

비중심 카이제곱분포의 동질성검정¹⁾

황형태²⁾, 오희정³⁾

요약

공통의 자유도를 갖는 k 개의 비중심 카이제곱분포들의 동질성을 검정하기 위하여 우선 적당한 형태의 검정방법을 제시하였다. 통상적인 방법대로, 제시된 검정방법이 주어진 유의수준을 만족시키도록 하기 위해서는, 귀무가설하에서 제 1종의 오류의 확률을 최대화하는 모수의 최소 우호적 위치(Least favorable configuration)가 유도되었으며, 이에 따라서 주어진 유의수준을 충족하는 기각치를 도표화하였다.

1. 서 론

분산이 동일한 여러개의 정규모집단의 동질성이란 모평균들의 동질성으로 표현되며, 이에 대한 검정은 통상적인 일원분석법에서의 F -검정에 의해서 수행될 수 있을 것이다. 이 논문에서는 이와같이 모집단들의 동질성에 대한 검정을 공통의 자유도를 갖는 k 개의 비중심 카이제곱분포들에 대하여 수행하려고 한다. 자유도가 공통인 경우이므로 비중심 카이제곱분포들의 동질성이란 각 분포의 비중심도(noncentrality)들의 동질성으로 이해될 수 있다. 다수의 비중심 카이제곱분포에서의 비중심도들에 대한 비교문제는 convex mixture의 특별한 경우로 Gupta and Panchapakesan(1972)에 의해서 부분선택적 접근방법으로 다루어진 바 있다.

자유도가 r , 비중심도가 λ 인 비중심 카이제곱분포를 $\chi^2(r, \lambda)$ 로 표기하면 잘 알려진 바와 같이 $\chi^2(r, \lambda)$ 의 p.d.f.와 c.d.f.는 다음과 같이 주어진다.

$$\begin{aligned} f(x; r, \lambda) &= \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} g(x; r+2j) \\ F(x; r, \lambda) &= \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} G(x; r+2j) \end{aligned} \tag{1.1}$$

여기서, 함수 $g(\cdot; r+2j)$ 과 $G(\cdot; r+2j)$ 은 각각 자유도 $r+2j$ 인 카이제곱분포의 p.d.f.와 c.d.f.를 뜻한다.

1) 이 논문은 1998년도 단국대학교 대학연구비에 의해 연구되었음.

2) (140-714) 서울 용산구 한남동 산 8 단국대학교 전산통계학과 교수

3) (140-714) 서울 용산구 한남동 산 8 단국대학교 전산통계학과 박사과정

이제 $i=1, 2, \dots, k$ 에 대하여 확률변수 X_i 가 서로 독립적으로 $\chi^2(r, \lambda_i)$ 를 따라서 분포한다고 하자. 여기서, r 은 주어진 공통의 자유도이고 λ_i 들은 미지의 비중심도들이다. 여기서 우리가 검정하려고 하는 가설은 다음과 같이 표현된다.

$$H_0: \lambda_1 = \lambda_2 = \dots = \lambda_k \quad \text{대} \quad H_1: H_0 \text{가 아님} \quad (1.2)$$

이 가설을 검정하기 위한 가능도비검정(Likelihood ratio test)을 구하려고 시도하였지만, 분포의 수식적 표현이 복잡하여 이용 가능한 형태의 가능도비검정이 유도되지 못하였다. 따라서, 여기에서는 (1.2)에서 주어진 가설에 대한 특별한 형태의 검정방법을 제시하고, 제시된 형태의 검정방법이 주어진 유의수준 α 를 만족하도록 기각치(critical value)를 유도하려고 한다.

2. 검정방법의 제시

$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)}$ 를 X_1, X_2, \dots, X_k 의 순서통계량이라고 하자. 그러므로 $X_{(1)}$ 과 $X_{(k)}$ 는 각각 주어진 관측값들의 최소값과 최대값을 의미한다. 비중심 카이제곱분포는 자유도가 일정할 때 비중심도에 대하여 확률적으로 증가(stochastically increasing)하므로, $X_{(1)}$ 과 $X_{(k)}$ 를 비교함으로써 가설 (1.2)에 대한 검정을 수행하려고 한다.

즉, 모든 양의 실수 x 에 대하여 $h(x) > x$ 인 어떤 함수 h 가 있다고 하자. 이 때 가설 (1.2)에 대한 검정방법은 다음과 같이 제시된다.

$$\text{“}X_{(k)} \geq h(X_{(1)}) \text{이면 } H_0 \text{를 기각한다”} \quad (2.1)$$

예를 들어, 함수 h 를 $c > 1$ 인 상수 c 에 대하여 $h(x) = cx$ 로 하면 기각역의 형태는 “ $X_{(k)} / X_{(1)} \geq c$ ”가 될 것이고, 함수 h 를 $d > 0$ 인 상수 d 에 대하여 $h(x) = x + d$ 로 하면 기각역의 형태는 “ $X_{(k)} - X_{(1)} \geq d$ ”와 같이 될 것이다. 이제 우리는 검정방법 (2.1)이 주어진 유의수준 α 를 만족하도록하는 적당한 형태의 함수 h 를 모색하려고 한다.

3. 최소 우호적 위치

제 1종의 오류의 확률을 고려하기 위하여 $\lambda_1 = \lambda_2 = \dots = \lambda_k = \lambda$ 라고 하고, 이때 제 1종의 오류의 확률은 $P_\lambda(E)$ 로 표기하자. 그러면 각 X_i 가 $X_{(1)}$ 이 되는 경우들로 분할하여 여사건의 확률을 이용하면 $P_\lambda(E)$ 는 다음과 같이 됨을 알 수 있다.

$$P_\lambda(E) = 1 - k \int_0^\infty (F(h(x); r, \lambda) - F(x; r, \lambda))^{k-1} f(x; r, \lambda) dx \quad (3.1)$$

여기서, $P_\lambda(E)$ 를 최대화하는 λ 의 최소 우호적 위치(Least favorable configuration)를 구하기 위해서 다음의 Lemma 1과 Lemma 2가 필요하다. Lemma 2는 Chebyshev에 의하여 최초로 발견된 부등식의 한 형태로서 그 증명은 Hardy, Littlewood and Polya(1952)의 책을 참고할 수 있으므로 생략한다.

Lemma 1. 식 (1.1)에 주어진 $\chi^2(r, \lambda)$ 의 p.d.f. f 와 c.d.f. F 에 대하여 다음의 관계가 성립한다.

$$\begin{aligned} \frac{\partial}{\partial \lambda} F(x; r, \lambda) &= -2f(x; r+2, \lambda) \\ \text{증명. } \frac{\partial}{\partial \lambda} F(x; r, \lambda) &= \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} G(x; r+2(j+1)) - \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} G(x; r+2j) \\ &= \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} (G(x; r+2(j+1)) - G(x; r+2j)). \end{aligned} \quad (3.2)$$

그런데 부분적분을 적용하면,

$$\begin{aligned} G(x; r+2(j+1)) &= \int_0^x \frac{t^{\frac{x}{2}+j}}{\Gamma(\frac{r}{2}+j+1) 2^{\frac{x}{2}+j+1}} e^{-\frac{t}{2}} dt \\ &= -2g(x; r+2(j+1)) + G(x; r+2j) \end{aligned}$$

가 성립하게 된다. 따라서,

$$\begin{aligned} \frac{\partial}{\partial \lambda} F(x; r, \lambda) &= -2 \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} g(x; r+2(j+1)) \\ &= -2 \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} g(x; r+2+2j) \\ &= -2f(x; r+2, \lambda). \end{aligned}$$

□

Lemma 2. 양의 수열 $\langle a_j \rangle$, $\langle b_j \rangle$, $\langle d_j \rangle$ 에 대하여 수열 $\langle d_j \rangle$ 와 $\langle \frac{b_j}{a_j} \rangle$ 가 증가수열이고 다음의 모든 급수가 수렴하면 다음이 성립한다.

$$\frac{\sum b_j d_j}{\sum a_j d_j} \geq \frac{\sum b_j}{\sum a_j}.$$

이제 식 (3.1)의 $P_\lambda(E)$ 를 λ 에 대하여 미분하면,

$$\begin{aligned} & \frac{\partial}{\partial \lambda} P_\lambda(E) \\ &= -k(k-1) \int_0^\infty (F(h(x); r, \lambda) - F(x; r, \lambda))^{k-2} \left(\frac{\partial}{\partial \lambda} F(h(x); r, \lambda) - \frac{\partial}{\partial \lambda} F(x; r, \lambda) \right) f(x; r, \lambda) dx \\ &\quad - k \int_0^\infty (F(h(x); r, \lambda) - F(x; r, \lambda))^{k-1} \frac{\partial}{\partial \lambda} f(x; r, \lambda) dx. \end{aligned} \quad (3.3)$$

식 (3.3)의 우변의 2항에 부분적분을 적용해 보면 다음을 얻을 수 있다.

$$\begin{aligned} & \int_0^\infty (F(h(x); r, \lambda) - F(x; r, \lambda))^{k-1} \frac{\partial}{\partial \lambda} f(x; r, \lambda) dx \\ &= -(k-1) \int_0^\infty ((F(h(x); r, \lambda) - F(x; r, \lambda))^{k-2} (h'(x)f(h(x); r, \lambda) - f(x; r, \lambda))) \frac{\partial}{\partial \lambda} F(x; r, \lambda) dx \end{aligned} \quad (3.4)$$

따라서 식 (3.4)를 식 (3.3)에 적용하여 정리해보면,

$$\begin{aligned} \frac{\partial}{\partial \lambda} P_\lambda(E) &= -k(k-1) \int_0^\infty (F(h(x); r, \lambda) - F(x; r, \lambda))^{k-2} \\ &\quad \cdot (f(x; r, \lambda) \frac{\partial}{\partial \lambda} F(h(x); r, \lambda) - h'(x)f(h(x); r, \lambda) \frac{\partial}{\partial \lambda} F(x; r, \lambda)) dx \\ &= 2k(k-1) \int_0^\infty (F(h(x); r, \lambda) - F(x; r, \lambda))^{k-2} \\ &\quad \cdot (f(x; r, \lambda)f(h(x); r+2, \lambda) - h'(x)f(h(x); r, \lambda)f(x; r+2, \lambda)) dx. \end{aligned} \quad (3.5)$$

위의 식 (3.5)의 두 번째 등식은 Lemma 1의 결과를 이용하여 얻어진 것이다. 이 단계에서 $h(x) > x$ 인 구체적인 함수 h 를 선택해야 하는데, 1보다 큰 상수 c 에 대하여 $h(x) = cx$ 로 택하자. 이 때, 다음의 Lemma 3은 $\frac{\partial}{\partial \lambda} P_\lambda(E) \leq 0$ 임을 보이기 위해서 필요하다.

Lemma 3. $c > 1$ 일 때,

$$f(x; r, \lambda)f(cx; r+2, \lambda) - cf(cx; r, \lambda)f(x; r+2, \lambda) \leq 0$$

증명.

$$\begin{aligned}
& \frac{c f(cx; r, \lambda)}{f(cx; r+2, \lambda)} = \frac{\sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} \frac{c(cx)^{\frac{r}{2}+j-1}}{\Gamma(\frac{r}{2}+j) 2^{\frac{r}{2}+j}} e^{-\frac{cx}{2}}}{\sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} \frac{(cx)^{\frac{r}{2}+j}}{\Gamma(\frac{r}{2}+j+1) 2^{\frac{r}{2}+j+1}} e^{-\frac{cx}{2}}} \\
&= \frac{\sum_{j=0}^{\infty} b_j d_j}{\sum_{j=0}^{\infty} a_j d_j} \quad \text{라고 하자. 여기서, } j = 0, 1, 2, \dots \text{에 대하여} \\
&\quad a_j = \frac{e^{-\lambda} \lambda^j}{j!} \frac{x^{\frac{r}{2}+j}}{\Gamma(\frac{r}{2}+j+1) 2^{\frac{r}{2}+j+1}} e^{-\frac{cx}{2}} = \frac{e^{-\lambda} \lambda^j}{j!} g(x; r+2+2j) \\
&\quad b_j = \frac{e^{-\lambda} \lambda^j}{j!} \frac{x^{\frac{r}{2}+j-1}}{\Gamma(\frac{r}{2}+j) 2^{\frac{r}{2}+j}} e^{-\frac{cx}{2}} = \frac{e^{-\lambda} \lambda^j}{j!} g(x; r+2j) \\
&\quad d_j = c^j
\end{aligned}$$

를 뜻한다. 그런데, $\frac{b_j}{a_j} = \frac{r+2j}{x}$ 은 j 에 대하여 증가하고, $c > 1$ 일 때 d_j 도 j 에 대하여 증가하므로 Lemma 2에 의하여 다음과 같이 그 결과를 얻을 수 있다.

$$\frac{c f(cx; r, \lambda)}{f(cx; r+2, \lambda)} = \frac{\sum_{j=0}^{\infty} b_j d_j}{\sum_{j=0}^{\infty} a_j d_j} \geq \frac{\sum_{j=0}^{\infty} b_j}{\sum_{j=0}^{\infty} a_j} = \frac{f(x; r, \lambda)}{f(x; r+2, \lambda)}. \quad \square$$

Lemma 3을 식 (3.5)에 적용하면 $\frac{\partial}{\partial \lambda} P_\lambda(E) \leq 0$ 이므로 제 1종의 오류의 확률 $P_\lambda(E)$ 는 λ 의 감소함수이다. 따라서, $h(x) = cx$ ($c > 1$)로 할 때 제 1종의 오류를 최대화하는 모수의 최소 우호적 위치는 $\lambda_1 = \lambda_2 = \dots = \lambda_k = \lambda = 0$ 일 때임을 알 수 있는 것이다.

4. 맷 음

정리하면, 가설 (1.2)를 검정하기 위하여 선택된 기각역은 “ $X_{(k)} \geq cX_{(1)}$ ”이고($c > 1$), $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$ 일 때 제 1종 오류의 확률이 최대화됨을 보였다. 따라서, 주어진 유의수준

α 에 대하여 상수 c 의 선택은 다음의 방정식에 의하여 결정된다.

$$\int_0^\infty (G(cx; r) - G(x; r))^{k-1} g(x; r) dx = \frac{1}{k}(1 - \alpha). \quad (4.1)$$

여기서, 함수 $G(\cdot; r)$ 과 $g(\cdot; r)$ 은 앞에서와 같이 각각 자유도 r 인 카이제곱분포의 c.d.f.와 p.d.f.를 뜻한다.

표 4.1은 식 (4.1)에 의하여 계산된 상수 c 의 값을 상수 k, r, α 의 값들에 따라 도표화한 결과이다.

참고로, 함수 h 를 $h(x) = x + d$ ($d > 0$)으로 선택하는 경우를 검토하여 보았으나, 이 경우에는 오류의 최대확률이 d 의 값에 관계없이 항상 1이 됨을 보일 수 있게 되어 유의수준을 만족시킬 수 없으므로 고려할 수 없음을 밝힌다.

표 4.1 기각치 c 값

$\alpha = 0.05$ 일 때

$r \backslash k$	2	3	4	6	8	10
1	647.79	2869.56	6993.37	22160.54	46038.00	82916.97
2	39.00	88.91	143.51	264.34	405.84	547.38
3	15.44	27.90	39.78	61.55	83.23	104.13
4	9.60	15.39	20.45	29.52	37.70	44.93
6	5.82	8.40	10.26	13.62	16.22	18.47
8	4.43	5.98	7.20	8.96	10.44	11.59
10	3.72	4.81	5.65	6.90	7.80	8.62
15	2.86	3.52	3.98	4.66	5.17	5.53
20	2.46	2.94	3.26	3.73	4.06	4.34
30	2.07	2.39	2.61	2.89	3.11	3.26

$\alpha=0.10$ 일 때 (표4.1 계속)

$r \backslash k$	2	3	4	6	8	10
1	161.45	698.11	1670.72	5242.87	11105.91	20014.62
2	19.00	42.80	69.54	128.55	197.18	263.32
3	9.28	16.79	24.02	37.57	51.28	63.74
4	6.39	10.32	13.91	20.10	25.61	30.80
6	4.28	6.26	7.74	10.30	12.35	14.11
8	3.44	4.70	5.67	7.16	8.36	9.33
10	2.98	3.91	4.61	5.67	6.46	7.17
15	2.40	2.99	3.40	4.00	4.47	4.81
20	2.12	2.56	2.86	3.29	3.59	3.85
30	1.84	2.13	2.34	2.61	2.81	2.96

참고문헌

- [1] Gupta, S. S., and Panchapakesan, S.(1972). *On a class of subset selection procedures*, Annals of Mathematical Statistics, 43, 814-822.
- [2] Hardy, G. H., Littlewood, J. E., and Polya, G.(1952). *Inequalities*, 2nd edition, Cambridge University Press, Cambridge.