응용통계연구
1권 1호. 1987

다변량구조모형에서 최대우도추정

김 기 영*

요 약

다변량 정규모집단에서 공분산행렬의 m.l.e.를 유도하기 위하여 m.l.e.의 불변성에 기초를 둔 Anderson의 방법은 그 절차의 단순성이라는 측면에서 일반적으로 선호되고 있다. 본고는 공통요인분석모형이나 Simplex 구조모형과 같은 다변량구조모형에서 이 접근방법을 적용하여 요구되는 m.l.e.를 구하는 과정을 보이고 있다.

1. 개 설

p변량 정규모집단 $N_p(\mu, \Sigma)$로 부터 크기 n의 표본에 의거한 대수우도함수 $\ln L$은 위치모수(location parameter)를 그의 최대우도추정량(m.l.e.)으로 대치시켰을 경우 다음과 같이 표현된다.

$$\ln L \propto (-n/2)\ln|\Sigma| - (n/2)\mathrm{tr}(\Sigma^{-1}S)$$

여기서 $S=E/n$, E는 오차에 관한 Wishart 행렬이고 tr(·)는 정방(square)행렬의 trace를 나타낸다.

이로부터 공분산행렬 Σ의 m.l.e.를 유도하기 위한 여러가지 방법들 [1, 5, 6]중 Anderson [1]의 방법은 $\ln L$을 직접 Σ에 관하여 미분하는 대신 Σ^{-1}에 관하여 미분하며, 이는 근본적으로 m.l.e.의 불변성(invariance) 원리를 이용함으로써 미분처리의 과정을 훨씬 단순화시키고 있다.

이 연구는 Anderson의 접근방법을 확대적용하여 다중공통요인모형(Mutiple Common Factor Model)이나, Wiener Simplex 모형에 포함된 모수들의 m.l.e.를 유도하고 있다. 이와같은 처리절차는 인과모형(Causal Model), Markov Simplex 혹은 Circumplex 모형등 다양한 다변량구조모형의 경우에도 용이하게 적용될 수 있다.

2. 예비정리

다음의 보조정리를 통해 행렬조작 및 그의 미분과 관련하여 본고에 유용한 몇가지 기본

* 고려대학교 정치경제대학 통계학과, 서울 132.

공식을 증명없이 요약·정리한다.

〈보조정리〉

1) y를 행렬, $X=(x_{ij})$의 구성원소들의 어떤 스칼라함수, 그리고 X의 각 (i,j)원소, x_{ij}는 스칼라 z의 함수라 하자. 즉 $y=y(X)$, $x_{ij}=x_{ij}(z)$.

이때 X의 전치행렬(transpose matrix)을 X'으로 표시할 때,

$$\partial y/\partial z=\mathrm{tr}(\partial y/\partial X\cdot\partial X'/\partial z) \tag{2.1}$$

이고, 이는 연쇄규칙(chain rule)으로 알려져 있다.

2) 어떤 행렬의 (i,j)원소만 1이고 나머지 모든 원소가 0인 기초행렬(elementary matrix)을 $(el)_{ij}$라 하면,

$$\partial X/\partial x_{ij}=(el)_{ij} \tag{2.2}$$

3) 한 정방행렬 A의 역행렬 A^{-1}이 존재한다면

$$\partial\ln|A|/\partial A=(A')^{-1} \tag{2.3}$$

4) 상수의 행렬 B에 대하여

$$\partial\mathrm{tr}(BX)/\partial X=B' \tag{2.4}$$

5) 두 행렬 U, V가 행렬 X의 함수, 즉 $U=U(X)$, $V=V(X)$이고 U^{-1}가 존재한다면,

$$\partial U^{-1}/\partial X=-U^{-1}\cdot\partial U/\partial X\cdot U^{-1} \tag{2.5}$$

$$\partial(UV)/\partial x_{ij}=\partial U/\partial x_{ij}\cdot V+U\cdot\partial V/\partial x_{ij} \quad \blacksquare \tag{2.6}$$

이 보조정리를 이용하여 (1.1)의 $\ln L$을 Σ^{-1}에 관하여 미분하면 다음 결과를 얻는다.

$$\partial\ln L/\partial\Sigma^{-1}=(n/2)(\Sigma-S) \tag{2.7}$$

이 관계를 $\Sigma^{-1}=(\sigma^{ij})$의 각 원소별로 처리해 보면

$$\partial\ln L/\partial\sigma^{ij}=\mathrm{tr}\{\partial\ln L/\partial\Sigma^{-1}\cdot\partial\Sigma^{-1}/\partial\sigma^{ij}\}$$

이므로 다음 방정식

$$\mathrm{tr}\{\partial\Sigma^{-1}/\partial\sigma^{ij}(\Sigma-S)\}=0,\quad i,j=1,2,\cdots,p \tag{2.8}$$

을 풀어서 얻은 해가 Σ의 각 원소별 m.l.e.가 될 것이다.

여기에서

$$\mathrm{tr}\{X(el)_{ij}\}=x_{ji} \tag{2.9}$$

이므로 표본공분산행렬 $S=(s_{ij})$, $i,j=1,2,\cdots,p$에 대하여 식 (2.8)은

$$\mathrm{tr}\{(el)_{ij}(\Sigma-S)\}=\sigma_{ji}-s_{ji}=0 \tag{2.10}$$

으로 표현되므로 개별적 σ_{ij}의 m.l.e., $\hat{\sigma}_{ij}$는 곧 s_{ij}가 된다.

3. 共通要因分析模型에서의 M.L.E.

k공통요인모형하에서 귀무가설 H_{0k}는 일반적으로 "k개의 가설적 요인변량(hypothetical factor variate)이 존재하여 이들이 주어져 있다는 가정하에서 원래 반응변수(response variable)의 편(partial)공분산행렬은 대각행렬이 된다"는 것으로 규정된다. 이는 곧 원래 반응변수의 공분산행렬 Σ가 다음과 같이 분해되는 구조임을 나타내고 있다.

$$\Sigma_{p\times p}=D_r+\Lambda\Lambda' \tag{3.1}$$

단 Λ는 $p\times k$ 요인적재행렬(factor loading matrix)이고, 대각행렬 $D_r=\mathrm{diag}(r_1, r_2, \cdots, r_p)$의 i번째 대각원소 r_i, $i=1, 2, \cdots, p$는 i번째 반응변수의 특수분산(specific variance)이다.

이 모형에서 모수인 D_r과 Λ에 관한 최대우도추정량는 약간씩 다른 분포상의 가정하에서 여러 저자 [2,4]에 의하여 연구되었으며, 대부분의 처리절차는 모수에 관하여 우도함수를 직접 미분하는 통상적 방법에 의존하고 있다. 본 연구에서 고려되는 방법은 근본적으로 식 (2.8)의 관계에 그 기저를 두고 있다.

만약 크기 $k\times k$의 행렬$(I+\Lambda'D_r^{-1}\Lambda)^{-1}$이 존재한다면 Σ^{-1}는 다음과 같이 표현될 수 있다.

$$\Sigma^{-1}=D_r^{-1}-\Omega\Omega'$$

$$\text{단 } \Omega_{p\times k}=D_r^{-1}\Lambda\Gamma,\quad (I+\Lambda'D_r^{-1}\Lambda)^{-1}=\Gamma\Gamma' \tag{3.2}$$

여기에서 $\ln L$에 직접 Λ와 D_r을 적용하는 대신 $D_r^{-1}=(r^i)$와 $\Omega=(w_{ij})$에 관한 미분을 고려하면, 다음 관계

$$\partial\ln L/\partial(r^i)\propto\mathrm{tr}\{\Sigma-S)(el)_{ii}\}=\sigma_{ii}-s_{ii} \tag{3.3}$$

$$\partial\ln L/\partial w_{ij}\propto\mathrm{tr}\{(\Sigma-S)(\partial\Sigma^{-1}/\partial w_{ij})\} \tag{3.4}$$

가 얻어지고, 특히 식(3.3)은 H_{0k}하에서도 $\hat{\sigma}_{ii}=s_{ii}$, $i=1, 2, \cdots, p$, 즉 요인분석은 공분산중심의 분석임을 암시하고 있다. 한편 (3.4)의 경우 (2.6)을 이용하면,

$$\partial\Sigma^{-1}/\partial w_{ij}=-\{(el)_{ij})\Omega'+\Omega(el)_{ji}\}$$

가 되므로

$$\partial\ln L/\partial w_{ij}\propto\{(S-\Sigma)\Omega\}_{ij} \tag{3.5}$$

이고, 따라서 다음 결과가 얻어진다.

$$\partial\ln L/\partial\Omega\propto(S-\Sigma)\Omega \tag{3.6}$$

여기서 Σ의 m.l.e. $\hat{\Sigma}$를 $(\hat{D}_r+\Lambda\Lambda')$로 대치시키고 이를 (3.6)과 (3.2)의 관계를 이용하여 적절히 행렬조작하면 m.l.e.를 구하기 위하여 풀어야 할 방정식은 다음과 같게 된다.

$$(S-\Lambda\Lambda')\hat{D}_r^{-1}\Lambda\Lambda'=\Lambda\Lambda' \tag{3.7}$$

윗 식에서 Λ는 계수(rank)가 k인 크기 $p\times k$ 행렬이므로 $\Lambda'\Lambda$의 역행렬이 존재한다. 따라서 (3.7)의 좌우변에 $\Lambda(\Lambda'\Lambda)^{-1}$를 뒤에서 곱하면 (post-multiplication)

$$(S-\Lambda\Lambda')\hat{D}_r^{-1}\Lambda=\Lambda \tag{3.8}$$

, 혹은

$$S\hat{D}_r^{-1}\Lambda=\Lambda(I+\Lambda'\hat{D}_r^{-1}\Lambda) \tag{3.9}$$

가 얻어진다. 이 결과는 Lawley[4], Howe[2]에 의해 얻어진 것과 동일하며 식(3.9)의 표현은 $(I+\Lambda'\hat{D}_r^{-1}\Lambda)$의 크기 k가 일반적으로 반응변수의 수 p보다 훨씬 작으며$(k\ll p)$, 그의 역행렬이 존재한다.

4. Wiener Simplex 모형에서의 최대우도추정량

척도점(scale point) $q_1<q_2<\cdots<q_p$ 을 가진 실확률과정(real stochastic process)에 대하여 Wiener Simplex $X(q)$는 다음과 같은 특성을 가진다.

$$
\begin{aligned}
&E\{X(q_i)\}=\mu_i\\
&\mathrm{Var}\{X(q_i)\}=q_i\\
&\mathrm{Cov}\{X(q_i),X(q_j)\}=q_i,\quad i<j
\end{aligned}
\tag{4.1}
$$

이와같이 Wiener Simplex는 척도종속적인 모형이며 측정단위가 모두 동일할 경우에 사용될 수 있는 적절한 모형으로 알려져 있다. 이 모형에서 Σ는 (4.1)로 부터 p개의 모수만을 포함한 매우 특수한 구조를 가지게 된다.

$$
\Sigma=\begin{bmatrix}
q_1 & q_1 & \cdots & q_1 & q_1\\
q_1 & q_2 & \cdots & q_2 & q_2\\
\cdot & \cdot & \cdots & \cdot & \cdot\\
q_1 & q_2 & \cdots & q_{p-1} & q_{p-1}\\
q_1 & q_2 & \cdots & q_{p-1} & q_p
\end{bmatrix}
\tag{4.2}
$$

이와같은 Σ의 최대우도추정량를 구하는 과정을 단순화하기 위하여, $q'=(q_1,q_2,\cdots,q_p)$, $u'=(u_1,u_2,\cdots,u_p)$라 하고 $u_1=q_1$, $u_j=q_j-q_{j-1}$, $j=2,3,\cdots,p$와 같이 1 : 1 변환을 고려하자. 이 변환은 다음의 하측삼각행렬(lower triangular matrix), T를 이용하여 다음과 같이 표현될 수 있다.

$$u=T^{-1}q,\qquad q=Tu$$

, 단

$$
T=\begin{bmatrix}
1 & 0 & 0 & \cdots & 0\\
1 & 1 & 0 & \cdots & 0\\
1 & 1 & 1 & \cdots & 0\\
\cdot & \cdot & \cdot & \cdots & \cdot\\
1 & 1 & 1 & \cdots & 1
\end{bmatrix},\qquad
T^{-1}=\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 & 0\\
-1 & 1 & 0 & \cdots & 0 & 0\\
0 & -1 & 1 & \cdots & 0 & 0\\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot\\
0 & 0 & 0 & \cdots & -1 & 1
\end{bmatrix}
\tag{4.3}
$$

따라서 $D_u=\mathrm{diag}(u_1,u_2,\cdots,u_p)$라 할 때

$$
\begin{aligned}
\Sigma&=TD_uT'\\
&=\begin{bmatrix}
u_1 & u_1 & \cdots & u_1\\
u_1 & u_1+u_2 & \cdots & u_1+u_2\\
\cdot & \cdot & \cdots & \cdot\\
u_1 & u_1+u_2 & \cdots & u_1+u_2+\cdots+u_p
\end{bmatrix}
\end{aligned}
\tag{4.4}
$$

이고, (4.3)에서의 Jordan형식의 행렬, T^{-1}를 통하여

$$\Sigma^{-1}=(T')^{-1}\,D_u^{-1}\,T^{-1}$$

$$
=\begin{bmatrix}
1/u_1+1/u_2 & -1/u_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\
-1/u_2 & 1/u_2+1/u_3 & -1/u_3 & 0 & \cdots & 0 & 0 & 0 \\
0 & -1/u_3 & 1/u_3+1/u_4 & -1/u_4 & \cdots & 0 & 0 & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & \cdots & -1/u_{p-1} & 1/_{p-1}/u_p & -1/u_p \\
0 & 0 & 0 & 0 & \cdots & 0 & -1/u_p & 1/u_p
\end{bmatrix} \tag{4.5}
$$

즉 이는 삼각대각형(tri-diagonal) 행렬로서 아래와 같이 매우 간단한 구조를 가지고 있다.

$$
\begin{aligned}
\sigma^{ii} &= \begin{cases} 1/u_i+1/u_{i+1}, & i=1,2,\cdots,p-1 \\ 1/u_p, & i=p \end{cases} \\
\sigma^{i,i+k} &= \begin{cases} -1/u_{i+1}=\sigma^{i+k,i}, & k=1 \\ 0, & k>1,\ i=1,2,\cdots,p-1 \end{cases}
\end{aligned} \tag{4.6}
$$

이때

$$
\partial\Sigma^{-1}/\partial(u_i^{-1})=\begin{cases}(el)_{ii}, & ,\ i=1 \\ (el)_{ii}+(el)_{i-1,i-1}-(el)_{i-1,i}-(el)_{i,i-1}, & i=2,3,\cdots,p\end{cases} \tag{4.7}
$$

이므로 다음과 같은 p개의 방정식을 얻게 된다.

$$
\begin{aligned}
&\mathrm{tr}\{(el)_{11}(\hat{\Sigma}-S)\}=0 \\
&\mathrm{tr}[\{(el)_{ii}+(el)_{i-1,i-1}-(el)_{i-1,i}-(el)_{i,i-1}\}\{\hat{\Sigma}-S\}]=0, \\
&\qquad\qquad i=2,3,\cdots,p
\end{aligned} \tag{4.8}
$$

따라서 $\hat{\Sigma}$는 식(4.4)에 아래의 $\hat{u}_i$를 대입하여 쉽게 얻어진다.

$$
\begin{aligned}
&\hat{u}_1=s_{11} \\
&\hat{u}_i=s_{ii}+s_{i-1,i-1}-2s_{i-1,i},\quad i=1,2,3,\cdots,p
\end{aligned} \tag{4.9}
$$

References

(1) Anderson T.W. (1958). *An Introduction to Multivariate Statistical Analysis*, New York: John Wiley.

(2) Howe W.G. (1955). "Some Contribution to Factor Analysis", Oak Ridge National Laboratory, Oak Ridge, TN.

(3) Joreskog K.G. (1970). "A General Method for Analysis of Covariance Structure", *Biometrika 57*, 239~251.

(4) Lawley D.N. (1940). "The Estimation of Factor Loadings by the Method of Maximum Likelihood", *Proceedings of the Royal Society of Edinburgh 60*, 64~82.

(5) Smith D.W. (1978). "A Simplified Approach to the Maximum Likelihood Estimation of the Covariance Matrix", *The American Statistician* 32, 28~29.

(6) Watson G.S. (1964). "A Note on Maximum Likelihood", *Sankhya 26A*, 303~304.

Maximum Likelihood Estimation in Multivariate Structural Model

Kim, Kee Yong

ABSTRACT

For obtaining the m.l.e. of $\sum$ from p-variate Non-singular Normal parent, $N_p(\mu, \sum)$, Andersous' Procedure based on the invariance property of the m.l.e. seems to be generally Preferred in the view of its simplicity. This paper shows that his approach with respect to $\sum^{-1}$ rather than $\sum$ itself, he burther applicable to deriving the m.l.e. of parametersinvolved in the common factor model an dsimplex strucıure model as well.