2005 Proceedings of the Spring Conference, Korean Statistical Society

Generalized Inverse Matrix and SAS IML for Parameter Estimation in General Linear Model (선형 모형에 있어서 모수 추정을 위한 일반(화) 역행렬 및 SAS IML 이론에 관한 연구)

Kuey-Chung Choi[1]

1. 서론

주어진 n차 정방행렬 A에 대하여 $AB=BA=I_n$(단, I_n은 n차 단위행렬)인 B가 존재하면 B를 A의 역행렬이라 하고 $B=A^{-1}$로 표시하며, 이 때 A를 정칙행렬이라 한다.
일반적으로 미지모수 β에 관한 선형모형

$$y=X\beta+\epsilon \quad (1.1)$$

의 최소자승법에 의한 정규방정식 즉
연립 1차 방정식

$$Ab=c \quad (1.2)$$

의 해가 유일하게 존재하기 위한 필요충분조건은 $rank(A)=rank(A,c)$ 이다.
이때 유일한 해는 다음과 같이 써진다.

$$b=A^{-1}c$$

한편 A가 비정칙인 정방행렬이거나 정방행렬이 아닌 경우
(1.2) 의 해는 정할 수 없을 만큼(무수히) 많다(부정).
(1.1) 에서 X는 $n\times p$행렬이다. 최소제곱법에 의한 모수β의 추정량b는
아래 정규방정식

$$(X'X)b=X'y=c \ \ i.e. \ \ Ab=c$$

의 해와 같다. 여기서 $A=(X'X)$는 p차 대칭행렬이고 X'는 X의 전치행렬이다.
이 논문에서는 X를 $n\times p$행렬 $(p\neq n)$, $rank(A)=rank(X'X)=r\leq min\{n,p\}$
라 가정한다.
이 경우에 A의 역행렬 A^{-1}는 존재하지 않지만 다음과 같이 A^{-}를 정의 하므로서 정규방정식 (1.2)의 해를 구할수 있다.

(정의1) 비정칙인 정방행렬 A에 대하여 $AA^{-}A=A$ 인 A^{-}를 A의 일반역행렬이라 한다.
또한 $n\times p$행렬 $(n\neq p)$ X즉 정방행렬이 아닌 X에 대하여 $XX^{-}X=X$인 $p\times n$행렬 X^{-} 도 X의 일반역행렬이다.

$n\times p$행렬X에 대하여 만약 $rank(\mathrm{X})=p<n$(n:방정식의 개수,p:미지인 모수의 갯수)이면

1) Professor, Dept, of Computer Science & Statistics, Chosun University, Gwangju, 501-759, Korea,
E-mail : kjchoi@chosun.ac.kr

$A = X^tX$는 정칙행렬이므로 문제 될것이 없고
$rank(\mathrm{A}) = rank(X^tX) = r \leq n < p$ 인 경우 즉
비정칙인 정방행렬 $A = X^tX$와 $n \times p$행렬 X의 일반역행렬을 구하는 방법을 밝히고자 한다.
(1.2)에서
$AA^-Ab = AA^-c$
$Ab = A(A^-c)$
이므로 최소제곱 추정량 b 즉b에 관한 정규방정식의 해는
$b = A^-c$
이다.

2. 행렬의 분해

$rank(\mathrm{A}) = rank(X^tX) = r \leq n < p$ 인 경우:
p차 대칭행렬 A는 p개의 고유치를 갖는다. 그러나 $rank(\mathrm{A}) = r$이므로 0이 아닌 고유치는 r개 즉$\lambda_1 \cdots \lambda_r$개뿐이고 나머지 고유치 $p-r$개 즉 $\lambda_{r+1}, \cdots, \lambda_p$는 모두 0이다. $V = (v_1, \cdots, v_p)$를 고유(단위)벡타 $v_i (i = 1, \cdots, p)$들에 의한 행렬,
$V_1 = (v_1, \cdots, v_r)$을 0이 아닌 고유치에 대응한 고유(단위)벡터$v_i (i = 1, \cdots, r)$ 로 이루어진 행렬이라 두면
$VV^t = V^tV = I_p$ 즉 V는 p차 정규직교행렬이다. 그러나 $V_1^tV_1 = I_r$이나 $V_1V_1^t \neq I_p$이다.
(정리 2) p차 대칭행렬 A의 서로 다른 고유치에 대한 고유벡타들은 서로 직교 한다.
(증명) λ_i, λ_j를 A의 서로 다른 고유치라 하고 그에 대한 고유벡터를 각각 v_i, v_j라 하면

$$Av_i = \lambda_i v_i \tag{2.1}$$

$$Av_j = \lambda_j v_j \tag{2.2}$$

이다. A가 대칭행렬이므로 (2.1)식에서

$$(Av_i)^t = v_i^tA = \lambda_i v_i^t$$

을 얻고 이식의 양변의 뒤로부터 v_j를 곱하면

$$v_i^tAv_j = \lambda_i v_i^t v_j \tag{2.3}$$

이다. 또한 (2.2)식의 양변의 앞으로부터 v_i^t를 곱하면

$$v_i^tAv_j = \lambda_j v_i^t v_j \tag{2.4}$$

을 얻는다. (2.3),(2.4)식에서

$$\lambda_i v_i^t v_j - \lambda_j v_i^t v_j = (\lambda_i - \lambda_j) v_i^t v_j = 0$$

이다. $\lambda_i \neq \lambda_j$이므로 $v_i^tv_j = 0$이다. 그러므로 두 고유벡터 v_i와 v_j는 서로직교한다.

(정리 3) p차인 대칭행렬 A의 대각화는 가능하다.
(증명) A의 고유치를 $\lambda_1, \cdots, \lambda_p$라 하고 이에 대한 고유단위 벡터들을 $v_1, \cdots, v_p$라면 (정리2)에 의하여 서로 다른 고유치에 대한 고유벡터들은 서로 직교한다. 그러므로 행렬 $V = (v_1, \cdots, v_p)$는 정규직교행렬이다. 고유치와 고유벡터 정의

$Av_i = \lambda_i v_i \ (i=1,\cdots,p)$

에 의하여

$$AV = (Av_1,\cdots,Av_p) = (\lambda_1 v_1,\cdots,\lambda_p v_p) = (v_1,\cdots,v_p)\begin{pmatrix} \lambda_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \lambda_p \end{pmatrix}$$

$$= VD_\lambda \tag{2.5}$$

이다. 여기서 $D_\lambda = \begin{pmatrix} \lambda_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \lambda_p \end{pmatrix}$이다.

V가 정규직교행렬이므로 $V^t = V^{-1}$이며 (2.5)식으로부터 다음과 같이 대각화 할수 있다.

$$D_\lambda = V^{-1}AV = V^tAV$$

주의:정방행렬이지만 대칭이 아닌경우 행렬 A의대각화는 가능하지 않을 수 있다.

기호로 대칭행렬 A의 고유치에 대한 대각행렬을 다음과 같이 표시하기로 한다.

$$D_{1\lambda} = \begin{pmatrix} \lambda_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \lambda_r \end{pmatrix},\ D_\lambda = \begin{pmatrix} \lambda_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \lambda_p \end{pmatrix} = \begin{pmatrix} D_{1\lambda} & 0 \\ 0 & 0 \end{pmatrix}$$

(정리 4) p차 대칭행렬 A는 다음과 같이 분해된다.

1) $A = VD_\lambda V^t$

2) $A = V_1 D_{1\lambda} V_1^t$

(증명) 1) (정리 3)의 (2.5) 식 $AV = VD_\lambda$에서 V는 정규직교행렬 이므로

$$A = VD_\lambda V^{-1} = VD_\lambda V^t$$

이다.

2) 고유치와 고유벡터의 정의에 의하여

$$Av_i = \lambda_i v_i \ (i = 1,\cdots,r)$$

이므로

$$AV_1 = A(v_1 \cdots v_r) = (\lambda_1 v_1,\cdots,\lambda_r v_r)$$

$$= (v_1 \cdots v_r)\begin{pmatrix} \lambda_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \lambda_r \end{pmatrix} = V_1 D_{1\lambda}$$

을 얻으며 이식으로부터 $A = V_1 D_{1\lambda} V_1^t$가 유도된다.

A의 0이 아닌 고유치 $\lambda_1,\cdots,\lambda_r$를 써서 대각행렬을 나타내기로 한다.

$$D_{1\sqrt{\lambda}}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & & \bigcirc \\ & \ddots & \\ \bigcirc & & \frac{1}{\sqrt{\lambda_r}} \end{pmatrix},\ D_{\sqrt{\lambda}}^{-1} = \begin{pmatrix} D_{1\sqrt{\lambda}}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

$n\times p$ 행렬 X로부터

$$U = XVD_{\sqrt{\lambda}}^{-1} \tag{2.6}$$

$$U_1 = XV_1 D_{1\sqrt{\lambda}}^{-1} \tag{2.7}$$

라 두면 U는 $n\times p$ 행렬이고 U_1은 $n\times r$ 행렬이다.

(정리 4)와 $D_{\sqrt{\lambda}}{}^{-1}$의 정의에 의하여 다음을 얻는다

$$
\begin{aligned}
U^tU &= (XVD_{\sqrt{\lambda}}{}^{-1})^t(XVD_{\sqrt{\lambda}}{}^{-1}) \\
&= D_{\sqrt{\lambda}}{}^{-1}V^t(X^tX)VD_{\sqrt{\lambda}}{}^{-1} \\
&= D_{\sqrt{\lambda}}{}^{-1}(V^tAV)D_{\sqrt{\lambda}}{}^{-1} \\
&= D_{\sqrt{\lambda}}{}^{-1}D_{\lambda}D_{\sqrt{\lambda}}{}^{-1} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \neq I_p
\end{aligned}
$$

$$
\begin{aligned}
U_1^tU_1 &= (XV_1D_{1\sqrt{\lambda}}{}^{-1})^t(XV_1D_{1\sqrt{\lambda}}{}^{-1}) \\
&= D_{1\sqrt{\lambda}}{}^{-1}V_1^t(X^tX)V_1D_{1\sqrt{\lambda}}{}^{-1} \\
&= D_{1\sqrt{\lambda}}{}^{-1}D_{1\lambda}D_{1\sqrt{\lambda}}{}^{-1} = I_r
\end{aligned}
$$

(2.6), (2.7) 식으로부터 다음 정리를 얻는다.

(정리 5) (1) $X = UD_{\sqrt{\lambda}}V^t$

(2) $X = U_1D_{1\sqrt{\lambda}}V_1^t$

여기서 $D_{1\sqrt{\lambda}} = \begin{pmatrix} \sqrt{\lambda_1} & & \bigcirc \\ & \ddots & \\ \bigcirc & & \sqrt{\lambda_r} \end{pmatrix}$, $D_{\sqrt{\lambda}} = \begin{pmatrix} D_{1\sqrt{\lambda}} & 0 \\ 0 & 0 \end{pmatrix}$

이고 앞으로 계속 사용한다.

3. 일반 역행렬

$rank$ (A) $= p < n$인 경우는 A가 p차 정칙행렬이므로 역행렬 A^{-1}가 유일하게 존재한다. $n \neq p$인 $n \times p$ 행렬 X와 $rank$(A) $= r \leq n < p$인 대칭행렬 $A = X^tX$는 역행렬 X^{-1}와 A^{-1}는 존재하지 않으므로 일반 역행렬 X^-와 A^-를 구하는 방법을 밝히려 한다.

(정리 6)

1) $X^- = VD_{\sqrt{\lambda}}^{-1}U^t$ 는 X A의 Penrose 역행렬이다.

2) $X^- = V_1D_{1\sqrt{\lambda}}^{-1}U_1^t$ 는 X A의 Penrose 역행렬이다.

3) $X^- = VD_2^{-1}U^t$ 는 X 의 무수히 많은 일반역행렬이다.

단,$D_{1\lambda}^{-1} = \begin{pmatrix} \frac{1}{\lambda_1} & & \bigcirc \\ & \ddots & \\ \bigcirc & & \frac{1}{\lambda_r} \end{pmatrix}$, $D_{\lambda}^{-1} = \begin{pmatrix} D_{1\lambda}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$

$D_1^{-1} = \begin{pmatrix} D_{1\lambda}^{-1} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$, $D_2^{-1} = \begin{pmatrix} D_{1\sqrt{\lambda}}^{-1} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$

각행렬 H_{12} , H_{21} , H_{22} 의 첨자1,2는 행과 열의 차원으로 1=r, 2=p-r 이다.

(증명) 1) (정리 5)와 $D_{\sqrt{\lambda}}^{-1}$ 를 이용하면

$$XX^-X = (UD_{\sqrt{\lambda}}V^t)(VD_{\sqrt{\lambda}}^{-1}U^t)(UD_{\sqrt{\lambda}}V^t)$$

$$= U\left(D_{\sqrt{\lambda}} I_p D_{\sqrt{\lambda}}^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\right) D_{\sqrt{\lambda}} V^t$$
$$= U\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} D_{\sqrt{\lambda}} V^t$$
$$= U D_{\sqrt{\lambda}} V^t = X$$

을 얻는다.

2) (정리 5)의 (2)와 $V_1^t V_1 = I_r$ $U_1^t U_1 = I_r$ 를 이용하면

$$XX^- X = (U_1 D_{1\sqrt{\lambda}} V_1^t)(V_1 D_{1\sqrt{\lambda}}^{-1} U_1^t)(U_1 D_{1\sqrt{\lambda}} V_1^t)$$
$$= U_1 D_{1\sqrt{\lambda}} I_r D_{1\sqrt{\lambda}}^{-1} I_r D_{1\sqrt{\lambda}} V_1^t$$
$$= U_1 D_{1\sqrt{\lambda}} V_1^t = X$$

을 얻는다.

3) (정리 5)의 (1)과 $D_{\sqrt{\lambda}} D_2^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ 을 이용하면

$$XX^- X = (U D_{\sqrt{\lambda}} V^t)(V D_2^{-1} U^t)(U D_{\sqrt{\lambda}} V^t)$$
$$= U\left[D_{\sqrt{\lambda}} I_p D_2^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\right] D_{\sqrt{\lambda}} V^t$$
$$= U\left[\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} D_{\sqrt{\lambda}}\right] V^t$$
$$= U D_{\sqrt{\lambda}} V^t = X$$

을 얻는다.

(정리 7) 1) $A^- = V D_\lambda^{-1} V^t$는 A의 Penrose 역행렬이다.

2) $A^- = V_1 D_{1\lambda}^{-1} V_1^t$는 A의 Penrose 역행렬이다.

3) $A^- = V D_1^{-1} V^t$는 A의 무수히 많은 일반역행렬이다.

(증명) (정리 4)를 이용한다.

1) $AA^- A = (V D_\lambda V^t)(V D_\lambda^{-1} V^t)(V D_\lambda V^t)$
$$= V D_\lambda I_p D_\lambda^{-1} I_p D_\lambda V^t$$
$$= V(D_\lambda D_\lambda^{-1}) D_\lambda V^t$$
$$= V\begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} D_\lambda V^t$$
$$= V D_\lambda V^t = A$$

2) $AA^- A = (V_1 D_{1\lambda} V_1^t)(V_1 D_{1\lambda}^{-1} V_1^t)(V_1 D_{1\lambda} V_1^t)$
$$= V_1 D_{1\lambda} I_r D_{1\lambda}^{-1} I_r D_{1\lambda} V_1^t$$
$$= V_1 (D_{1\lambda} D_{1\lambda}^{-1}) D_{1\lambda} V_1^t$$
$$= V_1 I_r D_{1\lambda} V_1^t$$
$$= V_1 D_{1\lambda} V_1^t = A$$

3) $$\begin{aligned} AA^{-}A &= (V D_{\lambda} V^{t})(V D_{1}^{-1} V^{t})(V D_{\lambda} V^{t}) \\ &= V\ D_{\lambda} I_{p} D_{1}^{-1} I_{p} D_{\lambda} V^{t} \\ &= V\ (D_{\lambda} D_{1}^{-1}) D_{\lambda} V^{t} \\ &= V \left[\begin{pmatrix} I_{r} & H_{12}' \\ 0 & 0 \end{pmatrix} D_{\lambda} \right] V^{t} \\ &= V\ D_{\lambda} V^{t} = A \end{aligned}$$

참고문헌

허명회 (1995), 행렬의 이해와 계산, 자유아카데미, 서울.

박동권 (1995), 실험계획법, 자유아카데미, 서울.

Searle, S. R. (1982), Matrix Algebra Useful for Statistics, Wiley.

Strang, G. (1988), Linear Algebra and Its Applications, Third Edition,Harcourt Brace & Company.

Richarc Bellman (1997), Introduction to Matrix Analysis, 2nd ed. SIAM, Philadelphia, PA.

Biswa Nath Datta (1995), Numerical Algebra and Applications, Brooks/Cole, New York.

G. H. Golub and C. R, Johnson (1983), Matrix Analysis, Cambridge University Press, Baltimore, Maryland.

R. A Horn and C. R. Johnson (1985), Matrix Analysis, Cambridge University Press, Cambridge, UK.