

線型推測論

第04回 ガウス-マルコフの定理と最小二乗推定量(2)

2018/4/13

統計数理研究所

長島 健悟

前回のまとめ

- BLUEは良い推定量と考えられる
- 秤量問題の例の様に, 線形推定量の制約付き最適化によってBLUEを求めることができた
- しかし, Lagrangeの未定乗数法で解くのは結構面倒である...
- もっとよい方法はないだろうか?

別解法

- 最小二乗法
 - BLUEを求める単純なアルゴリズム
- ガウス-マルコフの定理
 - 誤差の三条件のもとで, 最小二乗推定量はBLUEである

最小二乗法

- 残差平方和を最小化するパラメータを求めるアルゴリズム
- 残差： $\hat{\epsilon}_i = y_i - \hat{E}[Y_i]$
 - $\hat{E}[Y_i]$ は $E[Y_i]$ の推定量をあらわす
- 残差平方和： $Q = \sum_{i=1}^n \hat{\epsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{E}[Y_i])^2$
 - 観測値と期待値の差が小さくなるほど残差平方和は小さい

単回帰モデルの場合

- モデル式： $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- Y_i の期待値： $E[Y_i] = \beta_0 + \beta_1 x_i$
- $E[Y_i]$ の推定量： $\hat{E}[Y_i] = b_0 + b_1 x_i$ 無数にある
- 残差平方和
 - $Q = \sum_{i=1}^n \{y_i - (b_0 + b_1 x_i)\}^2$
 - これを最小化する b_0, b_1 を求めれば、その値は最小二乗推定量 $\hat{\beta}_0, \hat{\beta}_1$ である

単回帰モデルの場合

- Q は下に凸なので, b_0, b_1 で偏微分した式 = 0の解が Q を最小化する値になる

- $$\frac{\partial Q}{\partial b_0} = -2 \sum_{i=1}^n \{y_i - (b_0 + b_1 x_i)\}$$

- $$\frac{\partial Q}{\partial b_1} = -2 \sum_{i=1}^n \{y_i - (b_0 + b_1 x_i)\} x_i$$

- $$-2 \sum_{i=1}^n \{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\} = 0$$

- $$-2 \sum_{i=1}^n \{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\} x_i = 0$$

単回帰モデルの場合

- $\Leftrightarrow \sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n x_i = 0$
- $\sum_{i=1}^n y_i x_i - \hat{\beta}_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 = 0$
- $\Leftrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - n^{-1}(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{\sum_{i=1}^n x_i^2 - n^{-1}(\sum_{i=1}^n x_i)^2}$
- $\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i}{n}$
- ちょっと面倒だが, 制約式を個別に考えなくても計算できる

最小二乗推定量の行列表現

- 次は, 一般線型モデル $Y = X\beta + \epsilon$ の場合の最小二乗推定量を考える
- 行列のまま解いてしまえば, 全ての一般線型モデルにおける最小二乗推定量の一般形がわかる
- 推定量の候補は $\mathbf{b} = (b_1, \dots, b_p)'$ と書く

練習問題

- 行列を用いて残差平方和 Q を定義せよ：

$$Q = ?$$

- $Q = (\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb})$

- 内積を思い出そう

- $\mathbf{a} = (a_1, a_2, a_3)'$

- $\mathbf{a}'\mathbf{a} = (a_1 \quad a_2 \quad a_3) \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \sum_{i=1}^3 a_i^2$

練習問題

• Q を**b**で偏微分せよ： $\frac{\partial Q}{\partial \mathbf{b}} = ?$

$$\bullet \frac{\partial Q}{\partial \mathbf{b}} = \frac{\partial \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}$$

$$\bullet \frac{\partial \mathbf{Y}'\mathbf{Y}}{\partial \mathbf{b}} = \mathbf{0}$$

$$\bullet \frac{\partial -2\mathbf{Y}'\mathbf{X}\mathbf{b}}{\partial \mathbf{b}} = -2 \frac{\partial \mathbf{b}'\mathbf{X}'\mathbf{Y}}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{Y}$$

$$\bullet \frac{\partial \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}}{\partial \mathbf{b}} = \frac{\partial \mathbf{b}'}{\partial \mathbf{b}} \mathbf{X}'\mathbf{X}\mathbf{b} + \frac{\partial \mathbf{b}'}{\partial \mathbf{b}} \mathbf{X}'\mathbf{X}\mathbf{b} = 2\mathbf{X}'\mathbf{X}\mathbf{b}$$

練習問題

- $\frac{\partial Q}{\partial \mathbf{b}} = 0$ を解いて最小二乗推定量の一般型 $\hat{\boldsymbol{\beta}}$ を求めよ
 - $-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$
 - $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$: 正規方程式
 - $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

おさらい：誤差の三条件

- $\epsilon_1, \dots, \epsilon_n$ は互いに無相関で何らかの分布に従う確率変数
- $E[\boldsymbol{\epsilon}] = \mathbf{0}$
- $\text{Var}[\epsilon_i] = \sigma^2 < \infty$
 - 系3-1 : $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$
 - 系3-2 : $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}$

最小二乗推定量の性質

- 最小二乗推定量の期待値 $E[\hat{\beta}]$ を求めよ

$$\begin{aligned} E[\hat{\beta}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta \\ &= \beta \end{aligned}$$

最小二乗推定量の性質

- 最小二乗推定量の分散 $\text{Cov}[\hat{\beta}]$ を求めよ

$$\begin{aligned}\text{Cov}[\hat{\beta}] &= \text{Cov}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Cov}[\mathbf{Y}]\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}' \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}\{(\mathbf{X}'\mathbf{X})^{-1}\}' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

最小二乗推定量のまとめ

- 一般線型モデル $Y = X\beta + \epsilon$ において
 - 最小二乗推定量は一般型
 $\hat{\beta} = (X'X)^{-1}X'Y$ を持つ
 - 最小二乗推定量は不偏推定量である
 $E[\hat{\beta}] = \beta$
 - 最小二乗推定量の分散は
 $\text{Cov}[\hat{\beta}] = \sigma^2(X'X)^{-1}$ である

ガウス-マルコフの定理

- 誤差の三条件のもとで, 最小二乗推定量はBLUEである
- ちょっと大変だが, これを証明してゆく
 - β の線型推定量を \mathbf{CY} とおく
 - $\mathbf{C} : p \times n$
 - 系3-1 : $E[\mathbf{Y}] = \mathbf{X}\beta$
 - 系3-2 : $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}$

不偏な線型推定量の条件

- $E[CY] = \beta$ を満たす必要がある
 - $E[CY] = CX\beta$
 - $CX\beta = \beta$ でなければならない
 - つまり $CX = I$ を満たす必要がある

最良な線型不偏推定量の条件

- $CX = I$ をみたし $\text{Cov}[CY]$ が最小となるような C を求めればよい
- $\text{Cov}[CY] = \sigma^2 CC'$
- 推定量の分散は CC' の対角要素だけに依存する; 制約の下で最小の CC' を求める

最良な線型不偏推定量の条件

$$\begin{aligned} \mathbf{CC}' &= [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}]' \\ &= [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}]' + [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\ &\quad + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}]' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

- $[\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} = \mathbf{CX} (\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$
($\because \mathbf{CX} = \mathbf{I}$)
- $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}]' = \mathbf{0}$

$$\mathbf{CC}' = [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] [\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}]' + (\mathbf{X}'\mathbf{X})^{-1}$$

最良な線型不偏推定量の条件

- $CC' = [C - (X'X)^{-1}X'] [C - (X'X)^{-1}X']' + (X'X)^{-1}$
- ここで, $[C - (X'X)^{-1}X'] [C - (X'X)^{-1}X']'$ の対角要素はかならず0以上になる
- $(X'X)^{-1}$ は定数である
- $[C - (X'X)^{-1}X'] [C - (X'X)^{-1}X']'$ の対角要素が0ならば最小値を取る
- $C = (X'X)^{-1}X'$ のとき $[C - (X'X)^{-1}X'] [C - (X'X)^{-1}X']' = 0$

最良線型不偏推定量

- 以上より
 - $CY = (X'X)^{-1}X'Y$ のときBLUEになる
 - これは最小二乗推定量に等しい

- 系4-1
 - $\hat{\beta}$ を最小二乗推定量とする
 - $a'\beta$ の推定量 $a'\hat{\beta}$ はBLUEである