비중심카이제곱분포 함수에 대한 효율적인 알고리즘

구 선 희†

요 약

비중심 χ^2분포의 누적분포 함수의 계산은 χ^2검정에서 검정력 계산에 요구된다. 본 논문에서는 중심 χ^2분포 함수를 통하여 비중심 χ^2분포 함수의 계산을 구하는 알고리즘을 제시하고 있으며 기존의 접근 방법에 의한 계산 결과와 비교하였다.

An Effective Algorithm for the Noncentral Chi-Squared Distribution Function

Son Hee Gu†

ABSTRACT

The evalution of the cumulative distribution function of the noncentral χ^2 distribution is required in approximate determination of the power of the χ^2 test. This article provides an algorithm for evaluating the noncentral χ^2 distribution function in terms of a single "central" χ^2 distribution function and compared various approximations.

키워드 : 알고리즘(Algorithm), 비중심 χ^2분포(Noncentral Chi-Squared Distribution), 확률 값(Probabillities Values)

1. 서 론

비중심 χ^2분포는 1928년 R. A. Fisher에 의해 다중 상관계수 R의 분포를 계산하는 과정에서 유도되었으며[4], Miller와 Park 등에 의하면 비중심 χ^2분포는 일반화된 Rayleigh 분포로서 Rayleigh-Rice 또는 Rice 분포라고 불리우고 있다[8,9]. 이와 같은 비중심 χ^2분포 함수의 계산을 통하여 범주형자료의 분석에 흔히 사용되는 적합도검정 및 독립성, 동질성검정 등과 같은 χ^2검정의 검정력을 구할 수 있다[2, 7]. 비중심 χ^2분포에서 분포 함수(distribution function)의 값을 계산하는 데에는 일반적으로 많은 시간이 소요될 뿐 아니라 정확성이 문제가 되고 있다. 이로 인하여 비중심 χ^2분포 함수의 계산에 대한 여러 접근 방법들이 제시되었다.

본 논문에서는 새로운 접근 방법으로 중심 χ^2분포 함수를 통하여 비중심 χ^2분포 함수의 계산을 구하는 알고리즘을 제시하고 있으며 기존의 접근 방법에 의한 계산 결과와 비교하였다.

† 정 회 원 : 전주대학교 교양학부 컴퓨터 강의전담전임강사
논문접수 : 2001년 1월 19일, 심사완료 : 2002년 5월 16일

2. 비중심 χ^2분포 함수

비중심 χ^2분포는 다음과 같이 정의한다.

정의

확률 변수 $X_1, X_2, \cdots, X_v$가 서로 독립이고 각각 표준정규분포를 따르며 $\delta_1, \delta_2, \cdots, \delta_v$가 상수일 때 $\sum_{j=1}^{v}(X_j+\delta_j)^2$의 분포를 자유도가 v이고 비중심모수가 $\lambda=\sum_{j=1}^{v}\delta_j^2$인 비중심 χ^2분포(non-central χ^2 distribution)라 하며 $\chi_v'^2(\lambda)$로 나타낸다.

비중심 χ^2분포의 확률밀도 함수의 유도 과정은 크게 Guenther에 의한 정의를 이용한 방법[5]과 Vaart에 의한 적률생성 함수(moment generating function : m.g.f)에 의한 방법이 있다[13].

정의를 이용한 방법으로는 다음과 같이 허수를 갖는 Bessel 함수로서 비중심 χ^2분포를 나타내는 Fisher가 제시한 확률밀도 함수가 있다[4].

$$p_{\chi^2}(x)=\frac{1}{2}\left(\frac{x}{\lambda}\right)^{\frac{1}{4}(v-2)} I_{\frac{1}{2}(v-2)}(\sqrt{\lambda}\sqrt{x}) \times \exp\left[-\frac{1}{2}(\lambda+x)\right] \quad (x>0) \tag{1}$$

여기서 $I_k(x)=\left(\frac{1}{2}x\right)^k \sum_{j=0}^{\infty} \frac{\left(\frac{1}{4}x^2\right)^j}{j!\Gamma(k+j+1)}$ 는 차수가 k 인 제1종의 수정된 Bessel 함수이다.

적률생성 함수를 이용한 방법으로는 포아송분포와 중심 χ^2분포로서 비중심 χ^2분포를 나타내는 Patnaik가 제시한 확률밀도 함수가 있다[10].

$$p_{\chi^2}(x)=\sum_{j=0}^{\infty}\left[\frac{\left(\frac{1}{2}\lambda\right)^j}{j!} e^{-\frac{1}{2}\lambda}\right] p_{\chi^2_{v+2j}}(x) = \frac{\exp\left\{-\frac{1}{2}(x+\lambda)\right\}}{2^{\frac{1}{2}v}} \sum_{j=0}^{\infty} \frac{(x)^{\frac{1}{2}v+j-1}\lambda^j}{2^{2j}\Gamma\left(\frac{1}{2}v+j\right) j!} \quad (x>0) \tag{2}$$

3. 비중심 χ^2분포 함수의 확률 계산

비중심 χ^2분포가 포아송분포와 중심 χ^2분포 형태를 통하여 제시되는 특성을 이용하였다[6]. 즉, Shea가 제시한 중심 χ^2분포 함수로부터 구한 확률을 가지고 비중심 χ^2분포에 순환적 알고리즘을 적용하는 접근 방법을 시도하였다.

Shea가 제시한 중심 χ^2분포는 다음과 같다[1,12].

$$F_v(x)=f_{\frac{v}{2}}\left(\frac{x}{2}\right)\sum_{s=0}^{\infty} C_s\left(\frac{x}{2},\frac{v}{2}\right) \tag{3}$$

여기서 $f_{\frac{v}{2}}\left(\frac{x}{2}\right)=\frac{e^{-\frac{x}{2}}\left(\frac{x}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}+1\right)}$ 이며,

$$C_s\left(\frac{x}{2},\frac{v}{2}\right)=\frac{\frac{x}{2}}{\left(\frac{v}{2}+s\right)} C_{s-1}\left(\frac{x}{2},\frac{v}{2}\right),$$

$s=1,2,3,\cdots$

$C_0\left(\frac{x}{2},\frac{v}{2}\right)=1$ 이다.

식 (2)로부터 비중심 χ^2분포는 포아송분포 $P(j)$와 중심 χ^2분포 $W(j)$의 곱으로 나타낼 수 있다.

$$\Pr(\chi_v'^2(\lambda)\le x)=\sum_{j=0}^{\infty} P(j)W(j) \tag{4}$$

여기서 $P(j)=\frac{\left(\frac{\lambda}{2}\right)^j e^{-\frac{\lambda}{2}}}{j!}$, $W(j)=\Pr[\chi^2_{v+2j}\le x]$이다.

$k>2$ 에 대하여

$$\Pr(\chi_k^2\le x)=-\frac{x^{\frac{k-2}{2}}e^{-\frac{x}{2}}}{\Gamma\left(\frac{k}{2}\right)2^{\frac{k-2}{2}}} + \Pr(\chi^2_{k-2}\le x)$$

으로 부터 $W(j)=-S(j)+W(j-1)$가 성립한다.

여기서 $S(j)=\frac{x^{\frac{v+2j-2}{2}}e^{-\frac{x}{2}}}{\Gamma\left(\frac{v+2j}{2}\right)2^{\frac{v+2j-2}{2}}}$, $j=0,1,2,\cdots$

이다.

또한 $S(j)=\frac{x}{v+2j-2}S(j-1)$, $j=1,2,\cdots$

와 $S(0)=\frac{x^{\frac{v-2}{2}}e^{-\frac{x}{2}}}{\Gamma\left(\frac{v}{2}\right)2^{\frac{v-2}{2}}}$

그리고 $W(0)=\Pr[\chi_v^2\le x]$로 부터 $W(j)$를 계산한다.

$$P(0)=e^{-\frac{\lambda}{2}} \text{ 와 } P(j)=\left(\frac{\lambda}{2j}\right)P(j-1),\ j=1,2,\cdots$$

로 부터

$$\Pr(\chi_v'^2(\lambda)\le x)=\sum_{j=0}^{\infty} P(j)W(j)$$

을 구할 수 있다.

알고리즘 1 비중심 χ^2분포의 확률 계산

단계 1. N : 식 (4)에서의 항의 수
v : 자유도
λ : 비중심모수
x : 백 분위수

단계 2. $P(0)=e^{-\frac{\lambda}{2}}$, $W(0)=\Pr[\chi_v^2\le x]$를 계산한다.

단계 3. $\ln S(0)=\left(\frac{v-2}{2}\right)\ln x-\frac{x}{2} -\left(\frac{v-2}{2}\right)\ln 2-\ln\Gamma\left(\frac{v}{2}\right)$

$S(0) = \exp[\ln S(0)]$

$SUM = P(0)\, W(0)$

단계 4. $j = 1, 2, \cdots, N$ 에 대하여

$$P(j) = \left(\frac{\lambda}{2j}\right) P(j-1)$$

$$S(j) = \left[\frac{x}{v+2j-2}\right] S(j-1)$$

$$W(j) = -S(j) + W(j-1)$$

$$SUM = SUM + P(j)\, W(j)$$

단계 5. $\Pr(\chi_v'^2(\lambda) \le x) = SUM$

단계 2에서의 $W(0) = \Pr[\chi_v^2 \le x]$의 계산은 Shea가 제시한 중심 χ^2분포를 적용한다.

알고리즘 2 중심 χ^2분포의 확률 계산

단계 1. n : 식 (3)에서의 항의 수
v : 자유도
x : 백 분위수

단계 2. $$\ln f\left(\frac{x}{2}\right) = -\frac{x}{2} + \left(\frac{v}{2}\right)\ln\left(\frac{x}{2}\right) - \ln\Gamma\left(\frac{v}{2}+1\right)$$

$$f\left(\frac{x}{2}\right) = \exp\left[\ln f\left(\frac{x}{2}\right)\right]$$

단계 3. $C(0) = 1$

단계 4. $s = 1, 2, \cdots, n$ 에 대하여

$$C(s) = \frac{\frac{x}{2}}{\left(\frac{v}{2}+s\right)} C(s-1)$$

단계 5. $SUM(1) = f\left(\frac{x}{2}\right) C(s)$

단계 6. $W(0) = \Pr[\chi_v^2 \le x] = SUM(1)$

앞표는 일부 자유도와 비중심모수에 대하여 Fisher와 Patnaik의 확률 값과 비교한 결과이다.

〈표 1〉 새로운 알고리즘과 기존의 접근 방법과의 확률 값 비교

자유도	비중심 모수	백 분위수	Fisher	Patnaik	new Algorithm
1	1	7.001	0.949979	0.949979	0.949979
1	1	11.201	0.990525	0.990525	0.990525
1	16	41.031	0.991880	0.991888	0.991880
1	25	43.771	0.946949	0.938768	0.946930
5	4	19.201	0.956998	0.957001	0.956995
10	1	19.641	0.943193	0.943193	0.943190
15	16	58.561	0.992082	0.992032	0.992054
25	1	38.601	0.943842	0.943841	0.943845
25	4	46.001	0.969655	0.969653	0.969701
25	25	86.001	0.994585	0.978828	0.994580

4. 결 론

본 논문은 중심 χ^2분포 함수를 통하여 비중심 χ^2분포 함수의 계산을 구하는 순환적 알고리즘 접근 방법을 제시하였으며, 기존 Fisher와 Patnaik의 접근 방법에 의한 계산값과 비교한 결과 확률 값이 거의 차이가 없음을 알 수 있다.

참 고 문 헌

[1] A. I. Abdel-Samad and S. K. Ashour, "On the computation of non-central chi-square distribution function," Communications in Statistics, Simulation, Vol.19(4), pp.1279-1291, 1990.

[2] P. J. Bickel and K. A. Doksum, "Mathematical Statistics : Basic ideas and selected topics," San Francisco : Holden-Day, 1977.

[3] T. J. Boardman and R. W. Kopitzke, "Probability and table values for statistical distributions," in Proceedings of the Statistical Computing Section, American Statistical Association, pp.81-86, 1975.

[4] R. A. Fisher, "The general sampling distribution of the multiple correlation coefficient," Proceedings of the Royal Society of London, Vol.121A, pp.654-673, 1928.

[5] W. C. Guenther, "Another derivation of the non-central chi-square distribution," Journal of the American Statistical Association, Vol.59, pp.957-960, 1964.

[6] N. L. Johnson, "On an extension of the connexion between Poisson and χ^2 distributions," Biometrika, Vol.46, pp.352-363, 1959.

[7] N. L. Johnson and S. Kotz, "Continuous Univariate Distributions," Boston : Houghton Mifflin, 1970.

[8] K. S. Miller and R. I. Bernstein and L. E. Blumenson, "Generalized Rayleigh Process," Quarterly of Applied Mathematics, Vol.16, pp.137-145, 1958.

[9] J. H. Park, "Moments of the generalized Rayleigh distribution," Quarterly of Applied Mathematics, Vol.19, pp. 45-49, 1961.

[10] P. B. Patnaik, "The non-central χ^2 and F-distribution and their applications," Biometrika, Vol.36, pp.303-332, 1949.

[11] H. O. Posten, "Computer algorithms for the classical distri-

bution functions," in Proceedings for the First International Conference on the Teaching of Statistics, Sheffield, U, K. : University of Sheffield Printing Unit, pp.313-330, 1982.

[12] B. L. Shea, "Chi-squared and incomplete Gamma integral," Applied Statistics, Statistical Algorithms, Vol.37, pp.466-473, 1988.

[13] H. R. Vaart, "A note on the derivation of the non-central chi-square density function," Statistica Neerlandica, Vol. 21, pp.99-100, 1967.

구 선 희

e-mail : gusonhee@hanmail.net

1986년 숙명여자대학교 수학과 졸업 (이학사)

1989년 숙명여자대학교 대학원 수학과 졸업(이학석사)

1996년 성균관대학교 대학원 통계학과 졸업(통계학박사)

1996년~현재 전주대학교 교양학부 컴퓨터 강의전담전임강사

관심분야 : 전산수학, 신경망이론