Approximations to the *t* Distribution

Bashar Zogheib[1] and Ali Elsaheli[2]

[1,2]Department of Mathematics and Natural sciences
American University of Kuwait
Salmiya, Kuwait

Abstract

In this paper, a few formulas are found to approximate the *t* cumulative distribution function. The formulas found in [5] to approximate the normal distributions are used to get good approximations for *t*. The derived formulas give good approximation for degrees of freedom greater than 3 and if the exact probabilities are less than 0.9.

Keywords: Cumulative distribution function, Approximation, Formula

1 Introduction

Cumulative distribution functions of continuous distributions are widely used in statistics, sciences and engineering. The *t*-distribution is one of the continuous distributions that arise frequently in statistical computations. It is a family of continuous probability distributions that is used to estimate population parameters when the sample size is small or when the population standard deviation is unknown. These probabilities are obtained by integrating the relevant probability density functions. However, these integrals do not have closed form solutions, and, therefore, numerical approximations are found in tables to approximate the *t* distribution. However, it is not quite practical to construct separate tables for all possible probabilities and degrees of freedom. For this reason, numerical integration is used to create simple formulas to approximate these integrals.

There are available a number of highly complex but accurate algorithms for computing such integrals, and there are also some simple ones that have also been proposed in literature. Which ones are preferable depends on the specific problem at hand, and the criteria are usually based on a critical balance between accuracy

and simplicity. In this paper, a new set of formulas will be presented to estimate the *t* cumulative distribution. These formulas are derived using a previous approximation of the normal distribution by [5] and the approximation of *t* by [3].

2 Proposed Formulas

The cumulative distribution function of *t* is given by

$$f(t/n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{t} \left(1 + \frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} dt$$

Where Γ is the gamma function and *n* are the degrees of freedom. This integral does not have a closed form solution and therefore, numerical techniques are needed to approximate it.

The cumulative density function of the normal distribution is given by $F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

A set of formulas for approximating the cumulative density function was obtained by [5]. The most accurate ones and simple in shape were

$$F_1(z) = \left(1 + e^{0.000345z^5 - 0.069547z^3 - 1.604326z}\right)^{-1} \tag{1}$$

$$F_2(z) = \left(1 + e^{0.0054 - 1.6101z\ -0.0674z^3}\right)^{-1} \tag{2}$$

$$F_3(z) = 1 - 0.5e^{-1.2z^3} \tag{3}$$

The following formula to approximate the probabilities for *t* distribution was proposed by [4].

$$pr[t_n < t] = F\left(t\left(1 - \frac{1}{4n}\right)\left(1 + \frac{1}{2n}t^2\right)^{-\frac{1}{2}}\right) \tag{4}$$

In this paper, equations (1) through (4) will be used to find approximations for the probabilities of the *t* distribution. The *z* variable in equations (1) through (3) will be replaced by

$z_1 = t\left(1 - \frac{1}{4n}\right)\left(1 + \frac{1}{2n}t^2\right)^{-\frac{1}{2}}$.   The following approximations for the *t* distribution are found

$$f_1(t/n) = \left(1 + e^{0.000345z_1^5 - 0.069547z_1^3 - 1.604326z_1}\right)^{-1} \tag{5}$$

$$f_2(t/n) = \left(1 + e^{0.0054 - 1.6101z_1 - 0.0674z_1^3}\right)^{-1} \tag{6}$$

$$f_3(t/n) = 1 - 0.5e^{-1.2z_1^3} \tag{7}$$

The accuracy of the proposed formulas is tested by comparing them to the results obtained from EXCEL using the function TDIST and to other existing approximations. Graphs 1 and 2 show the difference between probabilities obtained using formula 5 and the TDIST. The results are shown for degrees of freedom 10 and 6 at different values of t. Graph 1 shows that the maximum absolute error occurs at $t = 2.228$ and is equal to 0.001while the smallest absolute error occurs at $t = 1.372$ and is equal to 0.00008.
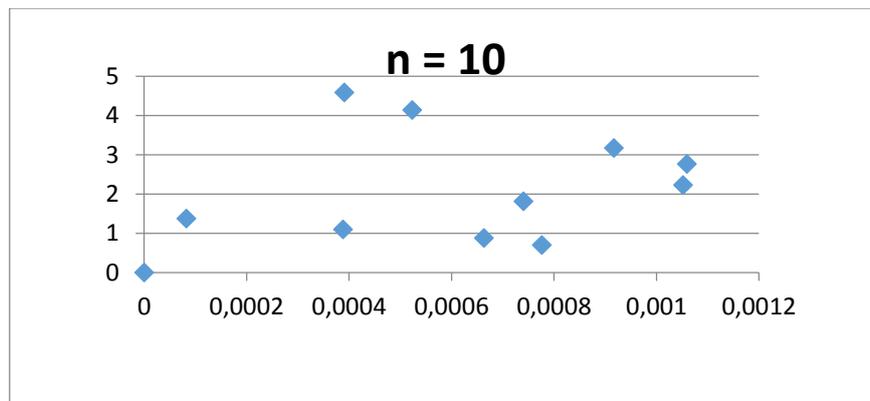
Graph 1: Absolute Error for n = 10

Graph 2 shows that the maximum absolute error occurs at $t = 3.143$ and is approximately equal to 0.0029 while the smallest absolute error occurs at $t = 1.134$ and is approximately equal to 0.0001.
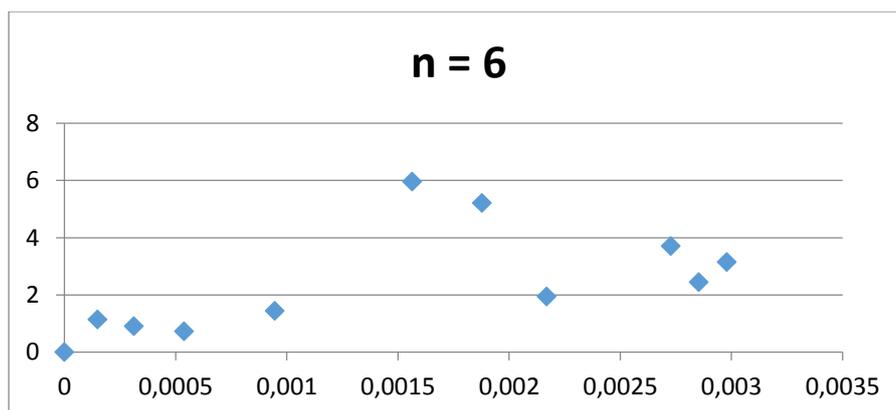
Graph 2: Absolute Error for n = 6

The maximum absolute error found by [4] at $t = 0.4$ for 3 degrees of freedom is 0.000582. In another approximation, [4] found the maximum error for 13 and 14 degrees of freedom at $t = 0.1$ to be 0.0007. It was found by [1] that the maximum absolute error at $t = 1$ for 3 degrees of freedom is 0.00495. A formula created by

[2] gives a maximum absolute error of 0.006982 at $t = 3.8$ for 3 degrees of freedom. Table 1 shows the maximum error obtained using proposed formula (5).

n = 3	n = 13	n = 14	n = 3	n = 3
t = 0.4	t = 0.1	t = 0.1	t = 1	t = 3.8
0.000078	0.00019	0.00020	0.0016	0.013

Table 1: Absolute Maximum Error

It is found that formula (5) gives better approximation than the formulas obtained by [1 & 4]. A slightly better approximation for 3 degrees of freedom at $t = 3.8$ was found by [2].
The results obtained by equation (6) are as accurate as those obtained by equation (5) for probabilities greater than 0.975 but lower in accuracy for smaller probabilities. Equation (7) is the simplest in shape but the approximations obtained are not as accurate as those obtained by equations (5) & (6).

3 Conclusion

In this paper, three approximations to the t cumulative distribution function are found. One of these approximations has a simple form but does not achieve sufficient accuracy, the other two are more complicated in form but achieve higher accuracy. Equation (5) gives the best approximation among the proposed three formulas. Moreover, it gives better results than some of those formulas found in the literature.

References

[1] J.R. Gleason, A note on a proposed student t approximation, *Computational Statistics & Data Analysis*, **34** (2000), 63-66.
http://dx.doi.org/10.1016/s0167-9473(99)00070-5

[2] B. Li and B. De Moor, A corrected normal approximation for student's t distribution, *Computational Statistics & Data Analysis*, **29** (1998), 213-216.
http://dx.doi.org/10.1016/s0167-9473(98)00065-6

[3] N.C. Severo and M. Zelen, Normal approximation to the chi-square and non-central F probability functions, *Biometrika*, (1960): 411-416.
http://dx.doi.org/10.1093/biomet/47.3-4.411

[4] R. Yerukala, N.K. Boiroju and M. K. Reddy, Approximations to the t-distribution, *International Journal of Statistika and Mathematika,* **8**.1 (2013), 19-21.

[5] B. Zogheib and M. Hlynka, *Approximations of the Standard Normal Distribution*. University of Windsor, Department of Mathematics and Statistics, 2009.