χ^2분포의 백분위수의 개선에 관한 연구
– Improving Percentile Points of χ^2 Distribution –

李 喜 春*

ABSTRACT

Generally there are three methods to derive an approximation formula: 1) approching standard normal distribution by appropriate changing variable 2) using standardization variable for expansion 3) deriving approximation formula by direct method.

This paper present correction terms having the form of $C_1 / \nu^{n/2} + C_2$ (n=1,2) with respect to $\chi^2_\alpha(\nu)$ distribution ($\nu \le 30$),which minimize the error by EDA method and least square method.

1. 근사식의 고찰

Fisher[7]는

$$\chi^2_\alpha(\nu) = (U_\alpha + \sqrt{2\nu-1})^2/2$$

단, U_α는 정규분포의 상측확률의 백분위수, ν는 자유도.

근사식을 제안하였는데 이것은 $\sqrt{2\chi^2}$이 근사정규분포를 따른다는 것에 기인한 것이다.

Wilson Hilferty[22]는 $(\chi^2/\nu)^{1/3}$이 근사정규분포를 한다는 사실에 의하여 다음과 같은 근사식을 제안하였다.

$$\chi^2_\alpha(\nu) = \nu\{1-2/9\nu + U_\alpha\sqrt{2/9\nu}\}^3$$

Cornish-Fisher[8]는 $(\chi^2_\alpha(\nu) - \nu)/\sqrt{2\nu}$에 대해 Cornish-Fisher전개식을 이용하여

$$\chi_\alpha{}^2(\nu) = \nu + \sqrt{\nu}(U_\alpha\sqrt{2}) + 2(U^2_\alpha - 1)/3 + (U^3_\alpha - 7U_\alpha)/9\sqrt{2\nu} - (6U^4_\alpha + 14U^2_\alpha - 32)/405\nu$$
$$+ (9U^5_\alpha - 256U^3_\alpha - 433U_\alpha)/4860\nu\sqrt{2\nu} + 0(\nu^{-3/2})$$

의 근사식을 제안하였다.

한편 Cornish-Fisher의 전개식과 χ^2의 변수변환을 이용하면

$$\chi^2_\alpha(\nu) = (U_\alpha + \sqrt{2\nu-1})^2 + (U^2_\alpha - 1)/6 + 0(\nu^{-1})$$

과 같은 근사식을 얻을 수 있는데 이것은 Fisher의 근사식과 유사한 형태를 하고 있으므로 수정 Fisher의 근사식이라 한다.

또한 χ^2의 변수변환과 Cornish-Fisher의 전개를 이용하면 Wilson-Hilferty 근사식을 얻을 수 있다.

---

* 상지대학교 통계학과
접수 : 1993년 10월 20일
확정 : 1993년 11월 6일

Takeuchi[26]는 Fisher의 근사식에 대해 다음과 같은 수정항을 제시하였다.

<table 1> Correction terms of Takeuchi

α	0.99	0.95	0.5	0.05	0.01
Cα	0.68	0.28	-0.17	0.25	0.78

Heyworth[10]는 Fisher의 근사식을 변형한

$$\chi^2_\alpha(\nu) = (U_\alpha + \sqrt{2\nu})^2/2$$

을 제안하였다.

Thompson[19]은 Wilson-Hilferty의 근사식에 대해 $1/\sqrt{\nu}$ 를 이용하여 다음과 같은 수정항을 제시하였다.

수정항 : $C/\sqrt{\nu}$

<Table 2> Correction terms of Thompson

α	C	α	C
0.995	0.233	0.250	0.039
0.99	0.157	0.100	0.056
0.975	0.067	0.050	0.035
0.95	0.011	0.025	-0.015
0.75	-0.046	0.005	-0.227
0.5	-0.013		

Severo-Zelen[18]은 Wilson-Hilforty의 근사식에 자유도를 변형하여

$$\chi^2_\alpha(\nu) = \nu\ \{1-2/9\nu+(U_\alpha-h_\nu)\sqrt{2/9\nu}\}^3$$

$$단,\quad h_\nu = -\frac{2}{3^2\nu}\left\{\frac{2^{-\frac{3}{2}}(U^2_\alpha-1)}{3\sqrt{\nu}} - \frac{(U^3_\alpha-3U_\alpha)}{4}\right\}$$

와 같은 근사식을 제시하였다.

한편 Zar[24]는 Severo Zelen과 유사한 방법으로

$$\chi^2_\alpha(\nu) = \nu\ \{1-2/9\nu+(U_\alpha-h_\nu') \times \sqrt{2/9\nu}\}^3$$

을 제안하였다.

$$단,\ h_\nu' = \{(9\nu+16)(U_\alpha-3U_\alpha)-24(U^2_\alpha-1)\sqrt{2\nu}\}\ /\ 486\nu^2$$

Gilbert[9]는 자유도가 $1 \le \nu \le 30$ 인 경우에 대해 다중회귀방법을 이용하여 몇가지 상측확률에 대해 아래와 같은 근사식을 제시하였다.

$$\chi^2 = a_0 + a_1\nu + a_2\nu + a_3 \ln\nu$$

< Table 3 > Gilbert's table ($1 \le \nu \le 30$)

coefficient \ α	0.05	0.01	0.001
a_0	2.521823	5.174627	9.205913
a_1	1.282189	1.402766	1.542489
a_2	-0.00211427	-0.00303260	-0.00411568
a_3	1.371169	1.858993	2.314035

Aroian[3]은 Gilbert와 유사하게 자유도 $30 \le \nu \le 50$ 에 대해 최소자승법을 이용하여 다음과 같은 근사식을 얻었다.

$$\chi^2_\alpha(\nu) = \nu + \sqrt{2\nu}t$$

단,

$$t(50\%) = -0.16636\alpha_3$$
$$t(40\%) = 0.25335 - 0.15567\alpha_3 - 0.012276\alpha_3^2$$
$$t(30\%) = 0.52440 - 0.12058\alpha_3 - 0.024541\alpha_3^2$$
$$t(25\%) = 0.67449 - 0.090613\alpha_3 - 0.030693\alpha_3^2$$
$$t(20\%) = 0.84162 - 0.048433\alpha_3 - 0.036788\alpha_3^2$$
$$t(10\%) = 1.28155 + 0.107033\alpha_3 - 0.04797\alpha_3^2$$
$$t(5\%) = 1.64485 + 0.28392\alpha_3 - 0.04902\alpha_3^2$$
$$t(2.5\%) = 1.95997 + 0.47228\alpha_3 - 0.04304\alpha_3^2$$
$$t(1\%) = 2.32635 - 0.73330\alpha_3 - 0.024957\alpha_3^2$$
$$t(0.5\%) = 2.5758 - 0.93600\alpha_3 - 0.00377\alpha_3^2$$
$$t(0.1\%) = 3.0903 - 1.4190\alpha_3 - 0.05667\alpha_3^2$$
$$t(0.01\%) = 3.7200 - 2.1260\alpha_3 + 0.17449\alpha_3^2$$

단, $\alpha_3 = \sqrt{8/\nu}$

만약 하측확률 $t(\alpha\%)$,$t(\alpha\%) = a_0 + a_1\alpha_3 + a_2\alpha_3^2$ 일때 상측확률에 대해서는 $t((1-\alpha)\%)$, $t((1-\alpha)\%) = -a_0 + a_1\alpha_3 - a_2\alpha_3^2$ 을 계산한다.

Hoaglin[11]은 자유도 $\nu \le 30$에 대해 EDA방법과 회귀선의 추정을 이용하여 우측확률과 상측확률에 대해 다음과 같은 근사식을 제시하였다.

χ^2 분포의 상측확률에 대해서는

$$\chi_\alpha^2\{1.00991\sqrt{\nu} + 1.95188(-\log_{10}\alpha)^{1/2} - 1.14485\}^2 ,$$

보다 정확한 근사도를 갖는식은

$$\chi_\alpha^2 = \{1.06807\sqrt{\nu} + 2.13161(-\log_{10}\alpha)^{1/2} - 0.04589\sqrt{\nu}(-\log_{10}\alpha)^{1/2} - 1.37266\}^2 ,$$

간단한 근사식으로는

$$\chi_{\alpha}^{2} = \{\sqrt{\nu} + 2(-\log_{10} \alpha)^{1/2} - 7/6\}^{2} .$$

을 제안하였으며 하측확률에 대해서는

$$\chi^{2}_{(1-\alpha)} = \{0.97657\sqrt{\nu} - 1.46049(-\log_{10} (1-\alpha)^{1/2}) + 0.59025\}^{2} ,$$

보다 좋은 근사식을 갖는 근사식은

$$\chi^{2}_{(1-\alpha)} = \{1.14309\sqrt{\nu} - 0.94591(-\log_{10} (1-\alpha))^{1/2} - 0.13138\sqrt{\nu}(-\log_{10}(1-\alpha))^{1/2} - 0.06198\}^{2},$$

간단한 근사식으로는

$$\chi_{\alpha}^{2} = \{\sqrt{\nu} - 2/3(-\log_{10} (1-\alpha))^{3} + 0.59025\}^{2} .$$

을 제안하였다.

2. 근사식의 평가

이상의 여러 근사식들을 비교 분석한 결과,상측확률에 따라 최소의 오차를 갖는 근사식이 다르다는 것을 알수 있었으며 대체적으로 Corrected Fisher, Zar의식이 작은 오차를 나타내고 있음을 알수 있었고, Fisher, Hoaglin, Cornish-Fisher등의 오차는 상대적으로 크게 나타났다.

< Table 4 > Approximation formula with least error

df	1%	5%	95%	99%
3	0.024169 ①	0.000287 ⑤	0.010979 ③	0.012866 ②
4	0.009163 ④	0.001154 ⑤	0.010154 ②	0.013195 ②
5	0.023783 ④	0.001070 ⑤	0.005817 ④	0.012098 ②
6	0.021158 ②	0.000770 ⑤	0.005838 ②	0.010735 ②
7	0.017240 ②	0.000512 ⑤	0.004519 ②	0.009472 ②
8	0.014408 ②	0.000197 ⑤	0.003554 ②	0.008353 ②
9	0.012212 ②	0.000089 ⑤	0.002818 ②	0.007443 ②
10	0.010503 ②	0.000242 ⑤	0.002261 ②	0.006638 ②
11	0.005380 ③	0.000429 ⑤	0.001818 ②	0.005956 ②
12	0.001857 ③	0.000656 ⑤	0.001473 ②	0.005392 ②
13	0.007232 ②	0.000829 ⑤	0.001187 ②	0.004907 ②
14	0.006488 ②	0.000886 ⑤	0.000961 ②	0.004496 ②
15	0.005949 ②	0.000989 ⑤	0.000770 ②	0.004127 ②
16	0.005435 ②	0.001041 ⑤	0.000611 ②	0.003803 ②
17	0.004920 ②	0.001134 ⑤	0.000485 ②	0.003525 ②
18	0.004570 ②	0.001214 ⑤	0.000379 ②	0.003275 ②
19	0.004291 ②	0.001245 ⑤	0.000266 ②	0.003055 ②
20	0.003955 ②	0.001295 ⑤	0.000186 ②	0.002859 ②

①:Wilson-Hilferty, ②:Zar, ③:Severo-Zehen, ④: Heyworth
⑤:Correct Fisher
* 번호는 최소의 오차인 근사식.

따라서 흔히 언급되는 상측확률인 α = 0.01,0.05,0.95,0.00 에대해 작은 오차를 갖는 근사식을 찾고자 했다.($\nu \le 20$)

위에서 언급된 여러 방법중 EDA방법과 선형변환및 최소자승법을 이용하여, 오차가 제일 크게 나타나는 Fisher의 근사식을 분석하여 아래와 같은 수정항을 찾았다.

< Table 5 > New correction terms

α	수정항 (1)식	수정항 (2)식
0.01	$0.4263/\nu + 0.8121$	$0.3820/\nu + 0.8205$
0.05	$-0.0154/\nu + 0.2863$	$-0.0102/\sqrt{\nu} + 0.2878$
0.95	$-0.3074/\nu + 0.2861$	$-0.3171/\nu + 0.2880$
0.99	$-1.6930/\nu + 0.6669$	$-1.6650/\nu + 0.6616$

위의 수정항에 대해 다른 근사식과의 오차의 비교는 <표 6>과 같다.

<Table 6> (1),(2) and other formula with least error

구분	1%	5%	95%	99%
df	오 차	오 차	오 차	오 차
3	.010708 ②	.000287 ⑤	.005166 ②	.004163 ②
4	.000145 ②	.000390 ②	.000551 ①	.000758 ②
5	.003933 ①	.000019 ①	.002624 ①	.000864 ①
6	.006196 ②	.000161 ②	.002877 ②	.002328 ②
7	.006547 ②	.000211 ②	.002528 ②	.002869 ②
8	.006128 ②	.000145 ②	.002040 ②	.003021 ②
9	.005361 ②	.000064 ②	.001520 ②	.002844 ②
10	.004365 ②	.000085 ②	.001031 ②	.002508 ②
11	.003189 ②	.000049 ②	.000573 ②	.002034 ②
12	.001857 ③	.000045 ②	.000164 ②	.001450 ②
13	.000755 ②	.000019 ①	.000211 ②	.000805 ②
14	.000463 ②	.000045 ①	.000543 ②	.000115 ②
15	.001754 ②	.000015 ①	.000343 ①	.000585 ②
16	.002635 ①	.000028 ①	.000112 ①	.001298 ②
17	.001705 ①	.000007 ①	.000091 ①	.001601 ①
18	.000694 ①	.000038 ①	.000275 ①	.000974 ①
19	.000324 ①	.000022 ①	.000266 ④	.000342 ①
20	.001216 ①	.000020 ②	.000186 ④	.000288 ①

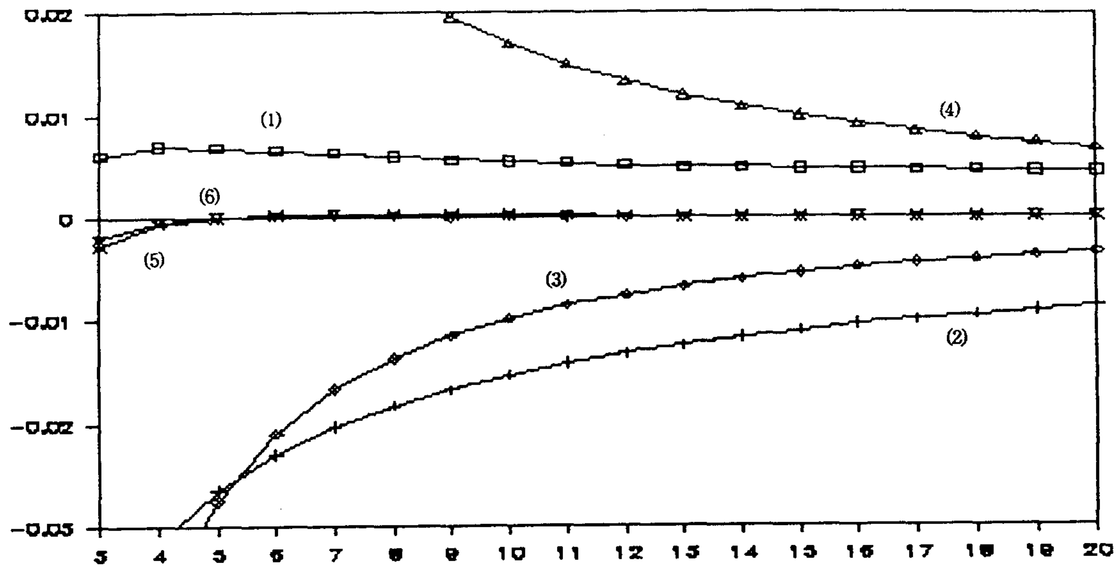
① : (1)식, ② : (2)식, ③ : Severo-Zehen, ④ : Zar

⑤ : Correct Fisher

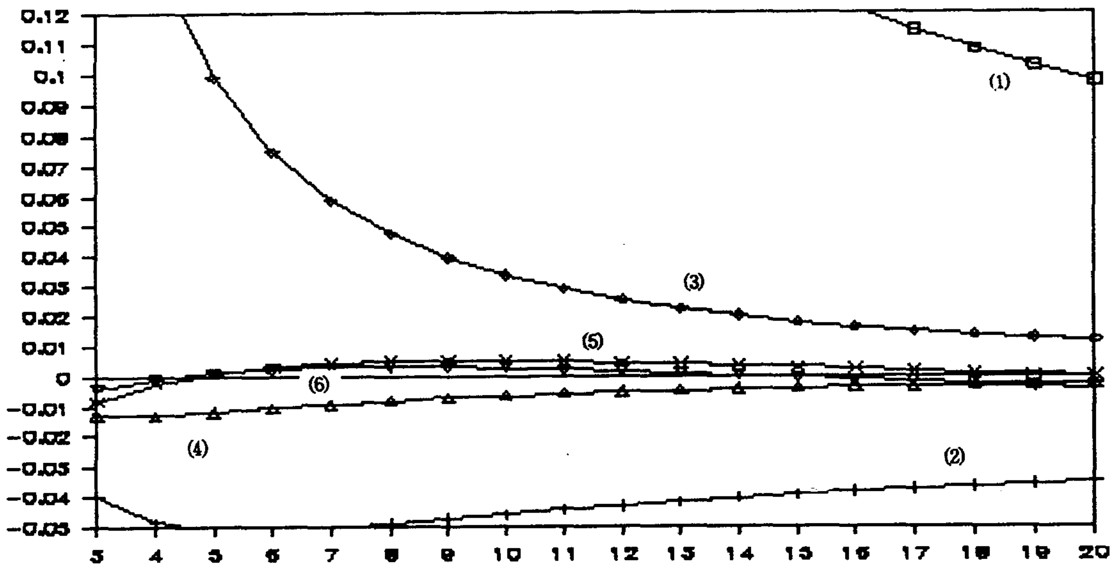
* 번호는 최소의 오차인 근사식.

위에 표에서 보는 바와같이 제시한 수정항은 α = 0.01,0.05,0.95,0.99의 상측 확률에서 위에서 언급한 근사식들 보다 대부분 오차가 작게 나타 났다.

α = 0.05



α = 0.99



(1):Takeuchi (2):Wilson-Hilferty (3):Cornish-Fisher (4):Zar (5):수정항 (1)식 (6):수정항 (2)식

< Fig.1> Graphical Comparison of Error

3. 근사식의 적용 및 검토

위에서 제시한 근사식들은 통계량의 검정에서 자유도가 소수로 나타나는 경우에 사용 되어 질수있다. 이제 제시된 수정항에대해 비심 χ^2의 근사식의 오차에는 얼마만큼 영향은 미치는가를 보이기 위해 Patnaik 이 제시한 비심 χ^2의 근사식과 Pearson이 제시한 비심 χ^2의 근사식

$$\chi_n'^2(\lambda) = C\chi^2_\nu$$

$$단,\ C = \frac{n+2\lambda}{n+\lambda}, \quad \nu = \frac{(n+\lambda)^2}{n+2\lambda}$$

$$\chi_n'^2(\lambda) = C x^2_{\nu} + b$$

$$단, \quad C = \frac{n+3\lambda}{n+2\lambda}, \qquad \nu = \frac{(n+2\lambda)^3}{(n+3\lambda)^2}, \qquad b = -\frac{\lambda^2}{n+3\lambda}$$

에 대해 오차의 영향을 분석한 결과 제시된 수정항들은 Patnaik, Pearson의 근사식과 오차의 차이가 없었다. 이것은 χ^2근사식을 이용한 다른 분포의 근사식에도 적용 할 수 있다는 것이다.

Reference

1. ABDEL-ATY, S.H, Approximate formulae for the percentage points and the probability integral of the noncentral χ^2 distribution, Biometrika 41 (1954).
2. ABRAMOWITZ, M and STEGUN LA, Handbook of Mathematical function, Dover publication (1972).
3. AROIAN, L.A, A new approximation to the levels of sigbificance of the chi-square distribution, Ann Math Statist 14 (1943).
4. B Stein, H, Close approximation of percentage points of the chi-square distribution and Poisson confidence limits,J.A.S.A 69 (1973).
5. BEST, D.J and ROBERTS, D.E, The percantage points of the χ^2-distribution, Appl Statist, 24 (1975).
6. DAWSON F.H, Alternatives to the use of the tabulated values of distribution in Statistical programmes, NATURE 356 (1975).
7. FISHER, R.A, Statistical Methods for Reseasch Workers, Edinburgh Oliver & Boyd (1921).
8. FISHER,R.A and CORNISH, E.A, Percantile points of distribution having known cumulants, Technometrics 2 (1960).
9. GILBERT, R.I, A Simple formulae for interpolating tables of χ^2, Biometrics 33 (1977).
10. HEYWORTH, M.R, Approximation to chi-square, The American statistician 30 (1976).
11. HOAGLIN, D.V, Direct Approximation for chi-square percentage points, JASA 72 (1977).
12. Johnson & KOTZ, Continuous univariate distribution, Wiely & Sons (1972).
13. KENDALL and STUART, THE Advanced theory of statistics, GRIFFIN LONDON (1977).
14. MARDIA, K.V and ZEMROCH, P.J, Table of the F and Related distribution with algorithm, Academic Press (1978).
15. PATNAIK J.H. The non-central χ^2 distribution and F-distribution and their applications. Biometrika, 34. 202-232, 1949.
16. PEARSON ,E.S. Note on an approximation to the the distribution of non-central χ^2. Biometrika, 46, 364, 1959.
17. SANKARAN, M. Approximate to the chi-square distribution. Biometrika 50, (1963).
18. SEVERO, N.C and ZELEN, M, Normal approximation to the Chi-square and noncentral F Probability functions, Biometrika 47 (1960).
19. THOMPSON, C.M, Table of percentage points of the χ^2 distribution, Biometrika 32 (1941).
20. TUKEY, J.W, Exploratory Data Analysis, Addison-Wesley publishing (1977).
21. WHITE, J.S, Table of normal percentile points, JASA 65 (1970).
22. WILSON, E.B and HILFERTY.M.M, The distridution of chi-square, Proc.not.acad sci 17 (1931).
23. ZAR, J.H, Biostatistical Analysis. Englewood cliffs, N.J Prentice-hall (1974).
24. ZAR, J.H,Approximation for the percentage points of the chi-squared distribution, Appl.Statist 27 (1978).
25. 統計數値標, 日本規格協會, (1972).
26. 竹內啓, 確率分布と統計解析, 日本規格協會, (1979).
27. 竹內啓, 確率分布の近似,教育出版,(1975).
28. 柴田義貞, 正規分布の特性と應用,東京大學 出版會, (1981).