2010年度 エコノメトリックスII & 上級エコノメトリックスII 第5回講義メモ(訂正版)

2010年10月29日

§3. 最小2乗推定量の性質

§3.1 統計量の基本概念

OLS 推定量など推定量は標本データ $(y_1, \cdots, y_n)$ の関数として表される。一般に標本データの関数 $T(\boldsymbol{y}) = T(y_1, \cdots . y_n)$ を $\boldsymbol{y}$ に基づく統計量と呼ぶ。

統計量に関しては次の十分統計量（sufficient statistic）という概念が重要である。

定義 3.1.1

いま、$y_i \sim i.i.d F(\theta),\ (i = 1, \ldots, n) \quad \theta \in \Theta$ とする。このとき、統計量 $T(\boldsymbol{y})$ がパラメータ θ の十分統計量であるとは T を条件としたときの $\boldsymbol{y} = (y_1, \cdots, y_n)'$ の条件付分布が θ に依存しないことである。

（例）離散型確率変数の場合

$y_i \sim i.i.d.\, Bin(1, p)$ $(i = 1, \ldots, n)$ とする。このとき、$y_1, \cdots, y_n$ の同時分布は

$$\mathrm{P}(y_1, \cdots, y_n) \text{ は} = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

で与えられる。

他方、$x \equiv n\bar{y} = \Sigma y_i$ の分布は、$x \sim Bin(n, p)$ より

$$\mathrm{P}(x) = {}_nC_x p^x (1-p)^{n-x}$$

となる。よって

$$\mathrm{P}(y_1, \cdots, y_n \mid x) = \frac{1}{{}_nC_x}$$

が得られ、パラメータ p に依存しないことがわかる。ゆえに x（または $\bar{y}$）は p の十分統計量である。

定理 3.1.2 分解定理

$y_1, \cdots, y_n$ を確率変数とし、$f(\boldsymbol{y}, \theta)$ を $y_1, \cdots, y_n$ の確率関数（確率密度関数）とする。このとき $T(\boldsymbol{y}) = T(y_1, \cdots, y_n)$ が θ の十分統計量であるための必要十分条件は、θ を含まない $\boldsymbol{y}$ の関数 $h(\boldsymbol{y})$ を用いて

$$f(\boldsymbol{y}, \theta) = g(T, \theta) h(\boldsymbol{y})$$

の形に分解できることである。

（例１）先程の $y_i \sim Bin(1, p)$ の例では $h(\boldsymbol{y}) = 1$(または $1/{}_nC_x$)。

（例２）$y_i \sim i.i.d.\, Poisson(\lambda)$ $(i = 1, \ldots, n)$ とする。このとき

$$f(y_1, \cdots, y_n, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{\Sigma y_i} e^{-n\lambda}}{\left(\prod_{i=1}^{n} (y_i!)\right)}$$

よって、$\bar{y} = \dfrac{1}{n} \sum_{i=1}^{n} y_i$ とすると

$$\begin{aligned} g(\bar{y}, \lambda) &= (\lambda^{\bar{y}} e^{-\lambda})^n \\ h(y_1, \cdots, y_n) &= \left(\prod_{i=1}^{n} (y_i!)\right)^{-1} \end{aligned}$$

によって

$$f(y_1, \cdots, y_n; \lambda) = g(\bar{y}, \lambda) h(y_1, \cdots, y_n)$$

と表されるので、$\bar{y}$ は λ の十分統計量である。

十分統計量に関しては次の Rao=Blackwell 定理が重要な性質を与えている。

定理 3.1.3 Rao=Blackwell 定理

$\hat{\theta}(y)$ を θ の推定量とし、$\theta^*(T)$ を十分統計量 $T = T(y)$ を条件としたときの $\hat{\theta}(y)$ の条件付期待値

$$\theta^*(T) = \mathrm{E}[\hat{\theta}(y) \mid T]$$

とする。このとき平均２乗誤差

$$\mathrm{MSE}(\hat{\theta}(y)) = \mathrm{E}[(\hat{\theta}(y) - \theta)^2]$$

をリスク関数とすると

$$\mathrm{E}[(\theta^*(T) - \theta)^2] \leq \mathrm{E}[(\hat{\theta}(y) - \theta)^2]$$

であり、等号成立は

$$\mathrm{P}(\hat{\theta}(y) = \theta^*(T)) = 1$$

のときのみである。

これより、$\hat{\theta}(y) = \theta^*(T)$ のケースを除いて $\theta^*(T)$ は小さいリスクをもつことがいえる。また $\hat{\theta}(y)$ が不偏推定量であるときには平均 2 乗誤差は分散に等しくなるので、$\theta^*(T)$ は最小分散を与えることになる。

§3.2 推定量の「良さ」とは

数理統計学では、推定量の「良さ」の基準として、以下の概念が用いられる。

定義 3.2.1

いま、$\boldsymbol{\theta}$ を $k \times 1$ のパラメータ (母数) ベクトル、$\hat{\boldsymbol{\theta}}_n$ をサンプルサイズ n の標本から求めた $\boldsymbol{\theta}$ の推定量とする。このとき、不偏性 (unbiasness)、一致性 (consistency)、効率性 (efficiency) は次のように定義される。

a). 不偏性　$\mathrm{E}(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\theta} \quad \forall_n$

b). 一致性　$\mathrm{plim}_{n\to\infty}\, \hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} \ \Big(\text{or } \lim_{n\to\infty} \mathrm{P}(\| \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \| > \epsilon) = 0 \qquad {}^\forall \epsilon > 0\Big)$

c). 効率性　$\tilde{\Theta}$ を不偏推定量の集合とする。もし、$\hat{\boldsymbol{\theta}}_n \in \tilde{\Theta}$, $\tilde{\boldsymbol{\theta}}_n \in \tilde{\Theta}$ であるとき、以下は同値。

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}_n) \le \mathrm{Var}(\tilde{\boldsymbol{\theta}}_n) \quad \Longleftrightarrow \quad \hat{\boldsymbol{\theta}}_n \text{は効率性を有する。}$$

$\boldsymbol{\beta}$ の OLS 推定量 $\hat{\boldsymbol{\beta}}$ はこの 3 つの基準から「良い」推定量であることが言える。以下では、詳細に見てゆくことにしよう。

§3.3 不偏性

いま、攪乱項 $\boldsymbol{u}$ について、$\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{X}) = \boldsymbol{0}$ を仮定する。OLS 推定量は、

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}$$

で与えられるから、

$$\mathrm{E}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \boldsymbol{\beta} + \underbrace{\mathrm{E}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \mid \boldsymbol{X})}_{\boldsymbol{0}} = \boldsymbol{\beta}$$

となり、不偏推定量であることが確認できる。

OLS 推定量 $\hat{\boldsymbol{\beta}}$ が偏りのある推定量となるケース (biased case) として、説明変数に従属変数のラグを含む場合を挙げることができる。いま、モデルを

$$y_t = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 y_{t-1} + u_t, \quad u_t \sim IID(0, \sigma^2) \quad (t = 1, \ldots, n)$$

としよう。このとき、$\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{y}_{-1}) = \boldsymbol{0}$ が成り立つだろうか。（ここで、$\boldsymbol{y}_{-1} = (y_0, y_1, \cdots, y_{n-1})'$ である。）

$$\mathrm{E}(u_t \mid y_{t-1}) = \mathrm{E}(u_t \mid \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 y_{t-2} + u_{t-1}) = 0$$

であるが、

$$\mathrm{E}(u_{t-1} \mid y_{t-1}) = \mathrm{E}(u_{t-1} \mid \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 y_{t-2} + u_{t-1}) = u_{t-1} \neq 0$$

になるため、

$$\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{y}_{-1}) \neq \boldsymbol{0}$$

が成立する。これより、$\mathrm{E}\left(\hat{\boldsymbol{\beta}} | \boldsymbol{y}_{-1}\right) \neq \boldsymbol{\beta}$ になる。

§3.4 一致性

次に、$\hat{\boldsymbol{\beta}}$ が一致推定量であることを確認しよう。$\hat{\boldsymbol{\beta}}$ がサンプルサイズ n からの推定量であることを強調するために、$\hat{\boldsymbol{\beta}}_n$ と表記しよう。

$$\hat{\boldsymbol{\beta}}_n = \underbrace{\boldsymbol{\beta}}_{const} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}$$

であるから、両辺の確率極限をとり、また

$$\begin{aligned} \operatorname*{plim}_{n\to\infty} A_n B_n &= \operatorname*{plim}_{n\to\infty} A_n \operatorname*{plim}_{n\to\infty} B_n \\ \operatorname*{plim}_{n\to\infty} \frac{A_n}{B_n} &= \frac{\mathrm{plim}_{n\to\infty} A_n}{\mathrm{plim}_{n\to\infty} B_n}, \qquad \text{when} \quad \operatorname*{plim}_{n\to\infty} B_n \neq 0 \end{aligned}$$

となることを用いると、

$$\begin{aligned} \operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\beta}}_n &= \boldsymbol{\beta} + \operatorname*{plim}_{n\to\infty} (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\ &= \boldsymbol{\beta} + \left(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \times \operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}'\boldsymbol{u} \end{aligned}$$

になるが、

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} = \underbrace{\boldsymbol{S}_{\boldsymbol{X}'\boldsymbol{X}}}_{\text{moment Matrix}} < +\infty$$

となることと、撹乱項に関する仮定 $\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{X}) = \boldsymbol{0}$ から、

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}'\boldsymbol{u} = \boldsymbol{0}$$

が成立するので、$\mathrm{plim}_{n\to\infty} (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} = \boldsymbol{0}$ になり、$\mathrm{plim}_{n\to\infty} \hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}$ が得られる。

§3.5 効率性

2 次形式 (quadratic form)

　効率性は推定量の分散に関する概念である。$\hat{\beta}$ の分散（分散共分散行列）は正方行列であり、かつ対称行列になっている。

　対称行列 $\boldsymbol{A}$ と $\boldsymbol{x} \neq \boldsymbol{0}$ に対して、$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ を 2 次形式とよぶ。簡単な例をみてみよう。いま

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \qquad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

とする。

$$\begin{aligned} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 & a_{12}x_1 + a_{22}x_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{12}x_1x_2 + a_{22}x_2^2 = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 \end{aligned}$$

このように、2 次形式は $\boldsymbol{x}$ の成分の 2 次式になっていることが分かる。

　2 次形式の値によって対称行列 $\boldsymbol{A}$ の性質を以下のように分類することができる。

定義 3.5.1

　いま $\boldsymbol{A}$ を $k \times k$ の実対称行列（実数値を要素とする対称行列）とする。$\forall \boldsymbol{x} \neq \boldsymbol{0}$ について、

$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ ならば $\boldsymbol{A}$ は正値定符号 (positive definite) 行列であるという。

$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ ならば $\boldsymbol{A}$ は非負値定符号 (positive semi-definite) 行列であるという。

$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$ ならば $\boldsymbol{A}$ は負値定符号 (negative definite) 行列であるという。

$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq 0$ ならば $\boldsymbol{A}$ は非正値定符号 (negative semi-definite) 行列であるという。

Note: **行列の固有値と 2 次形式の関係**

結果のみ述べると、$k \times k$ の実対称行列 $\boldsymbol{A}$ は適当な直交行列 $\boldsymbol{\Gamma}$ によって

$$\boldsymbol{A} = \boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}' \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_k \end{pmatrix} \boldsymbol{\Gamma}$$

に分解できる。ここで、λ_i $(i = 1, \ldots, k)$ は $\boldsymbol{A}$ の固有値である。

　いま、$\boldsymbol{A}$ の固有値のうち、正の個数を ℓ、負の個数を m とすると、

$$\begin{aligned} m = 0 &\iff \boldsymbol{A}\text{は非負値定符号} \\ \ell = k &\iff \boldsymbol{A}\text{は正値定符号} \end{aligned}$$

という関係が成立する。

OLSE の分散

では、$\hat{\boldsymbol{\beta}}$ の条件付分散を考えよう。

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}$$

であるから、

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= \mathrm{E}\Big[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mid \boldsymbol{X}\Big] \\ &= \mathrm{E}\Big[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \mid \boldsymbol{X}\Big] \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\underbrace{\mathrm{E}(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X})}_{\sigma^2\boldsymbol{I}}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\end{aligned}$$

ところで、

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}$$

であったから、

$$\boldsymbol{X}'\boldsymbol{X} = \overset{k\times k}{\sum_{t=1}^{n} \overset{k\times 1}{\boldsymbol{x}_t}\,\overset{1\times k}{\boldsymbol{x}_t'}}$$

と表すことができる。これより

$$\boldsymbol{X}'\boldsymbol{X} \to \infty \quad \text{as } n \to \infty$$

であることは容易に分かる。すなわち $\boldsymbol{X}'\boldsymbol{X} = O_p(n)$ である。

したがって、

$$\frac{1}{n}\boldsymbol{X}'\boldsymbol{X} \to \boldsymbol{S}_{\boldsymbol{X}'\boldsymbol{X}} < \infty$$

が得られるので、

$$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \frac{1}{n}\sigma^2 \times (\frac{1}{n}\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{1}{n}\sigma^2 S^{-1}_{\boldsymbol{X}'\boldsymbol{X}} \tag{3.5.1}$$

と表現できる。(3.4.1) 式より、$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})$ が (1) サンプルサイズ n、(2) 攪乱項の分散 σ^2、(3) $\boldsymbol{X}$ (non-stochastic の場合) または $\boldsymbol{S}_{\boldsymbol{X}'\boldsymbol{X}}$ (stochastic の場合)、に依存することが分かる。

定理 3.5.2 Gauss=Markov 定理

$\boldsymbol{\beta}$ の OLS 推定量 $\hat{\boldsymbol{\beta}}$ は線形不偏推定量の中で分散の下限になっている。

(証明のスケッチ)

$\boldsymbol{\beta}$ の線形不偏推定量を $\tilde{\boldsymbol{\beta}}$ とする。$\tilde{\boldsymbol{\beta}}$ は線形推定量であるから、

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{Z}\boldsymbol{y}, \qquad \text{where} \quad \boldsymbol{Z} = (\boldsymbol{D} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$$

で表現できる。次に、$\widetilde{\boldsymbol{\beta}}$ の期待値をとると、不偏推定量であるから、

$$\mathrm{E}(\widetilde{\boldsymbol{\beta}}) = \mathrm{E}(\boldsymbol{Dy}) + \mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

が成立しなければならない。$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ であるので、

$$\mathrm{E}(\boldsymbol{Dy}) = \mathrm{E}(\boldsymbol{DX})\boldsymbol{\beta} + \mathrm{E}(\boldsymbol{Du}) = \mathbf{0}$$

を満たしている必要がある。したがって、$\boldsymbol{DX} = \boldsymbol{O}$ でなくてはならない。
　この関係を使うと、

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} &= \boldsymbol{Zy} - \boldsymbol{\beta} \;=\; \boldsymbol{Dy} + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= \boldsymbol{DX\beta} + \boldsymbol{Du} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\
&= \boldsymbol{Du} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}
\end{aligned}$$

になるので、$\widetilde{\boldsymbol{\beta}}$ の分散は、

$$\begin{aligned}
\mathrm{Var}(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{D}, \boldsymbol{X}) &= \mathrm{E}\left[(\boldsymbol{Du} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u})(\boldsymbol{Du} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u})' \mid \boldsymbol{DX})\right] \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} + \sigma^2\boldsymbol{DD}'
\end{aligned}$$

として求められる。これより、$\mathrm{Var}(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{D}, \boldsymbol{X})$ が最小となるのは $\boldsymbol{D} = \boldsymbol{O}$ のときであることがわかる。
　というのも、

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{d}_1' \\ \vdots \\ \boldsymbol{d}_n' \end{pmatrix}$$

とおくと、$\boldsymbol{DD}'$ の対角要素が、

$$(\boldsymbol{DD}')_{ii} = \boldsymbol{d}_i'\boldsymbol{d}_i \geq 0$$

のように非負値を取るためである。 ∎