Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 98-50

A CORRECTED NORMAL APPROXIMATION FOR THE STUDENT'S t DISTRIBUTION

Baibing Li and Bart De Moor[2]

May 1998

[1]This paper is also available by anonymous ftp from ftp.esat.kuleuven.ac.be as file rep9850.ps.Z in the directory pub/SISTA/baibing/reports.

[2]ESAT/SISTA, K.U.Leuven, Kardinaal Mercierlaan 94, B-3001 Heverlee (Leuven), Belgium, tel.+32-16-32.18.55, fax +32-16-32.19.70, email: baibing.li@esat.kuleuven.ac.be.

[3]This is a revised version for the paper of rep9787.ps.Z. The proof of the theorem in this revised version has been omitted. See rep9787.ps.Z for the proof.

A corrected normal approximation for the Student's *t* distribution

Baibing Li and Bart De Moor

ESAT-SISTA, Dept. of Electrical Engineering, Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94, 3001 Leuven, Belgium
bart.demoor@esat.kuleuven.ac.be

Abstract

For the Student's *t* cumulative distribution function $F(x;n)$ with $n \geq 3$ degrees of freedom, a corrected normal approximation, $\Phi(\lambda x)$, is proposed as an extension of the well-known ordinary normal approximation $\Phi(x)$, where $\Phi(x)$ is the standard normal cumulative distribution and $\lambda=\lambda(x,n)$ is a shrinking factor ($0<\lambda<1$). This approximation has a theoretical error $O(1/n^2)$ uniformly in x. Numerical results show that it can give satisfactory accuracy for even very small n. Thus, it provides a competitive alternative because of its reasonable balance between accuracy and simplicity.

Keywords: approximation; normal distribution; Student's *t* distribution.

For the Student's *t* cumulative distribution function $F(x;n)$ with n degrees of freedom, a major concern raised by Johnson et al. (1995, p372) is how to quickly evaluate the value of $t_{n,\alpha}$ or $F(x;n)$. In this paper we propose a simple and accurate approximation of $F(x;n)$ as an alternative to various nomograms listed in Johnson et al. (1995).

It is well-known that the Student's *t* distribution has an ordinary normal approximation (ONA) for $n \geq 30$, i.e., $F(x;n) \approx \Phi(x)$. However, for small n which is the more important case in practice, this approximation can cause larger error. This paper suggests a natural modification of the ONA, in which a shrinking factor λ ($0<\lambda<1$) is introduced as follows:

$$F(x;n)\approx\Phi(\lambda x), \qquad \text{for } n\geq 3 \text{ with } \lambda=(4n+x^2-1)/(4n+2x^2).$$

For $n=1$ and 2, we suggest using the following exact formulae:

$$F(x;1)=1/2+(1/\pi)tan^{-1}x \quad \text{and} \quad F(x;2)=1/2+(x/2)(2+x^2)^{-1/2}.$$

We call this new approximation corrected normal approximation (CNA). The basic idea of the CNA lies in the fact that the *t* distribution has heavier tail than the normal distribution, i.e., for a large x, there exists x_0 $(0<x_0<x)$ such that $\varphi(x)<f(x;n)=\varphi(x_0)$. This heavy tail can cause relatively large errors when directly using the normal distribution to approximate *t*. The approximation error can be substantially reduced by introducing an appropriate shrinking factor λ $(0<\lambda<1)$ such that $\lambda x\approx x_0$. For this purpose, we use the Cornish-Fisher expansion and normalize it such that the derived shrinking factor $\lambda=(4n+x^2-1)/(4n+2x^2)$ satisfies $0<\lambda<1$.

It is easy to check that the first supiriority of the CNA is that it is always unimodal and non-negative. In contrast, the two important approximations based on direct asymptotic expansion, the Fisher approximation and Hotelling-Frankel approximation (see, Eq.(28.14) and Eq.(28.20) in Johnson et al., 1995), may produce non-unimodal distributions. In addition, the Fisher approximation sometimes gives negative densities.

Moreover, the CNA has a much simpler form in comparison with the Gentleman-Jenkins' empirical approximation which is regarded as one of the best approximation formulae (see, Eq. (28.52) in Johnson et al., 1995), and with the logistic approximation which is given in a piecewise formula for different n (George et al., 1986).

To evaluate the performances of above approximations, we use the maximum absolute error (*MAE*) over the range from 0 to 4 by increment of 0.001:

$$MAE=Max|F(x;n)-H(x)|, \qquad \text{for } x\in \mathcal{X}=\{0.001k \mid k=0,1,\cdots,4000\},$$

where $H(x)$ indexes any approximation. Each evaluation was carried out twice, once by MATLAB version 4.2c, and the other by SAS version 6.11. The numerical results show that both of them produce the same outcome as illustrated in Table I. Some conclusions based on the numerical results in Table I are in order.

TABLE I. Comparison of maximum absolute errors
for six approximations of t cumulative distribution function

n	CNA	Ordinary Normal	Logistic‡	Fisher#	Hotelling-Frankel*	Gentleman-Jenkins
3	7.0e-03†	4.9e-02	-	1.6e-02	4.8e-02	7.6e-03
4	3.2e-03	3.8e-02	-	6.1e-03	2.7e-02	1.1e-03
5	1.7e-03	3.0e-02	-	2.8e-03	1.6e-02	2.7e-04
6	9.3e-04	2.6e-02	7.3e-03	1.4e-03	9.8e-03	1.5e-04
7	5.6e-04	2.2e-02	2.6e-03	8.2e-04	6.1e-03	7.6e-05
8	3.5e-04	1.9e-02	2.2e-03	5.0e-04	3.9e-03	3.7e-05
9	2.3e-04	1.7e-02	4.8e-03	3.2e-04	2.5e-03	1.9e-05
10	1.6e-04	1.6e-02	6.8e-03	2.2e-04	1.7e-03	1.0e-05
11	1.1e-04	1.4e-02	8.5e-03	1.5e-04	1.1e-03	5.4e-06
12	7.7e-05	1.3e-02	1.2e-03	1.1e-04	7.7e-04	3.8e-06
13	5.6e-05	1.2e-02	3.7e-04	8.0e-05	5.3e-04	3.1e-06
14	4.1e-05	1.1e-02	9.7e-04	6.1e-05	3.7e-04	2.5e-06
15	3.0e-05	1.0e-02	1.8e-03	4.7e-05	2.6e-04	2.3e-06
17	1.8e-05	9.2e-03	9.3e-04	2.9e-05	1.3e-04	1.6e-06
20	1.1e-05	7.8e-03	5.1e-04	1.5e-05	5.3e-05	1.1e-06
23	7.7e-06	6.8e-03	4.8e-04	8.9e-06	2.3e-05	1.3e-06
25	6.5e-06	6.3e-03	7.9e-05	6.5e-06	1.4e-05	1.3e-06
30	4.5e-06	5.2e-03	-	3.2e-06	4.8e-06	1.2e-06
35	3.3e-06	4.5e-03	-	1.7e-06	1.9e-06	1.1e-06

† ae+d=$a\times10^{d}$

‡ The logistic approximation was given only for $6\le n\le 28$. See, George et al., 1986.

Expansion to the order n^{-3}. See, Eq. (28.14), Johnson et al., 1995.

* Expansion to the order n^{-4}. See, Eq. (28.20), Johnson et al., 1995.

(i) Compared with the ONA, the CNA significantly improves accuracy. More specifically, even at n=4, the CNA has a better accuracy than the ONA at n=30 (Note that the ONA is usually suggested for $n\ge30$. See, Rohatgi, 1976, §10.4).

(ii) Although the Fisher and the Hotelling-Frankel approximations have higher error orders than the CNA (see below) in theory, the numerical results show that the CNA is

superior to them for $n<25$ and is almost equivalent to them for $25\leq n\leq 35$. Therefore, the CNA is fairly good for the cases of small and moderate n. These cases are most important in practice.

(iii) The Gentleman-Jenkins approximation has the best results among all of the methods for $n\leq 35$. However, as an empirical approximation, its approximation error does not monotonically decrease as n increases. When $n>50$, the error decreases very slowly. In fact, the numerical results (not given in Table I) find that its accuracy is worse than the one of the CNA for $n\geq 54$.

To sum up, the CNA is accurate enough for practical application on one hand and it has a neat form as a natural extension of the ONA on the other hand. Hence, it has a reasonable balance between accuracy and simplicity.

Besides the numerical results given in Table I, we can prove that the CNA has following theoretical bound:

$$\sup_x \left|F(x;n)-\Phi(\lambda x)\right| = O(1/n^2) \qquad \text{for large } n \text{ and } \lambda=(4n+x^2-1)/(4n+2x^2),$$

while most approximations given in Johnson et al. (1995) fail to provide such uniform bound. The proof of the above result is similar to the one in Li & De Moor (1997) and is omitted here. It is available via anonymous ftp from ftp.esat.kuleuven.ac.be as file 97-87.doc.Z in the directory pub/SISTA/baibing/reports.

Acknowledgments

The authors of this paper are grateful for Dr. Xiangkang Yin's helpful comments on the earlier version of this paper. The first author of this paper is a research fellow (post-doctor) of K.U.Leuven and the second author is a senior Research Associate with the

FWO-Flanders. This work was supported by a Fellowship of K.U.Leuven (F97-72), Sustainable Mobility Programme-Project MD/01/24 (1997-2000) (Traffic Congestion Problems in Belgium: Mathematical Models, Analysis, Simulation, Control and Actions), Concerted Research Action GOA-MIPS, the FWO project G.0256.97: Numerical Algorithms for Subspace System Identification, the FWO Research Communities: ICCOS and Advanced Numerical Methods for Mathematical Modeling, and the Belgian Government: Interuniversity Attraction Poles (IUAP P4-02 and P4-24).

References

George, E.O., M. El-Saidi, and K. Singh, A generalized logistic approximation of the Student's *t* distribution, *Commun. Statist.-Simula*.,**15** (1986) 1199-1208.

Johnson, N.L., S. Kotz, and N. Balakrishnan, *Distributions in statistics: continuous univariate distributions*, Vol.2, 2nd edition, (Wiley, New York, 1995)

Li, B. and B. De Moor, Simple approximation relation between the symmetrical generalized logistic and the Student's *t* distributions, *Commun. Statist.-Simula*., **26** (1997), 1029-1039.

Rohatgi, V.K., *An introduction to probability theory and mathematical statistics*, (Wiley, New York, 1976)