Edgeworth and Cornish-Fisher Expansion for the Non-normal t

Hark Hwang*

ABSTRACT

Let $X_1, \ldots, X_n$ be a random sample from a distribution with cumulants $K_1, K_2, \ldots$. The statistic $t = \frac{\sqrt{n}(\bar{X} - K_1)}{S}$ has the well-known 'student' distribution with $\nu = n-1$ degrees of freedom if the X_i are normally distributed ($i.e., K_i = 0$ for $i \geq 3$). An Edgeworth series expansion for the distribution of t when the X_i are not normally distributed is obtained.

The form of this expansion is Prob $(t < x) =$ Prob $(t^\circ < x) + f(x)\Sigma P_i(x) / \sqrt{\nu}^{\,i}$ where t^0 is student's t, $P_i(x)$ is a polynomial of degree $3i-1$ whose coefficients are functions of the first $i+2$ cumulants, and $f(x) = \exp(-x^2/2) / \sqrt{2\pi}$.

The Edgeworth series is inverted to yield the Cornish-Fisher expansion $t_p = t_p^\circ + \Sigma Q_i(x) / \sqrt{\nu}^{\,i}$, where t_p, t_p°, and $x = x_p$ are the 100p percentile points of the non-normal t, "student's" t and the unit normal, respectively, and $Q_i(x)$ is a polynomial of degree $i+1$ in x whose coefficients are functions of the first $i+2$ cumulants.

Comparison between the values obtained by the computer simulation and by the approximate formula shows good agreement on each of the 100p percentile points.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with cumulants

* Korea Advanced Institute of Science

$K_1, K_2, \ldots$. The statistic $t=\frac{\sqrt{n}(\bar{X}-K_1)}{S}$ has the well-known 'student' distribution with $\nu=n-1$ degrees of freedom if the X_i are normally distributed (*i.e.*, $K_i=0$ for $i \geqslant 3$).

However, in a variety of cases, it is necessary to test for the mean of a population which does not come from normal distribution and $\bar{X}$ and S are no longer independent.

The effect of non-normality on t for some small sample sizes has been discussed by a number of writers but the results do not have much practical value due to the small sample numbers [7][8].

So of most general interest are the results for Edgeworth series because it indicates the variation from the student t distribution that one might expect to be associated with given non-normal values.

Gayen [4], starting with the first four terms of a Gram-Charlier series expansion of the probability density function of the population, thereby ignoring population cumulants of order greater than four, tabulated the approximate correction it is necessary to apply to t (up to 24 degrees of freedom) for non-zero values of the third and foruth order cumulants. Tiku [10] using Hermite and Laguerre polynomials has obtained the distribution of t in terms of population cumulants up to the eigth order.

With an Edgeworth series expansion for the non-normal distribution and inversion of the series to yield the Cornish Fisher expansion, this paper presents an approach to find the cumulative distribution function and the value of 100p percentile point of the non-normal t.

2. The Edgeworth Series Approximation

Let X and Y be two independent random variables with distribution function F(x) and G(y).

We consider "Charlier's Differential Form"

$$F(u)=\exp[\Sigma(K_j-K_j^*)(-D)^j/j!]G(u) \qquad (2.1)$$

where the sequence K_j and K_j^* are the cumulants of X and Y, respectively and D is the differential operator D=d/du. To simplify the notation let

$$\lambda_j = K_j - K_j^* \tag{2.2}$$

Then Charlier's differential form becomes

$$F(u) = \exp[\Sigma \lambda j (-D)^j / j!] G(u) \tag{2.3}$$

To put this in more managerable form, we expand the differential as

$$\exp[\Sigma \lambda j (-D)^j / j!] G(u) = \Sigma \mu_j (-D)^j / j! \tag{2.4}$$

where μ_j and λ_j have the recursive relationship [11].

Charlier's differential form may be expanded as

$$F(u) = G(u) + \sum_{j=1} \mu_j (-D)^j / j! G(u) \tag{2.5}$$

If $G(u)$ has a density function

$$g(u) = \frac{d}{du} G(u) \tag{2.6}$$

then

$$F(u) = G(u) - \sum_{j=0} \mu_{j+1} (-D)^{j+1} / (j+1)! \, g(u) \tag{2.7}$$

In many cases of interest, Y is a standard normal variable and the derivative of $g(u)$ may be expressed as Hermite polynomials.

The Hermite polynomials may be defined as

$$g(u) H_j(u) = (-D)^j g(u) \tag{2.8}$$

and hence

$$F(u) = G(u) - g(u) \sum_{j=1} \frac{\mu_j}{j!} H_{j-1}(u)$$

The Gram-Charlier approximation is obtained by turncating, e.g.,

$$F(u) \approx F_n(u) = G(u) - g(u) \sum_{j=1} \frac{\mu_j}{j!} H_{j-1} \tag{2.9}$$

In many applications, the cumulants of X depends on an auxiliary parameter λ and the expansion of $F(x)$ is regrouped as an asymptotic series in power of λ.

This is the so called Edgeworth series.

So the distribution function is of the form

$$F(Z)=G(Z)-A(Z)g(Z) \tag{2.10}$$

where $A(Z)=\sum_{i=1} P_i(Z)/\lambda^i$

and $P_i(Z)$ is a polynomial in Z.

A computer program using Algebraic Manipulation Package [12] is established to calculate the Edgeworth series approximation of non-normal t in terms of the population cumulants and the sample size N up to $(1/N)**2$ accuracy.

The input of the program are the first six cumulants of the non-normal t and the cumulants of the standard normal variable [5]. The output of the program has 57 terms and we have

$$\text{Prob }(t<Z)=\text{Prob }(t^0<Z)+g(Z)\sum_{i=1}^{4}P_i(Z)/(\sqrt{N})^i$$

where t^0 is student's t, $P_i(Z)$ is a polynomial of degree $3i-1$ whose coefficients are functions of the first $i+2$ cumulants.

The result is consistent with the approximation of the distribution function of student's t given by Fisher.

So we conclude that the distribution function of non-normal t is equal to that of the normal t plus adjustment factor.

These factors are tabulated at table (2.1)

For instance, the first adjustment factor is described as $\frac{1}{6\sqrt{N}}(2Z^2K_3+ZK_3)$

3. Inverse Transformation from Edgeworth Series

Let $X=X_p$ be the 100p percentile of the standard normal distribution $G(x)$, and let $Z=Z_p$ be the corresponding percentile point of a second distribution $F(z)$. Then

$$F(z)=G(x)=G(z+(x-z)) \tag{3.1}$$

By Taylor's theorem

$$F(z)=G(z)-g(z)B(z,x) \tag{3.2}$$

Table (2.1) **Coefficients for the Edgeworth Expansion of the Distribution Function of the Non-normal t**

ADJUST MENT FACTOR	DIVIS OR	PRODUCT OF	K_3	K_3^2	K_4	K_5	K_3K_4	K_3^3	K_6	K_3K_5	K_4^2	$K_4K_3^2$	K_3^4
I	$6\sqrt{n}$	Z^1	2										
		1	1										
		Z^5		−2									
II	36n	Z^3		−4	3								
		Z		6	−9								
		Z^8						40					
		Z^6	540				−180	140					
III	$6480\sqrt{n}^3$	Z^4	−1890			−324	1350	−1050					
		Z^2				−1296	4050	−2625					
		1				−162	675	−525					
		Z^{11}											−20
		Z^9		−540								180	−100
IV	38880 n^2	Z^7		4320	810		648				−135	−2160	1800
		Z^5		3240	−7290				−864	5184	2835	−16200	9000
		Z^3			17010		−3240		2160		−4455	6480	−900
		Z		8100	12150		−19440		6480		−14985	46980	−18900

where $B(z,x)=\sum_{i=1}(z-x)^i H_{i-1}(z)/i!$

and $H_{i-1}(z)$ is Hermite polynomial.

If $F(z)$ has an expansion of the type (2.10), then we may solve for x and z by equating (2.10) and (3.2). That is

$$\sum_{i=1} P_i(z)y^i=\sum_{i=1}(z-x)^i H_{i-1}(z)/i! \tag{3.3}$$

where $y=1/\lambda$

Giving $(z-x)$ in powers of y

$$z-x=a_0(z)+a_1(z)y+a_2(z)y^2+\ldots, \tag{3.4}$$

where a_i is polynomial of z.

To find z in terms of x, (3.4) is rearranged

$$z-x=b_0(y)+b_1(y)z+b_2(y)z^2+,\ldots, \tag{3.5}$$

where b_i is polynomial of y.

From (3. 5)

$$x+bo(y)=(1-b_1(y))z-b_2(y)z^2-b_3(y)z^3-,\ldots. \tag{3. 6}$$

The next step is to invert (3. 6) to express z in power of $(x+bo)$ and then the percentile point of x in terms of the percentile point of a normal population is obtained.

Our computer program established to carry out the above procedure gives 42 terms with the following form.

$$t_p=t_p^o+\Sigma Q_i(x)/\sqrt{\nu}^{\,i} \tag{3. 7}$$

where t_p, t_p^o and $x=x_p$ are the 100p percentile point of the non-normal t, student's t and the unit normal respectively and $\nu=N-1$.

$\Sigma Q_i(x)/\sqrt{\nu}^{\,i}$ can be treated as the adjustment factors due to the sampling from non-normal population and $Q_i(x)$ is a polynomial of degree $i+1$ in x whose coefficients are functions of the first $i+2$ cumulants.

The six terms among the result corresponding to t_p^o are in agreement with the expansion of t given by Fisher and Cornish [3] while the remaining 36 terms represent the four adjustment factors.

These factors are shown at table (3. 1).

Table (3. 1) **Coefficients for the Cornish-Fisher Expansion of the Percentile Points of the Non-normal t Distribtion**

ADJUSTMENT FACTOR	DIVISOR	PRODUCT OF	COEFFICIENTS OF K_3	K_4	K_3^2	K_5	K_3K_5	K_3^3	K_6	K_3K_5	K_2^4	$K_4K_4^2$	K_3^4
I	$6\sqrt{\nu}$	X^2	−2										
		1	−1										
II	72ν	X^3		−6	20								
		X		18	−5								
III	$6480\sqrt{\nu}^2$	X^4				324		−1110					
		X^2	−1080			1296	−5130	2675					
		1	270			162	−945	655					
IV	$155520\nu^2$	X^5		3240	−19440				3456	−24624	−4860	34560	3600
		X^3		−38880	−17820				−8640	−27216		144180	−82730
		X		−68040	−21060				−25920	65448	69660	−151200	49055

4. Discussion

Comparison between the values obtained by the computer simulation and by the approximation formula for each percentile point on two distribution, rectangular distribution and reduced log-weibull distribution is carried out to check how good our estimate is.

For each non-normal population and for each sample size, 999 values of the simulated t are obtained and then these t are rearranged in the increasing order. The 100p percentile point of t is defined as the $(p \times 1000)$th value in this ordered sample. The above simulation procedures are repeated 10 times and so we have 10 simulated values for each of the 100p percentile point.

There is good agreement between the results from the approximation formula (Eq. (3.7)) and the corresponding values from the simulation.

Each result from the approximation is in the range of three sigma limit calculated from the mean and standard deviation of the 10 simulated values.

The absolute value of the difference between these two estimations becomes smaller as the degrees of freedom increase as we expect.

The rectangular distribution generates t_p values much closer to the standard t than the reduced log-weibull which shows the remarkable difference around the tail area which we are mostly interested.

The negative kurtosis of a rectangular population does not give an effect on t for the practical purpose if the sample size becomes around 16 while the negative skewness of the reduced log-weibull causes an extreme effects to t even with the sample size 31.

This skewness in the parent population cause t to be skew in the opposite sense decreasing it in the other side, so we can make false decision of declaring the positive value of t significant while there are not and negative values not significant when they are [1].

REFERENCES

[1] Bartlett, M.S., "The Effect of Non-normality on the t Distribution," *Proceedings of the cambridge philosophical Society*, 31, 223-231. 1935.

[2] David, F.N., Kendall, M.G. and Barton, D.E., *Symmetric Functions and Allied Tables*, Cambridge University Press, 1966.

[3] Fisher, R.A., "The Percentile Points of Distributions Having known Cumulants," *Technometrics* 2, p. 209-225, 1960.

[4] Gayen, A.K., "The Distribution of 'Student's' *t* in Random Samples of any Size Drawn From Non-normal Universe," *Biometrika*, 36, 353-369, 1949.

[5] Hawng H., *The Non-normal t Distribution*, Ph.D. Thesis, University of Minnesota, 1975.

[6] Kendall, M.G. and Stuart, A., *The Advanced Theory of Statistics*, Vol. 1, Hafner Publishing Company, New York, 1958.

[7] Laderman, J., "The Distribution of Student's Ratio for Samples of Two Items Drawn From Non-normal Universe," *Annals of Mathematical Statistics*, 10, 376-379, 1939.

[8] Perlo, V., "On the Distribution of Student's Ratio for Samples of Three Drawn From a Rectangular Distribution," *Biometrika*, 61, p. 177, 1974.

[9] Ratcliffe, J.F., "The Effect on the t-Distribution of Non-normality in the Sampled Population," *Applied Statistics*, 17, 42-48, 1968.

[10] Tiku, M. L., "Approximation to Student's t Distribution in Terms of Hermite and Laguerre Polynomials," *The Journal of the Indian Mathematical Society*, Vol. 27, No. 2, pp. 91—102, 1963.

[11] White, J. S., *Moments, Cumulants and FORMAC*. Mathematics Department, Research Laboratories, General Motors Corporation, 1967.

[12] Zimmerman, C.D., "Extended Fortran Algebraic Manipulator with Applications to Linear Problems of Physics," Ph.D. Thesis, University of Minnesota, 1969.