確率統計学概論　解答例

2019.04.24

■　与えられたデータ (x_1, y_1)，(x_2, y_2)，$\cdots$，(x_n, y_n) に対して，関数

$$f(a,b) = \sum_{k=1}^{n} \{y_k - (a\,x_k + b)\}^2$$

が最小値を取る点 (a,b) を求めよ．ただし，データ x_1，x_2，$\cdots$，x_n の標準偏差 s_x は $s_x > 0$ であると仮定する．

（解） データ x_1，x_2，$\cdots$，x_n の平均を $\bar{x}$，データ y_1，y_2，$\cdots$，y_n の平均を $\bar{y}$，標準偏差を s_y，データ (x_1, y_1)，(x_2, y_2)，$\cdots$，(x_n, y_n) の共分散を s_{xy} とすると，

$$\sum_{k=1}^{n} x_k = n\,\bar{x}, \qquad \sum_{k=1}^{n} y_k = n\,\bar{y}, \qquad \sum_{k=1}^{n} x_k^2 = n\,(s_x^2 + \bar{x}^2),$$
$$\sum_{k=1}^{n} y_k^2 = n\,(s_y^2 + \bar{y}^2), \qquad \sum_{k=1}^{n} x_k\,y_k = n\,(s_{xy} + \bar{x}\,\bar{y})$$

が成り立つことに注意したい．$f(a,b)$ が点 (a_0, b_0) において極値を取るならば

$$\begin{aligned}
0 =& \frac{\partial f}{\partial a}(a_0, b_0) = \sum_{k=1}^{n} \left[2\,\{y_k - (a_0\,x_k + b_0)\}\,(-x_k)\right] = -2\sum_{k=1}^{n} y_k\,x_k + 2\,a_0 \sum_{k=1}^{n} x_k^2 + 2\,b_0 \sum_{k=1}^{n} x_k \\
=& 2\,n\,\left\{(s_x^2 + \bar{x}^2)\,a_0 + \bar{x}\,b_0 - (s_{xy} + \bar{x}\,\bar{y})\right\} = 2\,n\,\left(s_x^2\,a_0 - s_{xy}\right) + 2\,n\,\bar{x}\,\left(\bar{x}\,a_0 + b_0 - \bar{y}\right), \\
0 =& \frac{\partial f}{\partial b}(a_0, b_0) = \sum_{k=1}^{n} \left[2\,\{y_k - (a_0\,x_k + b_0)\}\,(-1)\right] = 2\,n\,\left(\bar{x}\,a_0 + b_0 - \bar{y}\right)
\end{aligned}$$

であるから，

$$a_0 = \frac{s_{xy}}{s_x^2}, \qquad b_0 = \bar{y} - \bar{x}\,a_0 = \frac{\bar{y}\,s_x^2 - \bar{x}\,s_{xy}}{s_x^2}$$

となる．

$$\frac{\partial^2 f}{\partial a^2}(a_0, b_0) = 2\sum_{k=1}^{n} x_k^2 = 2\,n\,(s_x^2 + \bar{x}^2), \qquad \frac{\partial^2 f}{\partial a \partial b}(a_0, b_0) = 2\sum_{k=1}^{n} x_k = 2\,n\,\bar{x}, \qquad \frac{\partial^2 f}{\partial b^2}(a_0, b_0) = 2\,n$$

より点 (a_0, b_0) での $f(a,b)$ のヘッセ行列は

$$H = \begin{pmatrix} 2\,n\,(s_x^2 + \bar{x}^2) & 2\,n\,\bar{x} \\ 2\,n\,\bar{x} & 2\,n \end{pmatrix}$$

であり，

$$\det H = 4\,n^2\,s_x^2 > 0, \qquad \mathrm{Tr}\,H = 2\,n\,(s_x^2 + \bar{x}^2 + 1) > 0$$

より $f(a,b)$ は点 (a_0, b_0) で極小値を取る．■