Journal of the Korean Data & Information Science Society
2012, **23**(3), 487–494

http://dx.doi.org/10.7465/jkdi.2012.23.3.487
한국데이터정보과학회지

사영에 의한 모수모형의 추정가능함수

최재성[1]

[1]계명대학교 통계학과

접수 2012년 4월 7일, 수정 2012년 4월 25일, 게재확정 2012년 5월 7일

요약

본 논문은 고정효과모형의 모수추정과 관련된 추정가능함수를 다루고 있다. 고정효과모형에서 추정가능한 모수들의 함수를 구하기 위한 방법으로 가우스-조르단 방법을 이용하고 있다. 이 방법으로 구해진 추정가능함수의 일반형을 이용하여 추정가능함수들의 한 기저집합을 구성하는 문제를 다루고 있다. 또한, 추정가능함수들로 구성된 한 기저집합을 모수벡터로 갖는 모형은 완전계수의 열행렬을 모형행렬로 갖는 모형으로 효과모형과 동치인 모형임을 보여주고 있다. 두 모형에서의 사영행렬들은 동일공간으로의 사영을 나타내므로 총변동량은 변함이 없으나 사영행렬에 따른 자유도 1인 고유벡터로의 변동량은 달라질 수 있음을 논의하고 있다.

주요용어: 고유벡터, 기저집합, 사영, 사영행렬, 추정가능함수.

1. 서론

실험자료를 분석하기 위한 고정효과 모형의 가정은 실험단위의 반응에 영향을 미치는 독립변수들이 고정요인들로 간주될 때 가능하다. 고정효과 모형에서 요인들의 수준효과를 나타내는 모수들은 대다수 추정가능한 (estimable)모수가 아니다. 자료로부터 유일하게 추정될 수 있는 모수의 수보다 더 많은 모수가 포함된 과다설정의 모형 (overspecified model)이거나 모형행렬이 완전계수가 아닌 모형 (less than full rank model)이 가정되는 경우에 모수 또는 모수들의 함수에 대한 추정가능성 (estimability)을 검토하게 된다.

가정된 선형모형이 완전계수모형이면 모형내 미지모수에 대한 유일 해를 구할 수 있고 해는 추정가능한 모수의 추정값이 된다. 불완전계수모형이면 추정가능하지 않은 모형내 미지모수들의 해가 무수히 많이 존재하게 되고 이 경우 해가 갖는 의미는 단순히 정규방정식의 선형방정식계가 일치하게 되는 가능한 한 값으로 간주된다. 그러나 추정가능함수의 추정값은 정규방정식의 선형방정식계에서 구해진 모수의 어떤 해를 사용하더라도 불변의 (invariant) 고정 값으로 주어진다. 추정가능함수에 대한 논의는 Searle (1971), Graybill (1976) 그리고 Milliken과 Johnson (1984) 등에서 살펴볼 수 있다. 최재성 (2011)은 확률모형의 가정하에 사영을 이용하여 분산성분을 구하는 방법을 제시하고 있다. 이외에, 최재성 (2008, 2010)은 실험 자료의 분석을 위한 선형모형으로 일반화 선형모형 및 혼합모형의 가정하에 모형내 모수를 추론하는 방법을 다루고 있다.

고정효과 모형의 적합방식에 따라 여러 유형의 추정가능함수가 다루어진다. 다양한 유형의 추정가능함수에 대한 구체적 논의는 Milliken과 Johnson (1984)에서 보여진다. 모형적합 방식에 따른 추정가능함수의 인지는 추정가능함수의 추론과 연관되어 있다. 추정가능함수가 아닌 모수의 추론은

[1] (704-701) 대구광역시 달서구 신당동 1000번지, 계명대학교, 통계학과, 교수. E-mail: jschoi@kmu.ac.kr

의미가 없고 추론의 타당성을 부여할 수 없다. 왜냐하면 모수추정공간에서 자료로부터 추정할 수 있는 모수가 아니기 때문이다.

본 논문에서는 고정효과 모형의 가정하에 자료분석을 할 때, 모형내 모수와 추정가능함수와의 관련성을 살펴보고 모형적합 방식에 따른 추정가능함수의 형태는 어떻게 결정되는 가를 논의한다. 그리고 추정가능함수가 모형행렬과 어떤 연관성을 갖는 가를 다루고자 한다. 또한, 사영의 측면에서 추정가능함수를 어떻게 규명할 것인 가를 논의하고 구체화 하고자 한다. 사영의 개념 및 사영행렬의 이용에 관한 자세한 논의는 Johnson과 Wichern (1988)에서 보여 진다. 사영은 벡터공간에서 정의되며 벡터 $\boldsymbol{y}$를 벡터 $\boldsymbol{x}$로의 사영은 $\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y}$로 정의된다. 벡터 $\boldsymbol{x}$나 벡터들의 집합인 행렬 $\boldsymbol{X}$에 의해 생성된 벡터공간으로의 사영이 행해질 때 사영이 행해지는 벡터공간을 사영공간이라 부르고 사영과 관련된 행렬을 사영행렬이라 한다.

2. 추정가능함수의 모형에 대한 가정

일반형의 추정가능함수를 규명하기 위한 모형으로 이원 고정효과 모형을 가정한다. 실험단위의 반응에 영향을 미치는 두개의 처리요인을 각기 T와 B라 두자. 요인 T의 수준 수는 t개이고 요인 B의 수준 수는 b개라 가정한다. 두 고정요인 T와 B의 수준결합으로 주어지는 처리조합의 수는 모두 tb개가 되고, 요인 T의 수준 i와 요인 B의 수준 j에서의 처리조합은 (i,j)로 표시된다. 요인 T의 i번째 수준의 효과를 τ_i $(i=1,2,\cdots,t)$라 둘 때, τ_i는 고정효과를 나타낸다. 요인 B의 j번째 수준의 효과를 β_j $(j=1,2,\cdots,b)$라 둘 때, β_j는 고정효과를 나타낸다. 처리조합 (i,j)가 n_{ij}개 실험단위들에 임의로 배정된다고 가정한다. 실험단위들이 거의 동질적이라 가정할 때 tb개의 처리조합은 실험단위들에 임의로 배정된다. 처리조합 (i,j)가 행해진 k번째 실험단위에서의 관측 값을 y_{ijk}라 둘 때 고정효과 모형은 다음과 같다.

$$y_{ijk}=\mu+\tau_i+\beta_j+\gamma_{ij}+\epsilon_{ijk} \tag{2.1}$$

단, μ는 전체평균이고 τ_i와 β_j는 각기 요인 T의 수준 i에서의 고정효과와 B의 수준 j에서의 고정효과를 나타낸다. γ_{ij}는 처리조합 (i,j)에서의 두 요인의 교호작용을 나타낸다. 오차항 ϵ_{ijk}는 평균이 0이고 분산이 σ_ϵ^2이라 가정한다. 모형내 추정되어야 할 미지모수들의 수는 $1+t+b+tb$개이고 추정방법으로 최소제곱법을 이용한다고 가정한다. 최소제곱법으로 주어지는 정규방정식의 수는 추정해야 할 모수의 수와 일치하나 모형내 미지모수들의 해를 얻기 위한 정규방정식들이 종속적이므로 해를 구하기 위한 제약식이 주어진다. 제약식으로

$$\sum_{i=1}^{t}\tau_i=0,\ \ \sum_{j=1}^{b}\beta_j=0 \text{ 이고 } \sum_{i=1}^{t}\gamma_{ij}=\sum_{j=1}^{b}\gamma_{ij}=0$$

임을 이용하여 정규방정식들에 대한 유일 해를 구할 수 있다. 또는 $\tau_t=0$, $\beta_b=0$, $\gamma_{tj}=\gamma_{ib}=0$, $(i=1,2,\cdots,t$ 이고 $j=1,2,\cdots,b)$를 이용하여 해를 구할 수 있다. 해를 구하기 위한 방법으로 두 가지 서로 다른 제약식을 이용할 때 모수에 대한 해들도 서로 다른 값을 취하게 된다. 그러나 추정가능함수는 서로 다른 값으로 주어진 해의 이용과 상관없이 항상 일정한 값을 나타내는 불변성의 성질을 갖고 있는 함수이다. 이러한 불변성의 성질을 갖는 추정가능함수는 모형행렬에 의해서 생성되는 모수추정공간에서 어떤 성질을 만족시키는 함수이다. 추정가능함수가 모수추정공간에서 무엇을 의미하고 있는 가를 알아보기 위해 식 (2.1)을 다음과 같이 행렬로 표현한다.

$$\boldsymbol{y}=\boldsymbol{j}\mu+\boldsymbol{X}_T\boldsymbol{\tau}+\boldsymbol{X}_B\boldsymbol{\beta}+\boldsymbol{X}_{TB}\boldsymbol{\gamma}+\boldsymbol{\epsilon} \tag{2.2}$$

단, $\boldsymbol{y}$는 $n\times 1$인 관측벡터이고 $n=\sum_{i=1}^{t}\sum_{j=1}^{b}n_{ij}$이다. $\boldsymbol{j}$는 모든 원소가 1인 $n\times 1$벡터이다. $\boldsymbol{X}_T$는 크기가 $n\times t$인 계수행렬, $\boldsymbol{X}_B$는 크기가 $n\times b$인 계수행렬이고 $\boldsymbol{X}_{TB}$는 크기가 $n\times tb$인 계수행렬을 나타낸다. $\boldsymbol{\epsilon}$은 $n\times 1$인 오차벡터이고 다변량 정규분포 $N(\boldsymbol{0},\sigma_\epsilon^2\boldsymbol{I}_n)$를 따른다고 가정한다. $\boldsymbol{\tau}$는 $t\times 1$인 모수벡터이고, $\boldsymbol{\beta}$는 $b\times 1$인 모수벡터이다. $\boldsymbol{\gamma}$는 교호작용을 나타내는 벡터이고 크기가 $tb\times 1$이다.

3. 추정가능함수의 일반형

행렬 모형식 (2.2)로부터 추정가능함수의 일반적 형태를 살펴보기로 한다. 식 (2.2)로부터 모형행렬 $\boldsymbol{X}=(\boldsymbol{j}\ \ \boldsymbol{X}_T\ \ \boldsymbol{X}_B\ \ \boldsymbol{X}_{TB})$인 분할행렬을 나타내고 $\boldsymbol{\theta}=(\mu\ \ \boldsymbol{\tau}\ \ \boldsymbol{\beta}\ \ \boldsymbol{\gamma})'$인 모수벡터라 두자. 모수벡터 $\boldsymbol{\theta}$의 한 선형함수 $\boldsymbol{q}'\boldsymbol{\theta}$가 추정가능함수이기만 하면 $\boldsymbol{q}=\boldsymbol{X}'\boldsymbol{X}\boldsymbol{h}$인 벡터 $\boldsymbol{h}$가 존재한다. 모형식 (2.2)의 추정가능함수들의 한 기저집합 (a basis set)은 $\boldsymbol{X}'X$를 기본 행 연산에 의한 기약가우스행렬로 변환시켰을 때 구해진다. 즉,

$$\boldsymbol{X}'\boldsymbol{X}=\begin{pmatrix}\boldsymbol{j}'j & \boldsymbol{j}'\boldsymbol{X}_T & \boldsymbol{j}'\boldsymbol{X}_B & \boldsymbol{j}'\boldsymbol{X}_{TB}\\ \boldsymbol{X}_T'\boldsymbol{j} & \boldsymbol{X}_T'\boldsymbol{X}_T & \boldsymbol{X}_T'\boldsymbol{X}_B & \boldsymbol{X}_T'\boldsymbol{X}_{TB}\\ \boldsymbol{X}_B'\boldsymbol{j} & \boldsymbol{X}_B'\boldsymbol{X}_T & \boldsymbol{X}_B'\boldsymbol{X}_B & \boldsymbol{X}_B'\boldsymbol{X}_{TB}\\ \boldsymbol{X}_{TB}'\boldsymbol{j} & \boldsymbol{X}_{TB}'\boldsymbol{X}_T & \boldsymbol{X}_{TB}'\boldsymbol{X}_B & \boldsymbol{X}_{TB}'\boldsymbol{X}_{TB}\end{pmatrix}$$

에 가우스-조르단 소거법을 적용하여 구해진 행렬의 0아닌 행벡터를 이용하여 추정가능함수의 일반형을 알 수 있다. 기약가우스 행렬의 0아닌 행벡터들은 선형적으로 독립이고 이들 벡터에 의해 생성된 행공간 (row space)의 차원은 기약가우스행렬에서 독립인 행벡터의 수와 일치한다.

고정효과모형의 모형행렬은 완전계수의 행렬이 아니므로 행렬의 열벡터 간에 선형종속인 벡터들로 구성된다. 따라서 개별 열벡터와 관련된 모수들은 추정가능하지 않음을 알 수 있다. 모형행렬 $\boldsymbol{X}$와 관련된 추정가능함수의 일반 형태와 추정가능한 함수의 개수에 관한 논의는 Milliken과 Johnson (1984), Elswick 등 (1991)에서 보여 진다. 추정가능함수의 일반 형태와 선형적으로 독립인 추정가능한 함수의 수를 파악하는 것은 추정가능한 함수에 관한 추론을 위해 반드시 다루어져야 할 부분이다.

완전계수가 아닌 모형행렬 $\boldsymbol{X}$의 구조와 특성에 따라 추정가능함수의 형태와 선형적으로 독립인 함수의 수도 달라진다. 식 (2.2)로부터

$$\begin{aligned}\boldsymbol{y}&=\boldsymbol{j}\mu+\boldsymbol{X}_T\boldsymbol{\tau}+\boldsymbol{X}_B\boldsymbol{\beta}+\boldsymbol{X}_{TB}\boldsymbol{\gamma}+\boldsymbol{\epsilon}\\ &=(\boldsymbol{j}\ \ \boldsymbol{X}_T\ \ \boldsymbol{X}_B\ \ \boldsymbol{X}_{TB})\begin{pmatrix}\mu\\ \boldsymbol{\tau}\\ \boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}+\boldsymbol{\epsilon} \qquad (3.1)\\ &=\boldsymbol{X}\boldsymbol{\theta}+\boldsymbol{\epsilon}\end{aligned}$$

이 성립함을 알 수 있다. 식에서 모형행렬 $\boldsymbol{X}$가 완전계수행렬이 아니면 모수벡터 $\boldsymbol{\theta}$는 추정가능한 모수들의 수보다 더 많은 모수를 갖는 벡터가 된다. 따라서 모수벡터 $\boldsymbol{\theta}$의 해를 구하기 위해 제약식을 이용하게 된다. 제약식으로 모수들의 합이 0이 되는 제약식 (sum-to-zero restrictions)을 이용하거나 일정 모수를 0으로 두는 제약식을 이용하여 모수를 추정하게 된다. 제약식의 유형에 따라 추정되는 모수의 추정값은 다른 값을 나타내고 또한 추정된 모수도 실제로 추정가능한 모수가 아니다. 그러므로 자료로부터 추정된 모수는 실제 어떤 모수인 가를 인식할 필요가 있다.

추정가능한 함수의 특성은 정규방정식의 해로 구해진 모수 값들이 일정하지 않더라도 추정가능함수의 추정 값은 항상 일정한 값으로 주어지는 불변성의 성질을 갖고 있다. 이러한 불변성의 성질을

갖는 벡터공간의 한 기저는 모형행렬 $\boldsymbol{X}$의 행벡터에 의해 생성된다. 이때 벡터공간의 기저를 구성하는 행벡터들의 수는 모형행렬 $\boldsymbol{X}$의 계수와 일치한다.

추정가능함수의 의미를 생각하기 위해 모형행렬 $\boldsymbol{X}$의 행공간을 $\Re$이라 두자. $\boldsymbol{X}$의 크기는 $n \times p$이고 계수는 $r\ (< p)$이라 가정한다. 단, $p = 1 + t + b + tb$개의 모수를 나타내고 있다. $\Re$은 $\boldsymbol{X}$의 n개 행벡터 중 서로 독립인 r개의 행벡터들로 구성된 행렬에 의해 생성된 공간이다. 따라서 $\Re$은 r개의 좌표축 상의 점들로 구성되는 공간이다. $\Re$내 임의의 좌표 점은 모두 추정가능한 함수의 모수점이다. $\Re$이 $\boldsymbol{X}$내 상호 독립인 r개의 행벡터로 생성된 공간이므로 사영행렬 $\boldsymbol{X}^{-}\boldsymbol{X}$에 의한 공간과 동일한 공간으로 인식된다. 그러므로 모수벡터 $\boldsymbol{\theta}$의 한 선형함수가 추정가능함수인 가의 확인은 그 함수를 공간 $\Re$로의 사영을 하였을 때 사영행렬 $\boldsymbol{X}^{-}\boldsymbol{X}$에 의해 확인가능하다. $\boldsymbol{l}'\boldsymbol{\theta}$를 추정가능한 함수라 두자. 이때 $\boldsymbol{l}$은 행공간 $\Re$내 한 점이어야 한다. 즉,

$$\boldsymbol{l}'\boldsymbol{X}^{-}\boldsymbol{X} = \boldsymbol{l}' \text{ 또는 } \boldsymbol{X}^{-}\boldsymbol{X}\boldsymbol{l} = \boldsymbol{l} \tag{3.2}$$

이다. $\boldsymbol{l}'\boldsymbol{\theta}$가 추정가능함수일 때 추정가능함수의 추정량은 $\boldsymbol{l}'\widehat{\boldsymbol{\theta}}$이고 $\widehat{\boldsymbol{\theta}}= \boldsymbol{X}^{-}\boldsymbol{y}$로 주어진다. $\boldsymbol{L} = [\boldsymbol{l}_1,\ \boldsymbol{l}_2, \cdots, \boldsymbol{l}_r]'$이라 두자. 단, $\boldsymbol{l}_i'\boldsymbol{\theta}(i = 1, 2, \cdots, r)$는 추정가능한 독립인 함수이고 $\boldsymbol{L}$은 추정가능한 함수들의 한 기저집합을 나타낸다. 따라서

$$\boldsymbol{L}\boldsymbol{X}^{-}\boldsymbol{X} = \boldsymbol{L} \text{ 또는 } \boldsymbol{X}^{-}\boldsymbol{X}\boldsymbol{L}' = \boldsymbol{L}' \tag{3.3}$$

이다. 관심의 모수벡터가 추정가능함수들의 벡터 $\boldsymbol{L}\boldsymbol{\theta}$일 때, 식 (3.1)은 다음과 같이 관심의 추정가능 함수벡터로 변환된다.

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}[(\boldsymbol{I} - \boldsymbol{L}^{-}\boldsymbol{L}) + \boldsymbol{L}^{-}\boldsymbol{L}]\boldsymbol{\theta} + \boldsymbol{\epsilon} \\ &= \boldsymbol{X}\boldsymbol{L}^{-}\boldsymbol{L}\boldsymbol{\theta} + \boldsymbol{\epsilon} \end{aligned} \tag{3.4}$$

로 변환된다. $\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{L}^{-}$라 두고 $\boldsymbol{\alpha} = \boldsymbol{L}\boldsymbol{\theta}$라 두면 식 (3.4)는

$$\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{3.5}$$

가 된다. 여기서 $\boldsymbol{Q}$는 $n \times r$인 완전계수행렬이고 $\boldsymbol{\alpha}$는 $r \times 1$인 추정가능함수의 한 기저집합을 나타내는 모수벡터이다. 모형행렬이 자료로부터 추정될 수 있는 모수의 수보다 많은 수의 모수를 포함하고 있는 모형의 계수행렬일 때 다양한 유형의 제약식을 이용하여 모수를 추정하게 된다. 실제로 추정가능한 모수의 일반형을 나타내는 추정가능함수들의 한 기저집합은 $\boldsymbol{X}'\boldsymbol{X}$의 기약가우스행렬로부터 구해진다. 기약가우스행렬을 구성하는 r개의 독립인 행벡터들로 이루어진 행렬 $\boldsymbol{L}$은 식 (3.3)을 만족하는 행렬이다.

관측벡터 $\boldsymbol{y}$를 식 (3.1)의 모형행렬 $\boldsymbol{X}$에 의해 생성된 사영공간으로의 추정은 $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$로 표현되는 사영이다. 식 (3.5)의 변환행렬 $\boldsymbol{Q}$는 모형행렬 $\boldsymbol{X}$와 동일한 사영공간을 이루고 있는 완전계수의 행렬이므로 $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$는 $\boldsymbol{Q}\boldsymbol{Q}^{-}\boldsymbol{y}$가 되고 두 사영행렬로 표현되는 사영공간은 동일한 벡터공간이 된다. 그러나 동일공간을 생성하는 두 사영행렬의 고유벡터들은 다르므로 기저집합내 추정가능함수들의 효과에 따른 변동량은 사영행렬 $\boldsymbol{Q}\boldsymbol{Q}^{-}$에서 고유근이 0 아닌 고유벡터로의 사영에 의한 사영까지의 거리로 계산된다.

사영공간으로의 사영에 따른 제곱거리는 사영행렬의 스펙트럴 분해를 이용하여 계산된다. 즉, $(\boldsymbol{Q}'\boldsymbol{Q})^{-} = \sum_{i=1}^{r} \lambda_i \boldsymbol{p_i}\boldsymbol{p_i}'$에 의해 자유도 1인 고유벡터 $\boldsymbol{p_i}$로의 사영에 이르는 거리를 계산하는 것이 가능하다. 단, λ_i는 양의 실수인 고유근을 나타내며 $\boldsymbol{p_i}$는 $\boldsymbol{Q}$의 열벡터에 의해 생성된 벡터공간의 i번째 축을 나타내는 r개 성분을 갖는 고유벡터이다.

모형행렬 X에 의해 생성된 공간으로의 사영을 나타내는 사영행렬 $\boldsymbol{XX}^-$와 완전계수의 열벡터로 구성된 행렬 $\boldsymbol{Q}$로 변환했을 때의 사영을 나타내는 사영행렬 $\boldsymbol{QQ}^-$는 동일한 벡터공간으로의 사영이다. 동일 자료의 효과에 따른 변동량의 분석에 있어서 사영행렬 간에 어떤 차이가 있고 어떻게 활용될 수 있는 가를 다음 예로 비교해 보고자 한다.

4. 불균형 자료의 예

다음 자료는 Milliken과 Johnson (1984)에서 사용된 불균형 자료이다. 행렬 모형식 (3.1)에 포함된 고정효과들의 추정가능함수에 대한 추정과 효과에 따른 변동량의 계산방법을 구체적으로 제공하기 위한 분석 자료이다. 표 4.1은 완전임의 설계구조에서 행해진 이원 처리구조의 실험으로부터 주어진 자료를 나타낸다. 표의 자료에서부터 요인 A는 A_1, A_2의 두 수준 이고 요인 B는 B_1, B_2, B_3의 세 수준이다. 따라서 처리조합은 이원구조의 6개로 두어진다. 실험에 이용될 실험단위들이 동질적이라 간주될 때 실험의 설계구조는 완전임의 설계구조가 된다.

표 4.1 불균형의 이원 실험 자료

요인 A ╲ 요인 B	B_1	B_2	B_3
A_1	19, 20, 21	24, 26	22, 25, 25
A_2	25, 27	21, 24, 24	31, 32, 33
합	112	119	168

위자료는 각 처리조합에서의 관측수가 동일하지 않으므로 불균형 자료로 분류된다. 자료의 행렬표현으로부터 관측벡터 $\boldsymbol{y}$는 크기가 16×1인 열벡터이다. 모형행렬 $\boldsymbol{X}$로의 사영은 $\boldsymbol{XX}^-\boldsymbol{y}$이고 사영까지의 거리제곱은 $\boldsymbol{y}'\boldsymbol{XX}^-\boldsymbol{y}$의 2차 형식으로 표현된다. 자료로부터 구해진 $\boldsymbol{y}'\boldsymbol{XX}^-\boldsymbol{y}$는 10,189이다. 모형식 (2.2)의 가정 하에 $\tau_2 = 0$, $\beta_3 = 0$, 그리고 $\gamma_{13} = \gamma_{21} = \gamma_{22} = \gamma_{23} = 0$의 제약조건하에 구해진 $\boldsymbol{\theta}$의 추정 값을 $\widehat{\boldsymbol{\theta}}_0$그리고 모수들의 합이 0이 되는 제약식에 의한 추정 값을 $\widehat{\boldsymbol{\theta}}_1$이라 두면 $\widehat{\boldsymbol{\theta}}_0$과 $\widehat{\boldsymbol{\theta}}_1$은 다음과 같다.

$$\widehat{\boldsymbol{\theta}}_0' = [32,\ -8, 0, -6,\ -9,\ 0,\ 2,\ 10,\ 0,\ 0,\ 0,\ 0]$$

이고,

$$\widehat{\boldsymbol{\theta}}_1' = [12.5,\ 4.75,\ 0.75, 2.83,\ 3.50,\ 6.17,\ -0.08,\ 4.25,\ 0.58, 2.92,\ -0.75,\ 5.58]$$

이다.

$\boldsymbol{X'X}$에 가우스-조르단 소거법을 이용하여 구한 기약가우스행렬의 6개 독립인 행벡터들로부터 구성되는 식 (3.3)의 행렬 $\boldsymbol{L}$을 구해보면 행렬 $\boldsymbol{L}$의 형태는 다음과 같이 주어진다.

$$\boldsymbol{L} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}$$

행렬 $\boldsymbol{L}$과 모수벡터 $\boldsymbol{\theta}$와의 곱으로 주어지는 식 (3.5)의 $\boldsymbol{\alpha}$는 일반형의 추정가능함수들로 구성된 한 기저집합을 나타내고 있음을 알 수 있다. $\boldsymbol{\alpha}$를 구해보면

$$\boldsymbol{\alpha} = \begin{pmatrix} \mu + \tau_2 + \beta_3 + \gamma_{23} \\ \tau_1 - \tau_2 + \gamma_{13} - \gamma_{23} \\ \beta_1 - \beta_3 + \gamma_{21} - \gamma_{23} \\ \beta_2 - \beta_3 + \gamma_{22} - \gamma_{23} \\ \gamma_{11} - \gamma_{13} - \gamma_{21} + \gamma_{23} \\ \gamma_{12} - \gamma_{13} - \gamma_{22} + \gamma_{23} \end{pmatrix}$$

이다. $\boldsymbol{\alpha}$의 추정량을 $\widehat{\boldsymbol{\alpha}}$라 두면 $\widehat{\boldsymbol{\alpha}} = \boldsymbol{Q}^{-}\boldsymbol{y}$이고 추정값은 $\widehat{\boldsymbol{\alpha}} = (32,\ -8, -6,\ -9,\ 2,\ 10)$로 주어진다. 이 값은 모수의 추정을 위해 0인 제약식을 이용했을 때의 0아닌 편의추정 값들과 동일함을 보여준다. 모형행렬 $\boldsymbol{X}$의 열벡터에 의해 생성된 사영공간으로의 사영에 의해 추정되는 $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\theta}} = \boldsymbol{Q}\widehat{\boldsymbol{\alpha}}$이다. $\boldsymbol{XX}^{-}$와 $\boldsymbol{QQ}^{-}$는 비록 동일한 사영공간을 나타내는 사영행렬이지만 각 사영행렬과 관련된 고유벡터로의 사영에 따른 변동량은 다른 양을 제공하게 된다. $\boldsymbol{XX}^{-}$와 $\boldsymbol{QQ}^{-}$가 동일한 벡터공간으로의 사영행렬임은 사영행렬 간의 차를 나타내는 행렬 $\boldsymbol{XX}^{-} - \boldsymbol{QQ}^{-}$의 대각합을 구해보면 0으로 주어진다. 즉, $\boldsymbol{XX}^{-} - \boldsymbol{QQ}^{-}$의 계수는 0인 영행렬임을 확인할 수 있다.

먼저 사영공간으로의 사영에 따른 양을 나타내는 $\boldsymbol{y}'\boldsymbol{XX}^{-}\boldsymbol{y}$나 $\boldsymbol{y}'\boldsymbol{QQ}^{-}\boldsymbol{y}$는 10,189로 동일한 값이다. 식 (3.5)의 $\boldsymbol{Q}$는 $n \times r$인 완전열계수 행렬이므로 표 4.1의 자료에서 주어지는 $\boldsymbol{Q}$는 16×6인 행렬을 나타낸다. $\boldsymbol{y}'\boldsymbol{QQ}^{-}\boldsymbol{y}$를 스펙트럴 분해에 따른 자유도 1인 변동량의 합으로 나타내 보기로 한다. 왜냐하면, $\boldsymbol{Q}$는 계수가 6인 완전열계수 행렬이므로 열벡터들은 상호독립이고 $\boldsymbol{QQ}^{-}\boldsymbol{y}$는 관측벡터 $\boldsymbol{y}$를 $\boldsymbol{Q}$의 열벡터로 생성된 벡터공간으로의 사영에 해당하기 때문이다. 이때 사영이 행해지는 위치는 직교하는 고유벡터들을 좌표축으로 하는 모수벡터 $\boldsymbol{\alpha}$로 주어진다. 각 고유벡터로의 사영을 생각할 때 사영까지의 제곱거리는 추정가능함수들의 효과에 따른 변동량을 나타내고 있다.

$\boldsymbol{Q}$의 열벡터로 생성된 벡터공간에서의 6개 고유근과 고유벡터는 소프트웨어 R언어로 구할 수 있다. 고유근들은 모두 0아닌 양수의 실수 값으로 주어지며 해당하는 고유벡터들을 구하게 된다. 이들 고유근과 고유벡터들을 $(\lambda_i, \boldsymbol{p}_i)$, $(i = 1, 2, \cdots, 6)$로 표시할 때 $(\lambda_i,\ \boldsymbol{p}_i)$에서의 변동량을 SS_{p_i}라 두자. 따라서 사영공간에서의 총제곱합은

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{QQ}^{-}\boldsymbol{y} &= \sum_{i=1}^{6} SS_{p_i} \\ &= 135.33 + 38.37 + 472.27 + 691.98 + 236.96 + 8614.10 \end{aligned}$$

의 합으로 구해진다. 예로써 $(\lambda_1, \boldsymbol{p}_1)$인 경우에 SS_{p_1}을 구해보기로 한다. $\lambda_1 = 3.436133$이고 $\boldsymbol{p}_1 = (0.228,\ -0.369,\ -0.361,\ -0.309,\ 0.556,\ 0.526)$이다. 따라서 고유벡터 $\boldsymbol{p}_1$으로의 사영을 나타내는 사영행렬을 $\boldsymbol{p}_1\boldsymbol{p}_1^{-}$라 두면 SS_{p_1}은 $\boldsymbol{y}'\boldsymbol{Q}(\lambda_1\boldsymbol{p}_1\boldsymbol{p}_1^{-})\boldsymbol{Q}'\boldsymbol{y}$에 의해 계산되고 그 값은 135.33이다. $\boldsymbol{y}'\boldsymbol{XX}^{-}\boldsymbol{y}$도 해당하는 고유근과 고유벡터들을 이용하여 자유도 1인 6개의 개별 변동량을 계산할 수 있으나 추정가능함수의 경우와는 다른 값들로 주어짐을 확인해 볼 수 있다.

5. 결론

본 논문은 이원 고정효과 모형의 가정하에 실험 자료를 분석할 때 사영의 관점에서추정가능함수와 관련된 변동량의 분석을 다루고 있다. 고정효과 모형의 모형행렬은 일반적으로 완전열계수 행렬이 아니기 때문에 모형내 개별모수들은 추정가능한 모수가 될 수 없고 또한 추론의 대상이 아니다.

모수들의 추정가능함수를 파악하기 위한 한 방법으로 가우스-조르단 소거법을 이용하고 있다. 불완전계수의 모형행렬을 가우스-조르단 방법에 의한 기약가우스 행렬로 변환하게 되면 모형행렬의 계수에 해당하는 상호 독립인 행벡터들을 얻게 된다. 이들 행벡터와 모수벡터와의 내적은 모형행렬의 계수에 해당하는 추정가능함수들을 나타내고 있음을 다루었다. 기약가우스 행렬로부터 구한 일반형의 추정가능함수들은 모형행렬의 계수에 해당하는 한 기저집합을 구성하고 있다. 한 기저집합내 추정가능함수를 모형내 모수들로 갖는 모형은 효과모형과 동치이며 완전계수의 행렬을 모형행렬로 취한다. 완전계수의 동치모형에서 주어진 모형행렬에 의해 생성된 사영공간은 효과모형의 모형행렬에 의해 생성된 벡터공간과 동일한 벡터공간을 구성한다. 동일한 벡터공간으로의 사영을 나타내는 사영행렬간의 차는 영행렬임을 보여주고 있다. 동일공간으로의 사영을 나타내는 두 사영행렬로 표현된 각 2차형식들은 각기 $\boldsymbol{y}'\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$와 $\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{Q}^{-}\boldsymbol{y}$로 계산되나 총변동량은 동일함을 보여준다. 자유도 1인 고유벡터로의 사영에 따른 변동량으로 총변동량이 분해될 때 사영행렬에 따라 다른 값으로 구해지고 있음을 논의하고 있다.

참고문헌

Choi, J. (2008). A marginal probability model for repeated polytomous response data. *Journal of the Korean Data & Information Science Society*, **19**, 577-585.

Choi, J. (2010). A mixed model for repeated split-plot data. *Journal of the Korean Data & Information Science Society*, **21**, 1-9.

Choi, J. (2011). Variance components in one-factor random model by projections. *Journal of the Korean Data & Information Science Society*, **22**, 381-387.

Elswick, K. R. Gennings, C. Jr., Chinchilli, M. V. and Dawson S. K. (1991). A simple approach for finding estimable functions in linear models. *The American Statistician*, **45**, 51-53.

Graybill, A. F. (1976). *Theory and application of the linear model*, Wadsworth Publishing Company, Inc., Belmont.

Johnson, A. R. and Wichern, W. D. (1988). *Applied multivariate statistical analysis*, 2nd edition, Prentice Hall, Inc., Englewood Cliffs.

Milliken, A. G. and Johnson, E. D. (1984). *Analysis of messy data*, Van Nostrand Reinhold, New York.

Searle, R. S. (1971). *Linear models*, John Wiley and Sons, Inc., New York.

Journal of the Korean Data &
Information Science Society
2012, **23**(3), 487–494

http://dx.doi.org/10.7465/jkdi.2012.23.3.487
한국데이터정보과학회지

Estimable functions of fixed-effects model by projections

Jaesung Choi[1]

[1]Department of Statistics, Keimyung University

Received 7 April 2012, revised 25 April 2012, accepted 7 May 2012

Abstract

This paper discusses a method for getting a basis set of estimable functions of model parameters in a two-way fixed effects model. Since the fixed effects model has more parameters than those that can be estimated, model parameters are not estimable. So it is not possible to make inferences for nonestimable functions of parameters. When the assumed model of matrix notation is reparameterized by the estimable functions in a basis set, it also discusses how to use projections for the estimation of estimable functions.

Keywords: Basis set, eigenvector, esitimable function, projection, projection matrix.

[1] Professor, Department of Statistics, Keimyung University, Daegu 704-701, Korea.
E-mail: jschoi@kmu.ac.kr