estimated by (3) and (4). When $\alpha = 0.01$, (2), (3) and (4) over-estimate the probability integral, (2) being the closer approximation.

[1] We are limited, for exact values of the probability integral of noncentral t, to the tables of Resnikoff and Lieberman [9]. These authors tabulate $P\{n^{-\frac{1}{2}}t \leqq x\}$ because the range for the argument $n^{-\frac{1}{2}}t$ is about the same for all n and δ. This, of course, makes tabulation more compact.

[2] The rather odd values of δ which appear in Table 1 are not strange if one understands the mechanism of Resnikoff-Lieberman tables! In these tables, once n is selected, the noncentrality parameters are determined through the relationship $\delta = (n + 1)^{\frac{1}{2}}K_p$ where K_p is the upper p-point of the unit normal distribution and K_p is determined from: $\Phi(K_p) = 1 - p$. Hence δ is given as a function of n and p and only tabulated for a suitable range of values of p. The reason for this construction depends on the original purpose for which these tables were constructed.

TABLE 1

Values of $\xi_i(n^{\frac{1}{2}}x)$, $i = 1, 2, 3$, *where* $n^{\frac{1}{2}}x$ *is the 90th percentile of* t

n	δ	x	ξ_1	ξ_2	ξ_3
9	2.132924	1.325	1.155	1.221	1.224
	4.052622	2.177	1.124	1.223	1.214
	7.356558	3.723	1.102	1.223	1.199
	9.772173	4.879	1.096	1.222	1.193
19	3.016410	1.073	1.223	1.263	1.263
	5.731273	1.811	1.206	1.266	1.264
	10.403744	3.131	1.191	1.267	1.261
	13.819939	4.113	1.186	1.266	1.259
29	3.694333	0.983	1.241	1.272	1.272
	7.019347	1.685	1.228	1.275	1.275
	12.741932	2.933	1.218	1.277	1.275
	16.925900	3.859	1.214	1.277	1.274
39	4.265848	0.935	1.255	1.281	1.281
	8.105244	1.618	1.240	1.280	1.280
	14.713116	2.829	1.232	1.282	1.281
	19.544345	3.726	1.229	1.282	1.280
49	4.769363	0.903	1.259	1.282	1.282
	9.061938	1.576	1.251	1.286	1.286
	16.449764	2.763	1.240	1.285	1.284
	21.851242	3.641	1.236	1.283	1.282

3. Noncentral F**.** If $X_1, \cdots, X_m$ are independently distributed and X_i is $N(\mu_i, 1)$, then the random variable $\chi'^2 = X_1^2 + \cdots + X_m^2$ is called a noncentral chi-square variable with m degrees of freedom and noncentrality parameter $\lambda = \mu_1^2 + \cdots + \mu_m^2$.

If χ'^2 has the noncentral chi-square distribution with m degrees of freedom and noncentrality parameter λ, and if χ^2, independently of χ'^2, follows the central chi-square distribution with n degrees of freedom, then the ratio

$$F = (\chi'^2/m)/(\chi^2/n)$$

has the topside noncentral F distribution [6] with m and n degrees of freedom respectively, and with noncentrality parameter λ.

3.1. *The* $\cosh^{-1}$ *transformation.* The first two moments of F are given by Patnaik [6] as follows:

$$\mu = n(m + \lambda)/[m(n - 2)],$$

$$\mu_2 = n^2[(m + \lambda)^2 + 2(m + 2\lambda)]/[m^2(n - 2)(n - 4)] - \mu^2.$$

It will be supposed throughout that $n > 4$. By eliminating the parameter λ between the second central moment and the mean, we find that the variance is

$$\mu_2 = 2\{[\mu + (n/m)]^2 - a^2\}/(n - 4),$$

where

$$a = n(m + n - 2)^{\frac{1}{2}}/m(n - 2)^{\frac{1}{2}}.$$

From (1), (with $c = 1$ and $K = a - (n/m)$), the variance-stabilizing transformation, $\tau(F)$, is obtained as

$$\tau(F) = \left(\frac{1}{2}n - 2\right)^{\frac{1}{2}} \cosh^{-1}\left[\frac{F + (n/m)}{a}\right].$$

It may be hoped that $\tau(F)$ is approximately normal with mean value

$$\tau(\mu) = \left(\frac{1}{2}n - 2\right)^{\frac{1}{2}} \cosh^{-1}\left[\frac{m + n + \lambda - 2}{(n - 2)^{\frac{1}{2}}(m + n - 2)^{\frac{1}{2}}}\right]$$

and with unit variance. Using the second derivative term in the Taylor series for expectations, the transformed random variable

$$\tau = \tau(F) = \left(\frac{1}{2}n - 2\right)^{\frac{1}{2}}\left\{\cosh^{-1}\left[\frac{1 + (m/n)F}{b}\right] - \cosh^{-1}\left[\frac{1 + (m/n)\mu}{b}\right]\right\} + \left[\frac{\mu + (n/m)}{n - 4}\right]\mu_2^{-\frac{1}{2}}, \tag{5}$$

where $b = (m + n - 2)^{\frac{1}{2}}/(n - 2)^{\frac{1}{2}}$, may be better approximated by the normal distribution with zero mean and unit variance.

However, the transformation $\cosh^{-1}\{[F + (n/m)]/a\}$ looks more innocent than it is: $\cosh^{-1}(x)$ is real only if $x \geqq 1$, i.e., in our case only if

$$F \geqq (n/m)\{[m/(n - 2) + 1]^{\frac{1}{2}} - 1\} > 0.$$

Thus, for small values of F, we get no sensible approximation at all[3]. If $m = n$, the lower limit for F is $\{[2(n - 1)]/(n - 2)\}^{\frac{1}{2}} - 1 \doteq 0.414$, so it is not only very small values of F which are affected.

Strictly speaking, therefore, this transformation is not well defined and hence should not be used to approximate the probability integral $P\{F \leqq x\}$. However, we have investigated the approximation of this by $P\{\tau(F) \leqq \tau(x)\}$. The values for m, n, λ and x are the same as those considered by Patnaik [6] and the approximation (5) is given in Table 2 and compared with the exact values as given by Patnaik [6]. As may be expected, this approximation is not very satisfactory.

3.2. *The Square Root Transformation.* It is well known [5] that $(2\chi^2)^{\frac{1}{2}}$ is approximately normal with mean $(2n - 1)^{\frac{1}{2}}$ and unit variance. Also [6], $(2\chi'^2)^{\frac{1}{2}}$ is approximately normal with mean $[2(m + \lambda) - (m + 2\lambda)/(m + \lambda)]^{\frac{1}{2}}$ and variance than $(m + 2\lambda)/(m + \lambda)$. Thus, to the extent that these approximations hold,

[3] However, it seems possible to remedy the situation as follows: It is clear that $\tau(F + \epsilon)$ is well-defined if $\epsilon \geqq a - (m/n)$. Hence consider a power series expansion of $\tau(F + \epsilon)$. Take mathematical expectations to find Var $[\tau(F + \epsilon)]$ as an ascending series in powers of λ^{-1}, where λ is the noncentrality parameter. Then it might perhaps be possible to select a value of ϵ for which $\tau(F + \epsilon)$ is well-defined and which will eliminate bias of $O(\lambda^{-1})$. Precisely this type of argument is used to derive the well-known transformation $(X + 3/8)^{\frac{1}{2}}$ from $X^{\frac{1}{2}}$ for the Poisson distribution. The details in the present problem might be overwhelming!

TABLE 2
Approximate and Exact Values of the Probability Integral $P\{F \leqq x\}$

m	n	λ	x	Approximation (5) $P\{\tau \leqq \tau(x)\}$	Approximation (6) $P\{\tau_1 \leqq \tau_1(x)\}$	Approximation (7) $P\{\tau_2 \leqq \tau_2(x)\}$	Exact $P\{F \leqq x\}$
3	10	4	3.708	0.734	0.743	0.750	0.745
		4	6.552	0.902	0.915	0.919	0.918
		16	3.708	0.274	0.205	0.202	0.206
		16	6.552	0.527	0.520	0.520	0.517
3	20	4	3.098	0.696	0.696	0.707	0.700
		4	4.938	0.881	0.888	0.889	0.887
		16	3.098	0.151	0.124	0.119	0.126
		16	4.938	0.357	0.346	0.349	0.347
5	10	6	3.326	0.716	0.730	0.734	0.731
		6	5.636	0.895	0.910	0.914	0.914
		24	3.326	0.232	0.158	0.155	0.158
		24	5.636	0.481	0.467	0.463	0.461
5	20	6	2.711	0.658	0.661	0.669	0.664
		6	4.103	0.861	0.870	0.872	0.870
		24	2.711	0.096	0.069	0.064	0.069
		24	4.103	0.264	0.244	0.244	0.245
8	10	9	3.072	0.698	0.715	0.716	0.714
		9	5.057	0.888	0.903	0.909	0.908
		36	3.072	0.197	0.118	0.117	0.119
		36	5.057	0.438	0.416	0.409	0.408
8	30	9	2.266	0.574	0.574	0.581	0.578
		9	3.173	0.806	0.813	0.815	0.813
		36	2.266	0.027	0.017	0.014	0.017
		36	3.173	0.105	0.087	0.085	0.088

$$(mF/n)^{\frac{1}{2}} = (2\chi'^2)^{\frac{1}{2}}/(2\chi^2)^{\frac{1}{2}}$$

is the ratio of two independent normal random variables.

From a theorem due to Fieller [2], if X and Y are normally and independently distributed with means m_x and m_y and standard deviations σ_x and σ_y respectively, then the function

$$R = (m_x V - m_y)/(\sigma_x^2 V^2 + \sigma_y^2)^{\frac{1}{2}}, \qquad \text{where } V = Y/X,$$

will be nearly normally distributed with zero mean and unit variance, provided the probability of X being negative is small.

Applying this theorem to the variable $(mF/n)^{\frac{1}{2}}$, it may be hoped that the transformed random variable

$$(6) \quad \tau_1 = \tau_1(F) = \frac{(2n-1)^{1/2}(mF/n)^{1/2} - [2(m+\lambda) - (m+2\lambda)/(m+\lambda)]^{1/2}}{[(mF/n) + (m+2\lambda)/(m+\lambda)]^{1/2}}$$

will approximately have the unit normal distribution.

To obtain the degree of accuracy of this transformation, the exact probabilities $P\{F \leqq x\}$ for given values of m, n, λ and x have been compared with those of the above approximation. This comparison is also shown in Table 2, and the closeness of approximation is very satisfactory.

3.3. *The Cube Root Transformation.* Another transformation for normalizing the noncentral F distribution is obtained in a similar way by using the fact [5] that $(\chi^2/n)^{1/3}$ is approximately normally distributed with mean $[1 - 2/(9n)]$ and variance $2/(9n)$ and that $(\chi'^2/r)^{1/3}$ is also approximately normally distributed [1] with mean $[1 - 2(1 + B)/(9r)]$ and variance $2(1 + B)/(9r)$, where $r = m + \lambda$ and $B = \lambda/r$. Thus

$$[mF/(m+\lambda)]^{1/3} = \frac{(\chi'^2/r)^{1/3}}{(\chi^2/n)^{1/3}}$$

is, insofar as these approximations hold, the ratio of two independent normal random variables, and thus it may be hoped that the transformed random variable

$$(7) \quad \tau_2 = \tau_2(F) = \frac{[1-2/(9n)][mF/(m+\lambda)]^{1/3} - \{1-[2(m+2\lambda)/9(m+\lambda)^2]\}}{\{(2/9n)[(mF/(m+\lambda)]^{2/3} + [2(m+2\lambda)/9(m+\lambda)^2]\}^{1/2}}$$

will approximately have the unit normal distribution. Table 2 shows the approximation of $P\{F \leqq x\}$ by means of (7).

By substituting $\lambda = 0$ in (7), this transformation reduces to the transformation of Paulson [7] for the central F distribution. From (6) it is seen that an alternative transformation for normalizing the central F distribution is

$$U = \frac{(2n-1)^{1/2}(mF/n)^{1/2} - (2m-1)^{1/2}}{[(mF/n)+1]^{1/2}}$$

If we put $m = 1$, then $F = \chi'^2/(\chi^2/n)$ reduces to the noncentral random variable t^2. We conjecture that (6) and (7) with $m = 1$, $F = t^2$, (hence $\lambda = \delta^2$), will transform t^2 approximately to the unit normal distribution.

Although it is known that the Wilson-Hilferty transformation $(\chi^2/n)^{1/3}$ and the Aty transformation $(\chi'^2/r)^{1/3}$ are both more nearly normally distributed than the Fisher transformation $(2\chi^2)^{1/2}$ and the Patnaik transformation $(2\chi'^2)^{1/2}$ respectively [1], [6], it can be seen from Table 2 that (6) is a better approximation than (7), at least for the values of m, n, λ and x considered. The reasons for this apparent inconsistency remain undiscovered. Furthermore, (6) is a better approximation than that given by Patnaik [6], at least for the values of m, n, λ, and x considered.

4. Acknowledgments. I wish to express my gratitude to Professor H. S. Steyn for suggesting the original problem that led to this study.

The approximations in the tables were computed on the "ZEBRA" digital computer of the National Physical Research Laboratory of the South African Council for Scientific and Industrial Research.

REFERENCES

[1] ABDEL-ATY, S. H., "Approximate formulae for the percentage points and the probability integral of the non-central χ^2 distribution," *Biometrika*, Vol. 41 (1954), pp. 538–540.

[2] FIELLER, E. C., "The distribution of the index in a normal bivariate population," *Biometrika*, Vol. 24 (1932), pp. 428–440.

[3] JOHNSON, N. L., "Systems of frequency curves generated by methods of translation," *Biometrika*, Vol. 36 (1949), pp. 149–176.

[4] JOHNSON, N. L. AND WELCH, B. L., "Applications of the non-central t-distribution," *Biometrika*, Vol. 31 (1940), pp. 362–389.

[5] KENDALL, M. G. AND STUART, A., *The Advanced Theory of Statistics*, I, 2nd Ed. Griffin and Co., London, 1958.

[6] PATNAIK, P. B., "The non-central χ^2-and F-distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.

[7] PAULSON, EDWARD, "An approximate normalization of the analysis of variance distribution," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 233–235.

[8] RAO, C. RADHAKRISHNA, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952.

[9] RESNIKOFF, GEORGE J. AND LIEBERMAN, GERALD J., *Tables of the Non-Central t-distribution*, Stanford University Press, Stanford, Cal., 1957.