

Estimable Functions of Fixed-Effects Model by Projections

Jaesung Choi^{a,1}

^aDepartment of statistics, Keimyung University

(Received March 20, 2014; Revised June 10, 2014; Accepted June 18, 2014)

Abstract

This paper deals with estimable functions of parameters of less than full rank linear model. In general, the parameters of an overspecified model are not uniquely determined by least squares solutions. It discusses how to formulate linear estimable functions as functions of parameters in the model and shows how to use projection matrices to check out whether a parameter or function of the parameters is estimable. It also presents a method to form a basis set of estimable functions using linearly independent characteristic vectors generating the row space of the model matrix.

Keywords: Estimable functions, overspecified model, characteristic vectors, projection matrices.

1. 서론

고정효과의 선형모형에서 모형내 모수들은 일반적으로 추정가능하지 않다. 이는 자료로부터 추정하고자 하는 모수들의 수보다 더 많은 모수가 모형에 포함되어 있거나 모수들의 관계가 종속적이기 때문이다. 즉, 실험상황에서 모형행렬이 불완전계수의 행렬이거나 초과설정의 모형(over specified model)일 때 최소제곱법에 의한 모수의 해가 많이 존재하게 된다.

모형에 포함되어 있는 모수가 자료로부터 추정될 수 없는 모수라면 그 모수의 추정값은 아무런 정보도 제공하지 않으므로 추정의 의미가 없다. 추정이 가능하지 않은 모수 또는 의미가 없는 모수의 추정은 자료분석에 도움이 되지 않으므로 분석모형으로부터 추정가능한 모수의 구성과 검증은 의미있는 추론을 위해 필요한 과정이다. 따라서 모형내 모수와 관련된 실제 추정가능한 함수의 형태를 파악하는 것이 필요하게 되며 그 모수에 대한 추론을 행하는 것이 의미가 있게 되므로 중요하다. 이와 관련한 추정가능함수에 관한 구체적 논의는 Searle (1971), Graybill (1976) 그리고 Milliken과 Johnson (1984) 등에서 살펴볼 수 있다.

관심있는 모수의 추정가능여부를 확인하기 위하여 가정된 선형모형으로부터 모수들의 어떤 함수가 추정가능함수인가를 판단하기 위한 이론과 방법에 대한 고찰과 함께 추정가능함수를 어떻게 구성해야 하는가에 관한 구체적 논의가 필요하다고 본다. Elswick, Gennings, Chinchilli와 Dawson (1991)은 불완전계수의 일반선형모형에서 가우스 소거법을 이용한 추정가능함수의 일반형을 제안하고 있다. Milliken과 Johnson (1984)은 모형적합방식과 관련된 네 가지 유형의 추정가능함수들을 언급하고 있다.

¹Department of Statistics, Keimyung University, 1095 Dalgubul-Daero, Dalseogu-Gu, Daegu 704-701, Korea.
E-mail: jschoi@kmu.ac.kr

본 연구논문은 이들 문헌에서의 논의와는 달리 고정효과의 선형모형내 모수들의 함수중에서 어떤 함수가 추정가능함수인 가를 판별하는 기법과 추정가능함수의 기본형태를 결정하는 방법을 벡터공간에서의 사영의 관점에서 제안하고자 한다. 또한, 사영이 행해지는 모수추정공간의 고유벡터가 추정가능함수의 구성에 어떻게 활용되는 가를 다루게 된다. 사영에 관한 기본개념과 활용에 관한 연구는 Johnson과 Wichern (1988) 그리고 Choi (2011, 2012) 등에서 살펴볼 수 있다.

2. 모형의 가정

추정가능함수의 논의를 위해 고정효과의 선형모형을 가정한다. 모형의 행렬표현식이 다음과 같다고 가정한다.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (2.1)$$

단, \mathbf{y} 는 $n \times 1$ 인 관측벡터이고 \mathbf{X} 는 원소가 0 또는 1로 구성되는 $n \times p$ 인 모형행렬로 계수가 $k (< p)$ 이다. $\boldsymbol{\beta}$ 는 $p \times 1$ 인 모수벡터이며 $\boldsymbol{\epsilon}$ 은 $n \times 1$ 인 오차벡터이다. $\boldsymbol{\epsilon}$ 에 대한 가정은 $E(\boldsymbol{\epsilon}) = \mathbf{0}$ 이고 $\text{var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$ 이다. $\boldsymbol{\beta}$ 에 대한 추정으로 최소제곱법을 이용한다. 최소제곱법의 적용으로부터 다음의 정규방정식

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \quad (2.2)$$

를 얻는다. 식 (2.2)의 $\mathbf{X}'\mathbf{X}$ 는 비정칙행렬이므로 $\mathbf{X}'\mathbf{X}$ 의 역행렬로 Moore-Penrose의 일반화역행렬 $(\mathbf{X}'\mathbf{X})^-$ 를 이용하게 된다. $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-$ 를 식 (2.2)의 양쪽에 곱해주면

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{y} \quad (2.3)$$

이다. 따라서, 일반화 역행렬을 이용한 해 벡터 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{y}$ 를 얻게 된다. $\hat{\boldsymbol{\beta}}$ 의 또 다른 해의 형태는 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{y} + (\mathbf{I} - (\mathbf{X}'\mathbf{X})^- (\mathbf{X}'\mathbf{X}))\mathbf{h}$ 로 주어진다. \mathbf{h} 는 임의의 실수벡터이다. 따라서 무수히 많은 해가 존재하게 된다. 모수벡터 $\boldsymbol{\beta}$ 의 한 선형함수 $l'\boldsymbol{\beta}$ 가 $\boldsymbol{\beta}$ 의 한 추정가능함수이기만하면 $l'\boldsymbol{\beta}$ 의 불편추정량이 존재한다. 즉, $E(\mathbf{a}'\mathbf{y}) = l'\boldsymbol{\beta}$ 되는 불편추정량 $\mathbf{a}'\mathbf{y}$ 가 존재한다. 불편추정량이 존재하는 모수벡터 $\boldsymbol{\beta}$ 의 선형함수 중 선형적으로 독립인 추정가능함수들의 수는 모형행렬 \mathbf{X} 의 계수와 동일하다. \mathbf{X} 의 계수가 $k (< p)$ 이므로 선형독립인 추정가능함수들을 $l_1'\boldsymbol{\beta}, l_2'\boldsymbol{\beta}, \dots, l_k'\boldsymbol{\beta}$ 라 두자. 이들 k 개 선형함수들을 행렬 \mathbf{L} 로 표현하면 $\mathbf{L} = (l_1, l_2, \dots, l_k)$ 인 완전열계수 행렬로 주어진다.

$\mathbf{L}'\boldsymbol{\beta}$ 는 선형적으로 독립인 k 개의 추정가능함수들의 집합이므로 추정가능함수들의 한 기저집합을 나타낸다. 기저 집합을 구성하는 문제를 생각해 본다.

3. 추정가능함수의 기저집합

고정효과의 선형모형에서 추정가능한 모수 또는 모수들의 함수인 추정가능함수들의 파악은 추론을 위해 필요하다. 이를 위해 가정된 모형으로부터 행렬표현식을 이용하거나 정규방정식을 이용하여 추정가능함수들의 한 기저집합을 유도할 수 있다. 추정가능함수의 한 기저집합을 모형식 (2.1)을 이용하여 유도해 보기로 한다. 모형행렬 \mathbf{X} 의 계수가 k 이므로 모형행렬의 변환을 통하여 구해진 완전계수의 모형행렬과 관련된 모수벡터는 선형적으로 독립인 추정가능함수들의 한 집합인 기저벡터를 나타낸다. 즉,

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= \mathbf{X}\mathbf{M}^- \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\epsilon} \end{aligned} \quad (3.1)$$

로 표현되는 계수가 k 인 $k \times p$ 행렬 M 으로 인해 XM^{-1} 는 완전계수의 행렬이다. 따라서 $M\beta$ 는 추정 가능함수의 한 기저집합을 나타내는 모수벡터이다. 여기서 M 은 X 의 선형적으로 독립인 k 개의 행벡터들로 구성된다. 다른 방법은 정규방정식을 나타내는 선형방정식계의 계수행렬 $X'X$ 에 가우스소거법을 이용하여 구해진 가우스행렬로부터 0 아닌 k 개의 행벡터들로 구성된 행렬을 R 이라 두면 $R\beta$ 는 한 기저집합이다. 추정가능함수의 정의로부터 $L'\beta$ 가 행렬모형식 (2.1)의 한 기저집합이면 $A'X = L'$ 되는 A' 가 존재한다. A' 는 계수가 k 인 $k \times n$ 인 행렬이다. A' 는 X 의 n 개 행벡터의 선형결합으로 k 개의 선형적으로 독립인 행벡터들을 구성하는 계수행렬이다. 즉,

$$A'X = (a_1, a_2, \dots, a_k)'X \tag{3.2}$$

에서 $A'X$ 의 각 행은 X 의 n 개 행벡터들의 선형결합으로 구성된다. X 의 n 개 행벡터들의 선형결합 중 k 개 독립인 선형결합을 얻는 한 방법은 X 의 n 개 행들에서 독립인 k 개 행들을 이용할 수 있다. 모형행렬 X 의 행공간은 X 의 n 개 행벡터들로 생성되는 p 차원의 벡터공간이므로 XX' 의 각 행은 X 의 p 개 행들의 선형결합이다. 행렬 XX' 에 가우스소거법을 적용하여 선형적으로 독립인 k 개 행들을 구성한다. XX' 의 가우스행렬에서 영아닌 행벡터는 X 의 행공간의 기저를 이루므로 A' 으로 나타낸다. 이 경우에 기저집합은

$$L'\beta = A'X\beta \tag{3.3}$$

로 주어진다. 주어진 행렬모형식으로부터 추정가능한 함수들의 기저집합을 구성하는 방법을 살펴왔다.

4. 사영행렬에 의한 추정가능함수의 검토

가정된 모형으로부터 추정가능한 모수들의 형태는 다양한 방식으로 결정되고 무수히 많은 기저집합이 있게 된다. 관심의 모수가 추정가능한 지 또는 모수들의 함수로 주어지는 하나 또는 둘 이상의 모수들이 추정가능한 함수들인 가를 판단하는 문제를 생각해 보자. 불완전계수의 모형행렬을 갖는 모형식 (2.1)의 추정가능함수들은 X 의 행공간을 구성하는 선형독립인 행벡터들의 선형함수이다. 따라서 추정가능함수의 계수벡터는 행공간의 좌표축을 나타내는 벡터들의 선형결합으로 주어짐을 알 수 있다. X 의 p 차원의 행벡터들로 생성되는 X 의 행공간은 $X'X$ 의 행공간과 동일하다. 계수가 k 이므로 X 의 행공간은 n 개 행벡터중 선형적으로 독립인 임의의 k 개 행벡터들로 구성되는 행공간을 생각할 수 있다. 따라서, k 차원의 행공간 내의 추정가능함수들은 p 차원의 벡터공간의 부분공간에 속해 있으므로 추정가능함수의 계수벡터는 p 개 행벡터들의 선형결합으로 주어지게 된다. 그러므로 k 차원의 부분공간에 속한 추정가능함수의 계수벡터는 p 차원의 행공간으로의 사영과 일치하게 된다. 즉, $L'\beta$ 가 추정가능함수이면

$$l'(X'X)^{-1}X'X = l' \tag{4.1}$$

임을 의미한다. l' 이 추정가능함수인가의 확인은 l' 이 주어지는 k 차원의 X 의 행공간으로의 사영을 이용할 수 있다. R 을 계수가 k 인 $k \times p$ 인 행렬이라 두자. l' 이 R 의 k 개 행벡터들의 선형결합으로 주어지면

$$l'R'(RR')^{-1}R = l' \tag{4.2}$$

이다. 즉, $X^{-1}X = R^{-1}R$ 이다. R_i 을 계수가 1인 $k \times p$ 인 행렬이라 두자. l' 이 R_i 의 k 개 행벡터들의 선형결합으로 주어지면

$$l'R_i^{-1}R_i = l' \tag{4.3}$$

Table 6.1. Two-way experimental data for Phospor, Glass-type experiment.

A	B								
	b ₁			b ₂			b ₃		
a ₁	9	12	13	14	15	16	1	6	7
a ₂	4	8	18	3	5	17	2	10	11

이다. 이 경우에

$$\mathbf{X}^- \mathbf{X} \mathbf{R}_l^- \mathbf{R}_l = \mathbf{R}_l^- \mathbf{R}_l \quad (4.4)$$

이고 \mathbf{R}_l 의 행공간은 \mathbf{X} 의 행공간의 부분공간임을 나타낸다. 이는 추정가능함수의 형태를 유형별로 분류할 수 있음을 의미한다. 가정된 고정효과의 선형모형에서 추정가능한 함수들의 일반형이 제시되어 있으나 추정가능한 함수가 되는 모수들의 계수를 설정하기가 쉽지 않다. 예를 들면, $\mathbf{l}'\boldsymbol{\beta}$ 가 추정가능함수이면 $\boldsymbol{\beta}$ 의 계수들로 구성되는 \mathbf{l}' 의 일반형은 $\mathbf{l}' = (l_1, l_2, \dots, l_p)$ 로 식 (4.1) 또는 식 (4.2)가 만족되는 상수를 결정할 필요가 있다.

5. 추정가능함수와 고유벡터

모형행렬 \mathbf{X} 의 행공간은 \mathbf{X} 의 계수 k 에 해당하는 선형적으로 독립인 행벡터들에 의해 생성되거나 $\mathbf{X}'\mathbf{X}$ 의 p 개 고유벡터에 의해 생성되는 벡터공간이다. 행공간은 또한 $\mathbf{X}^- \mathbf{X}$ 의 행벡터들로도 생성된다. 모형행렬 \mathbf{X} 와 관련된 추정가능함수의 한 기저집합은 \mathbf{X} 의 행공간에서 최소제곱추정해의 값에 불변인 양으로 추정되는 모수들이다. 따라서 한 기저집합 $\mathbf{L}'\boldsymbol{\beta}$ 의 \mathbf{L}' 은 \mathbf{X} 의 행공간에서

$$\mathbf{L}'\mathbf{X}^- \mathbf{X} = \mathbf{L}' \quad (5.1)$$

되는 성질을 만족해야 한다. 즉, \mathbf{L}' 은 \mathbf{X} 의 행공간을 구성하는 양의 고유근을 갖는 k 개 고유벡터들의 선형함수들로 구성되는 많은 완전행계수 행렬중 하나이다. \mathbf{L}' 이 k 개 고유벡터들의 선형함수들로 구성되므로 추정가능함수의 한 기저집합을 구성하는 방법으로 $\mathbf{X}'\mathbf{X}$ 의 고유벡터들을 활용할 수 있다. $\mathbf{X}'\mathbf{X}$ 의 0아닌 k 개 고유근 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ 에 해당하는 고유벡터를 $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k$ 라 두자. $\mathbf{l}'_i = (l_{i1}, l_{i2}, \dots, l_{ik})$ 이고 $\mathbf{l}'_i\boldsymbol{\beta}$ 가 추정가능함수일 때 $\mathbf{P}' = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k)'$ 이면 $\mathbf{c}'_i = (c_{i1}, c_{i2}, \dots, c_{ik}) \neq \mathbf{0}$ 에 대해

$$\mathbf{l}'_i = \mathbf{c}'_i \mathbf{P}' = c_{i1}\mathbf{p}'_1 + c_{i2}\mathbf{p}'_2 + \dots + c_{ik}\mathbf{p}'_k \quad (5.2)$$

가 성립한다. 기저집합을 나타내는 계수벡터의 행렬 \mathbf{L}' 은 선형적으로 독립인 k 개의 \mathbf{l}'_i 으로 구성됨을 알 수 있다.

6. 추정가능함수의 구성 예

가정된 선형모형의 행렬식으로부터 추정가능함수를 구성하는 예로 Hicks (1973)의 자료를 이용해 보기로 한다. Table 6.1은 광출력(light output)에 영향을 주는 두 유형의 페이스플레이트 유리(A)와 세 유형의 형광체(B)의 수준결합에서 수집된 광출력 자료를 나타내고 있다. 요인 A는 a_1, a_2 의 두 수준이고 요인 B는 b_1, b_2, b_3 의 세 수준이다. 여섯 개의 각 실험조건에서 세번의 관측이 행해지고 관측단위는 $\mu\mathbf{A}$ (microamperes)이다. Table 6.1의 자료에 대한 모형식 (2.1)의 모형행렬 \mathbf{X} 에서 반복되지 않

는 행벡터들을 순서대로 놓은 행렬을 \mathbf{X}_r 이라 두고 \mathbf{X}_r 에 가우스소거법을 이용하여 구한 가우스행렬을 \mathbf{G}_r 이라 두면

$$\mathbf{X}_r = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \text{이고, } \mathbf{G}_r = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

을 얻게 된다. 식 (3.1)의 완전행계수 행렬 \mathbf{M} 은 \mathbf{G}_r 의 행벡터가 아닌 행들로 구성된다. 따라서 추정가능함수의 한 기저집합은

$$\mathbf{M}\beta = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \mu + \alpha_2 + \beta_3 \\ \alpha_1 - \alpha_2 \\ \beta_1 - \beta_3 \\ \beta_2 - \beta_3 \end{pmatrix} \quad (6.1)$$

로 구해진다. 식 (2.2)의 $\mathbf{X}'\mathbf{X}$ 에 가우스소거법을 이용하여 구해지는 선형독립인 행들도 완전행계수 행렬 \mathbf{M} 을 나타내므로 식 (6.1)의 결과를 얻게 된다. $\mathbf{X}'\mathbf{X}$ 의 선형독립인 네개의 행벡터로 구성된 행렬

$$\mathbf{M}_1 = \begin{pmatrix} 18 & 9 & 9 & 6 & 6 & 6 \\ 9 & 9 & 0 & 3 & 3 & 3 \\ 6 & 3 & 3 & 6 & 0 & 0 \\ 6 & 3 & 3 & 0 & 6 & 0 \end{pmatrix}$$

으로부터 Gram-Schmidt의 직교화로 얻은 행렬

$$\mathbf{M}_{1s} = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.3 & 0.3 & 0.3 \\ 0 & 4.5 & -4.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & -2 & -2 \\ 0 & 0 & 0 & -2 & 4 & -2 \end{pmatrix}$$

를 이용하여 또 다른 기저함수의 집합을 얻을 수 있다. 즉,

$$\mathbf{M}_{1s}\beta = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.3 & 0.3 & 0.3 \\ 0 & 4.5 & -4.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & -2 & -2 \\ 0 & 0 & 0 & -2 & 4 & -2 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \mu + 0.5\alpha_1 + 0.5\alpha_2 + 0.3\beta_1 + 0.3\beta_2 + 0.3\beta_3 \\ 4.5\alpha_1 - 4.5\alpha_2 \\ 4\beta_1 - 2\beta_2 - 2\beta_3 \\ -2\beta_1 + 4\beta_2 - 2\beta_3 \end{pmatrix} \quad (6.2)$$

이다. 식 (5.2)는 $\mathbf{X}'\mathbf{X}$ 의 고유벡터들을 활용하여 추정가능함수를 구성하는 방식을 제시하고 있다. 자료로부터 $\mathbf{X}'\mathbf{X}$ 의 계수는 4이므로 양수인 네 고유근 $\lambda_1 = 33$, $\lambda_2 = 9$, $\lambda_3 = 6$, $\lambda_4 = 6$ 에 해당하는 고

유벡터들을 $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ 그리고 \mathbf{p}_4 라 둘 때, 이들 네개의 고유벡터로 구성된 행렬을 \mathbf{P} 로 나타내자. \mathbf{P} 는 6×4 인 행렬이다. $\mathbf{c}'_0 = (0, 0, 1, 1)$ 이라 두면, 식 (5.2)에 의한 \mathbf{l}'_0 는 $\mathbf{c}'_0 \mathbf{P}'$ 으로 표시되고

$$\mathbf{l}'_0 = \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.74 & 0.37 & 0.37 & 0.25 & 0.25 & 0.25 \\ 0 & 0.71 & -0.71 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.63 & 0.76 & -0.13 \\ 0 & 0 & 0 & -0.52 & -0.29 & 0.81 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -1.15 & 0.47 & 0.68 \end{pmatrix} \quad (6.3)$$

로 구해진다. 따라서, $\mathbf{l}'_0 \boldsymbol{\beta} = -1.15\beta_1 + 0.47\beta_2 + 0.68\beta_3$ 로 주어진다. 고유벡터들을 이용한 추정함수의 한 기저집합 $\mathbf{L}'\boldsymbol{\beta}$ 를 구하기 위해 서로 독립인 네 계수벡터를 \mathbf{c}_i , ($i = 1, 2, 3, 4$)라 두자. \mathbf{c}_i 들을 $\mathbf{c}'_1 = (1, 0, 0, 0)$, $\mathbf{c}'_2 = (1, 0, 1, 0)$, $\mathbf{c}'_3 = (0, 1, -1, 0)$, $\mathbf{c}'_4 = (0, 0, 0, 1)$ 이라 둔다. \mathbf{C}' 는 네 행벡터들로 구성된 행렬로 $\mathbf{L}'\boldsymbol{\beta}$ 는 $\mathbf{C}'\mathbf{P}'\boldsymbol{\beta}$ 로 주어진다. 즉,

$$\begin{aligned} \mathbf{L}'\boldsymbol{\beta} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.74 & 0.37 & 0.37 & 0.25 & 0.25 & 0.25 \\ 0 & 0.71 & -0.71 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.63 & 0.76 & -0.13 \\ 0 & 0 & 0 & -0.52 & -0.29 & 0.81 \end{pmatrix} \boldsymbol{\beta} \\ &= \begin{pmatrix} 0.74 & 0.37 & 0.37 & 0.25 & 0.25 & 0.25 \\ 0.74 & 0.37 & 0.37 & -0.39 & 1.01 & 0.12 \\ 0 & 0.71 & -0.71 & 0.63 & -0.76 & 0.13 \\ 0 & 0 & 0 & -0.52 & -0.29 & 0.81 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \\ &= \begin{pmatrix} 0.74\mu + 0.37\alpha_1 + 0.37\alpha_2 + 0.25\beta_1 + 0.25\beta_2 + 0.25\beta_3 \\ 0.74\mu + 0.37\alpha_1 + 0.37\alpha_2 - 0.39\beta_1 + 1.01\beta_2 + 0.12\beta_3 \\ 0.71\alpha_1 - 0.71\alpha_2 + 0.63\beta_1 - 0.76\beta_2 + 0.13\beta_3 \\ -0.52\beta_1 - 0.29\beta_2 + 0.81\beta_3 \end{pmatrix} \quad (6.4) \end{aligned}$$

이다. 한편, $\mathbf{l}'_0 \boldsymbol{\beta}$ 가 추정가능함수인가의 여부는 \mathbf{l}'_0 이 \mathbf{L}' 의 행공간내 벡터인가를 확인함으로써 알 수 있다. 즉, $\mathbf{l}'_0 = \mathbf{l}'_0 \mathbf{L}'^+ \mathbf{L}'$ 이므로 추정가능함수가 된다. $\mathbf{l}'_0 \boldsymbol{\beta}$ 가 추정가능함수이므로 모수벡터 $\boldsymbol{\beta}$ 의 최소제곱해를 이용하여 추정치를 구할 수 있고 $\mathbf{l}'_0 \boldsymbol{\beta}$ 와 관련된 가설검정이 가능하게 된다. 식 (3.3)에서 $\mathbf{A}'\mathbf{X}=\mathbf{L}'$ 되는 \mathbf{A}' 을 식 (6.4)의 \mathbf{L}' 을 이용하여 구해 본다. $\mathbf{A}' = (\mathbf{A}_1|\mathbf{A}_2)$ 라 둘 때, 부분행렬 \mathbf{A}_1 은

$$\mathbf{A}_1 = \begin{pmatrix} 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\ -0.06 & -0.06 & -0.06 & 0.17 & 0.17 & 0.17 & 0.02 & 0.02 & 0.02 \\ 0.18 & 0.18 & 0.18 & -0.05 & -0.05 & -0.05 & 0.10 & 0.10 & 0.10 \\ -0.09 & -0.09 & -0.09 & -0.05 & -0.05 & -0.05 & 0.13 & 0.13 & 0.13 \end{pmatrix}$$

이고 부분행렬 \mathbf{A}_2 는

$$\mathbf{A}_2 = \begin{pmatrix} 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\ -0.06 & -0.06 & -0.06 & 0.17 & 0.17 & 0.17 & 0.02 & 0.02 & 0.02 \\ 0.03 & 0.03 & 0.03 & -0.21 & -0.21 & -0.21 & -0.06 & -0.06 & -0.06 \\ -0.09 & -0.09 & -0.09 & -0.05 & -0.05 & -0.05 & 0.13 & 0.13 & 0.13 \end{pmatrix}$$

로 얻어진다. 추정가능함수의 한 기저집합을 구성하는 다양한 방법들에 의해서 많은 기저집합이 존재하게 된다. 이들 기저집합이 추정가능한 함수들의 집합인 가를 확인하기 위한 방법은 추정가능함수의 정의나 모수점이 존재하는 사영공간으로의 사영행렬을 이용할 수 있다.

사영공간으로의 사영행렬은 모형행렬 \mathbf{X} 에 의해 주어지는 $\mathbf{X}^{-}\mathbf{X}$, 추정가능함수의 한 기저집합을 나타내는 계수벡터 \mathbf{M} 또는 \mathbf{L}' 으로 주어지는 $\mathbf{M}^{-}\mathbf{M}$ 이나 $\mathbf{L}'^{-}\mathbf{L}'$ 중 하나를 이용하여 추정가능함수 또는 추정가능함수의 기저여부를 확인할 수 있다. 이들 사영행렬은 모두 동일한 행공간으로의 사영을 나타내는 사영행렬들이기 때문이다. 즉, 식 (6.2)의 \mathbf{M}_{1s} 가 추정가능함수의 한 기저집합일 때

$$\mathbf{M}_{1s} = \mathbf{M}_{1s}\mathbf{X}^{-}\mathbf{X} = \mathbf{M}_{1s}\mathbf{M}^{-}\mathbf{M} = \mathbf{M}_{1s}\mathbf{L}'^{-}\mathbf{L}' \quad (6.5)$$

이 성립함을 알 수 있다.

7. 결론

본 논문은 고정효과모형의 가정하에서 모형내 모수들의 추정가능한 함수를 다루고 있다. 일반적으로 자료로부터 추정할 수 있는 모수의 수보다 많은 모수를 포함하고 있는 모형인 경우에 모수의 최소제곱해가 유일하지 않기 때문에 모수의 어떤 형태가 추정가능한가에 관심을 갖게 된다. 이런 관점에서 모형내 모수들의 함수로 추정가능함수를 구성하는 방법과 관심의 모수가 추정가능함수인 가를 판단하는 기법으로 사영공간에서의 사영행렬을 어떻게 활용하는 가를 구체적으로 기술하고 있다.

추정가능함수의 한 기저집합을 구성하기 위한 방법으로 모형행렬 \mathbf{X} 또는 $\mathbf{X}'\mathbf{X}$ 의 가우스행렬을 이용하거나 $\mathbf{X}'\mathbf{X}$ 의 고유벡터를 이용하여 구성하는 방법을 제시하고 있다. 행공간의 기저를 구하는 많은 방법들로 인해 추정가능함수의 기저집합이 다양한 형태로 주어질 수 있음을 상세히 기술하고 있다. 또한, 추정가능함수들의 기저집합인 가를 확인하기 위한 방법으로 \mathbf{X} 의 행공간으로의 사영행렬을 이용할 수 있음을 다루고 있다.

추정가능함수의 기저집합을 나타내는 다양한 기법으로부터 사영행렬이 어떻게 달라지는 가를 구체적으로 보여주고 있으며 이들은 모두 동일한 행공간으로의 사영을 나타내는 사영행렬임을 논의하고 있다. 또한, 추정가능함수의 정의로부터 모형행렬 \mathbf{X} 의 행벡터들의 선형결합에 해당하는 계수벡터를 구하는 방법도 구체적으로 다루고 있다.

References

- Choi, J. S. (2011). Variance components in one-factor random model by projections, *Journal of the Korean Data & Information Science Society*, **22**, 381–387.
- Choi, J. S. (2012). Type II analysis by projections, *Journal of the Korean Data & Information Science Society*, **23**, 155–1163.
- Elswick, R. K., Gennings, Jr. C., Chinchilli, V. M. and Dawson, K. S. (1991). A simple approach for finding estimable functions in linear models, *The American Statistician*, **45**, 51–53.
- Graybill, F. A. (1976). *Theory and Application of the Linear Model*, Wadsworth Publishing Company, Inc., Belmont.
- Hicks, R. C. (1973). *Fundamental Concepts in the Design of Experiments*, 2nd edition, Holt, Rinehart and Winston, Inc., New York.
- Johnson, R. A. and Wichern, D. W. (1988). *Applied Multivariate Statistical Analysis*, 2nd edition, Prentice Hall, Inc., Englewood Cliffs.
- Milliken, G. A. and Johnson, D. E. (1984). *Analysis of Messy Data*, Van Nostrand Reinhold, New York.
- Searle, S. R. (1971). *Linear Models*, John Wiley and Sons, Inc., New York.

사영을 이용한 고정효과모형의 추정가능함수

최재성^{a,1}

^a계명대학교 통계학과

(2014년 3월 20일 접수, 2014년 6월 10일 수정, 2014년 6월 18일 채택)

요약

본 논문은 고정효과 모형에서 모수 또는 모수들의 선형함수로 추정가능한 함수를 다루고 있다. 추정할 수 있는 모수들의 수보다 더 많은 모수를 갖는 고정효과모형의 가정에서 관심모수가 추정가능한 모수가 아닌 경우에 최소제곱해는 유일하지 않다. 모형내 모수추정법으로 최소제곱법의 이용은 자료의 벡터공간에서 사영을 구하는 방법과 동일하므로 최소제곱해에 불변인 성질의 추정량을 갖는 추정가능함수의 형태를 사영의 관점에서 파악하고 구성하는 방법을 다루고 있다. 또한, 선형적으로 독립인 추정가능함수들의 기저집합을 구성하는 방법으로 사영공간의 고유벡터들을 활용할 수 있음을 논의하고 있다.

주요용어: 고유벡터, 고정효과, 사영, 최소제곱해, 추정가능함수.

¹(704-701) 대구광역시 달서구 신당동 1000번지, 계명대학교 통계학과. E-mail: office@kss.or.kr