한국통계학회 논문집 제1권 1호
1994년 12월 pp. 81-86

On an Approximation to the Distribution of Product of Independent Beta Variates[1)]

Hea-Jung Kim[2)]

Abstract

A Chi-square approximation to the distribution of product of independent Beta variates denoted by U is developed. The distribution is commonly used as a test criterion for the general linear hypothesis about the multivariate linear models. The approximation is obtained by fitting a logarithmic function of U to a Chi-square variate in terms of the first three moments. It is compared with the well known approximations due to Box(1949), Rao(1948), and Mudholkar and Trivedi(1980). It is found that the Chi-square approximation compares favorably with the other three approximations.

1. Introduction

Let a random matrix $T : p \times q$ follow the matrix T-distribution given by Kshirsagar (1960), then its density is expressible as

$$f(T) \propto \frac{|Q^{-1}|^{p/2}|P^{-1}|^{(m-q)/2}}{|P^{-1}+TQ^{-1}T'|^{m/2}}, \tag{1}$$

where $P : p \times p$ and $Q : q \times q$ are positive definite matrices, and $m > p+q-1$. From the distribution (1), Geisser(1965) defined a random variable

$$U \equiv \frac{|P^{-1}|}{|P^{-1}+TQ^{-1}T'|}, \quad 0 < U < 1, \tag{2}$$

as $U_{p,q,N}$, where $N = m-q$, and it is shown in Anderson(1984) that $U_{p,q,N}$ follows the distribution of a product of independent Beta variates :

$$U_{p,q,N} = \prod_{i=1}^{p} X_i, \tag{3}$$

where X_i , $i = 1,2,\cdots,p$ are independent Beta variables distributed according to $B(x_i;\alpha_i,\beta)$, where $\alpha_i = (N-i+1)/2$ and $\beta = q/2$.

Developments in multivariate statistics and applied probability often entail the distribution of

1) This work was supported by Dongguk University, Grant DGU-1994
2) Department of Statistics, Dongguk University, Seoul 100-715, KOREA.

$U(\equiv U_{p,q,N})$ either in small samples or asymptotically. Examples occur in commonly used likelihood ratio test statistic for multivariate analysis of variance(cf. Anderson, 1984 ; Press, 1980), in inference for multivariate regression models(cf. Box and Tiao, 1973 ; Press, 1982), and in the analysis of growth curves(cf. Geisser, 1980 ; Kabe, 1986). Several approximations to the distribution of U are available in the literature. Prominent among these are the asymptotic series approximation due to Box (1949), a F approximation due to Rao(1948), and a normal approximation by Mudholkar and Trivedi(1980). A brief survey of such approximations, with references, is in Anderson(1984).

In this paper we propose and assess a new approximation leading to Chi-square approximation not depending on excessive parameters, requiring only the first three moments of a logarithmic function of U.

2. New Approximation

Given a nonnegative random variable Y, we fit the distribution of Y to be almost distributed as AW^r, $r \neq 0$, where W has the $\chi^2_{(k)}$ distribution, and approximate distribution of $(Y/A)^{1/r}$ by an appropriate $\chi^2_{(k)}$, where the constants A, k, and r will be found by a method of matching moments. This method can be summarized as follows.

Let μ, μ_2', and μ_3' denote the first three moments of Y, and assume that they exist. If we match these with respective moments of AW^r we get three equations as

$$\begin{aligned} \mu &= 2^r A\, \Gamma(r+k/2)/\Gamma(k/2), \\ \mu_2' &= 4^r A^2\, \Gamma(2r+k/2)/\Gamma(k/2), \\ \mu_3' &= 8^r A^3\, \Gamma(3r+k/2)/\Gamma(k/2). \end{aligned} \tag{4}$$

Define

$$\begin{aligned} R_1 &= \mu_2'/\mu^2 = \Gamma(k/2)\, \Gamma(2r+k/2)/\{\Gamma(r+k/2)\}^2, \\ R_2 &= \mu_3'/\mu^3 = \{\Gamma(k/2)\}^3\, \Gamma(3r+k/2)/\{\Gamma(r+k/2)\}^3. \end{aligned} \tag{5}$$

These two equations can be solved for r and k; then A is obtained from the expression for μ. Given A, r, and k, the distribution of $(Y/A)^{1/r}$ may be approximated by the Chi-square distribution with k degrees of freedom. Now as $k \to \infty$, $(Y/A)^{1/r}$ may be approximated by a normal distribution, where the mean and variance of the approximating distribution depend on the first three moments of Y, so that Wilson-Hilferty type approximation by Mudholkar and Trivedi(1980) can be thought as a limiting case of the

Chi-square approximation.

For the new(Chi-square) approximation of U, we set $Y = -\log U$ so that

$$\Pr(U \geq U_0) = \Pr(Y \leq U^*_0) \tag{6}$$

for $U_0^* = -\log U_0$. Using the matrix T integral attained from the density (1) (cf. Johnson and Kotz, 1972) ;

$$\int \frac{|Q^{-1}|^{p/2} \; |P^{-1}|^{(m-q)/2}}{|P^{-1} + TQ^{-1}T'|^{m/2}} \, dT = \frac{\pi^{pq/2} \, \Gamma_q\{(m-p)/2\}}{\Gamma_q\{m/2\}}, \tag{7}$$

where $\Gamma_q(z) = \pi^{q(q-1)/4} \Gamma(z) \Gamma(z-1/2) \cdots \Gamma(z-q/2+1/2)$, the h-th moment of U is

$$EU^h = \prod_{i=1}^{p} \frac{\Gamma(\lambda_i/2+h) \, \Gamma((\lambda_i+q)/2)}{\Gamma(\lambda_i/2) \Gamma((\lambda_i+q)/2)+h)} \tag{8}$$

provided $h > -(N-p+1)/2$, where $\lambda_i = N-p+i$ and $N = m-q$. Since the cumulant generating function of Y has the relation ;

$$\Phi_Y(t) = \log E e^{-t\log U} = \log E\, U^{-t}, \tag{9}$$

it is easily obtained from (8) as

$$\Phi_Y(t) = \sum_{i=1}^{p} \log\Gamma(\lambda_i/2 - t) - \sum_{i=1}^{p} \log\Gamma((\lambda_i+q)/2 - t) + B, \tag{10}$$

where B is a constant not involving t. The first three cumulants κ_r, $r = 1,2,3$ of $Y = -\log U$ are obtained from (10) :

$$\begin{aligned}
\kappa_1 &= \sum_{i=1}^{p}\{\Psi^{(0)}\{(\lambda_i+q)/2\} - \Psi^{(0)}\{\lambda_i/2\}\}, \\
\kappa_2 &= \sum_{i=1}^{p}\{\Psi^{(1)}\{\lambda_i/2\} - \Psi^{(1)}\{(\lambda_i+q)/2\}\}, \\
\kappa_3 &= \sum_{i=1}^{p}\{\Psi^{(2)}\{(\lambda_i+q)/2\} - \Psi^{(2)}\{\lambda_i/2\}\}.
\end{aligned} \tag{11}$$

Here

$$\begin{aligned}
\Psi^{(s)}(z) &= \frac{\partial^{s+1}}{\partial z^{s+1}} \log\Gamma(z) \\
&= (-1)^{s+1} s! \sum_{j=0}^{\infty} \frac{1}{(z+j)^{s+1}}, \quad s = 0, 1, \cdots,
\end{aligned} \tag{12}$$

denotes the polygamma function(cf. Abramovitz and Stegun, 1972). For the special case where q is an even number, we see that the series (12) gives a simple expression for the τ-th cumulant of $-\log U$;

$$\kappa_\tau = 2^\tau(\tau-1)! \sum_{i=1}^{p} \sum_{j=0}^{q/2-1} \frac{1}{(N+2i-j+1)^\tau}, \quad \tau = 1,2,3. \tag{13}$$

In case of q odd the τ-th cumulant can be approximated by

$$\kappa_\tau = 2^\tau(\tau-1)!\sum_{i=1}^{p}\left[\sum_{j=0}^{(q-3)/2}\left\{\frac{1}{(N+2i-j+1)^\tau}+\frac{1}{2(N-j+q)^\tau}\right\}\right]. \tag{14}$$

Since the first three moments of Y are functions of the cumulants in (11), in fact $\mu=\kappa_1$, $\mu_2'=\kappa_2+\kappa_1^2$, and $\mu_3'=\kappa_3+3\kappa_1\kappa_2+\kappa_1^3$, substituting the first three moments of $Y=-\log U$ for those in (4) gives appropriate constants A, r, and k for the new approximation. Hence $(-\log U/A)^{1/r}$ may be taken to be approximately distributed as Chi-square with k degrees of freedom, i. e.

$$\Pr(U \geq U_0) \approx \begin{cases} F(U_0^*/A)^{1/r} & \text{for } r>0 \\ 1-F(U_0^*/A)^{1/r} & \text{for } r<0 \end{cases}, \tag{15}$$

where F denotes the cdf of $\chi^2_{(k)}$ distribution and $U_0^* = -\log U_0$.

Computer routines(as an example, SAS/IML) are available to calculate appropriate A, r, and k and then to calculate probabilities or significance points of $\chi^2_{(k)}$ even with noninteger degrees of freedom.

3. Numerical Comparisons

The goal of this section is to study the overall effectiveness of the Chi-square approximation for the distribution of U and to identify some situations where one would(and would not) expect good approximation. The performance of the approximation (CHI) is compared with the other approximations : (i) Rao's approximation(RAO); (ii) Box's approximation(BOX); (iii) normal approximation(NOR) by Mudholkar and Trivedi(1980). The comparisons are made by calculating the error of approximation for significance levels of .01 and .05 for N from 4 to 66, $p=3$, 7, and $q=2$, 6, 10.

Exact critical values U_0 of U for each p, q and N were obtained by using the correction factors in Table 1 of Anderson(1984, p.609) and the probabilities $\Pr(U \geq U_0)$ were computed using each of the four approximations. Box's approximation was used with only the first term approximation that has the remainder term $O(N^{-2})$ (cf. Anderson, 1984, p.317). Box's third term approximation with the remainder term $O(N^{-6})$ has almost the same pattern of the approximation errors as Rao's approximation. Thus we did not use it in our

Table 1. Errors of the Four Approximations for $\alpha = .05$ and $.01$:

Error=$\{(1-\alpha)$-approx.prob.$\} \times 10^5$

			$q=2$				$q=6$				$q=10$			
α	p	N	CHI	NOR	RAO	BOX	CHI	NOR	RAO	BOX	CHI	NOR	RAO	BOX
.05	3	4	-87	-28	-5	-1996	1	17	-390	-4257	1	26	-581	-4894
		8	-12	-37	-2	-399	0	-15	-38	-1702	-4	-9	-63	-3104
		12	-93	-35	4	-159	-2	1	3	-838	-1	-12	-23	-1866
		17	-8	-32	8	-66	7	-14	-6	-453	-6	-8	-8	-1102
		32	-7	-44	-2	-23	1	-20	-9	-143	-7	-28	-23	-394
		62	4	-36	5	0	5	-11	1	-36	-2	13	20	-90
.05	7	8	0	-2	-11	-4393	-1	44	-1503	-4817	6	67	-2243	-4970
		12	-1	-4	-1	-1824	-13	24	-142	-2828	-15	27	-311	-3822
		16	2	-16	-6	-924	2	18	-22	-1652	-8	-20	-117	-2621
		21	0	-22	-10	-498	10	10	0	-964	-21	-69	-99	-1733
		36	7	-28	-13	-161	11	17	20	-308	7	6	4	-629
		66	0	-8	8	-33	-13	-13	-7	-105	5	0	4	-196
.01	3	4	28	12	2	-586	-3	3	-143	-953	-59	5	-200	-997
		8	20	10	-3	-136	0	6	-10	-476	0	-1	-22	-767
		12	26	10	-2	-58	-9	11	2	-247	0	1	-7	-502
		17	-9	13	0	-26	-3	12	4	-132	2	-1	-7	-314
		32	-2	13	1	-7	-9	10	2	-39	-3	9	4	-106
		62	1	14	2	0	-1	9	2	-10	-1	1	-4	-37
.01	7	8	-1	9	2	-967	-110	-4	-471	-994	-165	2	-637	-999
		12	2	7	0	-515	-32	-2	-59	-711	-71	3	-104	-873
		16	-9	9	2	-275	1	-2	-19	-451	-30	7	-26	-647
		21	-7	7	-1	-152	5	3	-5	-272	-30	0	-12	-443
		36	-11	11	2	-45	-3	3	-1	-95	-18	-2	-5	-180
		66	0	12	4	-9	-5	4	1	-28	-5	7	6	-52

comparison. Errors of the four approximations for selected values from computations are given in Table 1.

Several points were noted from Table 1. The maximum error of the suggested approximation is less than .0017 ; in most cases the error is considerably less and the approximation is nowhere strikingly fool. The error for the other approximations are much larger, especially for small values of N. However, some exceptions are also noted from the table : (i) As mentioned in Anderson(1984), for $q=2$ Rao's approximation coincides with the exact distribution of U. Thus the small errors observed in Table 1 for this case are due to errors in the F probability estimates using SAS/IML and to rounding errors. (ii) For $p \geq 6$,

$q \geq 6$ and small values of N, the normal approximation performs better than the others. In spite of the exceptional cases, we can safely conclude that the Chi-square approximation compares favorably with the other three approximations, and hence it can be considered as an another useful approximation for the distribution of U.

References

[1] Abramovitz, M and Stegun, I.(1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C..

[2] Anderson, T. W.(1984). *An Introduction to Multivariate Analysis*, John Wiley Sons Inc.,New York.

[3] Box, G. E. P.(1949). "A general distribution theory for a class of likelihood ratio criteria," *Biometrika*, 36, 317-346.

[4] Box, G. E. P. and Tiao, G. C.(1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Company Inc., Massachusetts.

[5] Geisser, S.(1980). "Growth curve analysis," *Handbook of Statistics*, Vol. 2, ed. by P.R. Krishnaiah, North Holland Publishing Co., New York.

[6] Johnson, N. L. and Kotz, S. (1972). Distributions in Statistics : *Continuous Multivariate Distributions in Statistics*, John Wiley & Sons, New York.

[7] Kabe, D. G.(1986). "On a GMANOVA model likelihood ratio test criterion,"*Communications in Statistics : Theory and Methods*, 15, 3419-3427.

[8] Kshirsagar, A. M.(1960). "Some extensions of the multivariate t-distribution and the multivariate generalization of the distribution of the regression coefficients," *Proceedings of the Cambridge Philosophical Society*, Vol. 57, 80-85.

[9] Mudholkar, G. S. and Trivedi, M. C.(1980). "A normal approximation for the distribution of the likelihood ratio statistic in multivariate analysis of variance," *Biometrika*, 67, 485-488.

[10] Press, S. J.(1980). "Bayesian inference in MANOVA," *Handbook of Statistics*, Vol. 2, ed. by P. R. Krishnaiah, North Holland Publishing Co., New York.

[11] Press, S. J.(1982). *Applied Multivariate Analysis : Using Bayesian and Frequentist Methods of Inference*, Krieger Publishing Co., Florida.

[12] Rao, C. R.(1948). "Testing of significance in multivariate analysis," *Biometrika*, 35, 58-79.