通常最小二乗法 (Ordinary Least Squares: OLS)

T.Chiba

2020 年 5 月 24 日

概要

最小二乗法についてみていく.

1　最小二乗法 (1 変量)

本題に入る前に, リハビリもかねて 1 変量の最小二乗法を確認する. 線形回帰モデルは下記のような構造をしていると考えられる.

$$y = \alpha + \beta x + \varepsilon \tag{1}$$

x, y についての観測値 $\{(x_i, y_i) | i = 1, 2, \ldots, n\}$ がそれぞれ次のような構造を持っていると仮定する.

$$y_i = \alpha + \beta x_i + \varepsilon_i, i = 1, \ldots, n \tag{2}$$

誤差 ε の平方和を考える.

$$Q(\alpha, \beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \{y_i - (\alpha + \beta x_i)\}^2 \tag{3}$$

これを. 偏差平方和と呼ぶ.$Q(\alpha, \beta)$ を最小にするような $\alpha = \hat{\alpha}, \beta = \hat{\beta}$ を最小二乗推定量 (Least Squares Estimator: LSE) と呼ぶ.

$$\begin{cases} \frac{\partial Q(\alpha,\beta)}{\partial \alpha} = 0 \\ \frac{\partial Q(\alpha,\beta)}{\partial \beta} = 0 \end{cases} \tag{4}$$

を満たすような α, β を考える. 2 階の偏微分まで計算すると,

$$\frac{\partial Q}{\partial \alpha} = -2 \sum_{i=1}^{n} (y_i - \alpha - \beta x_i) \tag{5}$$

$$\frac{\partial^2 Q}{\partial \alpha^2} = 2n > 0 \tag{6}$$

$$\frac{\partial Q}{\partial \beta} = -2 \sum_{i=1}^{n} x_i (y_i - \alpha - \beta x_i) \tag{7}$$

$$\frac{\partial^2 Q}{\partial \beta^2} = 2 \sum_{i=1}^{n} x_i^2 >= 0 \tag{8}$$

$$\frac{\partial^2 Q}{\partial \alpha \partial \beta} = 0 \tag{9}$$

ヘッセ行列は,

$$H = \begin{bmatrix} \partial^2 Q / \partial \alpha^2 & \partial^2 Q / \partial \alpha \partial \beta \\ \partial^2 Q / \partial \alpha \partial \beta & \partial^2 Q / \partial \beta^2 \end{bmatrix} = \begin{bmatrix} 2n & 0 \\ 0 & 2 \sum x_i^2 \end{bmatrix} \tag{10}$$

いま, 任意の 2 次元ベクトル $\boldsymbol{t} = (t_1, t_2)'$ に対して二次形式,

$$\boldsymbol{t}' H \boldsymbol{t} = (t_1, t_2) \begin{pmatrix} 2n & 0 \\ 0 & 2 \sum x_i^2 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = 2 \left(n t_1^2 + t_2^2 \sum x_i^2 \right) \geq 0 \tag{11}$$

となり, これが 0 となるのは, $\boldsymbol{t} = \boldsymbol{0}$ のときのみ. よってヘッセ行列 H は正定値. つまり, 式 (4) を満たす, α, β が極小値であることが分かる. また, 図 1 の形から, 極小かつ最小であることも分かる.

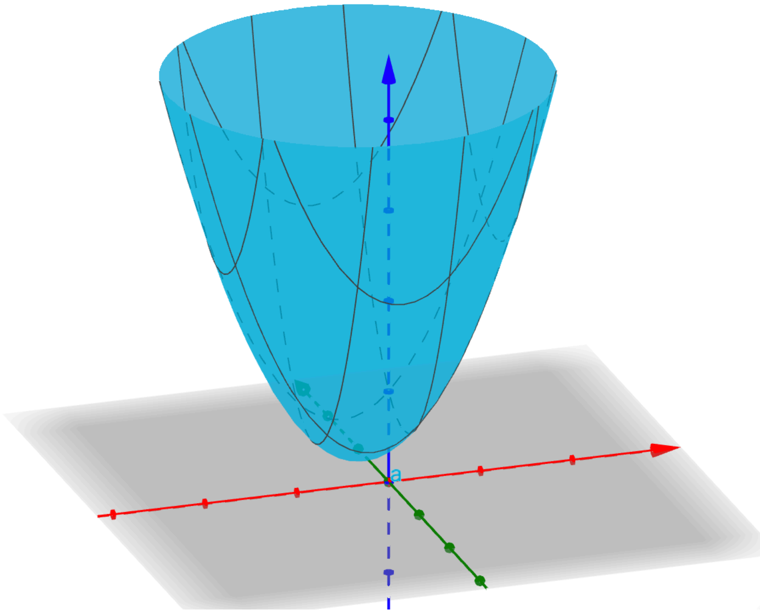

図 1　2 変数関数の二次関数のイメージ

式 (4) を解いていく.

$$\begin{cases} \frac{\partial Q}{\partial \alpha} = -2\sum(y_i - \alpha - \beta x_i) = -2\left\{\sum y_i - n\alpha - \beta\sum x_i\right\} = 0 \\ \frac{\partial Q}{\partial \beta} = -2\sum_{i=1}^{n} x_i(y_i - \alpha - \beta x_i) = -2\left\{\sum x_i y_i - \alpha\sum x_i - \beta\sum x_i^2\right\} = 0 \end{cases} \tag{12}$$

$$\Leftrightarrow \begin{cases} \alpha n + \beta\sum x_i = \sum y_i \\ \alpha\sum x_i + \beta\sum x_i^2 = \sum x_i y_i \end{cases} \tag{13}$$

付録 A

A.1　2 変数関数の極値

偏差平方和を最小にするような, α と β を求める際に,α, β に関する偏微分を 0 として, 推定量を求めた. しかし, 偏微分が 0 と置くだけでは, $Q(\alpha, \beta)$ の極値が求められるだけで, 最小値かどうかは分からない. テキストだと当たり前のように書かれているけど.

A.1.1　極大・極小

関数 $f(x, y)$ は連続かつ微分可能とする.

補題　境界点 [*1]でない極大点, 極小点においては

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0, \frac{\partial f}{\partial y}(x_0, y_0) = 0 \tag{14}$$

となる.

[*1] [5] p18 参照.

証明　もし $\frac{\partial f}{\partial x}(x_0, y_0) \neq 0$ ならば、微分可能性の定義[*2]

$$\begin{cases} f(x,y) - f(x_0,y_0) = \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)(y-y_0) + g(x,y), \\ \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \frac{g(x,y)}{|\boldsymbol{x}-\boldsymbol{x}_0|} = 0 \end{cases}$$

において, $y = y_0$ と置くと,

$$\begin{aligned} f(x,y_0) - f(x_0,y_0) &= \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + g(x,y_0) \\ &= \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0)\left\{1 + c \cdot \frac{g(x,y_0)}{x-x_0}\right\} \end{aligned}$$

右辺の $\{\}$ の中は x が x_0 に十分近いとき > 0 となる. 一方, $x - x_0$ は x が x_0 の近傍を動く時必ず符号を変えるため, $f(x_0, y_0)$ は極大値でも極小値でもないことが分かる. $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$ としても同様.

これにより, 連立方程式

$$\frac{\partial f}{\partial x}(x_0,y_0) = 0, \frac{\partial f}{\partial y}(x_0,y_0) = 0$$

を解けば, 境界点でない極値の候補者が全て見つかる. □

A.1.2　極大・極小の判定

極大値か極小値かの判定はヘッセ行列により行うことができる.

定義 (ヘッセ行列) 行列

$$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix}$$

を $f(x, y)$ に関するヘッセ行列という.

定理　$\frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial f}{\partial 0}(x_0, y_0) = 0$ を満たす点 x_0, y_0 においてヘッセ行列が[*3]

正定値　(x_0, y_0) は極小値

負定値　(x_0, y_0) は極大値

不定値　(x_0, y_0) は極値点ではない

A.1.3　正定値・負定値・不定値

$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ を任意の二次形式とする. ただし, $\boldsymbol{x} = (x_1, \ldots, x_n)'$.

非負定値 (nonnegative definite)　$\mathcal{R}^n$[*4]のあらゆる $\boldsymbol{x}$ に対して $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ であるとき. $\boldsymbol{x}$ の少なくとも 1 つの値, 具体的には $\boldsymbol{x} = \boldsymbol{0}$ に対して, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = 0$ であることに注意.

正定値 (positive definite)　$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ が非負定値かつ, 零ベクトル $\boldsymbol{0}$ に対してのみ $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = 0$ となる時. すなわち, $\boldsymbol{x} = \boldsymbol{0}$ を除く, あらゆる $\boldsymbol{x}$ に対して $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ のとき.

半正定値 (positive semidefinite)　非負定値であるが, 正定値でない二次形式. あらゆる $\boldsymbol{x} \in \mathcal{R}^n$ に対して, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ で, ある $\boldsymbol{x} \neq \boldsymbol{0}$ に対して $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = 0$ のとき.

非正定値, 負定値, 半負定値　それぞれ, $-\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ が, 非負定値, 正定値, 半正定値のとき.

[*2] $z = f(x, y)$ が点 $\boldsymbol{x}_0 = (x_0, y_0)$ で微分可能であるとは, ある定数 α, β があって,

$$\begin{cases} f(x,y) = f(x_0,y_0) + \alpha(x-x_0) + \beta(y-y_0) + g(x,y) \\ \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \frac{g(x,y)}{|\boldsymbol{x}-\boldsymbol{x}_0|} \end{cases} \qquad (\boldsymbol{x} = (x,y) \neq \boldsymbol{x}_0)$$

と表せること ([5] p153 参照)

[*3] 高校数学の美しい物語とか参考になる：`https://mathtrain.jp/hessian`

不定値 (indefinite)　非負定値でも, 非正定値でもない二次形式.

参考文献

[1] Tze Leung Lai, Haipeng Xing 著, 松原 望, 山村吉信 訳: 『ファイナンスのための統計学 -統計的アプローチによる評価と意思決定-』, 東京図書, 2016

[2] 鈴木 武, 山田 作太郎:『数理統計学 -基礎から学ぶデータ解析』, 内田老鶴圃, 1996

[3] 稲垣宣生:『数学シリーズ 数理統計学 (改訂版)』, 裳華房, 2003

[4] 江川博康:『弱点克服 大学生の微積分』東京図書, 2005

[5] 笠原晧司: 『サイエンスライブラリー 数学=12 微分積分学』, サイエンス社, 1974

[6] David A. Harville, (監訳) 伊里正夫 : 『統計のための行列代数 (上)』, 丸善出版, 2016

*4 記号 $\mathcal{R}^{m\times n}$ はその要素が全ての $m \times n$ 行列からなる線形空間を表す. 記号 $\mathcal{R}^n$ は全ての n 列次元ベクトルの集合.