Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 97-05

A SIMPLE APPROXIMATION RELATION BETWEEN THE SYMMETRIC GENERALIZED LOGISTIC AND THE STUDENT'S t DISTRIBUTION

Baibing Li and Bart De Moor[2]

January 1997

[1]This paper is also available by anonymous ftp from ftp.esat.kuleuven.ac.be as file rep9705.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.

A SIMPLE APPROXIMATION RELATION BETWEEN THE SYMMETRIC GENERALIZED LOGISTIC AND 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

Key Words: approximation distribution; Student's t distribution; symmetric generalized logistic distribution.

ABSTRACT

In this paper, we obtain a new approximation of the Student's *t* distribution by using the symmetric generalized logistic (SGL) distribution function. The error of this approximation is shown to be $O(1/n^2)$ where n is the degrees of freedom of the t distribution. In comparison to similar approximations by George and Ojo(1980) and George et al. (1986), this new approximation is simpler and more accurate. It is also shown that under some conditions, the *t* distribution is a good approximation of the SGL distribution. Therefore, the complicated expressions for the cumulants and moments of the SGL can be approximated by those of the *t* distribution. Finally, numerical results are given.

1. INTRODUCTION

It is of theoretical and practical interest to consider approximations for distributions. A large body of literature on approximations is summarized by Johnson et al. (1995) and Patel et al. (1976). These approximations often play important roles in statistical inference and computations. Generally speaking, a good approximation should have both satisfactory accuracy and simplicity to use in practice.

The *t* distribution is one of most frequently used distributions and its approximation has been discussed by several researchers (Mudholkar & George, 1978; Ling, 1978; Gentleman & Jenkins, 1968 etc.). Considering its importance, George & Ojo (1980) and

George et al. (1986) proposed two quite accurate approximations based on the logistic distribution.

However, the approximations proposed by George & Ojo (1980) and George et al. (1986) are quite complicated. For instance, in George & Ojo (1980), the parameters were determined by a set of complicated equations, and in George et al.(1986), the approximation distribution function was given by different formulae for different degrees of freedom of the *t* distribution. Moreover, no theoretical errors for their approximations were given to provide bounds on the accuracy of the approximations. Therefore, from the practical point of view, it is necessary to provide a new approximation which has simpler parameter formulae than those of George & Ojo (1980) and George et al. (1986), and meanwhile has accuracy not worse than them.

On the other hand, in recent years, a lot attention has been given to the theory and applications of the logistic distribution (for a comprehensive review, see Balakrishnan, 1992). It is shown that both the ordinary and generalized logistic distributions are very useful in practice. However, for the SGL distribution, its cumulants are generally expressed in infinite series (see, George & Ojo, 1980) which are quite inconvenient to use in practice. Thus, appropriate approximations are needed for numerical computations with either computer or hand calculator.

In this paper, for the *t* distribution with *n* degrees of freedom, we propose a new approximation based on the SGL distribution. The theoretical error of this approximation is shown to be $O(1/n^2)$. Compared with the approximations proposed by George & Ojo (1980) and George et al. (1986), it is much simpler in determining its parameters and is better for numerical results. Moreover, since the SGL can be also approximated by the *t* distribution, we use the moments and cumulants of the *t* distribution to approximate those of the SGL which are quite tedious for computation by George & Ojo (1980).

2. MAIN RESULTS

In this section, first we summarize the approximations by George & Ojo (1980) and George's (1986), then we propose a new one.

Consider the density of the *t* distribution with *n* degrees of freedom *(n≥2)*

$$f(x;n) = C_t[1 + x^2 / n]^{-(n+1)/2}$$

and cumulative distribution function $F(x;n)$, where the normalization coefficient

$$C_t = \Gamma\{(n+1)/2\} / \{(n\pi)^{1/2}\,\Gamma(n/2)\}$$

and $\Gamma(z)$ represents the gamma function. The SGL cumulative distribution function and its density function will be denoted by $L(x;r,s)$ and $l(x;r,s)$ respectively, where the density

$$l(x;r,s) = C_l\,[ch(rx/2)]^{-2s}$$

the normalization coefficient

$$C_l = 2^{-2s}\,r\Gamma(2s) / \{\Gamma(s)\}^2$$

and $ch(x) = [\exp(x) + \exp(-x)]/2$.

2.1. A summary of the approximations by George & Ojo and George et al.

For the *t* distribution, George & Ojo (1980) proposed following approximation

$$F(x;n) \approx L(x;r_1,s_1) \qquad \text{for } n \geq 9 \tag{1}$$

where $r_1 = [(n-2)M/n]^{1/2}$

$$M = \begin{cases} \pi^2/3 - \sum_{i=1}^{m-1} 2i^{-2} & \text{for } s_1 = m \text{ (positive integer)} \\ \pi^2 - \sum_{i=0}^{m-1} 8(2i+1)^{-2} & \text{for } s_1 = m + 1/2 \end{cases}$$

$$n' = \begin{cases} 4 + 5[\pi^2 - 6\sum_{i=1}^{m-1} i^{-2}]^2 / [\pi^4 - 90\sum_{i=1}^{m-1} i^{-4}] & \text{for } s_1 = m \\ 4 + 3[\pi^2 - 8\sum_{i=0}^{m-1}(2i+1)^{-2}]^2 / [\pi^4 - 96\sum_{i=0}^{m-1}(2i+1)^{-4}] & \text{for } s_1 = m + 1/2 \end{cases}$$

and n is the closest integer to n'.

To reduce the laborious computations, George & Ojo (1980) used least squares to approximate n' by $n'=3.25+5.5s_1$ and then, for $s_1=1, 1.5, 2, 2.5, 3, 3.5, 4$, they obtained the corresponding values $n=9, 12, 14, 17, 20, 23, 25$.

George et al. (1986) proposed another approximation

$$F(x;n) \approx L_2(x,r_2) \tag{2}$$

where (i) for $6 \leq n \leq 11$ and $r_2 = \pi[(n-2)/(3n)]^{1/2}$

$$L_2(x,r_2) = [1 + \exp(-r_2 x)]^{-1}$$

(ii) for $12 \leq n \leq 16$ and $r_2 = [(\pi^2 - 6)(n-2)/(3n)]^{1/2}$

$$L_2(x,r_2)=[1+3\exp(-r_2 x)]/[1+\exp(-r_2 x)]^3$$

(iii) for $17\le n\le 22$ and $r_2=[(2\pi^2-15)(n-2)/(6n)]^{1/2}$

$$L_2(x,r_2)=[1+5\exp(-r_2 x)+10\exp(-2r_2 x)]/[1+\exp(-r_2 x)]^5$$

(iv) for $23\le n\le 28$ and $r_2=[(6\pi^2-49)(n-2)/(18n)]^{1/2}$

$$L_2(x,r_2)=[1+7\exp(-r_2 x)+21\exp(-2r_2 x)$$
$$+35\exp(-3r_2 x)]/[1+\exp(-r_2 x)]^7$$

2.2. A simple approximation of the *t* distribution

In this subsection, we propose a new approximation for $F(x;n)$ and $f(x;n)$ by $L(x;r^*,s^*)$ and $l(x;r^*,s^*)$ through deriving two parameter formulae $r^*=r^*(n)$ and $s^*=s^*(n)$.

First, we have following the lemma.

Lemma. If $|x|\le n^{1/3}/3$, then

$$|f(x;n)-l(x;r^*,s^*)|\le(K+x^4+7x^6)f(x;n)/n^2,\qquad \text{for } n=2,3,\cdots$$

where K is a universal constant, $r^*=2[3(n+1)]^{1/2}/n$ and $s^*=n/6$.

The proof of the Lemma is given in APPENDIX.

The main results of this paper are given by the following two theorems.

Theorem 1. For $r^*=2[3(n+1)]^{1/2}/n$ and $s^*=n/6$, we have

$$\sup_x|f(x;n)-l(x;r^*,s^*)|=O(1/n^2),\qquad \text{for large } n. \tag{3}$$

Proof: Let $A=n^{1/3}/3$. First, from the Lemma

$$\sup_{|x|\le A}|f(x;n)-l(x;r^*,s^*)|\le\sup_{|x|\le A}[(K+x^4+7x^6)f(x;n)/n^2]$$

It is easily checked that

$$\sup_x[x^{2m}f(x;n)]<C,\qquad \text{for } m=0,2,3 \text{ and universal constant } C.$$

Hence $\sup_{|x|\le A}|f(x;n)-l(x;r^*,s^*)|=O(1/n^2)$

Second, noting that both $f(x;n)$ and $l(x;r,s)$ are symmetric density functions about the origin and strictly decreasing for $x>0$, we have

$$\sup_{|x|\ge A}|f(x;n)-l(x;r^*,s^*)|\le f(A;n)+l(A;r^*,s^*)$$

Therefore, it follows by noting that

$$f(A;n) = O(1/n^3) \text{ and } l(A;r^*,s^*) = O(1/n^3) \tag{4}$$

Theorem 2. For *r*=2[3(n+1)]$^{1/2}$/n* and *s*=n/6*, we have

$$\underset{x}{Sup}\left|F(x;n) - L(x;r^*,s^*)\right| = O(1/n^2), \quad \text{for large } n. \tag{5}$$

Proof: Let $A=n^{1/3}/3$. First, noting that both $F(x;n)$ and $L(x;r^*,s^*)$ are symmetric cumulative distribution functions about the origin, we have

$$\underset{|x|\geq A}{Sup}\left|F(x;n) - L(x;r^*,s^*)\right| \leq 1 - F(A;n) + 1 - L(A;r^*,s^*) \tag{6}$$

It is easily checked that

$$1 - L(x;r,s) \leq l(x;r,s)/[rs\tanh(rx/2)], \quad \text{for } x>0,\ r>0 \text{ and } s>0.$$

$$1 - F(x;n) \leq [(n+x^2)/(n-1)][1+n^{1/2}/x]f(x;n), \quad \text{for } x>0,\ n=2,3,\cdots$$

Thus, from the above inequalities and (4), we obtain

$$1 - F(A;n) = O(1/n^2) \text{ and } 1 - L(A;r^*,s^*) = O(1/n^2) \tag{7}$$

Second, from (6)

$$\underset{|x|\leq A}{Sup}\left|F(x;n) - L(x;r^*,s^*)\right|$$

$$\leq \left|\int_{-\infty}^{A}[f(x;n) - l(x;r^*,s^*)]dx\right| + \underset{|x|\leq A}{Sup}\left|\int_{A}^{x}[f(x;n) - l(x;r^*,s^*)]dx\right|$$

$$\leq 1 - F(A;n) + 1 - L(A;r^*,s^*) + \int_{-A}^{A}\left|f(x;n) - l(x;r^*,s^*)\right|dx \tag{8}$$

From the Lemma

$$\int_{-A}^{A}\left|f(x;n) - l(x;r^*,s^*)\right|dx \leq n^{-2}\int_{-\infty}^{+\infty}(K + x^4 + 7x^6)f(x;n)dx = O(1/n^2) \tag{9}$$

Therefore, it follows by combining (6)-(9).

Remarks: (i) Theorem 1-2 show that, through the two simple parameter formulae *r*=2[3(n+1)]$^{1/2}$/n* and *s*=n/6*, the Student's *t* distribution can be approximated by the SGL with the approximation error $O(1/n^2)$. This approximation is much simpler than those given by George & Ojo (1980) and George et al. (1986).

(ii) It can be seen from the above theorems and the proof of the Lemma, for $s=(n+c)/6$ (c is any fixed constant), the above results still hold. However, numerical results showed that $c=0$ is a satisfactory choice.

(iii) Similar to the proof of the Lemma in APPENDIX, for any fixed x, we can obtain

$$f(x;n) = \phi[\{(n+1)/n\}^{1/2} x] + O(1/n)$$

$$l(x;r^*,s^*) = \phi[\{(n+1)/n\}^{1/2} x] + O(1/n)$$

where $\phi(x)$ is the density of the standard normal distribution. Therefore, in the sense of the orders of the approximation errors, the SGL $L(x;r^*,s^*)$ is nearer to the t distribution $F(x;n)$ than the normal distribution.

2.3. An approximation of the SGL by the *t* distribution

Directly from the results in last subsection, we conclude that the SGL can also be approximated by the t distribution, that is,

$$l(x;1,s) \approx r^{-1} f(x/r;n) \qquad \text{for large } s \tag{10}$$

with $n=6s$ and $r=[3(6s+1)]^{1/2}/(3s)$.

Since the cumulants and moments of the t distribution are much simpler than those of the SGL, we can use the former to approximate the latter.

George & Ojo (1980) and George et al. (1986) gave the following formula for the pth (p is a positive integer) cumulant of $l(x;1,s)$

$$K_p = \begin{cases} 2[(p-1)!]\sum_{i=0}^{+\infty}(i+s)^{-p} & \text{for even } p \\ 0 & \text{for odd } p \end{cases} \tag{11}$$

To avoid tedious computations for K_p, we consider the pth cumulant K'_p of the Student's t as an approximation of K_p. For instance, if s is not small, we have

$$K_2 \approx K'_2 = r^2[6s/(6s-2)] = 2(6s+1)/[s(6s-2)] \tag{12}$$

The numerical results given in next section show that the above approximation is fairly good even for small s. Therefore, it provides a quite simple approach to approximately compute the cumulants and moments of the SGL for numerical computations with either computer or hand calculator in practice.

3. NUMERICAL COMPARISON

In this section, we show that the approximation given by Theorem 1 and 2 not only hold theoretically for large n but also are quite satisfactory for small n in numerical results.

In order to investigate the accuracy of the approximation (5), we use the Kolmogorov-Smirnov distance between *F(x;n)* and *L(x;r*,s*)* given by

$$d = \sup_{x \in R} \left| F(x;n) - L(x;r^*,s^*) \right|$$

However, since d is very difficult to compute, we consider

$$D = Max\left| F(x;n) - L(x;r^*,s^*) \right|, \quad \text{for } x \in \mathcal{X} = \{-4.0+0.01k \mid k=0,1,\cdots,800\}$$

which is a reasonable approximation of d.

The numerical results are given by Table I.

TABLE I. The distance between *F(x;n)* and *L(x;r*,s*)*

n	*D*	*n*	*D*	*n*	*D*
2	0.019605	15	0.000206	28	0.000073
3	0.008914	16	0.000181	29	0.000068
4	0.004809	17	0.000165	30	0.000064
5	0.002900	18	0.000151	31	0.000061
6	0.001892	19	0.000139	32	0.000057
7	0.001309	20	0.000128	33	0.000054
8	0.000948	21	0.000118	34	0.000052
9	0.000712	22	0.000109	35	0.000049
10	0.000551	23	0.000102	36	0.000047
11	0.000437	24	0.000095	37	0.000044
12	0.000353	25	0.000088	38	0.000042
13	0.000291	26	0.000083	39	0.000040
14	0.000243	27	0.000077	40	0.000038

For comparison, we list the numerical results of George & Ojo (1980) and George et al. (1986) in Table II. From George & Ojo (1980), the difference between *F(x;n)* and its approximation $L(x;r_1,s_1)$ given by (1) was computed as follows:

$$D_1 = Max\left| F(x;n) - L(x;r_1,s_1) \right|$$

for $x \in \mathcal{X}_I$ ={0.10, 0.20, 0.30, 0.50, 0.90, 1.25, 1.75, 2.5, 3.0, 4.0}

From George et al. (1986), the difference between *F(x;n)* and its approximation $L_2(x;r_2)$ given by (2) was computed as follows:

$$D_2 = Max\left| F(x;n) - L_2(x,r_2) \right|$$

for $x \in \mathcal{X}_2$={0.10, 0.30, 0.40, 0.80, 1.0, 1.30, 1.60, 1.90, 2.30, 2.70, 3.0, 3.40, 3.80, 4.0}

(Note that $F(x;n)$, $L(x,r_1,s_1)$ and $L_2(x;r_2)$ are symmetric about the origin).

TABLE II. The distance between $F(x;n)$ and $L(x,r_1,s_1)$, and $F(x;n)$ and $L_2(x;r_2)$

n	D_1	n	D_2
9	0.00894	9	0.004555
11	0.00499	10	0.005560
14	0.00094	13	0.000374
20	0.00048	14	0.000915
23	0.00037	18	0.000401
25	0.00009	19	0.000270
		23	0.000478
		29	0.000289
		38	0.000141

It can be seen from Table I and II that

(i)since $\mathcal{X}_1 \subset \mathcal{X}$ and $\mathcal{X}_2 \subset \mathcal{X}$, and the results in Table I are better than those in table II, the approximation proposed here is better than those by George & Ojo (1980) and George et al. (1986) in precision.

(ii)even for small n, the approximation proposed here is still quite good.

Finally, let $\Delta = K_2' - K_2$, where K'_2 and K_2 are given by (11) and (12). We give the results on the approximation of the *2*nd cumulant of the SGL in Table III.

TABLE III. The *2*nd cumulant of $L(x;1,s)$ and its approximation

s	K'_2	K_2	Δ	s	K'_2	K_2	Δ
2	1.30000	1.28987	1.01×10^{-2}	7	0.30714	0.30709	5.27×10^{-5}
3	0.79167	0.78987	1.80×10^{-3}	8	0.26630	0.26627	3.05×10^{-5}
4	0.56818	0.56765	5.36×10^{-4}	9	0.23504	0.23502	1.88×10^{-5}
5	0.44286	0.44265	2.11×10^{-4}	10	0.21035	0.21033	1.23×10^{-5}
6	0.36275	0.36265	9.93×10^{-5}	20	0.09754	0.09754	7.67×10^{-7}

Obviously, the approximation for the *2*nd cumulant of $L(x;1,s)$ is quite satisfactory even if the parameter s is relatively small.

APPENDIX. THE PROOF OF THE LEMMA

First, consider the ratio $R_1 = [ch(r*x/2)]^{-2s*} / [1 + x^2/n]^{-(n+1)/2}$

From Taylor's Theorem

$$ch(x) = 1 + x^2/2! + x^4/4! + \theta_0 x^6/5! \quad \text{for } |x| < 5/6,\ |\theta_0| < 1 \qquad \text{(A1)}$$

$$\log(1+x) = x - x^2/2 + \theta_0' x^3 \quad \text{for } |x| < 2/3,\ |\theta_0'| < 1 \qquad \text{(A2)}$$

Then, since $|x| \le n^{1/3}/3$ implies $|r*x/2| < 5/6$, from (A1) we obtain

$$ch(r*x/2) = 1 + (r*x/2)^2/2! + (r*x/2)^4/4! + \theta_1 (r*x/2)^6/5! \quad \text{for } |\theta_1| < 1$$

Let $y = (r*x/2)^2/2! + (r*x/2)^4/4! + \theta_1 (r*x/2)^6/5!$

$v = 2s*/[(n+1)/2] = 2n/[3(n+1)]$

Noting $|y| < 2/5$ and then using (A2), we obtain

$$[ch(r*x/2)]^{-v} = \exp[-v\log(1+y)] = \exp[-v(y - y^2/2 + \theta_2 y^3)]$$

$$= \exp[-x^2/n + (n+1)x^4/(2n^3) + 4\theta_3 (n+1)^2 x^6/n^5],\quad |\theta_i| < 1,\ i = 2,3. \quad \text{(A3)}$$

Moreover, noting that $|x| \le n^{1/3}/3$ implies $x^2/n < 1/9$ *(n≥2)* and then using (A2)

$$1 + x^2/n = \exp[x^2/n - x^4/(2n^2) + \theta_4 x^6/n^3] \quad \text{for } |\theta_4| < 1 \qquad \text{(A4)}$$

Thus, combing (A3) and (A4)

$$R_1 = \exp[0.375\theta_5 x^4/n^2 + 6.825\theta_6 x^6/n^2] \quad \text{for } |\theta_5| < 1,\ |\theta_6| < 1 \qquad \text{(A5)}$$

Second, for the t and SGL distributions, consider the ratio of the normalization coefficients of their densities

$$R_2 = C_l / C_t = [2^{-2s*} r*\Gamma(2s*)/\{\Gamma(s*)\}^2] / [\Gamma((n+1)/2)/\{(n\pi)^{1/2}\Gamma(n/2)\}]$$

From the Legender's Duplication Formula (Patel et al., 1976, p244)

$$R_2 = [3(1+1/n)]^{1/2} [\Gamma(n/2)\Gamma(n/6+1/2)] / [\Gamma(n/2+1/2)\Gamma(n/6)] \qquad \text{(A6)}$$

From the Stirling's Formula

$$\Gamma(z)/\Gamma(z+1/2) = [\{1 + 1/(2z)\}^{-2z}]^{1/2} (e/z)^{1/2} [1 + O(1/z^2)] \qquad \text{(A7)}$$

Substituting (A7) into (A6) and after some computations, we get

$$R_2 = [(1+1/n)^{-n+1}(1+3/n)^{n/3}]^{1/2} + O(1/n^2) = 1 + O(1/n^2)$$

Thus, there exists a universal constant *K>0* such that

$$R_2 = 1 + K\alpha / (2n^2) \qquad \text{for } |\alpha| < 1 \tag{A8}$$

Finally, noting the inequality $|\exp(z)| \le |z| \exp(|z|)$, $|R_1| < 1.06$ and from (A5) (A8)

$$|f(x;n) - l(x;r^*,s^*)| = f(x;n)|l(x;r^*,s^*) / f(x;n) - 1| = f(x;n)|R_1 R_2 - 1|$$

$$\le (K + x^4 + 7x^6) f(x;n) / n^2$$

This completes the proof.

ACKNOWLEDGEMENT

The authors of this paper are grateful for the referee’s helpful and detailed comments on the earlier version of this paper. The second author is a senior Research Associate with the FWO-Flanders (Fund for Scientific Research-Flanders). This work was supported by 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 Modelling, and the Belgian Government: Interuniversity Attraction Poles (IUAP P4-02 and P4-24).

BIBLIOGRAPGY

Balakrishnan, N. (1992). *Handbook of the Logistic Distribution*. New York: Marcel Dekker, Inc.

Gentleman, W.M. and Jenkins, M.A. (1968). “An approximation of Student’s *t*-distribution,” *Biometrika*, **53**, 571-572.

George, E.O. and Ojo, M.O. (1980). “On a generalization of the logistic distribution,” *Ann. Inst. Statist. Math.*, **32**, 161-169.

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

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Distributions in Statistics: Continuous Univariate Distributions,* Vol.2, 2nd edition, New York: Wiley.

Ling, R.F. (1978). “Accuracy of approximations for *t*, χ^2 and *F* probabilities,” *J. Ammer. Statist. Assoc*., **73**, 274-283.

Mudholkar, G.S. and George, E.O. (1978). “A remark on the shape of the logistic distribution,” *Biometrika*, **65**, 667-668.

Patel, J.K., Kapadia, C.H. and Owen, D.B. (1976). *Handbook of Statistical Distributions.* New York: Marcel Dekker, Inc.