The Annals of Statistics
1977, Vol. 5, No. 4, 803-807

ONE-SAMPLE t-TEST WHEN SAMPLING FROM A MIXTURE OF NORMAL DISTRIBUTIONS[1]

BY AUSTIN F. S. LEE AND JOHN GURLAND

Boston University and University of Wisconsin

For the one-sample t-test a new form of the exact distribution of the test statistic t^2 is obtained when sampling from a distribution which is a mixture of two normal distributions. A numerical example is provided to show that the size of the test can differ greatly when sampling from distributions having the same skewness and kurtosis. Contours of equal size are plotted for a particular case in a certain cross section of the parameter space.

1. Introduction. Many studies have been made to ascertain the extent of distortion of the t-test when the underlying population from which the sample has been obtained is not normal. The most recent work in this area and related references can be found in Tiku (1971), Sansing and Owen (1974), and Subrahmaniam, Subrahmaniam and Messeri (1975).

In the present paper the effect of nonnormality on the one-sample t-test for small samples is studied when the underlying distribution is a member of the class of mixtures of two normal distributions, possibly with different variances. This class of distributions is very rich in deviations from normality, and provides some insight into the behavior of t when sampling from such a population. By an approach different from that employed by Hyrenius (1949), the distribution of t is obtained in a form more convenient for investigating the effect of nonnormality on the t-test. This technique can also be used to investigate the two-sample t and related tests, as can be seen in Lee and D'Agostino (1976).

2. Distribution function of t^2. Let X be a random variable having a compound normal distribution with probability density function (pdf) f given by

$$f(x) = \pi\phi(x; \mu_1, \sigma_1^2) + (1-\pi)\phi(x; \mu_2, \sigma_2^2) \tag{2.1}$$

where $0 < \pi < 1$. Here $\phi(x; \mu_i, \sigma_i^2)$ denotes the probability density function of a normal random variable with mean μ_i and variance σ_i^2, $i = 1, 2$. Denote the mean and variance of X by μ and σ^2, respectively.

Suppose $X_1, X_2, \cdots, X_n$ is a random sample from the above distribution, and let $\bar{X}$ and S^2 be the mean and variance of the sample. The cumulative distribution function (cdf) of the statistic $t^2 = n\bar{X}^2/S^2$ can be written as

$$G(x) = \sum_{m=0}^{n} H(x\,|\,m)\binom{n}{m}\pi^m(1-\pi)^{n-m}\,, \qquad x > 0\,, \tag{2.2}$$

Received March 1974; revised January 1977.

[1] Research supported, in part, by Wisconsin Alumni Research Foundation; and, in part, by Air Force Office of Scientific Research.

AMS 1970 *subject classifications.* Primary 62E15; Secondary 62F05.

Key words and phrases. Mixture of two normal distributions, one-sample t-test, distribution of t, effect of nonnormality, equal probability contours.

where $H(x \mid m)$, $m = 0, 1, \cdots, n$, is the conditional cdf of t^2, given that m observations are drawn from $N(\mu_1, \sigma_1^2)$ population and $n - m$ observations are drawn from $N(\mu_2, \sigma_2^2)$ population.

THEOREM. *The conditional* cdf $H(x \mid m)$ *of* t^2, *as defined above, can be expressed as*

$$H(x \mid m) = P[d_1\chi_1^2(\delta_1^2)/(d_2\chi_1^2(\delta_2^2) + d_3\chi_{m-1}^2 + d_4\chi_{n-m-1}^2) < 1]\,, \qquad m = 1, 2, \cdots, n-1\,; \tag{2.3}$$

$$H(x \mid 0) = P[\chi_1^2(\delta^2)/\chi_{n-1}^2 < x/(n-1)]\,, \qquad \textit{with} \quad \delta^2 = n\mu_2^2/\sigma_2^2\,; \tag{2.4}$$

and

$$H(x \mid n) = P[\chi_1^2(\delta^2)/\chi_{n-1}^2 < x/(n-1)]\,, \qquad \textit{with} \quad \delta^2 = n\mu_1^2/\sigma_1^2\,, \tag{2.5}$$

where $\chi_\nu^2(\delta^2)$ *denotes a chi-square random variable with degrees of freedom* ν *and noncentrality parameter* δ^2.

The notations in (2.3) are defined as follows. Let $r = \sigma_2/\sigma_1$ and

$$B = \begin{bmatrix} (n-m)x/(n-1) - m & -r(m(n-m))^{\frac{1}{2}}(1 + x/(n-1)) \\ -r(m(n-m))^{\frac{1}{2}}(1 + x/(n-1)) & r^2(mx/(n-1) - (n-m)) \end{bmatrix}.$$

Notice that since the determinant $|B| < 0$, the two characteristic roots of B are opposite in sign. Then $-d_1, d_2$ are these roots with $d_1 > 0$, $d_2 > 0$; $d_3 = nx/(n-1)$; $d_4 = r^2nx/(n-1)$. Furthermore,

$$\delta_i = (u_{i1}m^{\frac{1}{2}}\mu_1 + u_{i2}(n-m)^{\frac{1}{2}}\mu_2/r)/\sigma_1\,, \qquad i = 1, 2\,,$$

where (u_{11}, u_{12}), and (u_{21}, u_{22}) are the normalized characteristic vectors of B corresponding to $-d_1$ and d_2, respectively.

The proof of this theorem is obvious, since, conditioning upon the number of observations from each of the underlying distributions, the means and variances of observations from each of the underlying distributions are independent.

For the particular case $r = 1$, that is, $\sigma_1 = \sigma_2$, the problem simplifies tremendously because then $H(x \mid m)$ becomes

$$H(x \mid m) = P[\chi_1^2(\delta_1^2)/\chi_{n-1}^2(\delta_2^2) < x/(n-1)]\,, \qquad m = 0, 1, \cdots, n\,, \tag{2.6}$$

where

$$\delta_1^2 = (m\mu_1 + (n-m)\mu_2)^2/(n\sigma_1^2)\,;$$

and

$$\delta_2^2 = m(n-m)(\mu_1 - \mu_1)^2/(n\sigma_1^2)\,.$$

This case has been studied by Subrahmanian, Subrahmaniam and Messeri (1975).

3. Effect of nonnormality on two-sided t-test.

3.1 *Computation of distribution function of* t^2. To compute $P[t^2 < x]$ from (2.2) requires mainly the value of $H(x \mid m)$ shown in (2.3), which is a distribution function of a ratio of two linear combinations of independent central and noncentral χ^2 variables. It has been shown by Kotz, Johnson and Boyd (1967) that the pdf of a linear combination of noncentral χ^2 variables can be expressed as an infinite series involving the pdf's of central χ^2 variables. Direct application of this result yields

$$H(x \mid m) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} a_{1k} a_{2j} I_{\rho_2/(\rho_1+\rho_2)}((1 + 2k)/2, (n - 1 + 2j)/2), \tag{3.1}$$

where $I_y(p, q)$ is an incomplete beta function, and the constants ρ_1 and ρ_2 are suitably chosen to facilitate the convergence of the series. The coefficients a_{1k} and a_{2j} can be computed recursively and the formulas are given in Kotz et al. (1967).

3.2. *Some illustrative examples.* Suppose that one is interested in testing the null hypothesis H_0: $\mu = 0$ against two-sided alternatives $\mu \neq 0$. When $X_1, \cdots, X_n$ is a random sample from a normal population, $P[|t| > t_{n-1;\alpha/2} \mid H_0] = \alpha$. On the other hand, if the sample is from a compound normal distribution with density given in (2.1), then the above equality will not generally hold, and the deviation from the nominal significance level α will depend on the values of the parameters π, μ_1, μ_2, σ_1, and σ_2.

By way of illustration consider three members of the family in (2.1) given by the three sets of values for π, μ_1, μ_2, σ_1, σ_2, in Table 1. Also appearing in the table are the corresponding values of μ, σ, λ_3, λ_4, λ_5, λ_6 where $\lambda_r = \kappa_r/\sigma^r$, and κ_r is the rth cumulant of (2.1). It should be noted that although these three distributions have the some μ, σ, λ_3 and λ_4, (i.e., the same mean, variance, skewness, and kurtosis), the values of λ_5 and λ_6 differ substantially.

TABLE 1
Three distributions considered for illustration on the deviations of the actual sizes of the t-test from the nominal level of significance

Distribution	μ	σ	λ_3	λ_4	λ_5	λ_6	π	μ_1	μ_2	σ_1	σ_2
A	0	1	−1	2.5	−4	3.1845	.7782	.2269	−.7960	.6661	1.4620
B	0	1	−1	2.5	−2	−2.5514	.6672	.2759	−.5530	.5198	1.4159
C	0	1	−1	2.5	0	−7.8008	.5504	.3342	−.4091	.2385	1.3603

TABLE 2
Actual sizes of the one-sample t-test when sampling from the distributions listed in Table 1

Sample size	Nominal level of significance	Distribution A	B	C
	.01	.0109	.0127	.0329
5	.05	.0528	.0592	.1074
	.10	.1049	.1145	.1719
	.01	.0135	.0171	.0349
10	.05	.0581	.0653	.0853
	.10	.1105	.1182	.1331

We have computed the probability $P[|t| > t_{n-1;\alpha/2} | H_0]$ for the three sets of values π, μ_1, μ_2, σ_1, σ_2 given in Table 1 and for $n = 5$, 10 and $\alpha = .01, .05, .10$. These probabilities are listed in Table 2, and it is quite obvious from the example that, for some nonnormal distributions with the same skewness λ_3 and kurtosis λ_4 the actual size of the t-test may be quite different. λ_3 and λ_4 provide only partial information about a distribution, but it is the whole structure of the nonnormal distributions which may effect the behavior of the t-test (cf. Pearson (1963)).

For illustration purposes, we consider a sample of size $n = 5$ and examine the behavior of the size of the test for samples from the family of nonnormal populations considered here. It is easy to see that, for given x, $P[t^2 < x]$ is a function of $r = \sigma_2/\sigma_1$, μ_1/σ_1, μ_2/σ_2 and π. Furthermore, under H_0, $\mu_2 = -\pi\mu_1/(1 - \pi)$. Consequently, the size of the test of H_0 depends only on r, μ_1/σ_1 and π. For a given value of π, it is therefore possible to obtain, through the distribution given in (2.2), contours in space of r, μ_1/σ_1 along which the size has a specified value

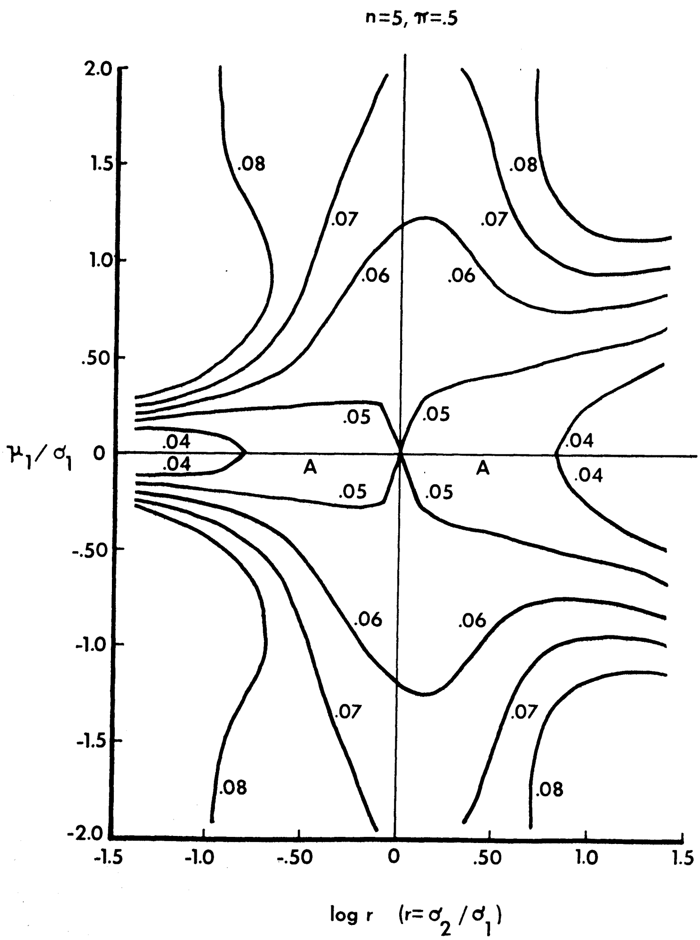

FIG. 1. Equal probability contours of $P(|t| > t_{4;.025}|\mu = 0)$ $n = 5$, $\pi = .5$.

α, say. In Figure 1, such contours are shown for a nominal 5% level t-test and with $\pi = .5$, and for $\alpha = .04, .05, .06, .07, .08$. From the gradient of the surface indicated by these contours, it is evident which regions in the plane afford specified control of size. For size between .04 and .05, for example, the points with coordinates ($\log r$, μ_1/σ_1) must lie in the region A in the figure. It is also apparent from the contours in the figure, that for larger values of r or $1/r$, the distortion of size is considerable, whereas for r in the neighborhood of 1, the size does not change so rapidly.

4. Conclusions. The exact distribution of t^2 has been obtained when sampling from a mixture of two normal distributions. A computational formula has also been provided, and some illustrative examples given as a direct application of the formula. From these examples it is evident how serious the distortion of the size of the test can be, especially for small samples. It is also evident that the size of the test may differ greatly when sampling from distributions having the same skewness and kurtosis.

Acknowledgment. The authors are grateful to the editor, associate editor and referees for their valuable suggestions to improve the presentation of this paper.

REFERENCES

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

[2] KOTZ, S., JOHNSON, N. L. and BOYD, D. W. (1967). Series representations of distributions of quadratic forms in normal variables II. Noncentral case. *Ann. Math. Statist.* **38** 838–848.

[3] LEE, A. F. S. and D'AGOSTINO, R. B. (1976). Levels of significance of some two-sample tests when observations are from compound normal distributions. *Comm. in Statist.* **A5(4)** 325–342.

[4] PEARSON, E. S. (1963). Some problems arising in approximating to probability distributions, using moments. *Biometrika* **50** 95–112.

[5] SANSING, R. C. and OWEN, D. B. (1974). The density of the t-statistic for nonnormal distributions. *Comm. in Statist.* **3(2)** 139–155.

[6] SUBRAHMANIAM, K., SUBRAHMANIAM, K. and MESSERI, J. Y. (1975). On the robustness of some tests of significance in sampling from a compound normal population. *J. Amer. Statist Assoc.* **70** 435–438.

[7] TIKU, M. L. (1971). Student's t distribution under nonnormal situations. *Austral. J. Statist.* **13** 142–148.

DEPARTMENT OF MATHEMATICS
BOSTON UNIVERSITY
264 BAY STATE ROAD
BOSTON, MASSACHUSETTS 02215

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON ST.
MADISON, WISCONSIN 53706