共分散行列이 單位行列과 같을 때의 假說檢定

李 鍾 厚

Testng the hypothesis a covariance matrix is equal to an identity matrix.

By
Jonghoo Lee

<目 次>

Abstract

Suppose $x_1 \cdots\cdots x_N$ are a set of N observations, x_α being drawn from a p variate normal distribution $N(\mu, \Lambda)$ and let the matrix of sums of squares and cross products of deviations about the mean be $V=\sum_{\alpha=1}^{n}(x_\alpha-\bar{x})'(x_\alpha-\bar{x})$.

Under the null hypothesis $H: \Lambda=I_p$, the likelihood ratio criterion λ is given by

$$\lambda=\frac{\max\limits_{\mu} L(I.\ \Lambda)}{\max\limits_{\mu.\ \Lambda} L(\mu.\ \Lambda)}=\left(\frac{e}{N}\right)^{\frac{1}{2}pN}|V|^{\frac{1}{2}N}e^{-\frac{1}{2}trV}$$

The h th moment of λ does not satisfy Box' conditions in the general theory of asymptotic expansions (see Ref [1]. P. 203).

In spite of the above fact, it can be shown that the asymptotic expansion of $-2\log\lambda$ is represented by the same form as Box and further more $-2\log\lambda$ will be distributed approximately as F distribution, comparing its cumulants to those of F-distribution.

1. 緖 論

母集團分布가 P變量正規分布 $N(\mu.\ \Lambda)$를 가질 때 크기 N의 任意標本을 $x_1 \cdots,\ x_N$라 하고 偏

差積和行列을 $V=\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})$라 두면 標本 $\boldsymbol{x}_1\cdots,\ \boldsymbol{x}_N$의 尤度函數는

$$L(\mu, \Lambda)=(2\pi)^{-\frac{p}{2}N}\ |\Lambda|^{-\frac{N}{2}}\exp\left\{-\frac{1}{2}\sum_{\alpha}^{N}(\boldsymbol{x}_\alpha-\mu)\Lambda^{-1}(\boldsymbol{x}_\alpha-\mu)'\right\} \tag{1}$$

임을 쉽게 알 수 있다.[1)]

共分散行列에 對한 歸無假說 $H:\Lambda=I_p$(單位行列)에 對한 LRC를 λ라 하고 λ에 關한 漸近分布를 考察하자.

LRC, λ의 값은 定義에 依하여

$$\lambda=\frac{\max_{\mu} L(\mu, I)}{\max_{\mu,\Lambda} \mathrm{L}(\mu, \Lambda)}=\frac{(2\pi)^{-\frac{1}{2}pN}\ e^{-\frac{1}{2}\Sigma(X\alpha-\bar{X})'(X\alpha-\bar{X})}}{(2\pi)^{-\frac{1}{2}pN}\left|\frac{1}{N}V\right|^{-\frac{1}{2}N}e^{-\frac{1}{2}pV}}$$

$$=\left(\frac{e}{N}\right)^{\frac{1}{2}pN}|V|^{\frac{1}{2}N}e^{-\frac{1}{2}trV} \tag{2}$$

이다.① λ의 h次 moment는 V가 Wishart分布 $W(\Lambda.\,p.\,n)$ $(n=N-1)$을 함으로

$$\mathcal{E}(\lambda^h)=\int\cdots\int\left(\frac{e^{\frac{1}{2}pN}}{N^{\frac{1}{2}pN}}|V|^{\frac{1}{2}N}\ e^{-\frac{1}{2}trV}\right)^h W(V|\Lambda.\,p.\,n)dV$$

$$=\frac{e^{\frac{1}{2}phN}}{N^{\frac{1}{2}phN}}\int\cdots\int|V|^{\frac{1}{2}Nh}e^{-\frac{1}{2}htrV}\ W(V|\Lambda.\,p.\,n)dV$$

그리고

$$|V|^{\frac{1}{2}Nh}\ e^{-\frac{1}{2}htrV}\ W(V|\Lambda.\,p.\,n)$$

$$=\frac{|V|^{\frac{1}{2}(n+Nh-p-1)}\ e^{\frac{1}{2}(tr\Lambda^{-1}V+trhV)}}{2^{\frac{1}{2}pn}\ \pi^{\frac{1}{4}p(p-1)}|\Lambda|^{\frac{1}{2}n}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}$$

$$=\frac{2^{\frac{1}{2}pNh}\ \Pi\ \Gamma[\frac{1}{2}(n+Nh+1-i)]}{|\Lambda^{-1}+hI|^{\frac{1}{2}(n+Nh)}|\Lambda|^{\frac{1}{2}n}\ \Pi\ \Gamma[\frac{1}{2}(n+1-i)]}$$

$$\frac{|\Lambda^{-1}+hI|^{\frac{1}{2}(n+Nh)}|V|^{\frac{1}{2}(n+Nh-p-1)}\ e^{-\frac{1}{2}tr(\Lambda^{-1}+hI)V}}{2^{\frac{1}{2}p(n+Nh)}\ \prod^{p}\Gamma[\frac{1}{2}(n+Nh+1-i)]\pi^{\frac{1}{4}p(p-1)}}$$

$$=\frac{2^{\frac{1}{2}pNh}|\Lambda|^{\frac{1}{2}Nh}\prod^{p}\ \Gamma[\frac{1}{2}(n+Nh+1-i)]}{|I+h\Lambda|^{\frac{1}{2}(n+Nh)}\prod^{p}\ \Gamma[\frac{1}{2}(n+1-i)]}\ W(V|(\Lambda^{-1}+hI)^{-1},n+Nh)$$

이므로 λ의 h次 moment는

$$\mathcal{E}(\lambda^h)=\left(\frac{2e}{N}\right)^{\frac{1}{2}phN}\frac{|\Lambda|^{\frac{1}{2}hN}\prod^{p}\ \Gamma[\frac{1}{2}(n+Nh+1-i)]}{|I+h\Lambda|^{\frac{1}{2}(n+Nh)}\prod^{p}\ \Gamma[\frac{1}{2}(n+1-i)]} \tag{3}$$

이다①. 그러므로 歸無假說 H가 眞이면 λ의 moment는

$$\mathcal{E}(\lambda^h)=\left(\frac{2e}{N}\right)^{\frac{1}{2}phN}\frac{\prod^{p}\ \Gamma[\frac{1}{2}(n+Nh+1-i)]}{(1+h)^{\frac{1}{2}p(n+Nh)}\prod^{p}\ \Gamma[\frac{1}{2}(n+1-i)]} \tag{4}$$

가 되어 Box의 尤度比統計量에 關한 定理를 適用할 수 없다. 卽 Box가 要求하는 moment에 關한 條件을 滿足하지 않는다.[5] 따라서 λ의 分布는 標本의 크기 N이 充分히 클 때 $-2\log\lambda$가 漸近的으로 自由度 $\frac{1}{2}p(p+1)$의 χ^2分布에 收斂한다는 Wilks의 定理만이 適用된다. 그러나 이 Wilks의 定理는 收斂이 너무 緩慢하다. 그러므로 여기서는 $-2\log\lambda$의 收斂이 빠른 漸近展開와 檢定에 關한 有意點의 近似값을 求하는 問題를 考察하였다.

2. 漸近分布

漸近展開를 特性函數와 cumulant 母函數를 利用하여 求한다.

(I) 假說 $H: \Lambda = I_p$((單位行列)가 眞일 때

$$M = -2\log\lambda \tag{5}$$

라 두면 M의 特性函數 $\phi(t)$는

$$\phi(t) = \mathcal{E}(e^{itM}) = \mathcal{E}(e^{-2it\log\lambda}) = \mathcal{E}(\lambda^{-2it})$$

$$= \left(\frac{2e}{N}\right)^{-ipNt}(1-2it)^{-\frac{1}{2}p(n-2itN)} \prod_{j=1}^{p} \frac{\Gamma[\frac{1}{2}(N-j)-iNt]}{\Gamma[\frac{1}{2}(N-j)]} \tag{6}$$

이다. cumulant 母函數는

$$\Phi(t) = \log\phi(t) = g(t) - g(0) \tag{7}$$

$$g(t) = \log\left[\left(\frac{2e}{N}\right)^{-ipNt}(1-2it)^{-\frac{1}{2}p(n-2iNt)} \prod_{j=1}^{p} \frac{\Gamma[\frac{1}{2}(N-j)-iNj]}{\Gamma[\frac{1}{2}(N-j)]}\right]$$

$$= -ipNt\log\left(\frac{2e}{N}\right) - \frac{1}{2}p(n-2iNt)\log(1-2it)$$

$$+\sum_{j=1}^{p}\left\{\log\Gamma\left[\frac{1}{2}(N-2itN)-\frac{j}{2}\right] - \log\Gamma\left[\frac{1}{2}(N-j)\right]\right\} \tag{8}$$

으로 두고 $g(0)$은 $\Phi(0)=0$이 되게 t와 獨立으로 取해지는 實數이다. Gamma函數의 漸近公式

$$\log\Gamma(x+h) = \log\sqrt{2\pi} + (x+h-\frac{1}{2})\log x - x$$

$$-\sum_{r=1}^{m}(-1)^r \frac{B_{r+1}(h)}{r(r+1)x^r} + R_{m+1}(x) \tag{9}$$

을 利用하여 $g(t)$를 整頓한다. 단, $R_{m+1}(x) = O(x^{-(m+1)})$, $B_r(h)$는 r次의 Bernoulli의 多項式이다.

$$g(t) = p\log\sqrt{2\pi} + \frac{1}{4}p(2N-p-3)\log\frac{N}{2} - \sum_{j=1}^{p}\log\Gamma\left[\frac{1}{2}(N-j)\right] - \frac{1}{2}pN$$

$$-\frac{1}{4}p(p+1)\log(1-2it) - \sum_{j=1}^{p}\sum_{r=1}^{m}(-1)^r \frac{B_{r+1}\left(-\frac{i}{2}\right)}{r(r+1)\left[\frac{N}{2}(1-2it)\right]^r} + R \tag{10}$$

$$R = \sum_{r=1}^{m} O(N^{-(m+1)})$$

$$\Phi(t) = Q - g(0) - \frac{f}{2}\log(1-2it) + \sum_{r=1}^{m}\alpha_r(1-2it)^{-r} + R \tag{11}$$

$$但 \quad \alpha_r = \sum_{j=1}^{p} (-1)^{r+1} \frac{2^r B_{r+1}\left(-\frac{j}{2}\right)}{r(r+1)N^r}, \qquad f = \frac{1}{2}p(p+1)$$

$$Q = p \log \sqrt{2\pi} + \frac{1}{4}p(2N-p-3)\log\frac{N}{2} - \sum_{j=1}^{p}\log \Gamma\left[\frac{1}{2}(N-j)\right] - \frac{1}{2}pN \tag{12}$$

이다. 또

$$\Phi(t) = -\frac{f}{2}\log(1-2it) + \sum_{r=1}^{m}\alpha_r[(1-2it)^{-r}-1] + R' \tag{13}$$

으로 두면 特性函數는

$$\phi(t) = e^{\Phi(t)}$$

$$= (1-2it)^{-\frac{1}{2}f}\exp\left[\sum_{r=1}\alpha_r(1-2it)^{-r} - \sum_{r=1}^{m}\alpha_r + R'\right] \tag{1 }$$

으로 表示되고 α_r의 값은 $B_{r+1}\left(-\frac{j}{2}\right)$, $1 \leqq j \leqq p$ 이므로 收斂하며 $N \to \infty$일 때 $\alpha_r \to 0$이 되어 Box의 定理와 같은 形이다. 따라서 다음의 定理를 얻는다.

［定理1］ 母集團이 p變量正規分布$N(\mu, \Lambda)$일 때 크기 N의 任意標本을 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$, $V = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \overline{\boldsymbol{x}})'(\boldsymbol{x}_\alpha - \overline{\boldsymbol{x}})$라 하고 共分散行列에 對한 假說 $H: \Lambda = I$(單位行列)에 對한 LRC를 λ, $M = -2\log\lambda$로 두면 H가 眞일 때 M의 漸近展開는 다음과 같다.

$$\begin{aligned} P(M \leqq M_o) = {} & P\{\chi^2{}_f \leqq M_o\} + \alpha_1\{P(x^2{}_{f+} \leqq M_o\} - P\{\chi^2{}_f \leqq M_o\}) \\ & + [\alpha_2(P\{\chi^2{}_{f+4} \leqq M_o\} - P\{x^2{}_f\{\chi^2{}_f \leqq M_o\}) \\ & + \frac{1}{2}\alpha^2{}_1(P\{\chi^2{}_{f+4} \leqq M_o\} - 2P\{\chi^2{}_{f+2} \leqq M_o\} + P\{\chi^2{}_f \leqq M_o\}] \\ & + \cdots + R'_{m+1}, \qquad R'_{m+1} = O(N^{-(m+1)}) \end{aligned} \tag{15}$$

단 $\chi_f{}^2$은 自由度 f의 χ^2分布의 確率變數이다.

(Ⅱ) Cumulant의 母函數 $\Phi(t) = \log \phi(t) = g(t) - g(0)$을 t에 關해서 展開하면

$$\Phi(t) = \sum_{j=1}^{m}\left[\frac{(it)^j}{j!}2^{j-1}(j-1)!\,f\right]\left\{1 + \sum_{r=1}^{\infty}\binom{j+r-1}{r}\frac{2r}{f}\right\} \tag{16}$$

이다. 그러므로 M의 j次 cumulant는

$$\kappa_j = 2^{j-1}(j-1)!f\left\{1 + jA_1 + \frac{j(j+1)}{2!}A_2 + \frac{j(j+1)(j+2)}{3!}A_3 + \cdots\right\} \tag{17}$$

$$A_r = 2r\alpha_r/f, \qquad f = \frac{1}{2}p(p+1)$$

$$\alpha_r = \sum_{j=1}^{p}(-1)^{r+1}\frac{2^r B_{r+1}\left(-\frac{j}{2}\right)}{r(r+1)\ N^r}, \quad A_1 = -\frac{2p^2+9p+N}{6(p+1)N}, \quad A_2 = \frac{p^2+5p+10}{6N^2}$$

$$A^2{}_1 = \frac{4p^4+36p^3+125p^2+198p+12}{36(p+1)^2N^2}, \quad A_2 - A^2{}_1 = \frac{2p^4+6p^3+p^2-48p-61}{36(p+1)^2N^2}$$

이 된다. 따라서 文獻(4)에서 다음의 結果를 얻는다.

(1) $A_2 - A^2{}_1 > 0$일 때

F를 自由度(f_1, f_2)인 F分布를 갖는 確率變數라 하면 bF(b는 常數)의 4個의 cumulant는 位數 $O(N^{-2})$을 無視할 때 M의 *cumulant*와 一致한다. 그러므로 M/b은 自由度(f_1, f_2)의 F分布에 漸近的으로 收斂한다.[4] 但

$$f_1=f, \quad f_2=\frac{f_1}{A_2-A_1^2}, \quad b=\frac{f_1}{1-A_1-f_1/f_2} \tag{19}$$

(2) $A_2-A^2_1<0$일 때

母數 $\left(\frac{f_1}{2}, \frac{f_2}{2}\right)$의 Beta分布를 갖는 確率變數를 X라 하면 bX(b는 常數)의 cumulant는 位數 $O(N^{-2})$을 無視하고 M의 cumulant와 4次까지 一致한다. 따라서 $\frac{f_2M}{f_1(b-M)}$은 自由度(f_1, f_2)의 F分布에 漸近한다.[4] 但

$$f_1=f, \quad f_2=\frac{f_1}{A^2_1-A_2}, \quad b=\frac{f_2}{1-A_1+2/f_2} \tag{20}$$

以上에서 다음의 定理를 얻는다.

[定理2] 母集團이 p變量正規分布$N(\mu, \Lambda)$일 때 共分散行列에 對한 假說 $H: \Lambda=I_p$(單位行列)에 對한 LRC를 λ, $M=-2\log\lambda$라 하고 $A_r=2r\alpha_r/f$, $\alpha_r=\sum_{j=1}^{p}(-1)^{r+1}\frac{2^r B_{r+1}\left(-\frac{j}{2}\right)}{r(r+1)\ N^r}$, $B_r(x)$는 r次의 Bernoulli 多項式, $f=\frac{1}{2}p(p+1)$이라 두면 假說 H가 眞일 때

(1) $A_2-A_1^2>0$ 이면 M/b는 自由度(f_1, f_2)인 F分布에 漸近的으로 收斂한다. 但

$$f_1=f, \quad f_2=\frac{f_1}{A_2-A^2_1}, \quad b=\frac{f_1}{1-A_1-f_1/f_2}$$

(2) $A_2-A_1^2<0$ 이면 $\frac{f_2M}{f_1(b-M)}$이 自由度(f_1, f_2)인 F分布에 漸近的으로 收斂한다. 但

$$f_1=f, \quad f_2=\frac{f_1}{A^2_1-A_2}, \quad b=\frac{f_2}{1-A_1+2/f_2}$$

3. 結　論

母集團이 p變量正規分布$N(\mu, \Lambda)$를 가질 때 共分散行列에 對한 歸無假說 $H: \Lambda=I$(單位行列)에 對한 LRC, λ의 分布는 假說이 眞일 때 λ의 moment가 Box의 條件을 滿足하지 않는다. 그러나 $M=-2\log\lambda$는 Box의 定理와 같은 形의 漸近展開가 可能함을 알았다. 따라서 Box의 條件은 充分條件이다. 그러므로 이 結果는 多變量分布論의 딴 問題에 對해서도 Box의 定理와 같은 形의 漸近展開가 可能함을 말한다.

參 考 文 獻

1. Anderson An introduction to multivariate statistical analysis. (1958) John Wiley & Sons.
2. G. E. P. Box, A general distribution theory for a class of likelihood criteria. Biometrika. (1949) 36. pp. 317～346.
3. H. Cramér, Mathematical methods of statistics. (1946) Princeton University
4. 李鍾厚, Box의 F分布에 依한 漸近論의 修訂에 對하 結果, 韓國海洋大學 論文集, (1976) 11 pp. 25～37.
5. 北川敏男, 多變量解析論(1966). 共立社.