

最小 2 乗推定量について

渡 部 重 明

I

われわれは次のような構造方程式を想定する.

$$y = X\beta + \varepsilon \quad (1)$$

β はパラメーターのベクトルであり, y , X はそれぞれ観測可能な量のベクトルおよび行列である. ε は攪乱変数のベクトルで, それは次の仮定をみたすものとする.

仮定 1. ε は平均値 0 の確率変数である.

$$E(\varepsilon) = 0$$

仮定 2. ε の分散行列は定数正直行列である.

$$V(\varepsilon) = E\{(\varepsilon - E(\varepsilon))(\varepsilon - E(\varepsilon))'\}$$

$$= \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{n1} & \cdots & v_{nn} \end{bmatrix}$$

われわれはパラメーター β の線型関数 $\theta = \tau'\beta$ を推定したい. そこで, 今 y の線型関数 $t'y$ を考える. $E(t'y) = \theta = \tau'\beta$ なる t が存在するとき, θ は推定可能 (*estimable*) であるといい, このとき $\hat{\theta} = t'y$ を θ の不偏推定量 (*unbiased estimator*) という. さらに, $V(t'y) = \text{minimum}$ がみたされるとき, $\hat{\theta} = t'y$ を最小分散不偏推定量 (*minimum variance unbiased estimator*) という.

y は n 次元ベクトル, X は $n \times m$ の行列である.

仮定 3. X の階数は n である.

われわれは, 観測値 y が二つの部分から成ると想定して構造方程式を構

最小 2 乗 推 定 量 について

成した。逆にいえば、 y を二つの空間に分割する。 $[x_1, \dots, x_n]$ の張る空間を推測空間といい、その推測空間の直交補空間を誤差空間という。 $x_1, \dots, x_n$ の線型結合によって、 y をもっともよく説明するようにその係数を決定したい。

定理 1. (Rao) $\theta = \tau' \beta$ が推定可能であるための必要十分条件は、 $\tau = X't$ となるような t が存在することである。

〔証明〕 $\theta = \tau' \beta$ が推定可能であるとする、

$$\theta = \tau' \beta = E(t'y) = t'E(y) = t'X\beta$$

最後の等式は仮定 1 による。故に、

$$\tau = X't$$

逆に、

$$\tau = X't$$

とすると、

$$\tau' \beta = t'X\beta = E(t'y)$$

すなわち、 $\theta = \tau' \beta$ は推定可能である。

〔証了〕

この定理は、構造方程式に基づいてパラメーターを推定するための基礎を与える。

一般に、 $\hat{\theta}$ を θ の推定量とすると、 $E(\hat{\theta}) = \theta$ ならば、 $\hat{\theta}$ を θ の不偏推定量という。さて、 $\theta = \tau' \beta$ の不偏推定量 $\hat{\theta} = t'y$ の集合の中から、 $V(t'y) = \text{minimum}$ となるようなものを求める。

$$\begin{aligned} t'y &= t'(X\beta + \varepsilon) \\ &= t'X\beta + t'\varepsilon \end{aligned}$$

であるから、仮定 1 により、

$$E(t'y) = t'X\beta$$

したがって仮定 2 から、

$$\begin{aligned} V(t'y) &= t'E(\varepsilon\varepsilon')t \\ &= t'Vt \end{aligned}$$

最小 2 乗 推 定 量 に つ い て

がえられる。推定可能とし、ラグランジュ乗数のベクトル λ を導入すると、問題は次のようになる。

$$\min = t' V t - 2\lambda'(X't - \tau)$$

そこで上の式を t と λ に関して微分し、それが 0 に等しいと置くと、

$$Vt = X\lambda$$

$$\tau = X't$$

がえられる。すなわち、

$$t = V^{-1}X\lambda \quad (2)$$

$$\tau = XV^{-1}X\lambda \quad (3)$$

のように表現できる。

結局、 $\theta = \tau'\beta$ の最小分散不偏推定量は、

$$\begin{aligned} \hat{\theta} &= t'y \\ &= \lambda'X'V^{-1}y \end{aligned}$$

で与えられる。以上のことをまとめておく。

定理 2. (Rao) $\theta = \tau'\beta$ の最小分散不偏推定量は、

$$\begin{aligned} \hat{\theta} &= t'y \\ &= \lambda'X'V^{-1}y \end{aligned} \quad (4)$$

で与えられる。

$\hat{\theta}$ は次のように書くことができる。

$$\begin{aligned} \hat{\theta} &= \lambda'X'V^{-1}y \\ &= \lambda'X'V^{-1}X\beta + \lambda'X'V^{-1}\varepsilon \\ &= \tau'\beta + t'\varepsilon \end{aligned} \quad (5)$$

構造方程式 (1) を考慮すると、(4) 式から、 $\hat{\theta}$ は β と ε の関数になることは明らかであるが、(5) 式から β の影響は τ を通じて、 ε の影響は t を通じて及ぶ。さらに、 $\hat{\theta}$ の分散は λ のみの関数となる。なぜならば、

$$\begin{aligned} V(\hat{\theta}) &= t' V t \\ &= \lambda'X'V^{-1} V V^{-1}X\lambda \\ &= \lambda'XV^{-1}X\lambda \end{aligned}$$

最小 2 乗 推 定 量 に つ い て

$$= \tau / \lambda$$

となるからである.

II

部分空間と射影に関してよく知られた事実を以下に要約しよう.

R を n 次元空間とし, M をその部分空間とする. 部分空間とは, α, β を任意の実数とすると, 任意の $x, y \in M$ に対して, $\alpha x + \beta y \in M$ なることをいう. 任意の $x \in M$ に対して, $x'y = 0$ となるような y の集合を M の直交補空間といい $M^\perp$ で表わす.

定理 3. 任意の $z \in R$ は

$$z = x + y, \quad x \in M, \quad y \in M^\perp$$

と一意的に分解される.

上の分解における x を z の M への射影 (Projection) という. それを $Pz = x$ と表わし, この P を射影行列という.

定理 4. P が R の M への射影行列であるための必要十分条件は,

$$(1) P^2 = P \text{ (冪等)}$$

$$(2) P' = P \text{ (対称)}$$

なることである.

射影はノルムの最小性 $\min_x \|z - x\|$ によって特徴づけられる.

定理 5. $z \in R$ とする.

$$\|z - \hat{x}\| = \min_{x \in M} \|z - x\|$$

とすると, $Pz = \hat{x}$ である.

[証明] 定理 3 による z の分解を,

$$z = \hat{x} + \hat{y}, \quad \hat{x} \in M, \quad \hat{y} \in M^\perp$$

とする. すると,

$$(z - \hat{x}) \in M^\perp, \quad (\hat{x} - x) \in M$$

であるから,

$$\|z - x\| = \| (z - \hat{x}) + (\hat{x} - x) \|$$

最小 2 乗 推 定 量 について

$$\begin{aligned} &= \|z - \hat{x}\| + \|\hat{x} - x\| \\ &\geq \|z - \hat{x}\| \end{aligned}$$

となる。そして、 $\|z - x\|$ は $x = \hat{x}$ なるとき最小となる。分解の一意性から、

$$Pz = \hat{x}$$

は明らかである。

〔証了〕

定理 6. P を M への射影行列とすると、 $I - P$ は $M^\perp$ への射影行列である。

〔証明〕 P は射影行列であるから、

$$(1) (I - P)^2 = I - 2P + P^2$$

$$= I - 2P + P$$

$$= I - P$$

$$(2) (I - P)' = I' - P'$$

$$= I - P$$

したがって、定理 4 から $(I - P)$ が射影行列であることがわかる。さらに、

$$(I - P)z = z - Pz$$

$$= z - x$$

$$= y$$

であるから、 $(I - P)$ は R の $M^\perp$ への射影行列であることがわかる。

〔証了〕

さて、 $\{x_1, \dots, x_n\}$ の張る空間を M としよう。そのとき射影行列 P の具体的表現が知りたい。 $x_1, \dots, x_n$ を並べてできる行列を X とすると、 M の任意の要素 x は、

$$x = b_1 x_1 + \dots + b_n x_n$$

と表現できる。すなわち、

$$x = Xb$$

と書くことができる。定理 5 より、 b は $\min \|z - Xb\|$ なるように決定さ

れる.

$$\begin{aligned} d &= (z - Xb)'(z - Xb) \\ &= z'z - 2b'X'z + b'X'Xb \end{aligned}$$

とする.

$$\frac{\partial d}{\partial b} = 0$$

とすると,

$$b = (X'X)^{-1}X'z$$

となる. したがって,

$$Pz = X(X'X)^{-1}X'z$$

故に,

$$P = X(X'X)^{-1}X'$$

なることがわかる. これが R の M への射影行列の具体的表現である. このことは定理 4 に訴えて直接確かめることもできる, 事実,

$$\begin{aligned} (1) \quad P^2 &= X(X'X)^{-1}X'X(X'X)^{-1}X' \\ &= X(X'X)^{-1}X' \end{aligned}$$

$$(2) \quad P' = X(X'X)^{-1}X' = P$$

なることから明らかであろう.

$$\min (z - Xb)'(z - Xb)$$

なるように決定される b を最小 2 乗推定量という, またその方法を最小 2 乗法という. すなわち, 最小 2 乗法とは, $[x_1, \dots, x_n]$ の張る空間への射影に他ならない. 構造方程式 (1) に最小 2 乗法を適用すると, その最小 2 乗法推定量は,

$$b = (X'X)^{-1}X'y \quad (6)$$

で与えられる. そして, $x \in M$ の表現 Xb の b に上の最小 2 乗推定量を代入したもの, すなわち, $X(X'X)^{-1}X'y$ は y の M への射影である.

さらに, 構造方程式 (1) に関して定理 6 を考慮すると,

$$(I - X(X'X)^{-1}X')$$

最小 2 乗 推 定 量 に つ い て

は $M^\perp$ への射影行列である. すなわち, これは誤差空間への射影行列である. 事実,

$$\begin{aligned} & (I - X(X'X)^{-1}X')y \\ &= y - X(X'X)^{-1}X'y \\ &= y - Xb \end{aligned}$$

であるから, 上のことは明らかである. $(I - X(X'X)^{-1}X')$ が射影行列なることは, 定理 4 を考慮すれば, 直接簡単に確かめることもできる.

III

さきに述べた最小 2 乗法に対して, $\min (y - Xb)'V^{-1}(y - Xb)$ なるように b を決定する方法を一般化された最小 2 乗法といい, そのようにして求められた b を一般化された最小 2 乗推定量という.

$$\begin{aligned} Q &= (y - Xb)'V^{-1}(y - Xb) \\ &= y'V^{-1}y - 2b'XV^{-1}y + b'X'V^{-1}Xb \end{aligned}$$

とする. Q を b で微分して 0 と置く.

$$2X'V^{-1}y = 2X'V^{-1}Xb$$

$$\therefore b = (X'V^{-1}X)^{-1}X'V^{-1}y \quad (7)$$

これが一般化された最小 2 乗推定量である.

定理 7. $\theta = \varepsilon'\beta$ の最小分散不偏推定量は $\hat{\theta} = \varepsilon'b$ で与えられる. b は一般化された最小 2 乗推定量である.

〔証明〕 定理 2 の (4) から, $\theta = \varepsilon'\beta$ の最小分散不偏推定量は,

$$\hat{\theta} = \lambda'X'V^{-1}y$$

で与えられる. λ はラグランジュ乗数のベクトルである.

$$\begin{aligned} \hat{\theta} &= \lambda'X'V^{-1}y \\ &= \lambda'X'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y \\ &= [\lambda'X'V^{-1}X][(X'V^{-1}X)^{-1}X'V^{-1}y] \end{aligned}$$

であるから, (3) と (7) により,

$$\hat{\theta} = \varepsilon'b$$

となる.

〔証了〕

最小 2 乗 推 定 量 に つ い て

構造方程式 (I) を考慮すると,

$$\begin{aligned} b &= (X' V^{-1} X)^{-1} X' V^{-1} y \\ &= (X' V^{-1} X)^{-1} X' V^{-1} (X\beta + \varepsilon) \\ &= \beta + (X' V^{-1} X)^{-1} X' V^{-1} \varepsilon \end{aligned}$$

$$\therefore E(b) = \beta$$

b は確かに β の不偏推定量である. b の分散行列は次のようになる.

$$\begin{aligned} V(b) &= E[(b - \beta)(b - \beta)'] \\ &= E[(X' V^{-1} X)^{-1} X' V^{-1} \varepsilon \varepsilon' V^{-1} X (X' V^{-1} X)^{-1}] \\ &= (X' V^{-1} X)^{-1} X' V^{-1} E(\varepsilon \varepsilon') V^{-1} X (X' V^{-1} X)^{-1} \\ &= (X' V^{-1} X)^{-1} \quad (8) \end{aligned}$$

次の関係式は明らかであろう.

$$\begin{aligned} y - Xb &= y - X(X' V^{-1} X)^{-1} X' V^{-1} y \\ &= [I - X(X' V^{-1} X)^{-1} X' V^{-1}] y \end{aligned}$$

であるから,

$$\begin{aligned} X' V^{-1} (y - Xb) &= X' V^{-1} [I - X(X' V^{-1} X)^{-1} X' V^{-1}] y \\ &= 0 \end{aligned}$$

さらに, 次の関係式から, 最小 2 乗推定量が誤差を最小ならしめるということがわかる.

$$\begin{aligned} \varepsilon' V^{-1} \varepsilon &= (y - X\hat{\beta})' V^{-1} (y - X\hat{\beta}) \\ &= [(y - Xb) + X(b - \hat{\beta})]' V^{-1} [(y - Xb) + X(b - \hat{\beta})] \\ &= (y - Xb)' V^{-1} (y - Xb) \\ &\quad + 2(b - \hat{\beta})' X' V^{-1} (y - Xb) \\ &\quad + [X(b - \hat{\beta})]' V^{-1} [X(b - \hat{\beta})] \\ &= (y - Xb)' V^{-1} (y - Xb) \\ &\quad + [X(b - \hat{\beta})]' V^{-1} [X(b - \hat{\beta})] \\ &\geq (y - Xb)' V^{-1} (y - Xb) \end{aligned}$$

最後の不等式は, 仮定 2 により V が正値行列であり, したがって V^{-1} が正値行列なることによる. なぜならば, 一般に, $x = Vy$ とすると,

最小 2 乗推定量について

$$\begin{aligned} X'V^{-1}x &= y'VV^{-1}Vy \\ &= y'Vy \geq 0 \end{aligned}$$

であるから、

β の不偏推定量は一般に y の線型関数である。それを、 $b_1 = Ay$ としよう。

$$\begin{aligned} E(b_1) &= E\{A(X\beta + \varepsilon)\} \\ &= AX\beta + AE(\varepsilon) \\ &= AX\beta \end{aligned}$$

であるが、 b_1 は β の不偏推定量であるから、

$$AX = I$$

でなければならない。これはまた十分条件でもある。すなわち、 $b_1 = Ay$ が β の不偏推定量であるための必要十分条件は、 $AX = I$ である。 b_1 の分散に関して次の不等式が成立する。

$$\begin{aligned} V(b_1) &= AE(\varepsilon\varepsilon')A' \\ &= AVA' \\ &= (X'V^{-1}X)^{-1} \\ &\quad + [A - (X'V^{-1}X)^{-1}X'V^{-1}]V[A - (X'V^{-1}X)^{-1}X'V^{-1}]' \\ &\geq (X'V^{-1}X)^{-1} = V(b) \end{aligned}$$

途中の不等式は V が正値行列なることに基づく。上の関係式から、一般化された最小 2 乗推定量は、 β の不偏推定量の中で分散がもっとも小さい推定量である。これは定理 7 の別証明である。

IV

最小 2 乗推定量はある意味の構造の不変性をもっている。それを以下に述べよう。簡単のために、 $V = \sigma^2 I$ と仮定する。さらに $C = X'X$, $d = X'y$ と表わすことにする。 β を次のように分割する。

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

それに対応して行列を分割する。

最小 2 乗推定量について

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \quad C^{-1} = \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$$

今の場合，最小 2 乗推定量の分散は， $V(b) = \sigma^2 C^{-1}$ である．そこで次の定理がなりたつことをみる．

定理 8. 最小 2 乗推定量 $b' = (b_1', b_2')$ から b_1 を除去した場合， b_2 の分散行列は，

$$V(b_2) = \sigma^2 C^{22}$$

となる．これはもとのシステムをそのまま縮小したものに他ならない．

〔証明〕正規方程式を分割した形で書くと，

$$C_{11}b_1 + C_{12}b_2 = d_1$$

$$C_{21}b_1 + C_{22}b_2 = d_2$$

この方程式から b_1 を消去すると，

$$(C_{22} - C_{21}C_{11}^{-1}C_{12})b_2 = d_2 - C_{21}C_{11}^{-1}d_1 \quad (9)$$

定義から明らかなように，

$$C_{22} - C_{21}C_{11}^{-1}C_{12} = (C^{22})^{-1}$$

であるから， $(C_{22} - C_{21}C_{11}^{-1}C_{12})$ は正値行列である．また，

$$d = X'y = X'(X\beta + \varepsilon) = C\beta + X'\varepsilon$$

$$\therefore E(d) = C\beta$$

したがって，

$$d_1 - E(d_1) = X_{11}'\varepsilon_1 + X_{21}'\varepsilon_2$$

$$d_2 - E(d_2) = X_{12}'\varepsilon_1 + X_{22}'\varepsilon_2$$

さらに定義により，

$$\begin{aligned} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{11} \end{bmatrix} &= \begin{bmatrix} X_{11}' & X_{21}' \\ X_{12}' & X_{22}' \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \\ &= \begin{bmatrix} X_{11}'X_{11} + X_{21}'X_{21} & X_{11}'X_{12} + X_{21}'X_{22} \\ X_{12}'X_{11} + X_{22}'X_{21} & X_{12}'X_{12} + X_{22}'X_{22} \end{bmatrix} \end{aligned}$$

以上の関係式を考慮すると，

最小 2 乗 推 定 量 に つ い て

$$\begin{aligned}
 & V(d_2 - C_{21}C_{11}^{-1}d_1) \\
 &= E\{X_{12}'\varepsilon_1 + X_{22}'\varepsilon_2 - C_{21}C_{11}^{-1}(X_{11}'\varepsilon_1 + X_{21}'\varepsilon_2)\} \\
 &\quad \cdot \{\varepsilon_1'X_{12} + \varepsilon_2'X_{22} - (\varepsilon_1'X_{11} + \varepsilon_2'X_{21})C_{11}^{-1}C_{12}\}' \\
 &= \sigma^2\{C_{22} - C_{21}C_{11}^{-1}C_{12} - C_{21}C_{11}^{-1}C_{12} + C_{21}C_{11}^{-1}C_{11}C_{11}^{-1}C_{12}\} \\
 &= \sigma^2\{C_{22} - C_{21}C_{11}^{-1}C_{12}\}
 \end{aligned}$$

となる. これと(9)式とから,

$$\begin{aligned}
 & (C_{22} - C_{21}C_{11}^{-1}C_{12})V(b_2)(C_{22} - C_{21}C_{11}^{-1}C_{12}) \\
 &= V(d_2 - C_{21}C_{11}^{-1}d_1) \\
 &= \sigma^2(C_{22} - C_{11}^{-1}C_{12}) \\
 \therefore V(b_2) &= \sigma^2(C_{22} - C_{21}C_{11}^{-1}C_{12})^{-1} \\
 &= \sigma^2(C^{22})^{-1}
 \end{aligned}$$

〔証了〕

文 献

1. Halmos, P.R., "Introduction to Hilbert Space", (Chelsea, 1957)
2. Plackett, R. L., A Historical note on the the method of Least Squares, *Biometrika*, 36 (1949)
3. Plackett, R.L., "Principles of Regression Analysis" (Oxford University Press, 1960)
4. Saaty, T. L. and J.Bram, "Nonlinear Mathematics" (McGraw-Hill, 1964)
5. Rao, C. R., Generalization of Markoff's Theorem and Tests of Linear Hypotheses, *Sankhya*, 7 (1945)