確率統計学　解答例

2014.06.03

■　データ (x_1, y_1)，(x_2, y_2)，$\cdots$，(x_n, y_n) が与えられ，分散 s_{xx} および s_{yy} は正であると仮定する．このとき，関数

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

を最小にする (a,b) を求めよ．

(解)　定義により

$$\begin{aligned}
s_{xx} =& \frac{1}{n}\sum_{k=1}^{n}\left(x_k^2 - 2\,\bar{x}\,x_k + \bar{x}^2\right) = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \frac{2\,\bar{x}}{n}\sum_{k=1}^{n} x_k + \bar{x}^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \bar{x}^2,\\
s_{yy} =& \frac{1}{n}\sum_{k=1}^{n}\left(y_k^2 - 2\,\bar{y}\,y_k + \bar{y}^2\right) = \frac{1}{n}\sum_{k=1}^{n} y_k^2 - \frac{2\,\bar{y}}{n}\sum_{k=1}^{n} y_k + \bar{y}^2 = \frac{1}{n}\sum_{k=1}^{n} y_k^2 - \bar{y}^2,\\
s_{xy} =& \frac{1}{n}\sum_{k=1}^{n}\left(x_k\,y_k - \bar{x}\,y_k - \bar{x}\,y_k + \bar{x}\,\bar{y}\right)\\
=& \frac{1}{n}\sum_{k=1}^{n} x_k\,y_k - \frac{\bar{x}}{n}\sum_{k=1}^{n} y_k - \frac{\bar{y}}{n}\sum_{k=1}^{n} x_k + \bar{x}\,\bar{y} = \frac{1}{n}\sum_{k=1}^{n} x_k\,y_k - \bar{x}\,\bar{y}
\end{aligned}$$

と表されることに注意すると，$f(a,b)$ は

$$\begin{aligned}
f(a,b) =& a^2\sum_{k=1}^{n} x_k^2 + 2\,a\,b\sum_{k=1}^{n} x_k + n\,b^2 - 2\,a\sum_{k=1}^{n} x_k\,y_k - 2\,b\sum_{k=1}^{n} y_k + \sum_{k=1}^{n} y_k^2\\
=& a^2\,n\,\left(s_{xx} + \bar{x}^2\right) + 2\,a\,b\,n\,\bar{x} + n\,b^2 - 2\,a\,n\,\left(s_{yx} + \bar{x}\,\bar{y}\right) - 2\,b\,n\,\bar{y} + n\,\left(s_{yy} + \bar{y}^2\right)
\end{aligned}$$

となる．$f(a,b)$ が最小値をとる (a,b) は

$$\begin{aligned}
0 = \frac{\partial f}{\partial a}(a,b) =& 2\,a\,n\,\left(s_{xx} + \bar{x}^2\right) + 2\,b\,n\,\bar{x} - 2\,n\,\left(s_{yx} + \bar{x}\,\bar{y}\right),\\
0 = \frac{\partial f}{\partial b}(a,b) =& 2\,a\,n\,\bar{x} + 2\,n\,b - 2\,n\,\bar{y}
\end{aligned}$$

をみたすので，

$$a = \frac{s_{xy}}{s_{xx}}, \qquad b = \bar{y} - a\,\bar{x}$$

である．■