A NORMAL APPROXIMATION TO THE F DISTRIBUTION

Daniel Furtado FERREIRA[1]

▪ *ABSTRACT: A new normal approximation,* $\Phi(z)$*, for the cumulative distribution function (c.d.f.) of the F distribution, F(x,* ν_1*,* ν_2*), with associated degrees of freedom,* ν_1 *and* ν_2*, is proposed for large* ν_2 *and fixed* ν_1*. The proposed approximation is compared to several others approximations such as a normal, ordinary chi-square, Scheffé-Tukey and Li and Martin approximations. The employed numerical analysis indicates that, for* $\nu_2/\nu_1 \geq 3$*, the accuracy of the proposed normal approximation is to at least the third decimal place for most small values of* ν_1*. This is a comparable accuracy that is achievable using Li and Martin (2002) approximation with shrinking factor approximation (SFA) using a chi-square c.d.f. with degrees of freedom* ν_1*. The advantage of the proposed normal approximation over SFA is that normal c.d.f. is easily obtained than that of chi-square.*

▪ *KEYWORDS: Approximation; F distribution; normal distribution.*

1 Introduction

As pointed out by Li and Martin (2002) among the normality based distributions, t, χ^2, and F, the derivation of approximations to the two-parameter F distribution is considered to be the most challenging. A great part of the proposed approximations for the distribution F with degrees of freedom ν_1 and ν_2 of the numerator and denominator, respectively, is based on the assumption that both degrees of freedom are large. If this assumption does not hold, the accuracy is not satisfactory.

Using the same argument of Li and De Moor (1999) that proposed a shrinkage factor approximation (SFA) to the Student's t distribution based on normal approximation, Li and Martin (2002) extended the SFA to the case of the F distribution based on a chi-square distribution. The proposed SFA was used to improve the Scheffé and Tukey (1944) approximation that is one of the most widely used approximations due to the accuracy. The Li and Martin (2002) SFA exhibits good behavior in the tail of the F distribution in contrast to the Scheffé-Tukey approximation.

The chi-square percentage points can be accurately obtained using the Wilson and Hilferty transformation (Wilson and Hilferty, 1931). Considering the 100α% point of a chi-square distribution with ν degrees of freedom ($\chi^2_{\alpha,\nu}$) and the 100α% point of a standard normal distribution. The Wilson-Hilferty transformation is:

[1] Universidade Federal de Lavras – UFLA, Departamento de Ciências Exatas – DEX, Caixa Postal 3037, CEP: 37200-000, Lavras, MG, Brazil. Pesquisador e Bolsista do CNPq. E-mail: *danielff@ufla.br*

$$\chi^2_{\alpha,\nu} \approx \nu\left[Z_\alpha\left(\frac{2}{9\nu}\right)^{0,5} + \left(1-\frac{2}{9\nu}\right)\right]^3, \qquad (1)$$

where Z_α is the 100α% upper quantile from the standard normal distribution.

Due to the simplicity and accuracy, the normal approximations are preferable used in practice than those of chi-square. Since the chi-square SFA of Li and Martin (2002) presented excellent accuracy, in this paper the chi-square SFA is extended to a normal SFA using the Wilson-Hilferty transformation. Numerical accuracy of the proposed normal SFA is investigated for several values of the degrees of freedom associated to the F distribution.

2 Normal shrinkage factor approximation

Consider the density of the F distribution:

$$f(x;\nu_1,\nu_2) = \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{\Gamma\left[(\nu_1+\nu_2)/2\right]}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} x^{[(\nu_1/2)-1]}\left(1+\nu_1 x/\nu_2\right)^{-(\nu_1+\nu_2)/2}, \quad x \geq 0, \qquad (2)$$

and the chi-square density function:

$$g(x;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{[(\nu/2)-1]}\exp(-x/2), \quad x \geq 0, \qquad (3)$$

where $\Gamma(\bullet)$ represents the gamma function of the argument. Let $F(x;\nu_1,\nu_2)$ and $G(x;\nu)$ denote the cumulative distribution functions, c.d.f.'s, of the F and chi-square distribution, respectively. Consider also the standard normal density:

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\exp(-z^2/2), \quad -\infty < z < \infty, \qquad (4)$$

and let $\Phi(z)$ denote the standard normal cumulative distribution function.

A brief revision of some of the existent approximation to the F c.d.f is outstanding as follows. First consider that the F distribution function has mean $E(x) = \mu = \nu_2/(\nu_2-2)$ for ν_2>2 and variance $\sigma^2 = \mu^2\left\{2(\nu_1+\nu_2-2)/\left[\nu_1(\nu_2-4)\right]\right\}$ for ν_2 > 4. Then a natural approximation for large ν_1 and ν_2 (Patel *et al.*, 1976) is:

$$F(x;\nu_1,\nu_2) \approx \Phi(z_1), \qquad (5)$$

where $z_1 = (x-\mu)/\sigma$, μ and σ defined above.

Other two normal approximations are considered by Johnson *et al.* (1995). The first one is the Fisher's approximation:

$$F(x;\nu_1,\nu_2) \approx \Phi(z_2), \qquad (6)$$

where:

$$z_2 = \frac{\sqrt{\left(1-\frac{1}{2\nu_2}\right)x} - \sqrt{1-\frac{1}{2\nu_1}}}{\sqrt{\frac{x}{2\nu_2}+\frac{1}{2\nu_1}}}. \tag{7}$$

The second is the Pulson's approximation:

$$F(x;\nu_1,\nu_2) \approx \Phi(z_3), \tag{8}$$

where:

$$z_3 = \frac{\left(1-\frac{2}{9\nu_2}\right)x^{1/3} - \left(1-\frac{2}{9\nu_1}\right)}{\sqrt{\frac{2x^{2/3}}{9\nu_2}+\frac{2}{9\nu_1}}}. \tag{9}$$

The chi-square is the often used asymptotic approximation to the F distribution (Johnson *et al*., 1995). The ordinary chi-square approximation is:

$$F(x;\nu_1,\nu_2) \approx G(\nu_1 x;\nu_1), \quad \text{for large } \nu_2 \text{ and fixed } \nu_1. \tag{10}$$

Scheffé and Tukey (1944) proposed an approximation to improve the ordinary chi-square approximation (10). The Scheffé-Tukey approximation is:

$$F(x;\nu_1,\nu_2) \approx G(\lambda_1\nu_1 x;\nu_1), \quad \text{for large } \nu_2 \text{ and fixed } \nu_1, \tag{11}$$

where λ_1 is given by:

$$\lambda_1 = \frac{2\nu_2+\nu_1-2}{2\nu_2+\nu_1 x}. \tag{12}$$

Finally as pointed out, Li and Martin (2002) extended the SFA of Li and De Moor (1999) to the Student's t distribution based on the normal distribution to the case of the F distribution to improve the Scheffé-Tukey approximation. The shrinkage factor approximation (SFA) is:

$$F(x;\nu_1,\nu_2) \approx G(\lambda_2\nu_1 x;\nu_1), \tag{13}$$

where the shrinkage factor λ_2 is given by:

$$\lambda_2 = \frac{2\nu_2+\nu_1 x/3+\nu_1-2}{2\nu_2+4\nu_1 x/3}. \tag{14}$$

Our approach uses the fact that the chi-square can be approximated by the normal distribution using the Wilson-Hilferty transformation. Thus, using the chi-square approximation of the F, given by (13) we can achieve a normal approximation of the F distribution directly. Following this ideas, it is easy to show that the chi-square c.d.f. can be approximate by:

$$G(x;\nu) \approx \Phi(z_4), \tag{15}$$

where:

$$z_4 = \frac{\sqrt[3]{\frac{x}{\nu}} - \left(1 - \frac{2}{9\nu}\right)}{\sqrt{\frac{2}{9\nu}}}. \tag{16}$$

Now the normal approximation to the chi-square SFA of Li and Martin (2002) can be proposed. The main aim is to obtain a chi-square approximation with shrinkage factor λ_2 of (14) and then obtain a normal approximation of the chi-square c.d.f. using the Wilson-Hilferty transformation (15). This can be represented by:

$$F(x;\nu_1,\nu_2) \approx G(\lambda_2\nu_1 x;\nu_1) \approx \Phi\left[z_4(\lambda_2\nu_1 x,\nu_1,\nu_2)\right].$$

Following this idea it is easy to show that our resulting normal SFA is:

$$F(x;\nu_1,\nu_2) \approx \Phi(z_5), \tag{17}$$

where:

$$z_5 = \frac{\sqrt[3]{\frac{(2\nu_2 + \nu_1 x/3 + \nu_1 - 2)x}{2\nu_2 + 4\nu_1 x/3}} - \left(1 - \frac{2}{9\nu_1}\right)}{\sqrt{\frac{2}{9\nu_1}}}. \tag{18}$$

3 Accuracy

The numerical accuracy of our proposed normal SFA (18) is investigated for several values of ν_1 and ν_2, satisfying the relations proposed by Li and Martin (2002), that is., $\nu_2/\nu_1 \geq 3$ and $\nu_1 \geq 3$. The maximum absolute error (MAE) is used to evaluate the performance of an arbitrary approximation function, H(x). The H(x) represents one of the seven approximations discussed above. The MAE is given by:

$$\text{MAE} = \underset{x \in \Im}{\text{Max}} \left|F(x;\nu_1,\nu_2) - H(x)\right|. \tag{19}$$

It was chosen the same set $\Im$ defined over the range of the random variable for which the cumulative probability lies between 0.0001 and 0.9999 in step of 0.01 propose by Li and Martin (2002).

The evaluations were performed using SAS version 8.0. The same set of values of ν_1 and ν_2 proposed by Li and Martin (2002) was used. The MAE for the approximation H(x) was obtained. The considered approximations were the ordinary normal (15), Fisher's normal (16), Paulson's normal (18), ordinary chi-square (10), Scheffé-Tukey's chi-square (11), chi-square shrinkage factor (13) and the normal shrinkage factor (17). The obtained MAE is given in Table 1. For the six first approximations evaluated by Li and Martin (2002) the same set of numerical results were obtained.

It can be seen from Table 1 that the accuracy of the normal SFA is to three decimals places for small ν_1 and to four decimal places for large ν_1. The accuracy of the approach does not increase as the ratio ν_2/ν_1 increases and depends on ν_1. Small ν_1 is associates to small accuracy and large ν_1 is associates to large accuracy. In comparison with the normality based approximations the accuracy of the normal SFA is greater in all cases than ordinary normal and Fisher's approximation and is practically the same of those of Paulson's approximation, and in particular when the ration ν_2/ν_1 is small for large ν_1 the normal SFA accuracy is greater. The normal SFA presents greater accuracy than the ordinary chi-square except for small ν_1 and large ν_2/ν_1. The normal SFA exhibits larger accuracy than Scheffé-Tukey approximation when the ratio ν_2/ν_1 is small, even for small value of ν_1. Since the accuracy of the proposed approximation practically does not increase as the ratio ν_2/ν_1 increases then the Scheffé-Tukey approximation exhibits great accuracy when the ratio ν_2/ν_1 is large.

Such approaches have great importance in scientific works involving some simplifications of theoretical asymptotic distributions and Monte Carlo simulations. Also, it worth emphasizing the academic importance of such probability distributions approaches, especially in the didactic aspects. This is true if we recognize that the F distribution is one of the most import distributions and has been extensively and intensively used in all areas of knowledge.

4 Example

To illustrate we consider the simple example to obtain the cumulative distribution function of the F distribution to $x = 2.16458$ with $\nu_1 = 10$ and $\nu_2 = 30$. The true value of the cumulative distribution function is 0.95. First we calculate z_5 using (18) by

$$z_5 = \frac{\sqrt[3]{\frac{(2\nu_2 + \nu_1 x/3 + \nu_1 - 2)x}{2\nu_2 + 4\nu_1 x/3}} - \left(1 - \frac{2}{9\nu_1}\right)}{\sqrt{\frac{2}{9\nu_1}}} = \sqrt[3]{\frac{\frac{2\times30+10\times2.16458+10-2}{2\times30+4\times10\times2.16458/3} - \left(1 - \frac{2}{9\times10}\right)}{\sqrt{\frac{2}{9\times10}}}} = 1.649339.$$

Now using a standard normal cumulative distribution function calculator we find

$$\Phi(1.649339) = 0.9504609\,.$$

We can see that the absolute error is 0.0004609, that is smaller than the MAE of 9.4×10^{-4} in the Table 1, in the line that has ν_1=10 and $\nu_1/\nu_2 = 3$.

Table 1 - Numerical maximum absolute error (MAE) for different approximations to the F cumulative distribution function

ν_1	ν_2/ν_1	Chi-Square SFA	Ordinary Normal	Fisher	Paulson	Ordinary Chi-Square	Scheffé and Tukey	Normal SFA
3	3	$2.4x10^{-3}$	$1.9x10^{-1}$	$3.1x10^{-2}$	$5.1x10^{-3}$	$7.3x10^{-2}$	$9.1x10^{-3}$	$5.6x10^{-3}$
3	5	$5.3x10^{-4}$	$1.6x10^{-1}$	$3.1x10^{-2}$	$5.3x10^{-3}$	$4.6x10^{-2}$	$3.4x10^{-3}$	$5.6x10^{-3}$
3	10	$5.9x10^{-5}$	$1.3x10^{-1}$	$3.2x10^{-2}$	$5.4x10^{-3}$	$2.4x10^{-2}$	$8.7x10^{-4}$	$5.5x10^{-3}$
3	30	$6.4x10^{-6}$	$1.2x10^{-1}$	$3.2x10^{-2}$	$5.5x10^{-3}$	$8.1x10^{-3}$	$9.8x10^{-5}$	$5.5x10^{-3}$
3	50	$2.3x10^{-6}$	$1.1x10^{-1}$	$3.2x10^{-2}$	$5.5x10^{-3}$	$4.9x10^{-3}$	$3.5x10^{-5}$	$5.5x10^{-3}$
3	100	$5.8x10^{-7}$	$1.1x10^{-1}$	$3.2x10^{-2}$	$5.5x10^{-3}$	$2.5x10^{-3}$	$8.9x10^{-6}$	$5.5x10^{-3}$
5	3	$1.1x10^{-3}$	$1.2x10^{-1}$	$1.7x10^{-2}$	$1.9x10^{-3}$	$6.5x10^{-2}$	$6.6x10^{-3}$	$2.3x10^{-3}$
5	5	$3.4x10^{-4}$	$1.1x10^{-1}$	$1.8x10^{-2}$	$2.0x10^{-3}$	$4.1x10^{-2}$	$2.5x10^{-3}$	$2.3x10^{-3}$
5	10	$8.8x10^{-5}$	$9.6x10^{-1}$	$1.8x10^{-2}$	$2.3x10^{-3}$	$2.1x10^{-2}$	$6.6x10^{-4}$	$2.5x10^{-3}$
5	30	$1.0x10^{-5}$	$8.8x10^{-2}$	$1.9x10^{-2}$	$2.5x10^{-3}$	$7.3x10^{-3}$	$7.7x10^{-5}$	$2.6x10^{-3}$
5	50	$3.7x10^{-6}$	$8.7x10^{-2}$	$2.0x10^{-2}$	$2.5x10^{-3}$	$4.4x10^{-3}$	$2.8x10^{-5}$	$2.6x10^{-3}$
5	100	$9.3x10^{-7}$	$8.6x10^{-2}$	$2.0x10^{-2}$	$2.6x10^{-3}$	$2.2x10^{-3}$	$7.0x10^{-6}$	$2.6x10^{-3}$
10	3	$9.4x10^{-4}$	$8.7x10^{-2}$	$8.8x10^{-3}$	$7.9x10^{-4}$	$5.6x10^{-2}$	$4.8x10^{-3}$	$9.4x10^{-4}$
10	5	$3.7x10^{-4}$	$7.6x10^{-2}$	$1.1x10^{-2}$	$9.5x10^{-4}$	$3.5x10^{-2}$	$1.9x10^{-3}$	$8.9x10^{-4}$
10	10	$1.0x10^{-4}$	$6.8x10^{-2}$	$1.3x10^{-2}$	$1.1x10^{-3}$	$1.8x10^{-2}$	$5.1x10^{-4}$	$1.2x10^{-3}$
10	30	$1.2x10^{-5}$	$6.2x10^{-2}$	$1.4x10^{-2}$	$1.2x10^{-3}$	$6.3x10^{-3}$	$5.9x10^{-5}$	$1.3x10^{-3}$
10	50	$4.3x10^{-6}$	$6.1x10^{-2}$	$1.4x10^{-2}$	$1.2x10^{-3}$	$3.8x10^{-3}$	$2.2x10^{-5}$	$1.3x10^{-3}$
10	100	$1.1x10^{-6}$	$6.0x10^{-2}$	$1.4x10^{-2}$	$1.2x10^{-3}$	$1.9x10^{-3}$	$5.5x10^{-6}$	$1.3x10^{-3}$
30	3	$8.8x10^{-4}$	$5.0x10^{-2}$	$4.9x10^{-3}$	$2.4x10^{-4}$	$4.7x10^{-2}$	$3.6x10^{-3}$	$5.5x10^{-4}$
30	5	$3.6x10^{-4}$	$4.4x10^{-2}$	$6.2x10^{-3}$	$2.9x10^{-4}$	$3.0x10^{-2}$	$1.4x10^{-3}$	$2.6x10^{-4}$
30	10	$1.0x10^{-4}$	$3.9x10^{-2}$	$7.3x10^{-3}$	$3.3x10^{-4}$	$1.6x10^{-2}$	$3.9x10^{-4}$	$2.9x10^{-4}$
30	30	$1.2x10^{-5}$	$3.6x10^{-2}$	$8.1x10^{-3}$	$3.7x10^{-4}$	$5.3x10^{-3}$	$4.6x10^{-5}$	$3.8x10^{-4}$
30	50	$4.4x10^{-6}$	$3.5x10^{-2}$	$8.3x10^{-3}$	$3.8x10^{-4}$	$3.2x10^{-3}$	$1.7x10^{-5}$	$3.9x10^{-4}$
30	100	$1.1x10^{-6}$	$3.5x10^{-2}$	$8.4x10^{-3}$	$3.9x10^{-4}$	$1.6x10^{-3}$	$4.2x10^{-6}$	$3.9x10^{-4}$
50	3	$8.5x10^{-4}$	$3.8x10^{-2}$	$3.8x10^{-3}$	$1.4x10^{-4}$	$4.4x10^{-2}$	$3.3x10^{-3}$	$6.3x10^{-4}$
50	5	$3.5x10^{-4}$	$3.4x10^{-2}$	$4.8x10^{-3}$	$1.7x10^{-4}$	$2.8x10^{-2}$	$1.3x10^{-3}$	$2.1x10^{-4}$
50	10	$9.8x10^{-5}$	$3.0x10^{-2}$	$5.7x10^{-3}$	$1.9x10^{-4}$	$1.5x10^{-2}$	$3.6x10^{-4}$	$1.3x10^{-4}$
50	30	$1.2x10^{-5}$	$2.8x10^{-2}$	$6.3x10^{-3}$	$2.2x10^{-4}$	$5.0x10^{-3}$	$4.2x10^{-5}$	$2.2x10^{-4}$
50	50	$4.3x10^{-6}$	$2.7x10^{-2}$	$6.4x10^{-3}$	$2.2x10^{-4}$	$3.0x10^{-3}$	$1.5x10^{-5}$	$2.3x10^{-4}$
50	100	$1.1x10^{-6}$	$2.7x10^{-2}$	$6.5x10^{-3}$	$2.3x10^{-4}$	$1.5x10^{-3}$	$3.9x10^{-6}$	$2.3x10^{-4}$

FERREIRA, D.F. Aproximação normal da distribuição F. *Rev. Bras. Biom.*, São Paulo, v.29, n.2, p.222-228, 2011.

- RESUMO: Uma nova aproximação, $\Phi(z)$, da função de distribuição (f.d.) da distribuição F, $F(x, \nu_1, \nu_2)$, com graus de liberdade associados, ν_1 e ν_2, foi proposta para grandes valores de ν_2 e valores fixos de ν_1. A aproximação proposta foi comparada com várias outras aproximações, tais como a normal, qui-quadrado ordinária, Scheffé-Tukey e aproximação de Li e Martin. As análises numéricas empregadas indicaram que, considerando $\nu_2/\nu_1 \geq 3$, a acurácia da aproximação normal proposta está na terceira casa decimal pelo menos para a maioria dos pequenos valores dos graus de liberdade ν_1. Esta é uma precisão semelhante à acurácia encontrada por Li e Martin (2002) utilizando a aproximação proposta por eles que considera a aproximação utilizando um fator de condensação (AFC) por meio da função de distribuição qui-quadrado com ν_1 graus de liberdade. A vantagem da aproximação normal sobre a ACF é fundamentada no fato de que a f.d. normal pode ser mais facilmente obtida do que a f.d. qui-quadrado.
- PALAVRAS-CHAVE: Aproximação; distribuição F; distribuição normal.

References

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Distributions in statistics:* continuous univariate distributions. 2.ed. New York: John Wiley, 1995. v.2.

LI, B.; De MOOR, B. A corrected normal approximation for the student's t distribution. *Comput. Stat. Data Anal.*, Amsterdam, v.29, p.213–216, 1999.

LI, B.; MARTIN, E. B. An approximation to the F distribution using the chi-square distribution. *Comput. Statist. Data Anal.*, Amsterdam, v.40, p.21–26, 2002.

PATEL, J. K.; KAPADIA, C. H.; OWEN, D. B. *Handbook of statistical distributions*. New York: Marcel Dekker, 1976.

SCHEFFÉ, H., TUKEY, J. W. A formula for sample sizes for population tolerance limits. *Ann. Math. Stat.*, Amsterdam, v.15, 217. 1944.

WILSON, E. B.; HILFERTY, M. M. The distribution of chi-square. *Proc. Nat. Acad. Sci.*, Washington, v.17, p.684-688, 1931.

Received in 09.03.2011

Approved after revised in 12.08.2011